Applied Econometrics for Health Policy

Matching

Lecturer: Li-Lin Liang, Ph.D.

Institute of Public Health, National Yang Ming Chiao Tung University

Contents

Evaluating programs when assignment not clear

- Random assignment
- Instrumental variables
- Regression discontinuity

All three methods produce estimates of the counterfactual through explicit program assignment rules.

- Difference-in-differences
- **Matching**

These two methods are used when the assignment rules are less clear, or when none of the three methods above is feasible.

Constructing an artificial comparison group

- Matching uses large data sets and statistical techniques to construct an artificial comparison group based on observed characteristics.
  - For each unit under treatment, it attempts to find a non-treatment unit (or set of nontreatment units) that has the most similar characteristics possible.
  - Those matched non-treatment units then become the comparison group that you use to estimate the counterfactual.
- Example: evaluating the impact of a job training program on income.
  - You have a data set, such as income and tax records, that contains both individuals that enrolled in the program and individuals that did not enroll.
  - The program does not have any clear assignment rules (such as randomized assignment or an eligibility index) that explain why some individuals enrolled in the program and others did not.

Constructing an artificial comparison group

- Finding a good match for each program participant requires the characteristics that explain that individual's decision to enroll in the program.
  - If use only a few characteristics to identify the matched comparison group, you run the risk of leaving out other important characteristics.
    - E.g. age, gender, and whether the individual has a secondary school diploma
  - If use a large number of relevant characteristics, or if each characteristic takes on many values, it will be hard to identify a match for each of the units in the treatment group. You run into the curse of dimensionality.
    - E.g. the above three, plus number of children, number of years of education, number of months unemployed, number of years of experience.

Exact matching on four characteristics

Treated units			
Age	Gender	Months unemployed	Secondary diploma
19	1	3	0
35	1	12	1
41	0	17	1
23	1	6	0
55	0	21	1
27	0	4	1
24	1	8	1
46	0	3	0
33	0	12	1
40	1	2	0

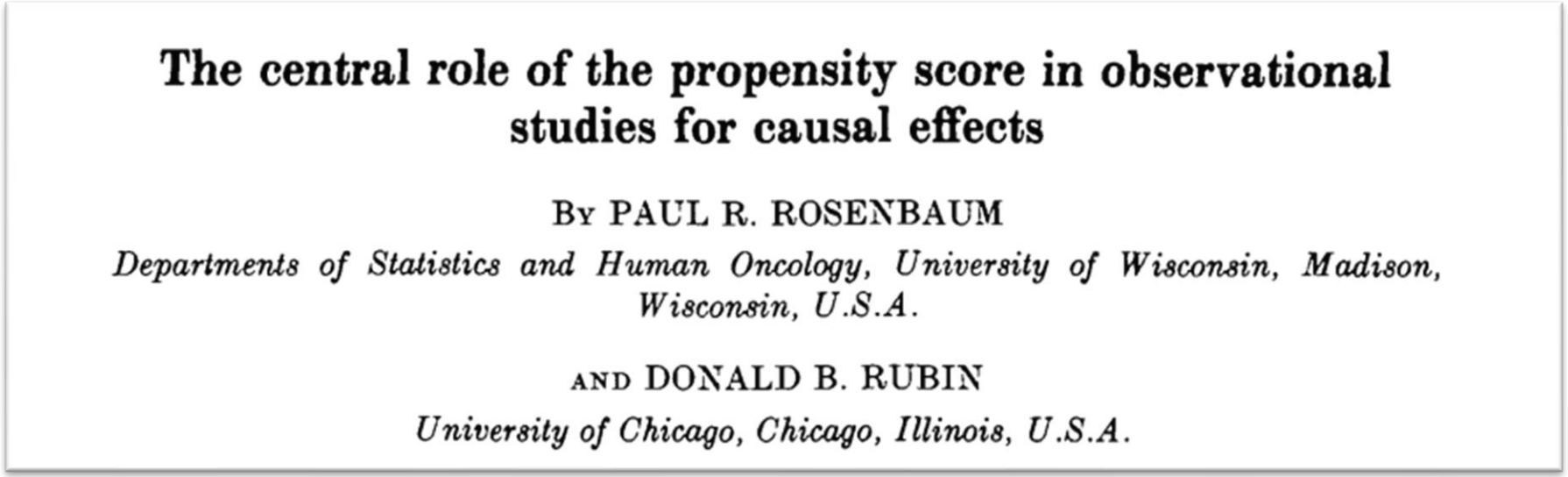
Untreated units			
Age	Gender	Months unemployed	Secondary diploma
24	1	8	1
38	0	1	0
58	1	7	1
21	0	2	1
34	1	20	0
41	0	17	1
46	0	9	0
41	0	11	1
19	1	3	0
27	0	4	0

Is this strategy feasible?

- This may not be feasible when
  - The sample is small
  - The set of covariates is large
  - Many of the covariates have many values or are continuous
- This is what we call the dimensionality problem
  - How many cells do we have
    - with 2 binary X variables? 2*2
    - with 3 binary X variables? 2*2*2
    - with K binary X variables? 2*2*2*2*….
    - with 2 variables that take on 7 values each? 7*7
- As the number of cells grows, we'll get lack of common support
  - cells containing only treated observations
  - cells containing only controls

A method to solve the dimensionality problem

- Rosenbaum and Rubin (1983) propose an equivalent and feasible estimation strategy based on the concept of Propensity Score.



The central role of the propensity score in observational studies for causal effects

BY PAUL R. ROSENBAUM

Departments of Statistics and Human Oncology, University of Wisconsin, Madison, Wisconsin, U.S.A.

