Propensity Score Matching

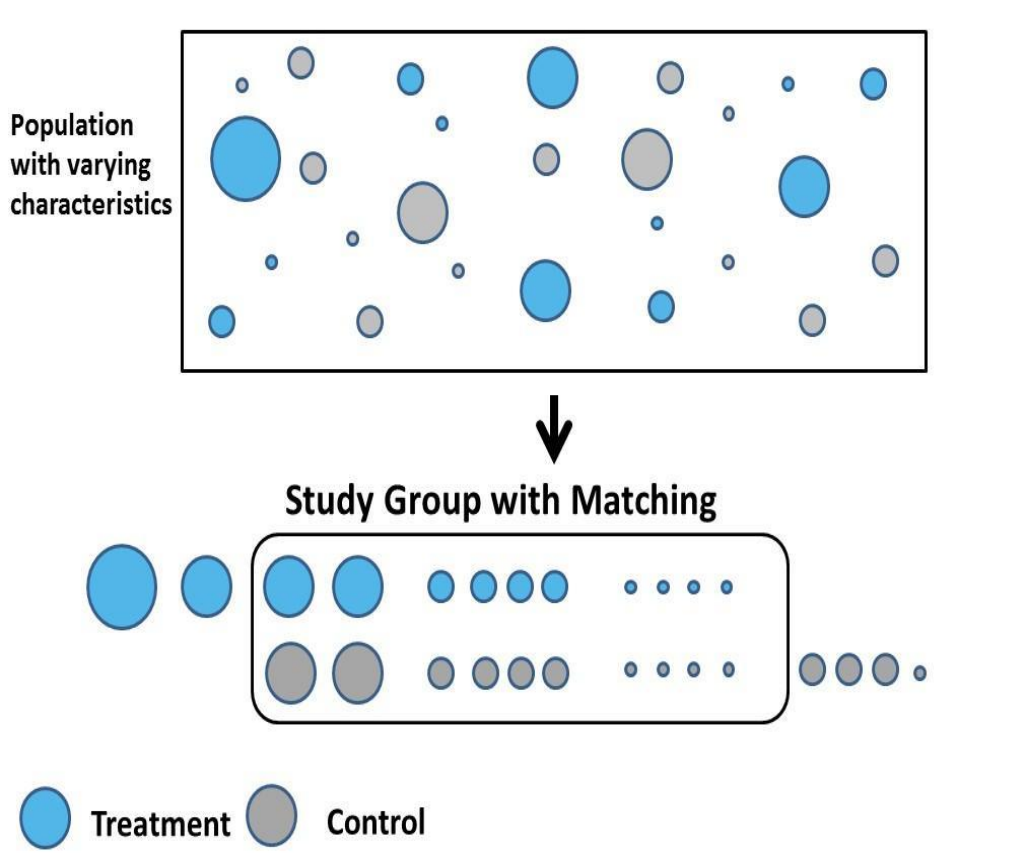
Propensity Scores

- Propensity Scores are first established in the seminal paper by Rosenbaum and Rubin (1983)
- The propensity score is defined as the probability of receiving the treatment of interest (vs the control treatment) conditional on measured participant covariates.
- Propensity scores can be used in observational studies to reduce bias and confounding when estimating treatment effects

Propensity-Score Matching (PSM)

- Propensity score matching: match treated and untreated observations on the estimated probability of being treated (propensity score).
- Most commonly used.
- Key assumption: participation is independent of outcomes conditional on matching variable
- Enables matching not just at the mean but balances the distribution of observed characteristics across treatment and control

Propensity Scores



https://stats.stackexchange.com/questions/553853/understanding-propensity-score-matching

