Propensity Score Matching in Observational Studies

Author: Arane Thavaneswaran
Corresponding Author: Lisa Lix

Date: April 22, 2008

Remediated for Accessibility: April 15, 2024

University of Manitoba

Manitoba Centre for Health Policy

Propensity Score Matching in Observational Studies

Introduction

Propensity scores are an alternative method to estimate the effect of receiving treatment when random assignment of treatments to subjects is not feasible. Propensity score matching (PSM) refers to the pairing of treatment and control units with similar values on the propensity score, and possibly other covariates, and the discarding of all unmatched units (Rubin, 2001). It is primarily used to compare two groups of subjects but can be applied to analyses of more than two groups.

History of PSM

The concept of PSM was first introduced by Rosenbaum and Rubin (1983) in a paper entitled "The Central Role of the Propensity Score in Observational Studies for Casual Effects."

Heckman (1997) also played a role in the development of propensity score matching methods. He focused on selection bias, with a primary emphasis on making casual inferences when there is non- random assignment. He later developed the difference-in-differences approach which has applications to PSM.

Statistical Definition

The estimated propensity score $e(x_i)$, for subject *i,(i = 1,…, N*) is the conditional probability of being assigned to a particular treatment given a vector of observed covariates x_i (Rosenbaum and Rubin, 1983):

$$e(\mathbf{x}_i) = \Pr(z_i = 1 \mid \mathbf{x}_i)$$

and

$$\Pr(z_1,\ldots,z_n \mid x_1,\ldots,x_n) = \prod_{i=1}^{N} e(x_i)^{z_i}\{1 - e(x_i)\}^{1-z_i}$$

where:

$z_i = 1$, for treatment

$z_i = 0$, for control

and

x_i, the vector of observed covariates for the i^th^ subject

In randomized studies, covariates are variables that are not affected by the allocation of treatments to subjects.

Since the propensity score is a probability, it ranges in value from 0 to 1.

To explain further, IF propensity score matching was used in a randomized experiment comparing two groups, then the propensity score for each participant in the study would be 0.50. This is because each participant would be randomly assigned to either the treatment or the control group with a 50% probability. In study designs where there is no randomization, such as in a quasi-experimental design, the propensity score must be estimated. Propensity score values are dependent on a vector of observed covariates that are associated with the receipt of treatment.

Generally, if a treated subject and a control subject have the same propensity score, the observed covariates are automatically controlled for. Therefore, any differences between the treatment and control groups will be accounted for and will not be as a result of the observed covariates.

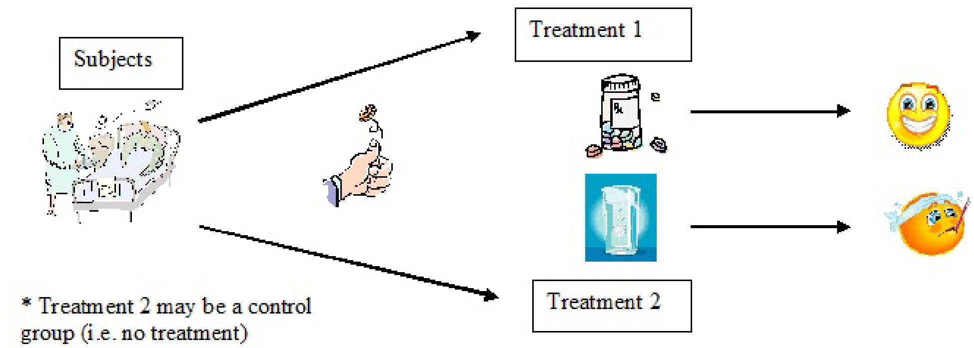
Why and When Do We Use Propensity Scores?

To make causal inferences, random selection of subjects and random allocation of the treatment to subjects is required. In observational studies random assignment to treatments is not possible. The primary limitation of an observational study is that there may be random selection of subjects but **not** random allocation of treatments to subjects. When there is a lack of randomization, casual inferences cannot be made because it is not possible to determine whether the difference in outcome between the treated and control (untreated) subjects is due to the treatment or differences between subjects on other characteristics. Subjects with certain characteristics may be more likely to receive treatment than others.

Study Designs and Control of Confounding Covariates

1.) Randomized Design

<u>Method:</u> This design uses random allocation of treatments to subjects. For two groups of subjects, randomization ensures that subjects are equally matched on all factors, usually with the simple flip of a coin.



Advantages:

- This method ensures that the two groups of subjects are matched equally on all factors even before determining what these factors may be.
- It is ideal for making casual inferences.
- It does not depend on conditioning on the observed covariates and can balance for both observed and unobserved covariates.

Disadvantages:

- Expensive
- Randomization may be infeasible or impractical because of ethical concerns.
- There are issues of generalizabilty of study designs:
  - Subjects may not be representative of the general population
  - Ideally, a design would include the random selection of subjects and random allocation of the treatments to subjects. In observational studies, there may be random selection of subjects but **not** random allocation of treatments to the subjects. Therefore, there is assignment bias which is when the researcher has no control over the assignment of treatments to subjects or over what variables are collected. Although causal inferences cannot be made from observational studies, they are less expensive and more generalizable to the general population than randomization.

2.) Quasi- experimental Design

Method: This method is implicated when randomization is often impractical or impossible and there is also no control over extraneous variables. A quasi-experimental design is created when the probability that a subject would have been treated is used to adjust for the estimate of the treatment effect.
For example, if you want to undertake a study that determines the effect of drinking an average of three beers a day on an individual's heart rate, it would be

unethical to use randomization. Subjects who drink an average of three beers a day are assigned to be the treatment group and those who do not drink any beer are assigned to the control group. In a quasi-experimental design, cause and effect relationships cannot be inferred because there is no randomization of treatments to the subjects or manipulation of variables.

