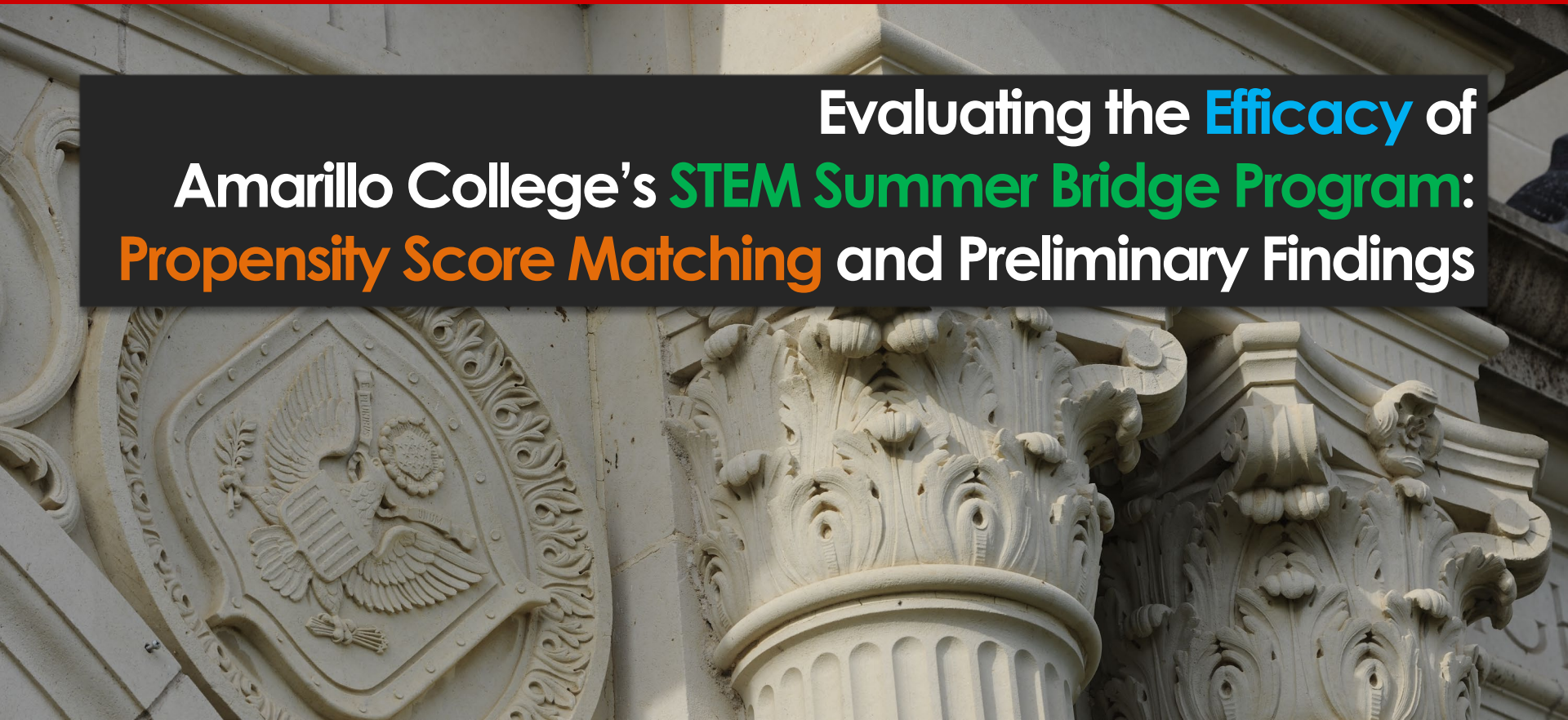


Evaluating the Efficacy of Amarillo College's STEM Summer Bridge Program: Propensity Score Matching and Preliminary Findings

BACKGROUND

Amarillo College (AC) received a 5-year DOE grant for their ***Innovating and Advancing in STEM Education*** project