Propensity Scores

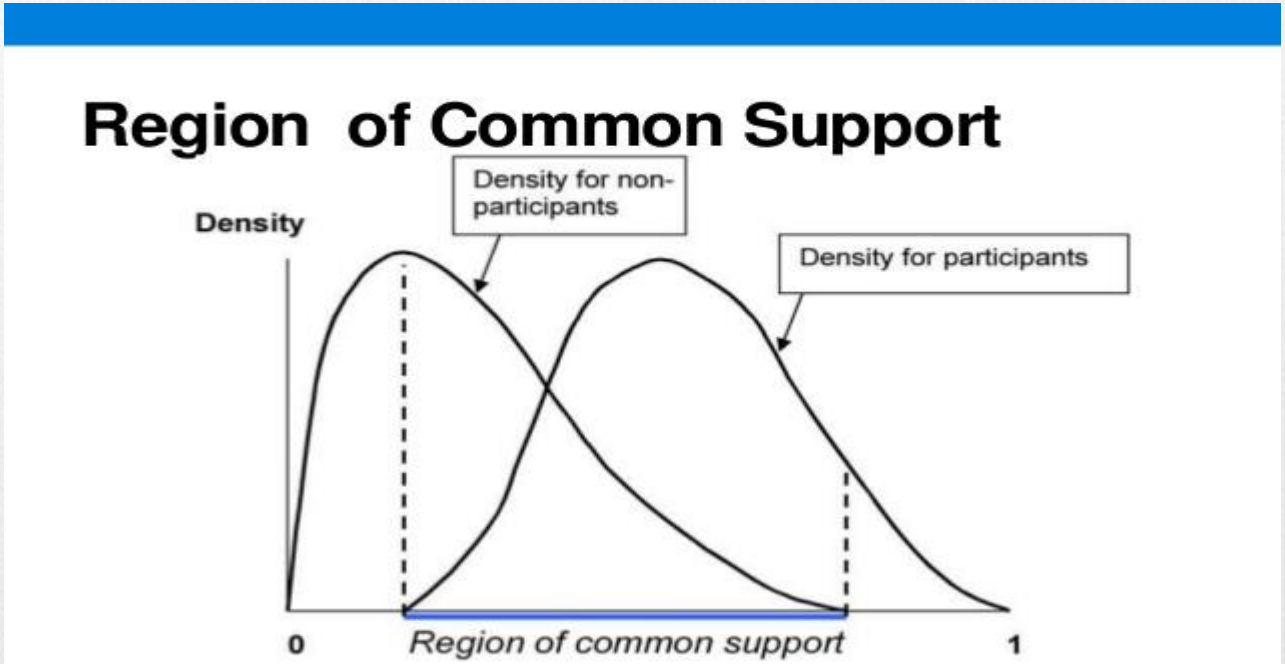
- Propensity score is a balancing score
  - ~~a subgroup of participants, all of whom have the same value~~ of the propensity score, the distribution of measured baseline covariates will be the same in treated and control participants in that subgroup. Thus, we can remove the effects of confounding by comparing outcomes between treated and control participants who share a similar value of the propensity score.
  - This balancing is analogous to that induced by randomization in RCTs, with the key difference being that conditioning on the propensity score balances measured covariates, whereas randomization ideally balances both measured and unmeasured covariates.

Steps in PS Matching

- Need representative and comparable data for both treatment (case) and comparison (control) groups.
  - Choose variables
- Estimation of the propensity score:
  - The propensity scores are usually unknown in observational studies, so it has to be estimated.
  - Usually, propensity scores are estimated by logistic regression models given the nature of the data.
  - Use predicted values from logit to generate propensity score p(xi) for all treatment and comparison group members

Steps in PS Matching

- Match Pairs:
  - Restrict sample to common support



Region of common support for propensity score between participants and non-participants must be **large** enough to find an adequate comparison group

Steps in PS Matching

- Match Pairs:
  - Restrict sample to common support (next figure)
  - Need to determine a tolerance limit: how different can control individuals or villages be and still be a match?
  - Nearest neighbors, nonlinear matching, multiple matches
- Once matches are made, we can conduct the analysis for the interested outcome

•**income**: Continuous variable (e.g., participant's income).

PS Matching Example

- A pharmaceutical company is conducting a **nonrandomized** clinical trial to demonstrate the efficacy of a new treatment **(Drug_X)** by comparing it to an existing treatment **(Drug_A).**
- Patients in the trial can choose the treatment that they prefer; otherwise, physicians assign each patient to a treatment.
- The data set Drugs contains:
  - PatientID: the patient identification number
  - Drug: the treatment group indicator (drug_X (1), drug_A (0))
  - Baseline variable measurements for individuals from both treated and control groups.
    - age: participant's age
    - gender: 1 Male, 0 Female
    - BMI
- Outcome: Blood Marker X

PS Matching Example

Drug	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	373	76.75	373	76.75
1	113	23.25	486	100.00

- We have 113 subjects treated with drug X (1), and want to identified same number of subjects in controls (treated with drug A, 0) with similar baseline character

PS Matching Example

- Pre-treatment (baseline) covariates distribution

Table of Drug by Gender

Drug	Gender 0	Gender 1	Total
0	171 45.84	202 54.16	373
1	49 43.36	64 56.64	113
Total	220	266	486