AND DONALD B. RUBIN

University of Chicago, Chicago, Illinois, U.S.A.

"The propensity score allows to convert the multidimensional setup of matching into a one-dimensional setup. In that way, it allows to reduce the dimensionality problem."

Propensity score matching: step 1

- **Step 1**: For each unit in the T group and in the pool of nonenrolled, compute the probability that this unit will enroll in the program (i.e. propensity score) based on baseline characteristics (the explanatory variables).
  - This score is a real number between 0 and 1 that summarizes the influence of all of the observed characteristics on the likelihood of enrolling in the program.
  - Do NOT use post treatment characteristics because they might have been affected by the program itself. This violates the requirement for a good estimate of the counterfactual: the C group must be similar in all aspects, except for the fact that the T group receives the treatment and the C group does not.

Propensity score matching: step 2

- **Step 2**: Match units in the T group with units in the pool of nonenrolled that have the closest propensity score.
  - The propensity score matching (PSM) method tries to mimic the randomized assignment to T and C groups by choosing for the C group those units that have similar propensities to the units in the T group.
  - Since PSM is not a randomized assignment method but tries to imitate one, it belongs to the category of quasi-experimental methods.

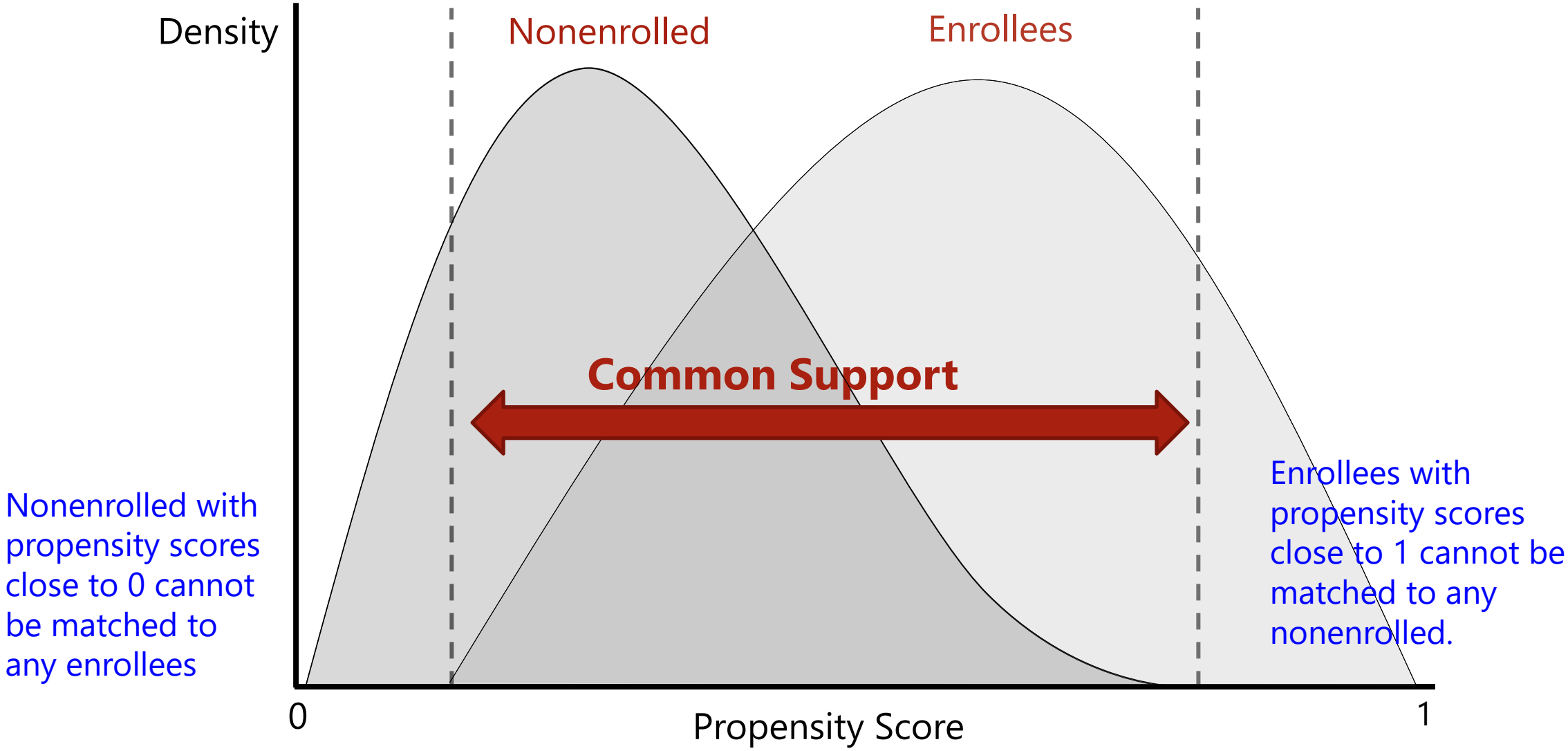
Propensity score matching: step 3 & 4

- **Step 3**: Restrict the sample to units for which common support appears in the propensity score distribution.
- **Step 4:** Compute the average difference in outcomes between the T units and their matched C units in the restricted sample. This produces the estimated impact of the program.
  - Those matched C units become the comparison group and are used to produce an estimate of the counterfactual.

Propensity score matching: common support

- A potential problem: lack of common support
  - Some enrolled units cannot find the units in the pool of nonenrolled that have similar propensity scores.
  - There may be a lack of common support, or lack of overlap, between the propensity scores of the treatment group and those of the pool of nonenrolled.
  - A lack of common support often appears at the extremes, or tails, of the distribution of propensity scores (see next slide). In this case, the matching procedure estimates the local average treatment effect (LATE) for observations on the common support.

