

STA305/1004 - Class 10

February 8, 2017

Today's class

- ▶ Estimating the propensity score
- ▶ The balancing property of the propensity score
- ▶ Assessing balance
- ▶ Ignorable treatment assignment and the propensity score
- ▶ Three methods that use the propensity score to reduce bias: matching; stratification; and regression adjustment

The propensity score

- ▶ Covariates are pre-treatment variables and take the same value for each unit no matter which treatment is applied.
- ▶ For example, pre-treatment blood pressure or pre-test reading level are not influenced by a treatment that would alter blood pressure or reading level.

The propensity score is

$$e(\mathbf{x}) = P(T = 1|\mathbf{x}),$$

where \mathbf{x} are observed covariates.

The i^{th} propensity score is the probability that a unit receives treatment given all the information, recorded as covariates, that is observed before the treatment.

The propensity score

In experiments the propensity scores are known. In observational studies they can be estimated using models such as logistic regression where the outcome is the treatment indicator and the predictors are all the confounding covariates.

The propensity score

- ▶ Consider a study that plans to use a doctor's medical records to compare two treatments ($T = 0$ and $T = 1$) given for a certain condition.
- ▶ Treatments were not assigned to patients randomly, but were based on various measured and unmeasured patient factors.

Logistic Regression

- ▶ The logistic regression model with one covariate x is:

$$\log(P(T_i = 1)/P(T_i = 0)) = \beta_0 + \beta_1 x_i$$

- ▶ The logistic regression model with k covariates x_1, x_2, \dots, x_k is

$$\log(P(T_i = 1)/P(T_i = 0)) = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik}$$

Parameter Estimates from Logistic Regression

- ▶ In a logistic model with one binary covariate the parameter estimate of β_1 is:

$$\frac{(P(T = 1|x = 1)/P(T = 0|x = 1))}{(P(T = 1|x = 0)/P(T = 0|x = 0))} = \exp(\beta_1)$$

- ▶ $\exp(\beta_1)$ is the odds ratio comparing those with $x = 1$ to those with $x = 0$.

Predicted probabilities from Logistic Regression

- In a logistic model with one binary covariate the predicted probabilities can be calculated using the fitted model:

$$\hat{p}_i = \frac{\exp(\hat{\beta}_0 + \hat{\beta}_1 x_{i1})}{1 + \exp(\hat{\beta}_0 + \hat{\beta}_1 x_{i1})}$$

Can be used as an estimate of the propensity Score for the i^{th} Subject-

The propensity score

- ▶ The patient factors that were measured are age (x_1), sex (x_2), and health status before treatment (x_3).
- ▶ The propensity score can be estimated for each patient by fitting a logistic regression model with treatment as the dependent variable and x_1, x_2, x_3 as the predictor variables.

$$\log \left(\frac{p_i}{1 - p_i} \right) = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \hat{\beta}_3 x_{i3},$$

where $p_i = P(T_i = 1)$.

The propensity score

- ▶ The predicted probabilities from the above equation are estimates of the propensity score for each patient.

$$\hat{p}_i = \frac{\exp(\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \hat{\beta}_3 x_{i3})}{1 + \exp(\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \hat{\beta}_3 x_{i3})}$$

The propensity score in Smoking Cessation Study

fits a logistic regression model of
Smoking Status.

The propensity score for each subject in smoking and weight gain study can be estimated by fitting a logistic regression model.

```
prop.model <- glm(qsmk ~ as.factor(sex) + as.factor(race) +  
                    age + as.factor(education.code) + smokeintensity +  
                    smokeyrs + as.factor(exercise) + as.factor(active) +  
                    wt71, family = binomial(), data = nhefshwdat)
```

#Summary of propensity score model
summary(prop.model)