Drug=0

Variable	N	Mean	Std Dev	Minimum	Maximum
Age	373	40.4048257	6.5791033	25.0000000	57.0000000
BMI	373	23.7532708	1.9807783	19.2200000	28.6100000

Drug=1

Variable	N	Mean	Std Dev	Minimum	Maximum
Age	113	36.3097345	5.5341144	26.0000000	49.0000000
BMI	113	24.4925664	1.8637974	20.3300000	28.3400000

Propensity Score Calculation

```
proc logistic data=drugs descending;
     class gender (ref="0")/ param=ref;
     model drug(event='1') = age gender bmi;
     output out=propensity_scores p=propensity;
run;
```

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-2.7230	1.5392	3.1296	0.0769
Age		1	-0.1122	0.0191	34.7050	<.0001
Gender	1	1	0.2065	0.2300	0.8057	0.3694
BMI		1	0.2369	0.0605	15.3435	<.0001

Propensity Score Calculation

$$\ln\left(\frac{p(D=1)}{1-p(D=1)}\right) = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$$

$$p(D=1) = \frac{1}{e^{-(\alpha+\sum \beta X)} + 1}$$

$\hat{\alpha}$ = -2.7230, $\widehat{\beta_1}$ = -0.1122 , $\widehat{\beta_2}$ =0.2065 , $\widehat{\beta_3}$= 0.2369

What is the probability for drug=1 for a person with:

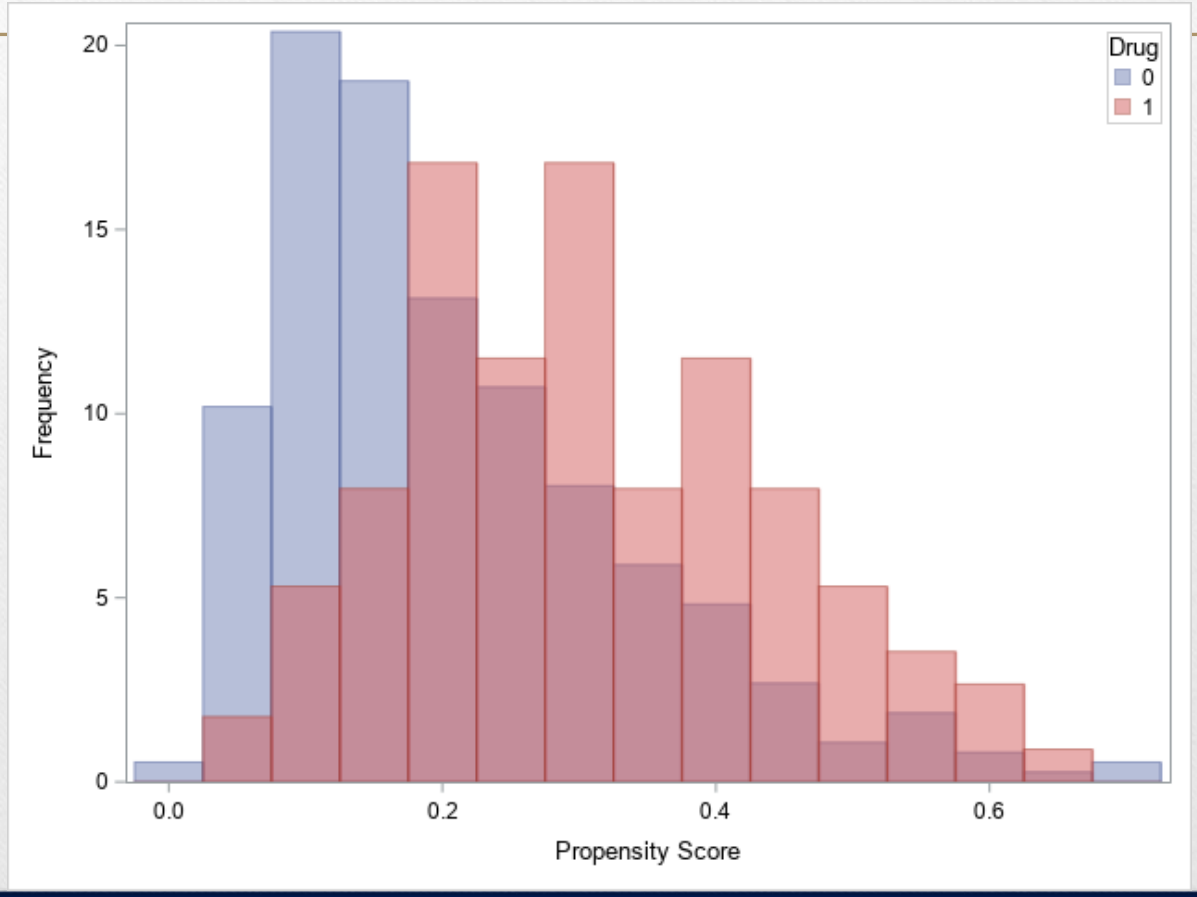
Age=36, Gender=0, BMI=25.53

$$\hat{p}(D=1) = \frac{1}{e^{-(-2.7230-0.1122\times 36+0.2369\times 25.53)}+1} = 0.328$$

Propensity Score Calculation

	Age	BMI	PatientID	Gender	Drug	Outcome	Response Value	Estimated Probability
1	36	25.53	1	0	0	107.928941	1	0.3280471281
2	45	22.48	4	0	0	116.6277167	1	0.0794621924
3	40	26.23	6	1	0	112.6523944	1	0.3113645874
4	39	24.01	7	1	0	108.4473984	1	0.2301761442
5	46	24.07	8	1	0	119.8637058	1	0.1214418787
6	34	24.29	9	0	0	106.3382589	1	0.3129779217
7	36	24.17	12	0	0	95.98429707	1	0.2613133163
8	43	24.35	13	0	0	116.4610955	1	0.144024228
9	40	21.52	14	1	0	104.6277972	1	0.129055637
10	39	23.27	15	0	0	106.6296312	1	0.1695148881
11	33	20.2	16	0	0	90.77976464	1	0.1620950125
12	27	23.71	18	0	0	91.84313858	1	0.4655872505
13	37	22.55	19	1	0	102.5080925	1	0.2093864299
14	39	22.36	20	0	0	97.36996579	1	0.1412911471
15	39	28.29	21	0	0	112.1322116	1	0.40129861

```
proc sgplot data=propensity_scores;
    histogram propensity / group=Drug transparency=0.5
binwidth=0.05;
    xaxis label="Propensity Score";
    yaxis label="Frequency";
    keylegend / position=topright location=inside across=1;
run;
```



PS Matching

Nearest-Neighbor Matching with a Caliper

- Separate the data for treatment and control
- It goes through each participant in the treated dataset and finds the nearest participant in the control dataset with a propensity score within a specified caliper range.
- Easy to understand and implement; Offers good results in practice; fast running time;