Advantages:

- Tend to be more generalizable and representative of real-world conditions than randomized experiments.
- Can be used to adjust for the estimate of the treatment effect in studies where randomization is not possible.
- Can control for confounding variables and extraneous variables. Extraneous variables are variables other than independent variables that influence the outcome. A confounding variable is an independent variable, whose effects on the dependent variable cannot be differentiated from the other independent variables because of its relation to both. For example, in a study where you want to know if being a male causes liver cancer, drinking would be a confounding variable.

Disadvantages:

- Primary drawback is that the estimates of the treatment effects may be affected by selection bias. Since there is nonrandom selection, the differences between the groups may be regarded as being because of the treatment effects when in fact it may be due to differences between the treatment and control groups.
- Since there is no randomization, casual inferences cannot be made.

3.) Matching Designs

Method: In this method, we match on observed characteristics that distinguish treatment and control groups to make the groups more similar.

Advantages:

- Matching ensures that any differences between the treatment and the control groups are **not** a result of differences on the matching variables.
- Useful in studies with small sample sizes because when there are only a few confounding variables, it is easy to match on one or more variables as opposed to matching on many variables, which is difficult.

Disadvantages:

- The effects of the matching variables on the outcome cannot be studied.
- If there isn't sufficient overlap between the two groups on the matching variables, then biases such as the regression toward the mean may occur.
- Assumes that all relevant covariates have been measured.

General Method for Calculating Propensity Scores

1.) Propensity scores are generally calculated using one of two methods: a) Logistic regression or b) Classification and Regression Tree Analysis.

a) **<u>Logistic regression</u>**: This is the most used method for estimating propensity scores. It is a model used to predict the probability that an event occurs.

$$\ln \frac{e(x_i)}{1-e(x_i)} = \ln \frac{\Pr(z_i = 1 \mid x_i)}{1 - \Pr(z_i = 1 \mid x_i)} = \alpha + \beta' x_i$$

where:

$$e(\mathbf{x}_i) = \Pr(z_i = 1 \mid \mathbf{x}_i)$$

$$e(X_i) = b_0 + b_1 X_1 + b_2 X_2 + b_3 X_3 + \ldots + b_i X_i$$

and

b_0 is the intercept

b_i is the regression coefficient

X_i, the treatment variables and covariates (random variables)

$\mathbf{x}_i$, observed value of variables

In logistic regression, the dependent variable is binary, Z_i=1 is the value for the treatment and the value for the control is Z_i =0.

b) **<u>Classification and Regression tree analysis (CART):</u>** This is a non-parametric decision tree method that can efficiently partition populations into homogenous subgroups (Lemon, Freidmann, Rakowski, 2003). It is not as widely used as logistic regression for estimating propensity scores because it is complex and more suitable for use by those with a statistical background.

2.) Adjustment for the estimated propensity scores is accomplished using one or a combination of the four main methods. (1) Stratification, (2) Matching, (3) Covariate/Regression adjustment, and (4) Weighting.

An Example: Estimating the Propensity Score

- $e(\mathbf{x}_i) = \Pr(T_i = 1 \mid x_{1i}, x_{2i}, \ldots x_{ki})$

  $$\hat{e}(\mathbf{x}_i) = \frac{1}{1 + e^{-(\hat{a} + S\, \hat{b}_k x_{ik})}}$$

  T_i= TRT where T_i =0, without treatment or T_i =1, with treatment
  x_{1i}= $\text{SEX}_{(1,2)}$ where 1= female and 2= male
  x_{2i}= $\text{AGE}_{(\text{CONTINUOUS})}$
  x_{3i}= $\text{HYP}_{(0,1)}$ where HYP=0, without hypertension or HYP=1, with hypertension

  So the model becomes,

  $$\hat{e}(\mathbf{x}_i) = \frac{1}{1 + e^{-(\hat{a} + \hat{b}_1 SEX_i + \hat{b}_2 AGE_i + \hat{b}_3 HYP_i)}}$$

- Suppose that we have the following parameters, which are usually estimated using maximum likelihood (ML) techniques:

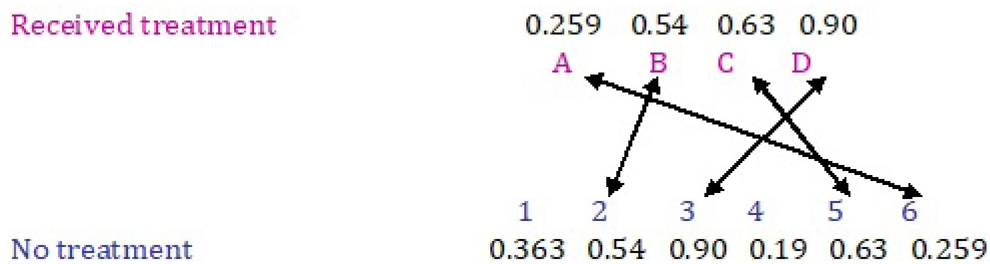
  $$\left.\begin{array}{l} \hat{a} = -3.9 \\ \hat{b}_1 = 0.63 \\ \hat{b}_2 = 0.025 \\ \hat{b}_3 = 0.343 \end{array}\right\} \quad \hat{e}(\mathbf{x}_i) = \frac{1}{1 + e^{-[-3.9 + 0.63(SEX_i) + 0.025(AGE_i) + 0.343(HYP_i)]}}$$

  Let’s say that particular subject is female (SEX=1), 75 (AGE=75) and has hypertension (HYP=1).

$$\hat{e}(\mathbf{x}_i) = \frac{1}{1 + e^{-[-3.9 + 0.63(1) + 0.03(75) + 0.34(1)]}} = \frac{1}{1 + e^{-(-1.05)}} = 0.259$$

Once we calculate the estimated propensity scores, we can match the treated subjects with subjects that have the same/similar propensity score but did not receive treatment. This example follows a 1-to-1 match:

Propensity Score Matching in Observational Studies



The unmatched subjects are discarded from the analysis.