Propensity score matching: common support



Three critical issues about matching

1. Matching methods can use only observed characteristics to construct a C group.
   - If there are any unobserved characteristics that affect whether a unit enrolls in the program and also affect the outcome, then the impact estimates obtained with the matched C group would be biased.
2. Matching must be done using only characteristics that are not affected by the program.
   - Matching solely based on postintervention characteristics is not recommended. If baseline data are not available, we can only use (usually few) characteristics that are unaffected by a program, such as age and gender.
   - If baseline data are available, we can match based on a richer set of characteristics, including the outcomes of interest (and combine it with DID).

Three critical issues about matching

3. The matching method's estimation results are only as good as the characteristics that are used for matching.
   - Even more important is to be able to match on the basis of characteristics that determine enrollment. The more we understand about the criteria used for participant selection, the better we will be able to construct the matched C group.

Combining matching with other methods

- Matching technique requires a significant amount of data and carries a significant risk of bias
- To reduce the bias, matching is often combined with other IE methods.
  - Matched difference-in-differences.
  - The synthetic control method.

Matched difference-in-differences

- When baseline data on outcomes are available, matching can be combined with DID.
  - PSM itself cannot account for unobserved characteristics that explain why a group chooses to enroll in a program and that also affect outcomes.
- Matched DID takes care of any unobserved characteristics that are constant across time between the two groups.
  - "Rural Roads and Local Market Development in Vietnam." *Journal of Development Studies* 47 (5): 709–34.
  - "Housing, Health, and Happiness." *American Economic Journal: Economic Policy* 1 (1): 75–105.

Matched difference-in-differences

- It is implemented as follows:
  1. Perform matching based on observed baseline characteristics.
  2. For each enrolled unit, compute the change in outcomes between the before and after periods (first difference).
  3. For each enrolled unit, compute the change in outcomes between the before and after periods for this unit's matched comparison (second difference).
  4. Subtract the second difference from the first difference; that is, apply the difference-in-differences method.
  5. Finally, average out those double differences.

The synthetic control method

- This method uses information about the characteristics of the treated unit and the untreated units to construct a "synthetic," or artificial, comparison unit by weighting each untreated unit in such a way that the synthetic comparison unit most closely resembles the treated unit.
  - This requires a long series of observations over time of the characteristics of both the treated unit and the untreated units.
  - This combination of comparison units into a synthetic unit provides a better comparison for the treated unit than any untreated unit individually.
  - "The Economic Costs of Conflict: A Case Study of the Basque Country." *American Economic Review*, 93(1), 113–132.

Limitations of the matching method

- Requires extensive data sets on large samples of units.
- There may be a lack of common support.
- Although matching helps control for observed background characteristics, we can never rule out **selection bias** that stems from unobserved characteristics.
  - By definition, we cannot incorporate unobserved characteristics in the calculation of the propensity score.
  - Must assume that no unobserved characteristics that simultaneously affect program participation and outcomes.
  - However, we cannot test the assumption.

Limitations of the matching method

- Matching alone is generally less robust than the other IE methods.
  - Randomized assignment, IV, and RDD do not require the untestable assumption that there are no such unobserved variables.
  - Matching methods are typically used when those IE methods are not possible.
- Ex-post matching is very risky when no baseline data are available on the outcome of interest or on background characteristics.
  - If an evaluation uses survey data that were collected ex post to infer what people's background characteristics were at baseline, and then use those inferred characteristics to perform matching, it may inadvertently match based on characteristics that were also affected by the program; in that case, the estimation result would be invalid.

When is the matching method most useful?

- Matching based on baseline background characteristics can be very useful when it is combined with other techniques.
  - E.g. DID + matching allows us to correct for differences between the groups that are fixed over time.
  - Matching is also more reliable when the program assignment rule and underlying variables are known, in which case matching can be performed on those variables.

Checklist: Matching

1. Matching relies on the assumption that enrolled and nonenrolled units are similar in terms of any unobserved variables that could affect both the probability of participating in the program and the outcome.
2. Is program participation determined by variables that cannot be observed?
   - This cannot be directly tested, so you will need to rely on theory, common sense, and good knowledge of the setting of the IE for guidance.

Checklist: Matching

3. Are the observed characteristics well balanced between matched sub-groups?
   - Compare the observed characteristics of each treatment and its matched comparison group of units at baseline.
4. Can a matched comparison unit be found for each treatment unit?
   - Check whether sufficient common support exists in the distribution of the propensity scores.
   - Small areas of common support indicate that enrolled and nonenrolled persons are very different, and that casts doubt as to whether matching is a credible method.

Keep in mind: Matching

Matching requires large samples and good quality data.

Matching at baseline can be very useful:

- Know the assignment rule and match based on it
- combine with other techniques (i.e. diff-in-diff)

Ex-post matching is risky:

- If there is no baseline, be careful!
- matching on endogenous ex-post variables gives **bad** results.

If there are *unobservable* characteristics and those unobservables influence participation: Selection bias!

Case study: estimating the impact of HISP using PSM

Case Study: HISP

- National Health Insurance Subsidy Program (HISP)
  - Subsidizes the purchase of health insurance for poor rural households.
  - Covers all expenses related to outpatient care and medicine for enrollees.
- Objective
  - Reduce health spending for poor households
- Many outcomes of interest
  - Here per capita yearly out-of-pocket health expenditures (refers to OOP hereafter)
- What is the effect of HISP (P) on OOP (Y) ?