Treatment = { Stop Smoking
logistic Smoking
Regression model

The propensity score in Smoking Cessation Study

	Estimate	Std. Error	z value
(Intercept)	-2.401228039	0.484016356	-4.9610473
as.factor(sex)1	-0.499080121	0.146530691	-3.4059767
as.factor(race)1	-0.778222994	0.207031619	-3.7589572
age	0.046207220	0.009889326	4.6724338
as.factor(education.code)2	-0.065716379	0.196122828	-0.3350777
as.factor(education.code)3	0.052634524	0.175523000	0.2998725
as.factor(education.code)4	0.108653058	0.269190883	0.4036283
as.factor(education.code)5	0.466164550	0.224105901	2.0801083
smokeintensity	-0.026527450	0.005664293	-4.6832762
smokeyrs	-0.028491730	0.010008629	-2.8467165
as.factor(exercise)1	0.359556747	0.178603430	2.0131570
as.factor(exercise)2	0.422771538	0.185656969	2.2771649
as.factor(active)1	0.044927909	0.131555137	0.3415139
as.factor(active)2	0.158150602	0.213435405	0.7409764
wt71	0.006099273	0.004368231	1.3962800
	Pr(> z)		
(Intercept)	7.011411e-07		
as.factor(sex)1	6.592780e-04		
as.factor(race)1	1.706230e-04		
age	2.976515e-06		
as.factor(education.code)2	7.375665e-01		
as.factor(education.code)3	7.642744e-01		

How do we build a propensity score model?

- ▶ Usual tool is logistic regression model for the treatment allocation decision – We therefore want to consider including any variables that have a relationship to the treatment decision (i.e. precede it in time, and are relevant) – No information is included on the actual treatment received, or on the outcome(s).

Ten commandments of Propensity Model Development

1. Thou shalt value parsimony.
2. Thou shalt examine thy predictors for collinearity.
3. Thou shalt test all thy predictors for statistical significance.
4. Thou shalt have ten times as many predictors as subjects.
5. Thou shalt examine thy regression coefficients
6. Thou shalt perform bootstrap analyses to assess shrinkage.
7. Thou shalt perform regression diagnostics and examine residuals with care.
8. Thou shalt hold out a sample of thy data for cross validation.
9. Thou shalt perform external validation on a new sample of data.
10. Thou shalt ignore commandments 1 through 9 and instead ensure that the model adequately balances covariates.

Regression
Course
Don't
worry
about
X's

Propensity model development

1. Diagnostics for the successful prediction of probabilities and parameter estimates underlying those probabilities
2. Diagnostics for the successful design of observational studies based on estimated propensity scores.

In propensity score model development the second point is important, but the first is not important .

→ Key issues

Propensity model development

- ▶ All covariates that subject matter experts (and subjects) judge important when selecting treatments.
- ▶ All covariates that relate to treatment and outcome, including any covariate that improves prediction (of exposure group).
- ▶ As much “signal” as possible.

Propensity score in smoking cessation study

The propensity score for each subject is \hat{p}_i is the predicted probability of quitting smoking from the logistic regression model. The predicted probabilities are obtained using `predict()`.

#Propensity scores for each subject

```
p.qsmk.obs <- predict(prop.model, type = "response")
```

predict.glm

$$\hat{p}_i = \frac{\exp(\hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \dots)}{1 + \exp(\hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \dots)}$$

Propensity score in smoking cessation study

Subject	Quit Smoking	Estimated Propensity Score
1	0	0.12
2	0	0.16
3	0	0.16
4	0	0.31
5	0	0.32
6	0	0.17
7	0	0.24
8	0	0.26
9	0	0.30
10	0	0.29
11	1	0.26
12	0	0.19

Subject 1's estimated probability of quitting smoking is 0.12 (so the estimated probability of not quitting smoking is $1 - 0.12 = 0.82$) and subject 11's estimated probability of quitting smoking (propensity score) is 0.26 (so the estimated probability of not quitting smoking is $1 - .26 = 0.74$).

Propensity score in smoking cessation study

response = "term"

```
p1 <- predict.glm(prop.model)[1] #predicted value for the first subject  
p1
```

1
-1.955973

$$\hat{\beta}_0 + \hat{\beta}_1 x_{1L} + \hat{\beta}_2 x_{2L} + \dots$$

```
exp(p1)/(1+exp(p1))
```

1
0.1239035

```
# use type="response" to get predicted  
predict.glm(prop.model, type = "response")[1]
```

1
0.1239035

corresponds
to first
subject in
dataset.

The balancing property of the propensity score

The balancing property of the propensity score says that treated ($T = 1$) and control ($T = 0$) subjects with the same propensity score $e(\mathbf{x})$ have the same distribution of the observed covariates, \mathbf{x} ,

$$P(\mathbf{x}|T = 1, e(\mathbf{x})) = P(\mathbf{x}|T = 0, e(\mathbf{x}))$$

or

$$T \perp \mathbf{x} | e(\mathbf{x}).$$

This means that treatment is independent of the observed covariates conditional on the propensity score.