PS Matching

Optimal matching

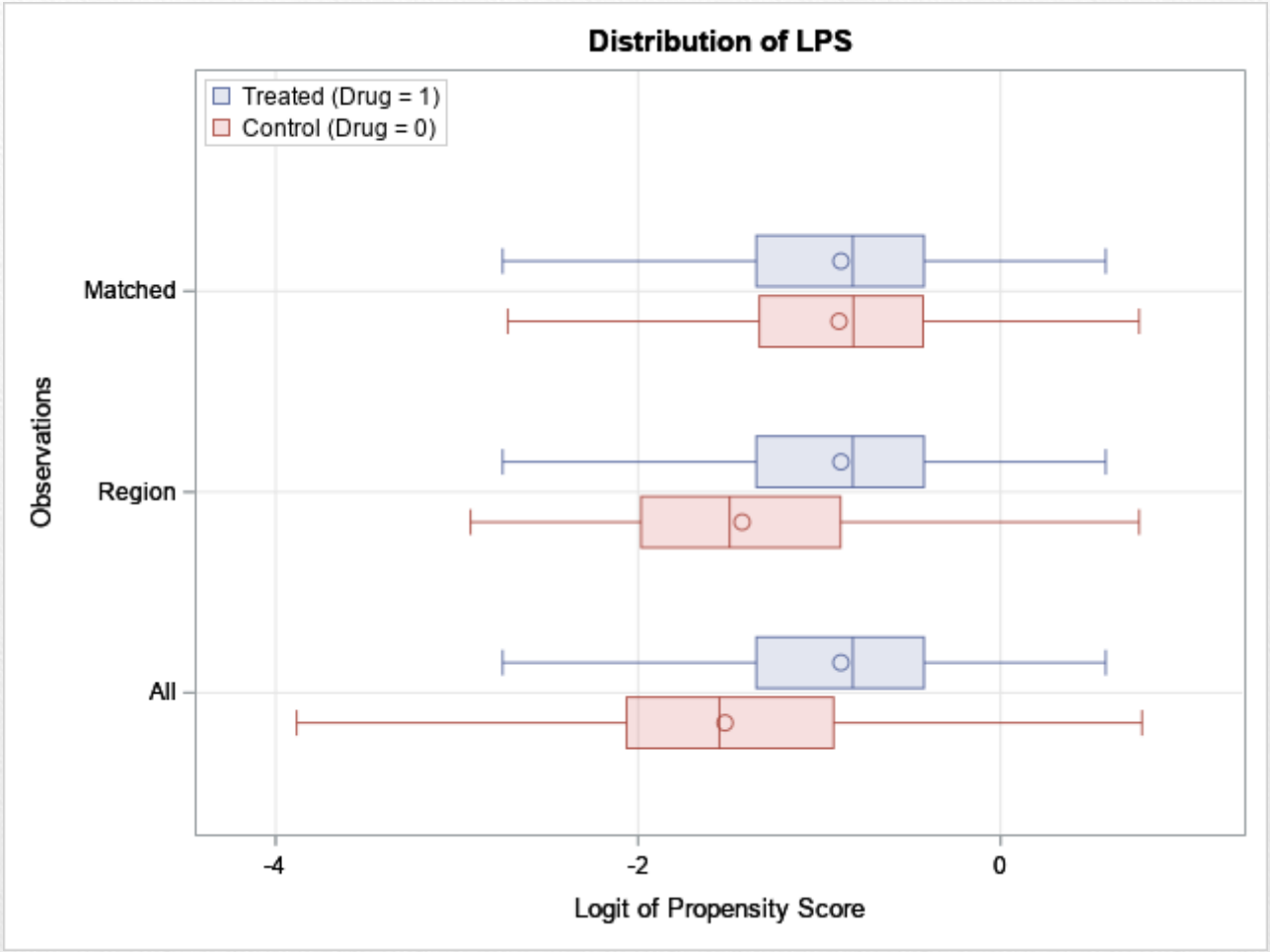
- To minimize the total distance for the overall population
- Offers the "best" matching results overall; Runs reasonably fast; Implementation is not easy; Not readily to extend to n-cube matching (n>2)