Case Study: HISP

- Operational rules
  - Introduced HIPS as a pilot program to 100 rural villages
    - Promotional campaigns to encourage enrollment
  - Threshold: reduce OOP by at least $10 on average within in two years
  - Targeting: eligibility based on a poverty index
- Rich data
  - Round 0: baseline survey of all 4,960 households in pilot villages before piloting
  - Round 1: data collected on the same 4,960 households at the end of the two-year pilot.

Case Study: HISP

- One of the outputs:
  - A total of 2,965 (out of 4,960) households enrolled in HISP
- If outcome is a decrease of **$10** or more, then scale up nationally

The impact of HISP based on PSM

- MoH wants you to estimate the impact of HISP by applying matching methods. Your aim is to select a group of nonenrolled households that look similar to the enrolled households based on baseline observed characteristics.

The impact of HISP based on PSM

- Class exercise:

  Use “HISP survey.dta”. To perform PSM, you need to transform data from a long to wide format. Your data set includes an indicator for the enrolled household (*enrolled*=1) and for the post-period (*round*=1).

  1. Please estimate the propensity score (probability of enrollment) based on all socioeconomic household characteristics in the HISP survey.
  2. Check the distribution of propensity scores.
  3. Please estimate the program effect by using the PSM alone.
  4. Please estimate the program effect by using matched DID analysis.

	health_expenditures	Out of pocket health expenditure (per capita per year)
Control variables	age_hh	Age of the head of the household (in years)
	age_sp	Age of the spouse (in years)
	educ_hh	Education of the head of household (completed years of schooling)
	educ_sp	Education of the spouse (completed years of schooling)
	indigenous	Head of household speaks an indigenous language (0=no, 1=yes)
	female_hh	Head of the household is a woman (0=no, 1=yes)
	hhsize	Number of household members (at baseline)
	dirtfloor	Home has a dirt floor at baseline (0=no, 1=yes)
	bathroom	Home with private bathroom at baseline (0=no, 1=yes)
	land	Number of hectares of land owned by household at baseline
	hospital_distance	Distance to closest hospital
Other variables	locality_identifier	Locality identifier
	household_identifier	Unique household identifier
	round	Survey round (0 = baseline; 1 = follow-up)
	enrolled	Household enrolled in HISP (0=no, 1=yes)
	enrolled_rp	Household enrolled in HISP under the randomized promotion scenario (0=no, 1=yes)
	eligible	Household eligible to enroll in HISP (0=no, 1=yes)
	treatment_locality	Household is located in treatment villages (0=no, 1=yes)
	promotion_locality	Household is located in locality randomly assigned promotion of HISP (0=no, 1=yes)
	poverty_index	Poverty index 20-100
	hospital	HH member visited hospital in the past year (0=no, 1=yes), used in power calculations

probit enrolled age_hh age_sp educ_hh educ_sp female_hh indigenous hhsize dirtfloor bathroom land hospital_distance

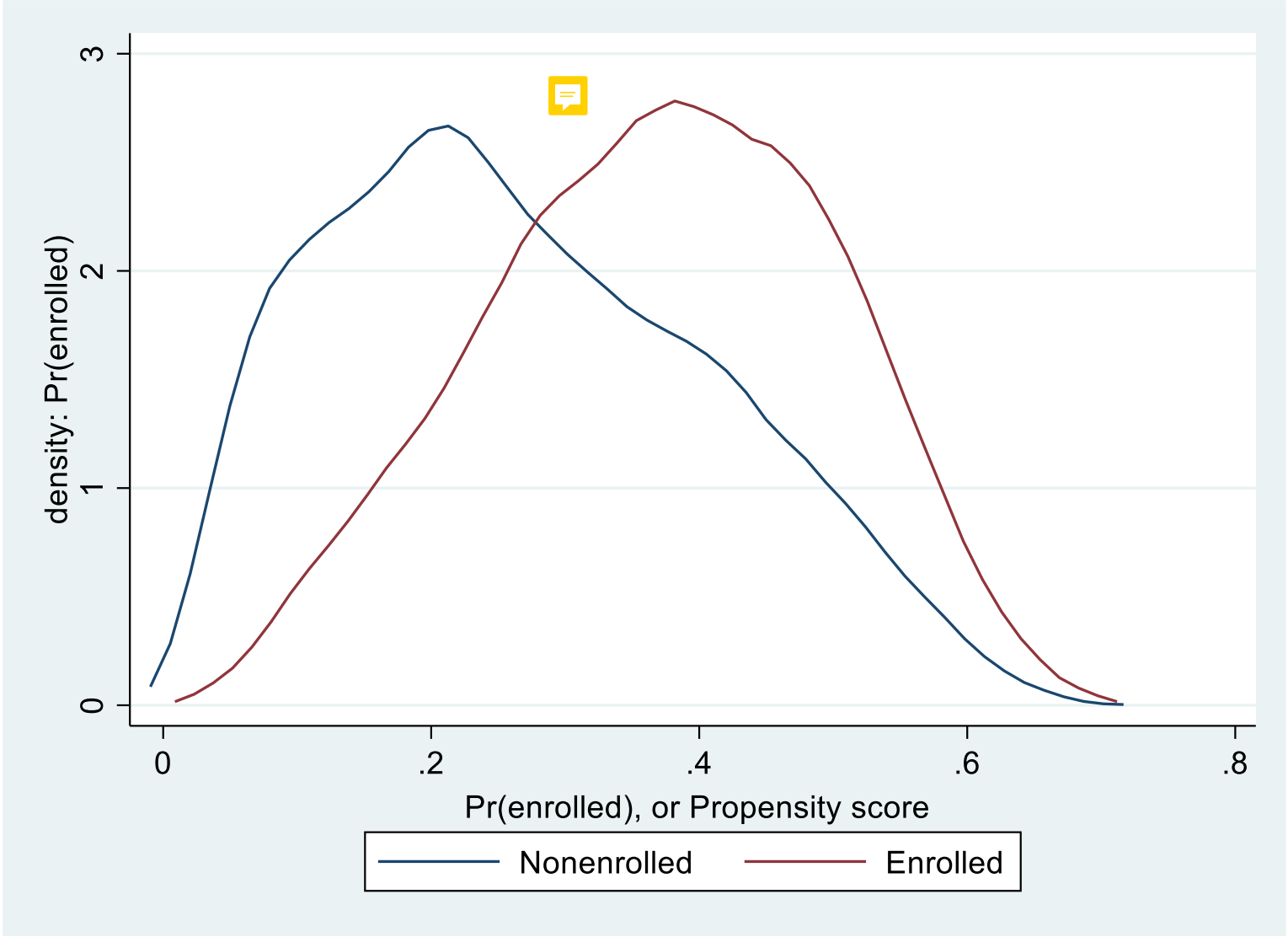
Baseline Characteristics	**Estimated Coefficient** *Probit Regression, Prob Enrolled=1*
Head's age (years)	-0.013**
Spouse's age (years)	-0.008**
Head's education (years)	-0.022**
Spouse's education (years)	-0.016*
Head is female=1	-0.020
Indigenous=1	0.161**
Number of household members	0.119**
Dirt floor=1	0.376**
Bathroom=1	-0.124**
Hectares of Land	-0.028**
Distance to Hospital (km)	0.002**
Constant	-0.497**