Propensity Score Matching in Observational Studies

Using Logistic Regression to Estimate Propensity Scores

- Consider including interactions and polynomial effects
  - Don't need to err on the side of parsimonious model

- Use C-statistic to guide model selection – does model discriminate between treatment and control groups?
  - But model discrimination does not have any relationship with bias reduction via propensity score adjustment

Using CART to Estimate Propensity Scores

- It is not as widely used as logistic regression for estimating propensity scores because it may not be as readily understood.

- It does not make any distributional assumptions about the explanatory variables, nor does it assume a linear relationship between the treatment and covariates.

An Example: Aspirin Use and Mortality

Example taken from Love, TE. (2004). Propensity Scores: Helping Non-Statisticians Get the Message. Presented at the Joint Statistical Meetings, Toronto.

- 6174 consecutive adults undergoing stress echocardiography for evaluation of known or suspected coronary disease
- 2310 (37%) were taking aspirin (treatment)
- Main outcome: all-cause mortality
- Median follow-up: 3.1 years
- Unadjusted analyses:
  - 4.5% of aspirin patients died and 4.5% of non-aspirin patients died
  - Hazard ratio: 1.08 (0.85, 1.39)

- 31 covariates were included in the LR model:
  - Demographics (age, sex)
  - Cardiovascular risk factors
  - Use of other medications
  - Ejection fraction
  - Exercise capacity
  - Heart rate recovery
  - Echocardiographic ischemia

Baseline Characteristics According to Aspirin Use (before matching)

Variable	**Aspirin* (n=2310**	**No Aspirin* (n=3864)**	**P value**
Age, years	62 (11)	56 (12)	< .001
Body mass index, kg/m^2	29 (5)	30 (7)	< .001
Ejection fraction, %	50 (9)	53 (7)	< .001
Resting heart rate, beats/min	74 (13)	79 (14)	< .001
Resting systolic BP, mm Hg	141 (21)	138 (20)	< .001
Resting diastolic BP, mm Hg	85 (11)	86 (11)	.04
Heart rate recovery, beats/min	28 (11)	30 (12)	< .001
Peak exercise cap, men (METS)	8.6 (2.4)	9.1 (2.6)	< .001
Peak exercise cap, women	6.6 (2.0)	7.3 (2.1)	< .001

* Cells contain mean (SD)

Baseline Characteristics By Aspirin Use (in %) (before matching)

Variable	**Aspirin (n=2310**	**No Aspirin (n=3864)**	**P value**
Men	77.0	56.1	< .001
Clinical history: diabetes	16.8	11.2	< .001
hypertension	53.0	40.6	< .001
prior coronary artery disease	69.7	20.1	< .001
congestive heart failure	5.5	4.6	.12
Medication use: Beta-blocker	35.1	14.2	< .001
ACE inhibitor	13.0	11.4	< .001

- Baseline characteristics appear very dissimilar: 25 of 31 covariates have p < .001, 28 of 31 have p < .05.
- Aspirin user covariates indicate higher mortality risk.

Propensity Score Matching in Observational Studies

- After propensity score analysis:
  - Aspirin use now associated with reduced mortality:
    Hazard ratio: 0.67 (0.51, 0.87)

Baseline Characteristics According to Aspirin Use (after matching)

Variable	**Aspirin* (n=2310**	**No Aspirin* (n=3864)**	**P value**
Age, years	60 (11)	61 (11)	.16
Body mass index, kg/m^2	29 (6)	29 (6)	.83
Ejection fraction, %	51 (8)	51 (9)	.65
Resting heart rate, beats/min	77 (13)	76 (14)	.13
Resting systolic BP, mm Hg	141 (21)	141 (21)	.68
Resting diastolic BP, mm Hg	85 (11)	86 (11)	.57
Heart rate recovery, beats/min	28 (12)	28 (11)	.82
Peak exercise cap, men (METS)	8.7 (2.5)	8.3 (2.5)	.01
Peak exercise cap, women	6.5 (2.0)	6.7 (2.0)	.13

* Cells contain mean (SD)

Baseline Characteristics By Aspirin Use (in %) (after matching)

Variable	**Aspirin (n=2310**	**No Aspirin (n=3864)**	**P value**
Men	70.4	72.1	.33
Clinical history: diabetes	15.0	15.3	.83
hypertension	50.3	51.7	.46
prior coronary artery disease	48.3	48.8	.79
congestive heart failure	5.8	6.6	.43
Medication use: Beta-blocker	26.1	26.5	.79
ACE inhibitor	15.5	15.8	.79

- Baseline characteristics similar in matched users and non-users.
- 30 of 31 covariates show NS difference between matched users and non-users. [Peak exercise capacity for men is p=.01]

Methods of Adjustment for Propensity Score Matching (How do we use them?)

The four primary methods of adjustment are:

1.) Stratification or Subclassification
2.) Matching
3.) Covariate/Regression adjustment
4.) Weighting

Once the propensity scores are estimated, these methods can be used to estimate the treatment effect after adjusting for differences between the treatment groups. Both stratification and matching are used to adjust for the covariate **before** calculating the treatment effect. In contrast, regression adjustment is used **while** determining the treatment effect. These methods allow us to estimate the treatment effects after adjusting for differences between the treatment and control groups but are regarded as impractical in situations when there are a large number of covariates or strata. In contrast, propensity scores provide a scalar summary of all the covariate information and there is no limit on the number of covariates for adjustment.

1.) Stratification or Subclassification

Method:
1.) In stratification, the estimated propensity score is used to stratify the subjects into homogenous subclasses, with similar propensity scores. Each subclass consists of relatively the same number of subjects.