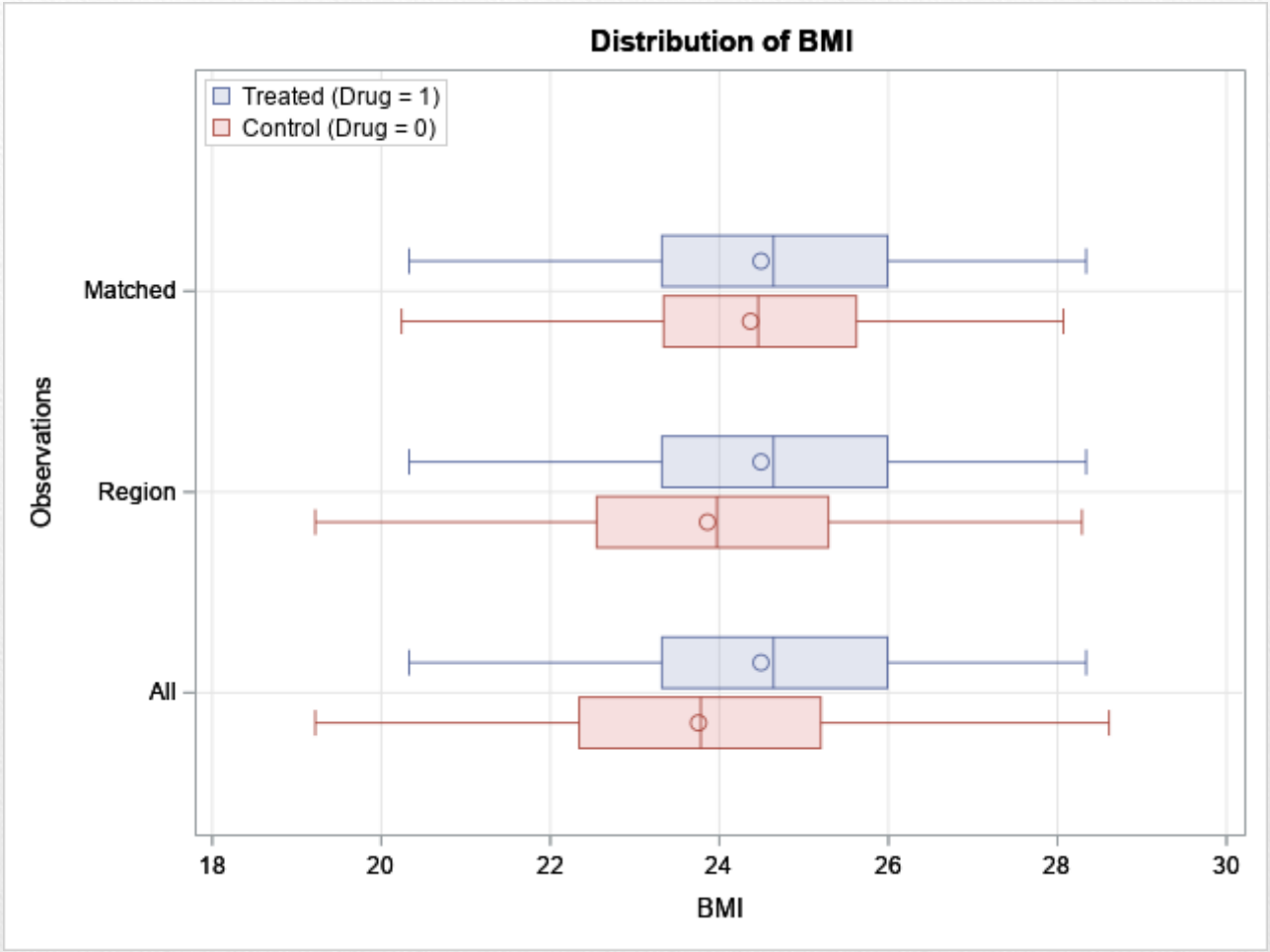
PS Matching

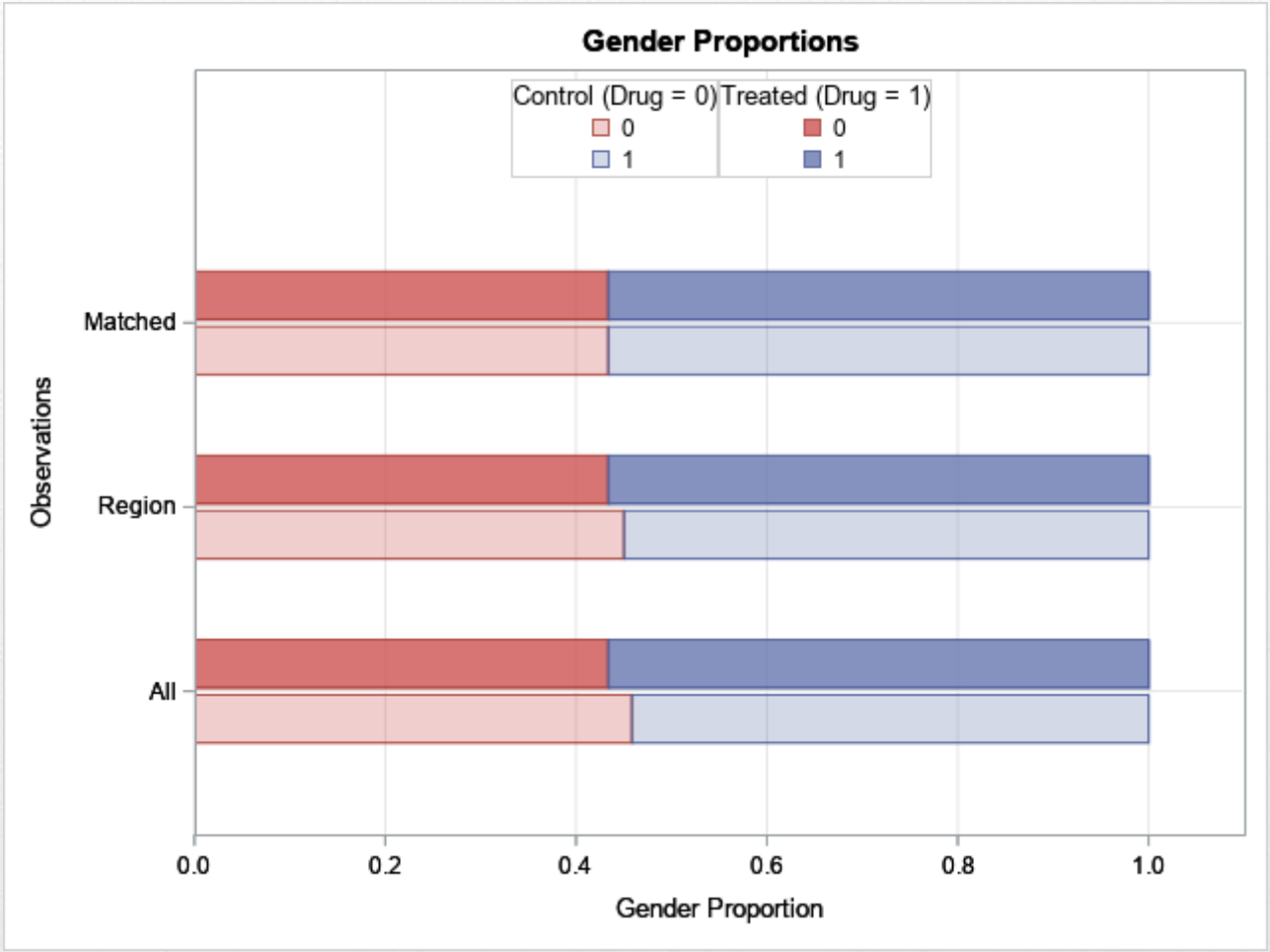
```
proc psmatch data=drugs region=cs;
   class Drug Gender;
   psmodel Drug(Treated='1')= Gender Age BMI;
   match method=optimal(k=1) exact=Gender
distance=lps caliper=0.25;
   assess lps allcov / weight=none plots=(barchart
boxplot);
   output out(obs=match)=Outgs lps=_Lps
matchid=_MatchID;
run;
```