Significant level: * = 5%. ** = 1%.

Matching for HISP: common support

predict pscore
(option pr assumed; Pr(enrolled))
kdensity pscore if enrolled ==1, gen(take1 den1)
kdensity pscore if enrolled ==0, gen(take0 den0)
twoway (line den0 take0) (line den1 take1)

Note: “take1” stores Pr(enrolled), and “den1” stores the density estimate for the enrolled. Similarly, “take0” and “den0” are for the nonenrolled.



PSM: nearest neighbourhood matching

- Stata locates, for each enrolled household, the nonenrolled household that has the closest propensity score to the enrolled household.

```
psmatch2 enrolled $controls, outcome (health_expenditures1) common
```

imposes a common support by dropping T units whose pscore is higher than the max. or less than the min. pscore of the C units.

Variable	Sample	Treated	Controls	Difference	S.E.	T-stat
health_expendi~1	Unmatched	7.83977335	20.70746	-12.8676866	.226604141	-56.78
	ATT	7.83977335	17.8088716	-9.96909828	.263484213	-37.84

Before matching

This is outcome at follow-up

Note: S.E. does not take into account that the propensity score is estimated.

```
sum health_expenditures1 if enrolled==1 & _weight!=. & _support==1
```

Variable	Obs	Mean	Std. dev.	Min	Max
health_exp~1	2,964	7.839773	7.995814	0	87.38017

PSM: nearest neighbourhood matching

- Stata locates, for each enrolled household, the nonenrolled household that has the closest propensity score to the enrolled household.

psmatch2 enrolled $controls, outcome (health_expenditures1) common

Variable	Sample	Treated	Controls	Difference	S.E.	T-stat
health_expendi~1	Unmatched	7.83977335	20.70746	-12.8676866	.226604141	-56.78
	ATT	7.83977335	17.8088716	-9.96909828	.263484213	-37.84

Note: S.E. does not take into account that the propensity score is estimated.

sum health_expenditures1 if enrolled==0 & _weight!=. & _support==1 [fweight=_weight]

Variable	Obs	Mean	Std. dev.	Min	Max
health_exp~1	2,964	17.80887	8.029909	8.246194	83.97215

PSM: nearest neighbourhood matching

- psmatch2 creates a number of variables (using the default: nearest neighbor):

  _treated: 0 for control observations and 1 for treatment observations.

  _support: 1 if the observation is on the common support; 0 if the observation is off the support.

  _pscore: the estimated propensity score.

  _outcome_variable: for every treatment observation stores the value of the matched outcome.

  _weight: the frequency with which the observation is used as a match

  _id: a new identifier created for all observations.

  _n1: for every treatment observation, it stores the observation number of the matched control observation. Do not forget to sort by _id

  _nn: for every treatment observation, it stores the number of matched control observations.

PSM: nearest neighbourhood matching

- psmatch2 creates a number of variables:

  _pdif: abs(pscore-pscore[nearest neighbor])

- Use _n1 to check _pdif
  - list _id _treated _weight _pscore _n1 _pdif if _id==7369

	_id	_treated	_weight	_pscore	_n1	_pdif
7369.	7369	Treated	1	.21885996	2904	.00001771

  - list _id _treated _weight _pscore _n1 _pdif if _id==2904

	_id	_treated	_weight	_pscore	_n1	_pdif
2904.	2904	Untreated	1	.21884225	.	.