The balancing property of the propensity score

The balancing property says that if two units, i and j , are paired, one of whom is treated, $T_i + T_j = 1$, so that they have the same value of the propensity score $e(x_i) = e(x_j)$, then they may have different values of the observed covariate,

equal propensity score $x_i \neq x_j$,

but in this pair the specific value of the observed covariate will be unrelated to the treatment assignment since

$$P(x|T=1, e(x)) = P(x|T=0, e(x))$$

is the distribution of Covariate
Values
among treated
 $T=1$

\tilde{x}_i - Multidimensional
age, sex, wt 1971,
race, educ, --.

prob of
treatment.

The balancing property of the propensity score

The propensity scores for subject's 10 and 18 in the smoking cessation study are

	Quit Smoking	Estimated Propensity Score
10	0 No quit	0.2941244
18	1 quit	0.3197956

The difference between the two subject's propensity scores are $0.32 - 0.29 = 0.03$. This could be set as a "caliper" or "tolerance" for what are considered equal propensity scores.

The covariates for each subject are

	age	sex	race	edu	smkint	smkyrs	exer	active	wt1971	qsmk
10	43	0	0	2	20	25	2	1	62.26	0
18	48	1	0	3	2	30	1	1	62.03	1

results in a match on 9 covariates!

The balancing property of the propensity score

- ▶ If many pairs are formed way then the the distribution of the observed covariates will look about the same in the treated and control groups.
- ▶ Individuals in matched pairs will typically have different values of x .
- ▶ It is difficult to match on 9 covariates at once, it is easy to match on one covariate, the propensity score $e(x)$, and matching on $e(x)$ will tend to balance all 9 covariates.

How can the degree of balance in the covariate distributions between treated and control units be assessed?

The balancing property of the propensity score

If the smoking cessation and smoking groups are balanced using the propensity score then both observed and unobserved covariates will have similar distributions in the two groups. Thus, this observational study has been turned into a randomized study by using propensity score methods.



Respond at PollEv.com/nathantaback

Text **NATHANTABACK** to 37607 once to join, then **A or B**

True

A

14

False

B

24

Total Results: 0

Figure 1:

What is the difference in using Propensity Scores to form two groups vs. using Randomization to form two groups?

In an observational study using Propensity Scores to form two similar groups all the observed covariates will be balanced, but the unobserved covariates may not be balanced.

Assessing balance

- ▶ The difference in average covariate values by treatment status, scaled by their sample standard deviation. This provides a scale-free way to assess the differences.
- ▶ As a rule-of-thumb, when treatment groups have important covariates that are more than one-quarter or one-half of a standard deviation apart, simple regression methods are unreliable for removing biases associated with differences in covariates (Imbens and Rubin (2015)).

Assessing balance

If \bar{x}_t, s_t^2 are the mean and variance of a covariate in the treated group and \bar{x}_c, s_c^2 are the mean and variance of a covariate in the control group then the pooled variance is

Can be derived from
pooled SD formula
 $\text{w} \& n_t = n_c = \frac{n}{2}$ $\sqrt{\frac{s_t^2 + s_c^2}{2}}$.

Similar to pooled
SD, but it's
different

The absolute pooled standardized difference is,

$$\frac{100 \times |\bar{x}_t - \bar{x}_c|}{\sqrt{\frac{s_t^2 + s_c^2}{2}}}.$$

Similar to the
two sample
t-statistic but
different.

Key point: It's not to
look for sig. difference
but to look for unbalance.

Assessing balance

The absolute pooled standardized difference between the groups can be calculated for all the covariates using the function MatchBalance in the library Matching.

library(Matching)

mb <- MatchBalance(qsmk ~ as.factor(sex) + as.factor(race) +
age + as.factor(education.code) +
smokeintensity + smokeyrs +
as.factor(exercise) +
as.factor(active) + wt71, data=nhefshwdat, nboots=1

logistic regression model.

If the absolute value of the standardized mean difference is greater than 10% then this indicates a serious imbalance. For example, sex has an absolute standardized mean difference of $| -16.022 | = 16.022$ indicating serious imbalance between the groups in males and females.

Assessing balance in the smoking cessation study

before matching using the PS.

$$-27.714 \\ > 10$$

Output from MatchBalance().