2.) The treated and untreated subjects are then compared:

- One approach: The treatment effect is estimated within each stratum and then the treatment effects for all strata are combined to estimate the overall treatment effect.
- Another approach: Logistic regression, including propensity score strata as a covariate in the model.

3.) According to Cochran (1968), using five strata will eliminate more than 90% of the covariate bias. Once the subjects are divided into quintiles, the treatment effect is estimated within each stratum and then the treatment effects with each of the stratum are combined to estimate the overall treatment effect.

4.) An ANOVA (Analysis of Variance) model containing the quintile main effect, treatment main effect and the treatment*quintile interaction effect is used.

5.) If the *p*-value is less than α for the treatment main effect, it indicates that there is an imbalance between the treated and control subjects for that variable.

OR

1.) In some cases, the estimated propensity score is used to assign the subjects to fewer than five strata.

2.) Initially, two subclasses are formed using a median split of the propensity scores. A two group *t*-test is then performed to test for a difference between the treatment and control groups within each subclass on the propensity scores. If the difference is statistically significant, then each of the subclasses is split at the median into two more subclasses. The process is repeated until there are two or more control and treatment subjects within each of the newly formed subclasses and the *t*-statistic exceeds 2.5.

3.) For each of the newly formed subclasses, a test for equality of means for each covariate, each covariate squared, and the 2-way interaction of covariates is performed. If the *t*-statistic exceeds 2.0 in any of the newly formed subclasses, then it is included in the propensity score model.

4.) The entire process is repeated until most of the significant t-statistics are removed indicating that there is good balance between the treatment and control groups. (King, 2008)

<u>Limitations:</u> According to Cochran (1968), as the number of covariates increases, the number of strata or subgroups increases exponentially. This consequentially makes it more challenging to create strata that allow for comparison between treated and untreated subjects.

2.) Matching

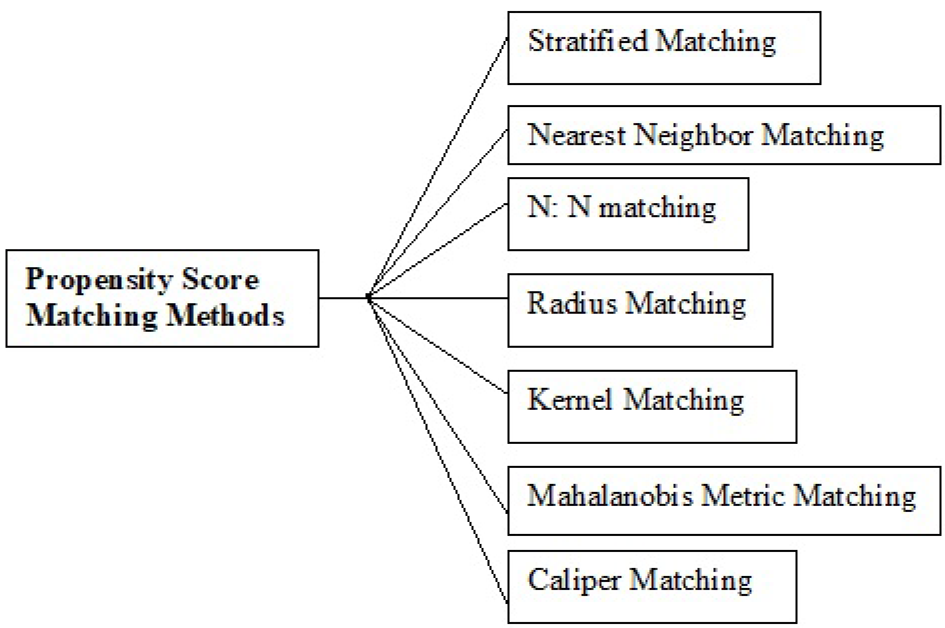
<u>Method</u>**:**
- Best used when there is a much larger number of control (untreated) subjects than treated subjects.
- The treatment and the control groups are matched on the estimated propensity score.
- Eliminates subjects who are not able to be matched.
- Comparison of groups in a matched analysis requires appropriate statistical tests for
  matched data.
- Automated matching programs are available (e.g. SAS)
- We can simultaneously control for all covariates by matching on a single scalar variable.

<u>Limitations:</u>
- The effects of the matching variables on the outcome cannot be studied.
- If there isn't sufficient overlap between the two groups on the matching variables, then biases such as regression effect toward the mean may occur.

- Assumes that all relevant covariates have been measured.
- The issue of whether matching on propensity score is applied with or without replacement is often disregarded. If matching with replacement is implemented, then there will be a greater number of matched pair sets (control subjects with treatment subjects). However, matching with replacement has its limitations in that a control subject may become a part of many matched pair sets.
- Inexact Matches
- Incomplete Matching:
  - Consider both matching and stratification or regression adjustment methods
  - Match using multivariate distance with calipers instead of matching just on propensity score
  - Match on logit of propensity score instead of on raw propensity scores

Propensity Score Matching Methods:

Once researchers obtain an estimated propensity score, an appropriate matching technique is implemented. Below are seven of the primary types of propensity score matching:



Many of the matching methods incorporate the caliper method to improve the quality of matching.

• Stratified Matching:
- The propensity scores are classified into intervals based on the range of values. Each interval consists of treatment and control subjects that on average, have equivalent propensity scores.
- The differences between the outcomes of the treatment and the control group are calculated to obtain the average treatment effect. It is an average of the outcomes of a treatment per block weighted by the distribution of treated subjects across the blocks.
- According to Cochran (1968), using five strata or grouping the sample into quintiles will eliminate more than 90- 95% of the covariate bias.