  - display .21885996- .21884225
    .00001771

PSM: nearest neighbourhood matching

- Two different ways to estimate the standard error of ATT
  1. Use bootstrapping.

     set seed 100

     bootstrap r(att) : psmatch2 enrolled $controls, outcome(health_expenditures1)

```
Bootstrap results                                       Number of obs = 9,913
                                                        Replications  =    50

      Command: psmatch2 enrolled age_hh age_sp educ_hh educ_sp female_hh indigen
                  hospital_distance, out(health_expenditures1)
        _bs_1: r(att)
```

	Observed coefficient	Bootstrap std. err.	z	P>|z|	Normal-based [95% conf.	interval]
_bs_1	-9.969098	.2534896	-39.33	0.000	-10.46593	-9.472268

PSM: nearest neighbourhood matching

- Two different ways to estimate the standard error of ATT

  2. Use linear regression.

```
reg health_expenditures1 enrolled [fweight=_weight]
```

Automatically restrict sample to matched units

The sample size: 2 times the number of treated units matched (2,964)

Source	SS	df	MS	Number of obs	=	5,928
				F(1, 5926)	=	2293.94
Model	147285.488	1	147285.488	Prob > F	=	0.0000
Residual	380486.195	5,926	64.2062428	R-squared	=	0.2791
				Adj R-squared	=	0.2789
Total	527771.683	5,927	89.0453321	Root MSE	=	8.0129

health_exp~1	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
enrolled	-9.969098	.2081443	-47.90	0.000	-10.37714	-9.56106
_cons	17.80887	.1471802	121.00	0.000	17.52034	18.0974

PSM: nearest neighbourhood matching

- Options for psmatch2: logistic regression

use logit instead of the default probit to estimate the propensity score.

psmatch2 enrolled $controls, outcome (health_expenditures1) common logit

Logistic regression

Number of obs = **9,913**
LR chi2(**11**) = **1071.23**
Prob > chi2 = **0.0000**
Pseudo R2 = **0.0886**

Log likelihood = **-5511.4708**

enrolled	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
age_hh	**-.022572**	**.0030448**	**-7.41**	**0.000**	**-.0285397**	**-.0166043**
age_sp	**-.0133585**	**.0035547**	**-3.76**	**0.000**	**-.0203256**	**-.0063915**

Variable	Sample	Treated	Controls	Difference	S.E.	T-stat
health_expendi~1	Unmatched	7.83977335	20.70746	-12.8676866	.226604141	-56.78
	ATT	7.83977335	18.1533395	-10.3135661	.280009557	-36.83

PSM: choice of matching estimator

- Options for psmatch2: choice of matching estimator

  - neighbor(integer): number of neighbors used to calculate the matched outcome. Defaults to 1. Default matching method is single nearest-neighbor (without caliper).
  - noreplacement: perform 1-to-1 matching without replacement. Nearest neigbor propensity score matching only. Default is with replacement.
  - radius: perform radius matching within the specified radius given by caliper.
  - caliper(real): value for maximum distance of controls. Use to perform nearest neighbor(s) within caliper, radius matching and Mahalanobis 1-to-1 matching.
  - kernel: perform kernel matching.

PSM: choice of matching estimator

- Options for psmatch2: choice of matching estimator

psmatch2 enrolled $controls, outcome (health_expenditures1) common neighbor (3)

_id	psmatch2: Identifier (ID)
_n1	psmatch2: ID of nearest neighbor nr. 1
_n2	psmatch2: ID of nearest neighbor nr. 2
_n3	psmatch2: ID of nearest neighbor nr. 3
_nn	psmatch2: # matched neighbors

psmatch2 enrolled $controls, outcome (health_expenditures1) common radius caliper(0.01)

sum _weight if _treated==0

Variable	Obs	Mean	Std. dev.	Min	Max
_weight	6,933	.4275206	.3087136	.007874	3.243391

Weight of matched controls. Stata used all controls. For treated units, weight =1

PSM: nearest neighbourhood matching

1. How many nearest neighbors should we use?
   - Matching just one nearest neighbor minimizes bias at the cost of larger variance (of estimated effect).
   - Matching using additional nearest neighbors increase the bias but decreases the variance.
2. Matching with or without replacement?
   - Matching with replacement keeps bias low at the cost of larger variance.
   - Matching without replacement keeps variance low at the cost of potential bias.

PSM: balance checking

pstest $controls, both

By default considers balancing for the ATT (Average Treatment Effect on the Treated), where the treated are the reference group.

Requires comparability to be assessed both before and after matching. Default is only after matching.

Variable	Unmatched / Matched	Mean Treated	Mean Control	%bias	%reduct |bias|	t-test t	t-test p>|t|
age_hh	U	41.657	48.138	-44.3		-19.70	0.000
	M	41.657	41.699	-0.3	99.3	-0.12	0.903
age_sp	U	36.836	41.612	-38.7		-17.27	0.000
	M	36.836	37.028	-1.6	96.0	-0.65	0.514
educ_hh	U	2.9712	2.7745	7.2		3.26	0.001
	M	2.9712	2.8969	2.7	62.2	1.07	0.287
educ_sp	U	2.7033	2.581	4.8		2.19	0.028
	M	2.7033	2.6699	1.3	72.7	0.50	0.615

t-tests for equality of means in the treated and non-treated groups. For good balancing, these should be non significant after matching.

The standardized bias before and after matching (formulae from Rosenbaum and Rubin, 1985): this should be less than 5% after matching

PSM: balance checking

- What can we do to improve the matching?