Propensity Score Information											
	Treated (Drug = 1)					Control (Drug = 0)					Treated - Control
Observations	N	Mean	Standard Deviation	Minimum	Maximum	N	Mean	Standard Deviation	Minimum	Maximum	Mean Difference
All	113	0.3108	0.1325	0.0602	0.6411	373	0.2088	0.1320	0.0202	0.6858	0.1020
Region	113	0.3108	0.1325	0.0602	0.6411	351	0.2176	0.1267	0.0510	0.6824	0.0932
Matched	113	0.3108	0.1325	0.0602	0.6411	113	0.3082	0.1310	0.0619	0.6824	0.0025

Matching Information	
Distance Metric	Logit of Propensity Score
Method	Optimal Fixed Ratio Matching
Control/Treated Ratio	1
Caliper (Logit PS)	0.191862
Matched Sets	113
Matched Obs (Treated)	113
Matched Obs (Control)	113
Total Absolute Difference	2.941871


Distribution of LPS
Treated (Drug = 1)
Control (Drug = 0)
Observations
Matched
Region
All
-4
-2
0
Logit of Propensity Score


Distribution of BMI
Treated (Drug = 1)
Control (Drug = 0)
Observations
Matched
Region
All
18
20
22
24
26
28
30
BMI


Gender Proportions
Control (Drug = 0)
0
1
Treated (Drug = 1)
0
1
Observations
Matched
Region
All
0.0
0.2
0.4
0.6
0.8
1.0
Gender Proportion

Obs	Age	BMI	PatientID	Gender	Drug	Outcome	_PS_	_Lps	_MATCHWGT_	_MatchID
1	49	23.24	213	0	0	117.0096641	0.06187	-2.71892	1	1
2	44	20.75	89	0	1	107.2506957	0.06023	-2.74745	1	1
3	43	20.55	141	0	0	103.2190828	0.06401	-2.68256	1	2
4	46	22.22	323	0	1	121.5133179	0.06763	-2.62375	1	2
5	45	22.08	420	1	0	110.9529866	0.08801	-2.33814	1	3
6	49	23.96	217	1	1	134.3381789	0.08772	-2.34185	1	3
7	40	20.57	290	0	0	94.74375008	0.08778	-2.34104	1	4
8	41	21.11	234	0	1	105.8242763	0.08904	-2.32538	1	4
9	45	23.76	473	0	0	111.6422494	0.10464	-2.14670	1	5
10	46	24.17	320	0	1	122.6538071	0.10323	-2.16184	1	5

Obs	Age	BMI	PatientID	Gender	Drug	Outcome	_PS_	_Lps	_MATCHWGT_	_MatchID
1	36	25.53	1	0	0	107.928941	0.32804	-0.71706	1	69
189	35	25.03	288	0	1	112.5413859	0.32668	-0.72324	1	69