• Nearest Neighbor Matching:
- In this method, the absolute difference between the estimated propensity scores for the control and treatment groups is minimized.
- The control and treatment subjects are randomly ordered. Then the first treated subject is selected along with a control subject with a propensity score closest in value to it.

$$C(P_i) = \min_j |P_i - P_j|$$

where:

$C(P_i)$ represents the group of control subjects j matched to treated subjects i (on the estimated propensity score)

P_i is the estimated propensity score for the treated subjects i

P_j is the estimated propensity score for the control subjects j

• N: N Matching:
- In this method, control and treatment subjects are randomly ordered but the first **n** treatments are matched to **n** control subjects with the closest propensity score. The commonly used matches are 1:1, 1: N or N: 1 match.

• Radius Matching
- In this method, every treated subject is matched with a corresponding control subject that is within a predefined interval of the treatment subject's propensity score. Since each of the treatment subjects must be matched with a control subject with a given interval, only a certain number of comparisons will be available.

• Kernel Matching:
- In this method, every treated subject is matched with the weighted average of the control subjects. The weights are inversely proportional to the distance between the treated and control group's propensity scores.

• Mahalanobis Metric Matching:

- In this method, the subjects are ordered randomly and then the distance between the treated and control subjects is calculated. The distance is:

$$D_{ij} = \sqrt{(x_i - y_j)^T S^{-1} (x_i - y_j)}$$

where: S^{-1} is the sample covariance matrix of matching variables from the control subjects.
x_i and y_j are the matching variable values including the propensity score where i represents the treated subjects and j the control subjects

- The treatment and control subjects are matched based on the smallest Mahalanobis distance. The process is repeated until each treatment subject is matched and then the unmatched control subjects are removed.
- If a treated subject doesn't have a control subject with a similar propensity score, then reliable causal inferences cannot be made without the use of extrapolation. Therefore, such units are generally removed from the analysis.
- Mahalanobis matching after propensity score matching in observational studies is regarded as the equivalent of blocking in randomized experimental designs.

• <u>Caliper Matching:</u>
- In this method, a pre-determined range of values is defined, usually within one-quarter of the standard error ($0.25s$) of the estimated propensity. Any values that fall outside that range are removed (Sianesi, 2002).

The range is: $|P_i - P_j| < e$

where: P_i is the estimated propensity score for the treated subjects i
P_j is the estimated propensity score for the control subjects j
e is the pre-determined range of values

Comparing the Propensity Score Matching Methods:

There is no one method that has been deemed the most appropriate or effective although each method works more effectively when given certain circumstances.

- Matching with replacement is more effective when the control data set is small.
- 2 to 1 matching is more appropriate when dealing with a large control data set.
- Stratified matching is useful in situations when we infer that there are unobserved effects in matching and since stratification groups subjects with similar propensity scores together, then it is presumed that the unobserved effects disappear.
- Kernel, Mahalanobis and radius matching are more suitable when dealing with large, asymmetrically distributed control data sets (Baser, 2006).

The following table compares the bias and variance increases and decreases associated with using each of the propensity score matching methods (Baser, 2006):

Types of Propensity Score Matching	Bias	Variance
Nearest Neighbor (NN) - 2:1 Matching / 1:1 Matching	(+) / (-)	(-) / (+)
Nearest Neighbor (NN) - With / Without Caliper	(-) / (+)	(+) / (-)
Mahalanobis Matching (MM) - With / Without Caliper	(-) / (+)	(+) / (-)
Kernel Matching (KM) - Small / Large	(-) / (+)	(+) / (-)
Kernel Matching (KM) - NN Matching/ Radius Matching	(-) / (+)	(+) / (-)
Kernel Matching (KM) - KM Matching or MM Matching / NN Matching	(+) / (-)	(+) / (-)