***** (V3) age *****

before matching:

mean treatment..... 46.174

mean control..... 42.788

std mean diff..... 27.714

NB: some output is omitted ...

Standardized mean
diff. before matching.
indicates Serious
imbalance in age.

If the absolute value of the standardized mean difference is greater than 10% then this indicates a serious imbalance. Age has an absolute standardized mean difference of 27.714 indicating serious imbalance between the groups in age.

27.714

Assessing balance in the smoking cessation study

```
***** (V2) as.factor(race)1 *****
before matching:
mean treatment..... 0.08933
mean control..... 0.14617
std mean diff..... -19.905

mean raw eQQ diff.... 0.057072
med raw eQQ diff.... 0
max raw eQQ diff.... 1

mean eCDF diff..... 0.028422
med eCDF diff..... 0.028422
max eCDF diff..... 0.056844

var ratio (Tr/Co).... 0.65287
T-test p-value..... 0.0012863
```

indicates Serious
imbalance in race
variable.

Ignore this
part of the
output

Assessing Balance in the smoking cessation study

```
***** (V14) wt71 *****
```

before matching:

mean treatment..... 72.355

mean control..... 70.303

std mean diff..... 13.13

mean raw eQQ diff.... 2.1872

med raw eQQ diff.... 2.04

max raw eQQ diff.... 14.75

mean eCDF diff..... 0.032352

med eCDF diff..... 0.032386

max eCDF diff..... 0.07

var ratio (Tr/Co).... 1.0606

T-test p-value..... 0.022421

KS Bootstrap p-value.. 0.1

KS Naive p-value..... 0.10646

KS Statistic..... 0.07

serious imbalance.

Propensity scores and ignorable treatment assignment

Assume that the treatment assignment T is strongly ignorable. This means that

$$P(T|Y(0), Y(1), \mathbf{x}) = P(T|\mathbf{x}),$$

or

$$T \perp Y(0), Y(1) | \mathbf{x}.$$

It may be difficult to find a treated and control unit that are closely matched for every one of the many covariates in x , but it is easy to match on one variable, the propensity score, $e(\mathbf{x})$, and doing that will create treated and control groups that have similar distributions for all the covariates.

Propensity scores and ignorable treatment assignment

Ignorable treatment assignment implies that

$$P(T|Y(0), Y(1), e(\mathbf{x})) = P(T|e(\mathbf{x})),$$

or

$$T \perp Y(0), Y(1) | e(\mathbf{x}).$$

This means that the scalar propensity score $e(\mathbf{x})$ may be used in place of the many covariates in \mathbf{x} .

Can use the ps to balance, for example,
9 Covariates!

Propensity scores and ignorable treatment assignment

- ▶ The propensity score can be used in place of many covariates.
- ▶ If treatment assignment is strongly ignorable then propensity score methods will produce unbiased results of the treatment effects.
- ▶ In the smoking cessation study what does it mean for treatment assignment to be ignorable?
 - ▶ The potential outcomes for weight gain in the smoking cessation (treated) and smoking (control) groups are independent conditional on the propensity score.
 - ▶ The treatment assignment mechanism has been reconstructed using the propensity score.

→ $Y = \text{Weight gain}$

$T = \text{Smoking}$

T is independent of $Y(0)$ and $Y(1)$ given a fixed value of the propensity score.

$0 = \text{Smoking}$
 $1 = \text{Non-Smoking}$

Propensity scores and ignorable treatment assignment

- ▶ Suppose a critic came along and claimed that the study did not measure an important covariate (e.g., spouse is a smoker) so the study is in no position to claim that the smoking cessation group and the smoking groups are comparable.
- ▶ This criticism could be dismissed in a randomized experiment — randomization does tend to balance unobserved covariates — but the criticism cannot be dismissed in an observational study.
- ▶ This difference in the unobserved covariate, the critic continues, is the real reason outcomes differ in the treated and control groups: it is not an effect caused by the treatment, but rather a failure on the part of the investigators to measure and control imbalances in the unobserved covariate.
- ▶ The sensitivity of an observational study to bias from an unmeasured covariate is the magnitude of the departure from the model that would need to be present to materially alter the study's conclusions.
- ▶ There are statistical methods to measure how sensitive an observational study is to this type of bias.

If the smoking cessation study were a randomized study comparing weight gain in smokers versus quitters then a valid criticism is that the treatment effect could be due to an unobserved covariate?