1. Change the matching method
    - In the NN method, all treated units find a match. However, some of these matches are fairly poor because for some treated units the nearest neighbor may have a very different propensity score
    - Caliper and radius matching (among others) offer a solution to this problem.
2. Change the propensity score model and re-do the matching
    - Add or remove covariates selected to predict the propensity score
    - For not well balanced variables: include higher order terms (e.g., squared values) and/or interactions (guidelines in Caliendo and Kopeining, 2006 and Dehejia and Wahba, 1999)

PSM + DID

- Manually compute the matched difference-in-differences (see do file: "Stata command_Propensity score matching")

Variable	Obs	Mean	Std. dev.	Min	Max
matchedDD	2,964	-9.421269	10.12336	-67.73042	67.7352

- Use regression to compute matched difference-in-differences and standard error on DID (see do file)

Source	SS	df	MS		
Model	131542.778	1	131542.778	Prob > F	0.0000
Residual	319049.757	5,926	53.8389734	R-squared	0.2919
Total	450592.535	5,927	76.0237109	Root MSE	7.3375

	Number of obs =	5,928
	F(1, 5926) =	2443.26
	Prob > F =	0.0000
	R-squared =	0.2919
	Adj R-squared =	0.2918
	Root MSE =	7.3375

diff	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
enrolled	-9.421269	.1906006	-49.43	0.000	-9.794916	-9.047622
_cons	2.771348	.134775	20.56	0.000	2.50714	3.035556

Supplementary: ATE, selection bias, and unconfoundedness

Potential outcomes

(In the perfect world: we observe both states)

ID	D	Y	Y(0)	Y(1)
1	0	21	21	22
2	1	31	16	31
…	…	…	…	…
n	1	15	15	15

But in reality we only observed bule numbers

ID	D	Y	Y(0)	Y(1)
1	0	21	21	22
2	1	31	16	31
…	…	…	…	…
n	1	15	15	15

(In reality)

ID	D	Y(0)	Y(1)
1	0	21	.
2	1	.	31
…	…	…	…
n	1	.	15

$\delta_i = Y_i(1) - Y_i(0)$

Individual treatment effect

D=1 Treated

D=0 Untreated

Y(0): population-level random variable for outcome Y in control state.

Y(1): population-level random variable for outcome Y in treatment state.

δ: individual-level causal effect of the treatment.

Potential outcomes

- There are two potential outcomes for each unit:

a) is exposed and participates in the program [$(Y_i|\ D_i = 1)$ or Y_{i1} or $Y_i(1)$]

b) is not exposed to the program and does not participate [$(Y_i|D_i = 0)$ or Y_{i0} or $Y_i(0)$].

- For each unit *i* in the population, the causal effect of the program is theoretically determined by a simple difference between the potential outcome with the program and the potential outcome without the program:

$$\delta_i = Y_{i1} - Y_{i0} = Y_i(1) - Y_i(0)$$ ⬅ Individual treatment effect

- The basic problem of causal inference is that we cannot observe the same unit in both states of the world at the same time, so it is impossible to observe program effects for each unit.

Average treatment effect

- Since causal effects cannot be measured for each unit *i*, let us go back to the population (*I*) to see how to identify the average treatment effect (ATE).

- At the population level, the average treatment effect is the difference between the expected value of the outcome when the population is exposed to the program and the expected value of the outcome when the population is not exposed to the program:

$$ATE_I = E_I(Y_{i1} - Y_{i0}) = E_I(Y_{i1}) - E_I(Y_{i0})$$

- We are interested in using a sample generated through an IE design to estimate the ATE for the population.

Average treatment effect

- One potential estimator that can help infer the average treatment effect for the population is to take the difference in the average outcome of units in a sample exposed to the program and the average outcome of units in a sample not exposed to the program:

Unknown: $$ATE_I = E_I(Y_{i1} - Y_{i0}) = E_I(Y_{i1}) - E_I(Y_{i0})$$

Estimator: $$\delta = (\bar{Y} \mid D = 1) - (\bar{Y} \mid D = 0)$$

- While ATE_I is for the whole population, the *estimator* is applied to an observed sample of units. The estimator produces an *estimate* of the ATE by taking the difference between the mean sample outcomes among units exposed to the program and the mean sample outcomes among units not exposed to the program.

Independence of potential outcomes

- This estimator of the ATE is only consistent, when specific conditions are met.
  - Consistency means, based on the sample, the estimator tends to generate accurate estimates of the average treatment effect for the population.
- The main condition is that exposure to the program should be independent of the distribution of potential outcomes, known as conditional independence.
- Selection bias is one case when the condition of independence of potential outcomes does not hold. A selection bias may arise when the units that participate in the program are different or react differently to the program than units that do not participate in the program.

Independence of potential outcomes

- This condition has two parts:

1. The average outcome of program beneficiaries and nonbeneficiaries should be the same if neither of them was exposed to the program. Formally, $(\overline{Y_1}|D = 1) = (\overline{Y_1}|D = 0)$.

2. The average outcome of program beneficiaries and nonbeneficiaries should be the same if they were both exposed to the program. Formally, $(\overline{Y_0}|D = 1) = (\overline{Y_0}|D = 0)$.

Independence of potential outcomes

- In other words, when the condition of independence of potential outcomes holds:

1. No baseline differences would be expected between the group exposed to the program and the group not exposed to the program; and
2. Both groups should react to the program in the same way if they are exposed to it.

Independence of potential outcomes

- If this is the case, the average causal effect of the program over the population can be estimated from the difference in average outcomes between the sample units exposed to the program and the sample units not exposed to the program.
- This means that we can replace the theoretical and unmeasurable treatment effect of the program on a specific individual unit *i* with the estimated average treatment effect of the program for a sample of units drawn from a population of such units.