- ***Overcharging Goal***: Improve students' *academic achievement* and *persistence*, ultimately enhancing their ability to complete a *STEM degree* and *transfer to a 4-year institution*

Amarillo College®

BACKGROUND

Amarillo College (AC) received a 5-year DOE grant for their ***Innovating and Advancing in STEM Education*** project

- ***Overcharging Goal***: Improve students' *academic achievement* and *persistence*, ultimately enhancing their ability to complete a *STEM degree* and *transfer to a 4-year institution*

- #1: Develop a work-based learning system (15 courses)
- #2: Update technology skills instruction
- *#3: Develop a STEM Scholars Program*
- #4: Strengthen articulation between AC and West Texas A&M University

BACKGROUND

Initiative #3: Develop a STEM Scholars Program – "***Summer Bridge Program***"

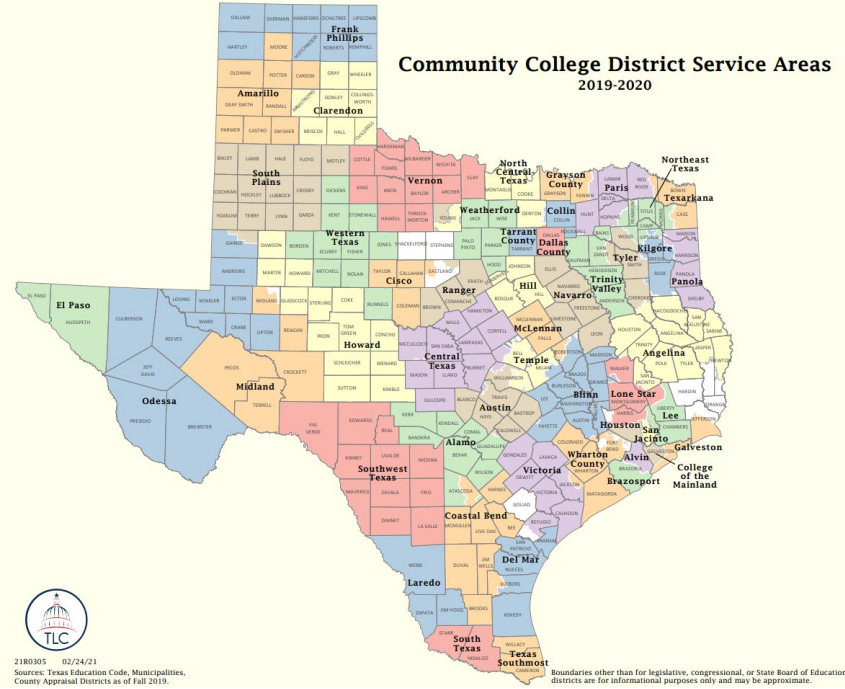
- For 1st year students: Development ***Seminars*** + math/science ***Bootcamps***
- For 2nd year students: ***STEM Research*** with WT A&M or TTU + ***Coaching Services***

BACKGROUND

Amarillo College

- Urban college offering:
  - 140+ transfer and technical programs
  - 10 STEM AS programs

- 9,159 students (in 2020)
  - 44% Hispanic
  - 70% first-generation (82% of Hispanic)
  - 51% low-income (57% of Hispanic)
  - 59% part-time
  - 770 STEM majors (342 Hispanic)