SAS Macro for Propensity Score Matching:

```
/* Define the library for study data */
LIBNAME study
'C:\Projects\SUGI_29\DataSetX';
/* ************************************** */
/* Perform the Logistic Regression */
/* Calculate and save propensity score */
/* Propensity score name = PROB */
/* Output file = STUDY.AllPropen */
/* ************************************** */
PROC LOGISTIC DATA = study.contra descend;
MODEL treatment = covariate_1 covariate_2 covariate_3 ...
                    covariate_n;
/ SELECTION = STEPWISE RISKLIMITS;
   LACKFIT RSQUARE PARMLABEL;
 OUTPUT OUT=study.AllPropen prob=prob ;
RUN;


/* ****************************************** */
/* ****************************************** */
/* Matching Macro */
/* ****************************************** */
/* ****************************************** */
%MACRO OneToManyMTCH (
 Lib, /* Library Name */
 Dataset, /* Data set of all patients */
 depend, /* Dependent variable that indicates Case or Control */
 /* Code 1 for Cases, 0 for Controls */
 SiteN, /* Site/Hospital ID */
 PatientN, /* Patient ID */
 matches, /* Output data set of matched pairs */
 NoContrls); /* Number of controls to match to each case */


/* ********************* */
/* Macro to Create the Case and Control Data sets */
/* ********************* */
%MACRO INITCC(CaseAndCtrls,digits);
data tcases (drop=cprob)
tctrl (drop=aprob) ;
set &CaseAndCtrls. ;
/* Create the data set of Controls */
if &depend. = 0 and prob ne . then
```

```
do;
cprob = Round(prob,&digits.);
Cmatch = 0;
Length RandNum 8;
RandNum=ranuni(1234567);
Label RandNum='Uniform Randomization Score';
output tctrl;
end;
/* Create the data set of Cases */
else if &depend. = 1 and prob ne . then
do;
Cmatch = 0;
aprob =Round(prob,&digits.);
output tcases;
end;
run;
%SORTCC;
%MEND INITCC;
/* ********************* */
/* Macro to sort the Cases and Controls data set */
/* ********************* */
%MACRO SORTCC;
proc sort data=tcases out=&LIB..Scase;
by prob;
run;
proc sort data=tctrl out=&LIB..Scontrol;
by prob randnum;
run;
%MEND SORTCC;


/* ********************* */
/* Macro to Perform the Match */
/* ********************* */
%MACRO MATCH (MATCHED,DIGITS);
data &lib..&matched. (drop=Cmatch randnum aprob cprob start oldi curctrl
matched);
/* select the cases data set */
set &lib..SCase ;
curob + 1; Posters
matchto = curob;
if curob = 1 then do;
start = 1;
oldi = 1;
end;
/* select the controls data set */
```

```
DO i = start to n;
set &lib..Scontrol point = i nobs = n;
if i gt n then goto startovr;
if _Error_ = 1 then abort;
curctrl = i;
/* output control if match found */
if aprob = cprob then
do;
Cmatch = 1;
output &lib..&matched.;
matched = curctrl;
goto found;
end;
/* exit do loop if out of potential matches */
else if cprob gt aprob then
goto nextcase;
startovr: if i gt n then
goto nextcase;
END; /* end of DO LOOP */
/* If no match was found, put pointer back*/
nextcase:
if Cmatch=0 then start = oldi;
/* If a match was found, output case and increment pointer */
found:
if Cmatch = 1 then do;
oldi = matched + 1;
start = matched + 1;
set &lib..SCase point = curob;
output &lib..&matched.;
end;
retain oldi start;
if _Error_=1 then _Error_=0;
run;
/* get files of unmatched cases and controls */
proc sort data=&lib..scase out=sumcase;
by &SiteN. &PatientN.;
run;
proc sort data=&lib..scontrol out=sumcontrol;
by &SiteN. &PatientN.;
run;
proc sort data=&lib..&matched. out=smatched (keep=&SiteN. &PatientN.
matchto);
by &SiteN. &PatientN.;
run;
data tcases (drop=matchto);
merge sumcase(in=a) smatched;
```

```
by &SiteN. &PatientN.;
if a and matchto = . ;
cmatch = 0;
aprob =Round(prob,&digits.);
run;
data tctrl (drop=matchto);
merge sumcontrol(in=a) smatched;
by &SiteN. &PatientN.;
if a and matchto = . ;
cmatch = 0;
cprob = Round(prob,&digits.);
run;
%SORTCC
%MEND MATCH;


/* ********************* */
/* Macro to call Macro MATCH for each of the 8-digit to 1-digit matches */
/* ********************* */
%MACRO CallMATCH;
/* Do a 8-digit match */
%MATCH(Match8,.0000001);
/* Do a 7-digit match on remaining unmatched*/
%MATCH(Match7,.000001);
/* Do a 6-digit match on remaining unmatched*/
%MATCH(Match6,.00001);
/* Do a 5-digit match on remaining unmatched*/
%MATCH(Match5,.0001);
/* Do a 4-digit match on remaining unmatched */
%MATCH(Match4,.001);
/* Do a 3-digit match on remaining unmatched */
%MATCH(Match3,.01);
/* Do a 2-digit match on remaining unmatched */
%MATCH(Match2,.1);
/* Do a 1-digit match on remaining unmatched */
%MATCH(Match1,.1);
%MEND CallMATCH;


/* ********************* */
/* Macro to Merge all the matched files into one file */
/* ********************* */
%MACRO MergeFiles(MatchNo);
data &matches.&MatchNo. (drop = matchto);
set &lib..match8(in=a) &lib..match7(in=b) &lib..match6(in=c)
&lib..match5(in=d)
```

```
&lib..match4(in=e)
&lib..match3(in=f) &lib..match2(in=g) &lib..match1(in=h);
if a then match_&MatchNo. = matchto;
if b then match_&MatchNo. = matchto + 10000;
if c then match_&MatchNo. = matchto + 100000;
if d then match_&MatchNo. = matchto + 1000000;
if e then match_&MatchNo. = matchto + 10000000;
if f then match_&MatchNo. = matchto + 100000000;
if g then match_&MatchNo. = matchto + 1000000000;
if h then match_&MatchNo. = matchto + 10000000000;
run;
%MEND MergeFiles;


/* ******************************** */
/* ******************************** */
/* Perform the initial 1:1 Match */
/* ******************************** */
/* ******************************** */
/* Create file of cases and controls */
%INITCC(&LIB..&dataset.,.00000001);
/* Perform the 8-digit to 1-digit matches */
%CallMATCH;
/* Merge all the matches files into one file */
%MergeFiles(1)


/* ********************************** */
/* ********************************** */
/* Perform the remaining 1:N Matches */
/* ********************************** */
/* ********************************** */
%IF &NoContrls. gt 1 %Then %DO;
%DO i = 2 %TO &NoContrls.;
%let Lasti=%eval(&i. - 1);


/* ********** */
/* Start with Cases from the last Matched Cases file and the remaining Un-
Matched */
/* Controls. NOTE: The Unmatched Controls file (Scontrol) is created at end
of the */
/* previous match */

/* Select the Matched Cases from the last Matched File */
data &LIB..Scase;
```

```
  set &matches.&Lasti.;
  where &Depend. = 1;
run;


/* ********** */
/* Perform the 8-1 digit matches between Matched Cases and the
Unmatched Controls */
%CallMATCH;

/* ********** */
/* Merge the 8-digit to 1-digit matches files into one file */
%MergeFiles(&i.)
%DO m = 1 %TO &Lasti.;
data &matches.&i.;
  set &matches.&i.;
  if &Depend.=0 then Match_&m. = .;
run;
%END;
/* ********** */
/* Determine which OLD Controls correspond to the kept Cases */
%DO c = 1 %TO &Lasti.;
   /* Select the KEPT Cases */
   proc sort data=&matches.&i. out=skeepcases (keep = Match_&c.);
         by Match_&c.;
         where &Depend. = 1;
   run;
   /* Get the OLD Controls */
   proc sort data = &matches.&Lasti. out = soldcontrols&c.;
         by Match_&c.;
         where &Depend. = 0 and Match_&c. ne . ;
   run;
   /* Get the OLD Controls that correspond to the kept Cases */
   data keepcontrols&c.;
     merge skeepcases (in = a) soldcontrols&c. (in = b);
     by Match_&c.;
     if a;
run;
%END;

/* ********** */
/* Combine all the OLD Controls into one file */
data keepcontrols;
set keepcontrols1 (obs=0);
run;
%DO k = 1 %TO &Lasti.;
```

```
  data keepcontrols;
   set keepcontrols keepcontrols&k.;
run;
%END;

/* ********** */
/* Append the OLD matched Controls to the new file of matched cases and
controls */
data &matches.&i.;
 set &matches.&i. keepcontrols;
run;

/* ********** */
/* If there are more matches to be made, add the previously matched, but
not kept, */
/* controls back into the pool of unmatched controls */
%if &i. lt &NoContrls. %then %do;
   %DO z = 1 %TO &Lasti.;

/* Select all the KEPT Cases */
proc sort data=&matches.&i. out=skeepcases (keep = Match_&z.);
 by Match_&z.;
 where &Depend. = 1;
run;

/* Select all the OLD Controls */
proc sort data = &matches.&Lasti. out = soldcontrols&z.;
 by Match_&z.;
 where &Depend. = 0 and Match_&z. ne .;
run;

/* Keep the OLD Controls that correspond to the NOT KEPT Cases */
/* Drop the previuos Match_X variable */
data AddBackControls&z. (drop = Match_&z.);
 merge skeepcases (in = a) soldcontrols&z. (in = b);
 by Match_&z.;
 if b and not a;
run;
%END; /* End DO */

/* Drop the previuos Match_X variable */
data &LIB..Scontrol (drop = Match_&lasti. );
 set &LIB..Scontrol;
run;

/* Append */
```

```
 %DO y = 1 %TO &Lasti.;
   data &LIB..Scontrol;
        set &LIB..Scontrol AddBackControls&y.;
   run;
      %END; /* End DO */
       %end; /* End IF */
  %END; /* End Main DO */
%END; /* End Main IF */


/* ************************************** */
/* ************************************** */
/* Save the final matched pairs data set */
/* ************************************** */
/* ************************************** */
/* Sort file by Treatment Variable */

proc sort data=&matches.&NoContrls. out = &lib..&matches.;
by &depend.;
run;
%MEND OneToManyMTCH;
```