Analysis the outcome after matching

```
proc ttest data=outgs1;
class drug;
var outcome;
run;
```

Variable: Outcome

Drug	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
0		113	103.6	10.2924	0.9682	76.9237	130.9
1		113	109.9	11.3743	1.0700	83.5712	134.4
Diff (1-2)	Pooled		-6.3514	10.8468	1.4430		
Diff (1-2)	Satterthwaite		-6.3514		1.4430		

Drug	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
0		103.6	101.7	105.5	10.2924	9.1030	11.8422
1		109.9	107.8	112.1	11.3743	10.0599	13.0870
Diff (1-2)	Pooled	-6.3514	-9.1951	-3.5078	10.8468	9.9287	11.9536
Diff (1-2)	Satterthwaite	-6.3514	-9.1953	-3.5076			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	224	-4.40	<.0001
Satterthwaite	Unequal	221.8	-4.40	<.0001

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	112	112	1.22	0.2916

Practical issues

- Matching vs. Covariance adjustment modeling

Matching: always reduce the bias; no worry about the true regression equation; easy post-matching analysis; restricted to common support

Covariance adjustment modeling: has to guess the true regression equation (prone to bias); apply to the full range of the data; may lead to smaller variance estimation

Limitations and new advances

- Limitation of PS method

  - Rely on a unverifiable assumption: strongly ignorable treatment assignment given the observed covariates

  Unlike the randomized studies, it has no control over the unobserved confounders

  One possible solution is to use sensitivity analysis to evaluate to what degree the results will change given a hypothesized unknown covariate

Limitations and new advances

- Need substantial overlap between the treated and the control groups, otherwise, it may result in significant loss of the data in analysis

One possible solution is to use regression-like technique to extrapolate; however, such extrapolation might not be reliable

Limitations and new advances

- Apply propensity score in longitudinal studies

  construct time-dependent propensity score

  - sequential matching

  - inverse-probability-of treatment weighted (IPTW) estimator

PSM vs Randomization

- Randomization does not require the *untestable* assumption of independence conditional on observables
- PSM requires large samples and good data:
  - Ideally, the same data source is used for participants and non-participants
  - Participants and non-participants have access to similar institutions and markets, and
  - The data include *X* variables capable of identifying program participation and outcomes.

Design	When to use	Advantages	Disadvantages
Randomization	▫ Whenever feasible ▫ When there is variation at the individual or community level	▫ Gold standard ▫ Most powerful	▫ Not always feasible ▫ Not always ethical
Matching	▫ When other methods are not possible	▫ Overcomes observed differences between treatment and comparison	▫ Assumes no unobserved differences (often implausible)

Reference

- D'Agostino (1998), "Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group", Stat. Med. 17, 2265-2281.
- Dehejie & Wahba (1999), "Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs", JASA, 94, 1053-1062.
- Dergis (1988), "Solving non-bipartite matching problem via shortest path techniques", Annals of Operations Research, 13, 225-261.
- Joffe & Rosenbaum (1999), "Propensity scores", American J. of Epi., 150, 327-333.
- Imbens (2004), "Nonparametric estimation of average treatment effects under exogeneity", Review of economics and statistics, 86, 4-29.

Reference

- Lalonde (1986), "Evaluating the econometric evaluations of training programs", American Economic Review, 76, 604-620.
- Lu (2005), "Propensity score matching with time dependent covariates", Biometrics, 61, 721-728.
- Lu, Zanutto, Hornik & Rosenbaum (2001), "Matching with doses in an observational study of a media campaign against drug abuse".
- Robins, Hernan & Brumback (2000), "Marginal structural models and causal inference in epidemiology", Epidemiology, 11, 550-560.
- Rosenbaum (1987), "Model-based direct adjustment", JASA, 82, 387-394.

Reference

- Rosenbaum (1989), "Optimal matching for observational studies", JASA, 84, 1024-1032.
- Rosenbaum (2002), Observatioal Studies, 2nd Edition, Springer.
- Rosenbaum & Rubin (1983), "The central role of the propensity score in observational studies for causal effect", Biometrika, 70, 41-55.
- Rosenbaum & Rubin (1985), "The bias due to incomplete matching", Biometrics, 41, 103-116.
- Rubin (2005), "Causal inference using potential outcomes: Design, Modeling, Decision", JASA, 100, 322-331.