Respond at PollEv.com/nathantaback



Text **NATHANTABACK** to 37607 once to join, then **A or B**

True

A

13

False

B

21

Total Results: 0

C If the Study was randomized then
Figure 2:
the unobserved Covariates Should be balanced!

Using the propensity score to reduce bias

The three most common techniques that use the propensity score are

1. matching,
 2. stratification (also called subclassification)
 3. regression adjustment.
-
- ▶ Each of these techniques is a way to make an adjustment for covariates prior to (matching and stratification) or while (stratification and regression adjustment) calculating the treatment effect.
 - ▶ With all three techniques, the propensity score is calculated the same way, but once it is estimated it is applied differently.

Maimonides' Rule

- ▶ Educators are very interested in studying the effect of class size on learning.
- ▶ Does smaller class size cause students to achieve higher math and verbal scores?
- ▶ Angrist and Lavy (1999) published an unusual study of the effects of class size on academic achievement.
- ▶ Causal effects of class size on pupil achievement is difficult to measure. The twelfth century Rabbinic scholar Maimonides interpreted the the Talmud's discussion of class size as:
- ▶ "Twenty-five children may be put in charge of one teacher. If the number in the class exceeds twenty-five but is not more than forty, he should have an assistant to help with instruction. If there are more than forty, two teachers must be appointed."

Maimonides' Rule

- ▶ Since 1969 the rule has been used to determine class size in Israeli public schools.
- ▶ Class size is usually determined by other factors such as wealth of a community, special needs of students, etc.
- ▶ If adherence to Maimonides' rule were perfectly rigid, then what would separate a school with a single class of size 40 from the same school with two classes whose average size is 20.5 is the enrollment of a single student.

Number of children in grade 5	40	80	120
Class size with one extra student	20.5	27	30.25

Maimonides' Rule

- ▶ Angrist and Lavy matched schools where the number of grade 5 students are 31-40 to schools where the number of grade 5 students are 41-50.
- ▶ 86 matched pairs of two schools were formed, matching to minimize total absolute difference in percentage disadvantaged.
- ▶ It's plausible that whether or not a few more students enrol in the fifth grade is a haphazard event.
- ▶ This is an example of natural experiment where students were haphazardly (randomly) assigned to small or large grade 5 classes.
- ▶ It was haphazard because it depended only on the number of grade 5 children at a school.

Maimonides' Rule

From Rosenbaum, 2010, pg.9

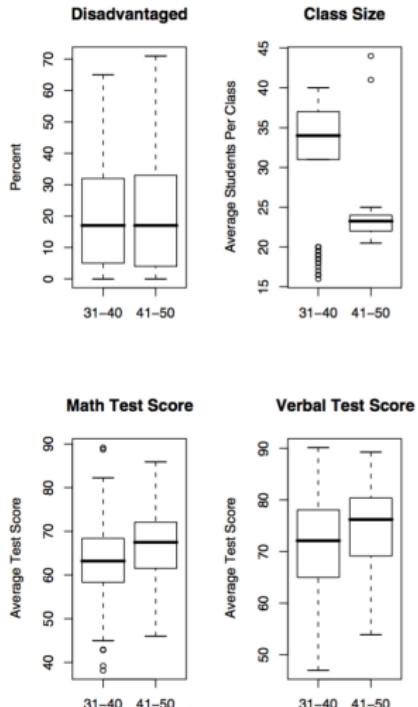


Fig. 1.1 Eighty-six pairs of two Israeli schools, one with between 31 and 40 students in the fifth grade, the other with between 41 and 50 students in the fifth grade, matched for percentage of students in the school classified as disadvantaged. The figure shows that the percentage of disadvantaged students is balanced, that imperfect adherence to Maimonides' rule has yielded substantially different average class sizes, and test scores were higher in the group of schools with predominantly smaller class sizes.