3.) Regression/Covariate Adjustment

Method:

- In order to determine whether regression adjustment is an appropriate method, there must be a substantial overlap between the treated and control groups. Additionally, the difference between the means of the propensity scores, the ratio of the variances, and the ratio of the covariate’s residuals between the two treatment groups are calculated. The difference between the means of the propensity scores must be relatively small and the ratios must be close to one.
- The propensity score is included as a covariate in a regression model, in addition to the treatment variable, to adjust for the estimate of the treatment effect. There may be additional covariates included in the model. Both treatment and the propensity scores are regarded as independent variables in the analysis. The estimated treatment effect is,

$$\hat{t} = (\overline{Y_t} - \overline{Y_c}) - b(\overline{X_t} - \overline{X_c})$$

- This method uses the actual propensity score whereas the other two methods use the estimated propensity score and match or stratify based on a similarity in propensity score values.

Limitations:

- It requires an adequate amount of overlap between the treatment and control groups. If there is a substantial difference between the covariate distributions then regression adjustment is not very effective. This is because the covariance would adjust the results to apply to the mean value of the dependent variable which would not reflect on the individual values of each group’s dependent variable, if they are substantially different. If any of the following 3 conditions are not satisfied, then covariance adjustment will be regarded as unreliable because of a lack of overlap between the treatment and control groups:
    1. There should be a small difference between the mean propensity scores for the treatment groups ($0.5s$) unless the covariate distributions are approximately symmetric and have the same variance and the sample sizes are about the same.
    2. The ratio of the variances of the propensity score between the two groups should be approximately equal to one. If the ratio is not close to one, then the bias may be inaccurately corrected for.
    3. After adjusting for the propensity score, the ratio of the variances of the covariate’s residuals between the two groups must be approximately equal to one.

- It can't be used to determine whether the model was effectively adjusted for differences between the groups.
- When studying rare occurrences, a restricted number of covariates are available.

Why Not Do Regression Adjustment with All Covariates Instead?

- Both methods should lead to the same conclusions (Rubin, 1979)
- Advantage of a two-step process:
  - Can fit a more complicated propensity score model with interactions and polynomial terms.
  - Goal is to obtain the best estimated probability of treatment assignment; therefore one is not concerned with over-parameterizing the model
  - Can fit a simpler model when propensity score is include.

4.) Weighting

Method:

- Weighting on the propensity score is not implemented as commonly as the other methods of adjustment.
- In propensity score weighting, the treated and control observations are re-weighted in order to make them more representative of the population.
- The weight of a treated subject is defined as the inverse of its propensity score:

$$w_i = \frac{1}{\hat{e}(\mathbf{x}_i)}$$

- The weight of a control subject is defined as the inverse of one minus its propensity score:

$$w_i = \frac{1}{1 - \hat{e}(\mathbf{x}_i)}$$

Limitations:

- If the estimated propensity scores are close to zero or one, then weighting often produces unrealistic weights for the control and treatment subjects.

What is the Best Method of Adjustment?

According to Rosenbaum and Rubin (1983), the propensity score can be used in observational studies to reduce bias through the methods of adjustment. Each method comes with its strengths and limitations so there is no general consensus on which one is the most effective or preferable.