Selection bias

- To see when the difference in average outcomes between the treatment and comparison groups can consistently estimate the average treatment effect, we can rewrite the average treatment effect of the program over the population as:

$$ATE_I = E_I(Y_{i1}) - E_I(Y_{i0}) = E_I(Y_{i1}|D=1) - E_I(Y_{i0}|D=0)$$

$$= E_I(Y_{i1}|D=1) - \boldsymbol{E_I(Y_{i0}|D=1)} + \boldsymbol{E_I(Y_{i0}|D=1)} - E_I(Y_{i0}|D=0)$$

$$= E_I\big((Y_{i1}-Y_{i0})|D=1\big) + [E_I(Y_{i0}|D=1) - E_I(Y_{i0}|D=0)]$$

= (Average treatment effect on the treated) + (Selection bias)

Selection bias

$$ATE_I = E_I(Y_{i1}) - E_I(Y_{i0}) = E_I(Y_{i1}|D=1) - E_I(Y_{i0}|D=0)$$

$$= E_I\big((Y_{i1} - Y_{i0})|D=1\big) + [E_I(Y_{i0}|D=1) - E_I(Y_{i0}|D=0)]$$

= (Average treatment effect on the treated) + (Selection bias)

- The difference in average outcomes between a group exposed to the program and a group not exposed to the program is the sum of the average treatment effect on the treated (i.e. on those participating in the program) plus the selection bias.
- The selection bias is zero when there is *no difference in average* Y_{i0} *between those who did and did not receive the program.*

Selection bias

$$ATE_I = E_I(Y_{i1}) - E_I(Y_{i0}) = E_I(Y_{i1}|D=1) - E_I(Y_{i0}|D=0)$$

$$= E_I\big((Y_{i1} - Y_{i0})|D=1\big) + [E_I(Y_{i0}|D=1) - E_I(Y_{i0}|D=0)]$$

= (Average treatment effect on the treated) + (Selection bias)

- When there is no selection bias, the difference in average outcomes between groups provides a consistent estimate of the average treatment effect in the population. This is achieved under the conditional independence assumption, in which case:

$$ATE_I = E_I((Y_{i1} - Y_{i0})|D=1) = ATT_I$$

Average treatment effect on the treated (ATT)

ID	D	Y(0)	Y(1)
1	0	21	.
2	1		31
…	…	…	…
n	1		15

$$ATT = E[\delta \mid D=1]$$
$$= E[Y(1)-Y(0) \mid D=1]$$
$$= E[Y(1) \mid D=1] - E[Y(0) \mid D=1]$$

Problem: We don't observe the control state for the treatment group.

Average treatment effect on the nontreated (ATNT)

ID	D	Y(0)	Y(1)
1	0	21	
2	1	.	31
...	...	...	...
n	1	.	15

$$\text{ATNT} = E[\delta \mid D=0]$$
$$= E[Y(1)-Y(0)|D=0]$$
$$= E[Y(1)|D=0]-E[Y(0)|D=0]$$

Problem: We don't observe the treatment state for the control group.

The evaluation problem

Usually refers to no treatment

Y_{0i} , Y_{1i} → Outcome of i under treatment 0 and under treatment 1

$D_i \in \{0, 1\}$ → Treatment indicator

$Y_i = D_i\, Y_{1i} + (1 - D_i)\, Y_{0i}$ → Observed outcome of i

$$Y_i = Y_{0i} + (Y_{1i} - Y_{0i})\, D_i$$

Causal effect on Y of treatment 1 relative to treatment 0 for i

X_i → Set of observed characteristics of i

Which parameter?

- ATT $= E(Y_1 - Y_0 \mid D{=}1) = E(Y_1 \mid D{=}1) - \mathbf{E(Y_0 \mid D{=}1)}$
- ATNT $= E(Y_1 - Y_0 \mid D{=}0) = \mathbf{E(Y_1 \mid D{=}0)} - E(Y_0 \mid D{=}0)$
- ATE $= E(Y_1 - Y_0) = \text{ATT}\cdot P(D{=}1) + \text{ATNT}\cdot P(D{=}0)$

Need to invoke (untestable) assumptions to identify **average unobserved counterfactuals.**

Two conditions for matching estimator

- To ensure that the matching estimators consistently estimate the treatment effects, we assume:

1. Unconfoundedness: assignment to treatment is independent of the outcomes, conditional on the covariates X.

$$(Y(0); Y(1)) \perp\!\!\!\perp D \mid X$$

2. Common support condition: the probability of assignment is bounded away from zero and one.

$$0 < \Pr(D=1|X) < 1$$

Unconfoundedness

- Unconfoundedness is the major identifying assumption
  - Also termed selection on observables, or **conditional independence**
- Intuition:
  - If the decision to take the treatment is purely random for individuals with similar values of the pre-treatment variables, then we could use the average outcome of some similar individuals who were not exposed to the treatment as the counterfactuals.
    - For each *i*, matching estimators impute the missing outcome by finding other individuals in the data whose covariates are similar but who were exposed to the other treatment.
    - In this way, differences in outcomes of this well selected control group can be attributed to the treatment.

Unconfoundedness in PSM

Definition

The propensity score $p(X)$ is the conditional probability of receiving the treatment given the pre-treatment variables D:

$$p(X) = Pr\{D = 1 | X\} = EX\{D | X\}$$

Lemma 1

If $p(X)$ is the propensity score, then $D \perp X \mid p(X)$

- Given the propensity score, the pre-treatment variables are balanced between beneficiaries and non- beneficiaries

Lemma 2

$Y1, Y0 \perp D \mid X \Rightarrow Y1, Y0 \perp D \mid p(X)$

- Suppose that assignment to treatment is unconfounded given the pre-treatment variables X. Then assignment to treatment is unconfounded given the propensity score $p(X)$.

Unconfoundedness in PSM

- The *balancing property* of the propensity score (Lemma 1) ensures that:
  - Observations with the same propensity score have the same distribution of observable covariates independently of treatment status; and
  - For a given propensity score, assignment to treatment is "random" and therefore treatment and control units are observationally identical on average.

Warning: Matching STILL doesn't allow to control for selection bias that arises when the assignment to the treatment is done **on the basis of non-observables.**