Propensity score matching

- ▶ In the Maimonides rule study assignment to a small/large was haphazard/random.
- ▶ If there is no opportunity to take advantage of this type of treatment assignment then we can calculate the propensity score and use this to match.
- ▶ For each unit we have a propensity score.
- ▶ Randomly select a treated subject.
- ▶ Match to a control subject with closest propensity score (within some limit or "calipers").
- ▶ Eliminate both units from the pool of subjects until there is no acceptable match.
- ▶ It's not always possible to match every unit treated to a unit that is not treated.

algorithm for matching using PS-

Propensity score matching

```
prop.model <- glm(qsmk ~ as.factor(sex) + as.factor(race) +
                    age + as.factor(education.code) +
                    smokeintensity + smokeyrs +
                    as.factor(exercise) + as.factor(active) +
                    wt71, family = binomial(),
                    data = nhefshwdat)

X <- prop.model$fitted; Y <- nhefshwdat$wt82_71; Tr <- nhefshwdat$qsmk

library(Matching)
rr <- Match(Y=Y, Tr=Tr, X=X, M=1); summary(rr)
```

Se of the estimate — difference in wt.gain between smokers and non-smokers.

(Estimate... 2.9342 — p-Value for the difference.
AI SE..... 0.5838
T-stat..... 5.026
p.val..... 5.0087e-07

Original number of observations.....	1566	not possible to match each subject
Original number of treated obs.....	403	
Matched number of observations.....	403	
Matched number of observations (unweighted).	1009	

Propensity score matching

After matching on covariates the treatment effect (difference in weight gain between the group that stopped smoking and the group that did not stop smoking) is 2.93 with a p-value of 0 (5.0087e-07) and 95% confidence interval (1.84, 4.02).

$$2.93 \pm 1.96 (0.5838)$$
$$t_{1009} \approx z$$

Propensity score matching -check covariate balance

Now, let's check covariate balance.

```
MatchBalance(qsmk ~ as.factor(sex) + as.factor(race) +  
             age + as.factor(education.code) +  
             smokeintensity + smokeyrs +  
             as.factor(exercise) +  
             as.factor(active) + wt71, data=nhefshwdat,  
             match.out=rr, nboots=10)
```

Propensity score matching -check covariate balance

***** (V1) as.factor(sex)1 *****

	Before Matching	After Matching
mean treatment.....	0.45409	0.45409
mean control.....	0.53396	0.45331
std mean diff.....	-16.022	0.15703 < 10

NB: some output omitted

Matching improved Covariate
balance for sex, age

***** (V3) age *****

	Before Matching	After Matching
mean treatment.....	46.174	46.174
mean control.....	42.788	46.595
std mean diff.....	27.714	{-3.4504} < 10

NB: some output omitted

Sex has an absolute standardized difference of 16 before matching and 0.16 after matching, and the absolute standardized difference of age has shifted from 27.71 to -3.45.

In theory PS should balance the observed covariates, but it may not be the case in every dataset.

Propensity score matching -check covariate balance

How does this compare to not adjusting for imbalance?

#Unadjusted t-test

```
t.test(nhefshwdat$wt82_71[as.factor(nhefshwdat$qsmk)==1],  
       nhefshwdat$wt82_71[as.factor(nhefshwdat$qsmk)==0],  
       var.equal=T)
```

Two Sample t-test

```
data: nhefshwdat$wt82_71[as.factor(nhefshwdat$qsmk) == 1] and nhefshwdat$wt82_71[as.factor(nhefshwdat$qsmk) == 0]  
t = 5.6322, df = 1564, p-value = 2.106e-08  
alternative hypothesis: true difference in means is not equal to 0  
95 percent confidence interval:  
 1.655796 3.425367  
sample estimates:  
mean of x mean of y  
4.525079 1.984498
```

*not everyone was matched
and matching adjusts for covariates*

The unadjusted treatment effect is 2.54 with a p-value of 0. So, both analyses lead to the same conclusion that stopping to smoke leads to a significant weight gain. Although the weight gain in the matched propensity score analysis is 0.39Kg higher.

Propensity score subclassification/stratification

Propensity scores permit subclassification on multiple covariates simultaneously. One advantage of this method is that the whole sample is used and not just matched sets.

Cochran (1968) showed that creating five strata removes 90 per cent of the bias due to the stratifying variable or covariate.

Rosenbaum and Rubin holds for stratification based on the propensity score. Stratification on the propensity score balances all covariates that are used to estimate the propensity score, and often five strata based on the propensity score will remove over 90 per cent of the bias in each of these covariates.

Stratification

The following data were selected from data supplied to the U. S. Surgeon General's Committee from three of the studies in which comparisons of the death rates of men with different smoking habits were made (Cochran, 1968).

The table shows the unadjusted death rates per 1,000 person-years.