Of the three methods of adjustment, propensity score matching has been considered the *most statistically efficient* method of integrating propensity scores. Stratification and matching on the estimated propensity score are both successful at achieving balance in the covariates between the control and treatment subjects. However, matching has been proven to be more effective in reducing the imbalance between treated and untreated subjects as well as in reducing treatment-selection bias than stratification (Austin, 2007). Moreover, since the covariate distributions of the treatment and control groups become closely matched when matched on the propensity score, they will be more similar than if from a random sample. Therefore, the variance of the estimated treatment effect will be lower for the matched pairs than the variance for subjects obtained from a random sample (Rosenbaum & Rubin, 1983). Implementation of this adjustment method does however have limitations in that it requires a large number of control variables, and the unmatched subjects are discarded from the analysis (Newgard, 2004).

The most commonly used adjustment method in clinical literature is covariate/regression adjustment on the propensity score (Austin & Mamdani, 2005). It is not as precise in reducing bias and it should only be implemented if certain conditions are satisfied. Generally, if there is no substantial overlap between the covariate distributions of the treatment and control group, then regression adjustment is not very effective in adjusting for differences. Rubin (1979) showed that its implementation under insufficient conditions may increase the expected square bias if the covariate matrices in the treated and control groups are unequal or if the variances between the two groups largely vary.

Comparing PSM with Hard Matching

- PSM is more suitable when dealing with a large number of covariates whereas hard matching is more appropriate when dealing with a small number of covariates.
- Both methods control for observed covariates and do not account for bias resulting from the unobserved covariates that may affect whether a subject receives treatment or not.
- PSM and matching both produce similar results when matching on a small number of covariates.

Propensity Score Matching in Observational Studies

Limitations of PSM

- This method requires **large samples**
- Since the propensity scores are obtained from observational data, there is **no randomization**. Therefore, the matching will only control for the differences on the observed variables and there may be some bias resulting from the unobserved covariates that could affect whether subjects receive treatment or not. To elaborate, if only conveniently available covariates such as age and gender are used, and other relevant covariates aren't accounted for, then bias may occur.
- In order to be effective in providing strong support of casual inference, there must be **substantial overlap** between the groups on the propensity scores. This method will not be useful if subjects with a high propensity score were treated and those with a low propensity score were untreated.
- There is no gold standard with respect to which variables should be included in the propensity score model. Sometimes researchers include variables that predict the treatment assignment, others include only the variables associated with both the treatment and outcome and others include any variables that could be potentially related to the outcome.

Characteristics of a Good PSM

- Matching is based on variables that can be accurately and reliably measured.
- Substantial overlap between the groups on the propensity scores.
- Model adequately balances covariates of the treated and untreated subjects.
- It adjusts for selection bias and minimizes group differences across many variables.
- It does not use only conveniently available covariates such as age and gender.
- Sensitivity analysis is a recommended part of the process:
- Choosing variables and adjusting for propensity scores is based on:
  - Logic
  - Theory
  - Empirical Evidence

References

American Psychological Association. (2001). Publication manual of the American Psychological Association (5th ed.). Washington, DC: Author.

Austin, P. C., & Mamdani, M. M. (2006). A comparison of propensity score methods: A case-study estimating the effectiveness of post-AMI statin use. Statistics in Medicine, 25(12), 2084-2106.

Austin, P. C. (2007). A critical appraisal of propensity-score matching in the medical literature between 1996 and 2003. Statistics in Medicine. doi: 10.1002/sim.3150.

Baser, O. (2006). Too Much Ado about Propensity Score Models? Comparing Methods of Propensity Score Matching. Value in Health: The Journal of the International Society for Pharmacoeconomics and Outcomes Research, 9(6), 377-385.

Bootzin, R. R., & McKnight, P. E. (2006). Strengthening Research Methodology: Psychological Measurement and Evaluation (1st ed.). Washington, DC: American Psychological Association.

D'Agostino, R. B.,Jr. (1998). Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. Statistics in Medicine, 17(19), 2265-2281.

Little, R. J., & Rubin, D. B. (2000). Causal effects in clinical and epidemiological studies via potential outcomes: Concepts and analytical approaches. Annual Review of Public Health, 21, 121-145.

Love, T.E. (2008). Reducing the Impact of Selection Bias with Propensity Scores. 7th International Conference on Health Policy Statistics [ICHPS], 18 January 2008. Cleveland, Ohio, USA: Center for Health Care Research and Policy, Case Western University at MetroHealth Medical Center.

Newgard, C. D., Hedges, J. R., Arthur, M., & Mullins, R. J. (2004). Advanced Statistics: The Propensity Score- A Method for Estimating Treatment Effect in Observational Research. Academic Emergency Medicine: official journal of the Society for Academic Emergency Medicine, 11(9), 953-961.

Oakes, J. M., & Kaufman, J. S. (2006). Methods in Social Epidemiology (1st ed.). San Francisco, CA: Jossey-Bass.

Parsons, L. S. (2004). Performing a 1:N Case- Control Match on Propensity Score. 29th annual SAS Users Group International. Retrieved from http://www2.sas.com/proceedings/sugi29/165-29.pdf

Rosenbaum, P.R. & Rubin, D.B. (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. Biometrika, 70(1), 41-55.

Rosnow, R. L., & Hantula, D. A. (2006). Advances in Social & Organizational Psychology: A Tribute to Ralph Rosnow. Mahwah, N.J.: Lawrence Erlbaum Associates.

Rubin, D.B. (1979). Using multivariate matched sampling and regression adjustment to control bias in observational studies. Journal of the American Statistical Association, 74, 318-324.

Rubin, D.B. (2001). Using Propensity Scores to Help Design Observational Studies: Application to the Tobacco Litigation. Health Services & Outcomes Research Methodology, 2, 169-188.

Sianesi, B. (2001). Implementing Propensity Score Matching Estimators with STATA [PowerPoint]. London, England: University College London and Institute for Fiscal Studies.