Smoking group	Canadian	British	U.S.
Non-smokers	20.2	11.3	13.5
Cigarettes only	20.5	14.1	13.5
Cigars, pipes	35.5	20.7	17.4

Conclusion: urge the cigar and pipe smokers to give up smoking and if they lack the strength of will to do so, they should switch to cigarettes.

Stratification

- ▶ Are there other variables in which the three groups of smokers may differ, that (i) are related to the probability of dying; and (ii) are clearly not themselves affected by smoking habits?
- ▶ The regression of probability of dying on age for men over 40 is a concave upwards curve, the slope rising more and more steeply as age advances. The mean ages for each group in the previous table are as follows.

Smoking group	Canadian	British	U.S.
Non-smokers	54.9	49.1	57.0
Cigarettes only	50.5	49.8	53.2
Cigars, pipes	65.9	55.7	59.7

Stratification

- ▶ The table shows the adjusted death rates obtained when the age distributions were divided into 9 subclasses.
- ▶ The results are similar for different numbers of subclasses.

Smoking group	Canadian	British	U.S.
Non-smokers	20.2	11.3	13.5
Cigarettes only	29.5	14.8	21.2
Cigars, pipes	19.8	11.0	13.7

bras was
due to
age!

Compare to the unadjusted death rates

Smoking group	Canadian	British	U.S.
Non-smokers	20.2	11.3	13.5
Cigarettes only	20.5	14.1	13.5
Cigars, pipes	35.5	20.7	17.4

Cochran (1968) showed that creating 5 or more strata removes 90% of the bias due to the stratifying variable.

Propensity score subclassification/stratification

```
prop.model <- glm(qsmk ~ as.factor(sex) + as.factor(race) +
                    age + as.factor(education.code) +
                    smokeintensity + smokeyrs +
                    as.factor(exercise) + as.factor(active) +
                    wt71, family = binomial(),
                    data = nhefshwdat)

p.qsmk.obs <- predict(prop.model, type = "response")
strat <- quantile(p.qsmk.obs, probs = c(.2,.4,.6,.8)) break up
strat1 <- p.qsmk.obs <= strat[1] ps into quintiles.
propmodel1 <- glm(wt82_71[strat1]~qsmk[strat1], data=nhefshwdat)
summary(propmodel1)

strat2 <- p.qsmk.obs > strat[1] & p.qsmk.obs <= strat[2]
propmodel2 <- glm(wt82_71[strat2]~qsmk[strat2], data=nhefshwdat)
summary(propmodel2)

strat3 <- p.qsmk.obs > strat[2] & p.qsmk.obs <= strat[3]
propmodel3 <- glm(wt82_71[strat3]~qsmk[strat3], data=nhefshwdat)
summary(propmodel3)

strat4 <- p.qsmk.obs > strat[3] & p.qsmk.obs <= strat[4]
propmodel4 <- glm(wt82_71[strat4]~qsmk[strat4], data=nhefshwdat)
summary(propmodel4)

strat5 <- p.qsmk.obs > strat[4]
propmodel5 <- glm(wt82_71[strat5]~qsmk[strat5], data=nhefshwdat)
summary(propmodel5)
```

Propensity score subclassification/stratification

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3.582883	0.4463651	8.026799	2.055049e-14
qsmk[strat1]	1.571867	1.2204794	1.287909	1.987319e-01

effect
of Smoking
in the
first quintile
effect in
2nd quintile

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	2.700017	0.4466046	6.045654	4.258361e-09
qsmk[strat2]	5.054241	1.0286540	4.913451	1.449627e-06

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	2.121351	0.5384292	3.939888	0.0001007005
qsmk[strat3]	3.726930	1.0519470	3.542888	0.0004564504

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.9552357	0.5130865	1.861744	6.358234e-02
qsmk[strat4]	3.8711676	0.9463872	4.090469	5.488916e-05

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.2892809	0.5878199	-0.4921251	0.62297810
qsmk[strat5]	2.0550465	0.9192030	2.2356829	0.02608187

effect
in 5th
quintile

Propensity score subclassification/stratification

In summary the 5 quintiles produced treatment effects

Estimate (se)	P-value	PS Quintile
1.57 (1.22)	0.199	1
5.05 (1.03)	0.00	2
3.73 (1.05)	0.00	3
3.87 (0.95)	0.00	4
2.06 (0.92)	0.03	5

The overall treatment effect is 3.26, which can be obtained by averaging the estimates within each stratum. This is a larger estimate compared to the treatment effect obtained by matching. The treatment effect can also be estimated by fitting a linear regression model for the change in weight on the treatment variable and the quintiles of the estimated propensity score.

Propensity score subclassification/stratification

```
attach(nhefshwdat)
#create a variable to describe subclass to include in the model
stratvar <- numeric(length(qsmk))
for (i in 1:length(qsmk))
{
  if (strat1[i]==T) {stratvar[i] <- 1}
  else
    if (strat2[i]==T) {stratvar[i] <- 2}
  else
    if (strat3[i]==T) {stratvar[i] <- 3}
  else
    if (strat4[i]==T) {stratvar[i] <- 4}
  else stratvar[i] <- 5
}
stratmodel <- glm(wt82_71~qsmk+as.factor(stratvar), data=nhefshwdat)
summary(stratmodel)$coef
```

add Stratification variable
to regression
to get treatment effect

Propensity score subclassification/stratification

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3.3564796	0.4373149	7.6752008	2.894022e-14
qsmk	3.2645028	0.4542610	7.1864036	1.027063e-12
as.factor(stratvar)2	-0.3191003	0.6134512	-0.5201723	6.030173e-01
as.factor(stratvar)3	-1.1139815	0.6157083	-1.8092683	7.060174e-02
as.factor(stratvar)4	-2.2229271	0.6172504	-3.6013378	3.265030e-04
as.factor(stratvar)5	-4.1403625	0.6255644	-6.6186033	4.971650e-11

	2.5 %	97.5 %
(Intercept)	2.499358	4.21360105
qsmk	2.374168	4.15483802
as.factor(stratvar)2	-1.521443	0.88324196
as.factor(stratvar)3	-2.320748	0.09278456
as.factor(stratvar)4	-3.432716	-1.01313860
as.factor(stratvar)5	-5.366446	-2.91427883

The linear regression yields the same treatment effect as averaging the estimates, but also provides an estimate of standard error, p-value, and confidence interval for the treatment effect.

Propensity score subclassification/stratification

We can investigate covariate balance within subclasses. In practice this should occur prior to looking at the outcome data. The number of subjects and average propensity score (shown in brackets) within each treatment group by subclass is shown in the table below.

Subclass	Smoking Cessation	No smoking cessation
1	42 (0.14)	272 (0.12)
2	59 (0.2)	254 (0.19)
3	82 (0.24)	231 (0.24)
4	92 (0.31)	221 (0.3)
5	128 (0.43)	185 (0.41)

For example, the percentage of males in each subclass are:

Subclass	Smoking Cessation	No Smoking Cessation
1	28.57%	22.79%
2	44.07%	43.31%
3	54.88%	46.32%
4	55.43%	59.73%
5	67.19%	70.81%

Multivariate adjustment using the propensity score

- ▶ Another method for using the propensity score to adjust for bias is to use the propensity score itself as a predictor along with the treatment indicator.
- ▶ The treatment effect is adjusted by the propensity score.

```
prop.model.adj <- glm(wt82_71 ~ qsmk + p.qsmk.obs, data = nhefshwdat)
summary(prop.model.adj)$coef
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	5.560244	0.5090376	10.923052	8.104078e-27
qsmk	3.397278	0.4559641	7.450757	1.528381e-13
p.qsmk.obs	-14.751531	1.8846521	-7.827190	9.128614e-15

```
confint(prop.model.adj)
```

	2.5 %	97.5 %
(Intercept)	4.562548	6.557939
qsmk	2.503604	4.290951
p.qsmk.obs	-18.445381	-11.057680

*treatment effect
adjusted for
PS
95% CI*

The treatment effect is similar to the stratification method.

Comparing the three methods

The three propensity score methods yield similar results for the treatment effect.

Method	Average Treatment Effect	95% Confidence Interval
Matched	2.93	1.8 - 4.0
Stratified	3.26	1.7 - 3.4
Regression	3.40	2.5 - 4.3
Unadjusted	2.54	1.7 - 3.4

The unadjusted analysis (two-sample t-test) underestimates the treatment effect by approximately 1kg.

Summary

- ① Used the PS in three different ways to calculate the effect of treatment in an observational study.
- ② Three methods : Matching, Stratification, direct regression adjustment.
- ③ Check that Covariates are actually balanced using PS methods.

(4)

If covariate were not balanced
using a method then treatment
difference might be biased.