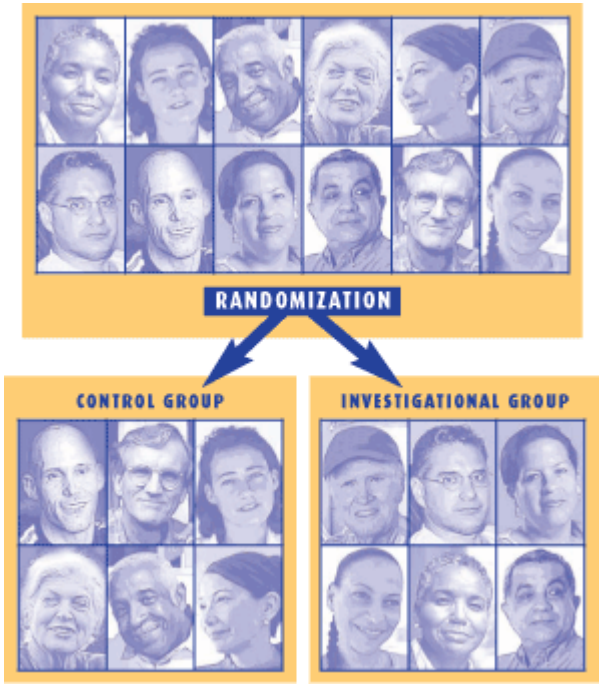
THE TASK

TTU performs ***summative program evaluation*** examining the effects of the *Summer Bridge Program* on student outcomes

- *4 cohorts*: AY22-23, AY23-24, AY24-25, AY25-26;
  24 program participants in the first cohort
- *Key outcomes*: GPA, dropout, STEM persistence, degree completion (any degree or STEM), transfer to a 4-year institution, etc.

BUT, THERE ARE ALWAYS CHALLENGES

In an <u>ideal</u> setting,



- Students are *<u>randomly</u>* assigned to the *Treatment (Intervention)* condition and the *Control* condition
  - By the virtue of ***random assignment***, the two condition groups are "comparable" at baseline
  - Thus, we can make a ***causal inference*** that any observed differences between the two groups are solely due to the *Treatment*

BUT, THERE ARE ALWAYS CHALLENGES

In reality,

- Students were *not randomized*. Rather, they "elected" to participate in the *Summer Bridge Program*
  - *Program participants* and *non-participants* may be considerably dissimilar in some personal characteristics — ***selection bias***
  - Thus, when the two groups show a difference in outcomes, this could be due to the *program*, personal factors, *or both*

BUT, THERE ARE ALWAYS CHALLENGES

In reality,

- Students were *not randomized*. Rather, they "elected" to participate in the *Summer Bridge Program*

- **24** program participants vs. **650** non-participants! (1 : 27)

BUT, THERE ARE ALWAYS CHALLENGES

In reality,

- Students were *not randomized*. Rather, they "elected" to participate in the *Summer Bridge Program*
- **24** program participants vs. **650** non-participants! (1 : 27)
- Required to address ***WWC Standards*** for *baseline equivalence*

0.00 ≤ ES Difference ≤ 0.05	0.05 < ES Difference ≤ 0.25	ES Difference > 0.25
Satisfies baseline equivalence	Statistical adjustment required to satisfy baseline equivalence	Does not satisfy baseline equivalence

CHALLENGE ACCEPTED.

PSM

A solution: Propensity score methods

- In real settings, it is often *infeasible* or *unethical* to randomly assign people into different (*Treatment* and *Control*) conditions
    - New drug testing for acute cancer

- In such case, ***propensity score methods*** are useful to account for possible *selection bias* and thereby allow us for addressing questions of *causal inference*

PSM

Propensity score (Rosenbaum & Rubin, 1983)

- "How *likely* does a person receive or select the treatment (*T*) *given* his/her personal characteristics (*X*) at baseline?"

$$P_i\left(T_i \middle| X_i\right)$$

PSM

Propensity score (Rosenbaum & Rubin, 1983)

- PS exists ***both*** in *randomized* trials and in *observational* studies
- In *randomized* trials, the "true" PS is ***known and equal*** for all individuals (e.g., 0.5 in coin toss)

PSM

Propensity score (Rosenbaum & Rubin, 1983)

- In *observational* studies, the "true" PS is *unknown*

$$\hat{P}_i\left(T_i \middle| X_i\right)$$

  - People already in the *Treatment* and *Control* conditions
  - PS is *estimated* for each person using his/her "actual" *treatment status* (*T*) and values on the *covariates* (*X*) measured at baseline
  - PSM utilize this *conditional probability* to "recreate" a situation that *would have been expected* in a *randomized* trial

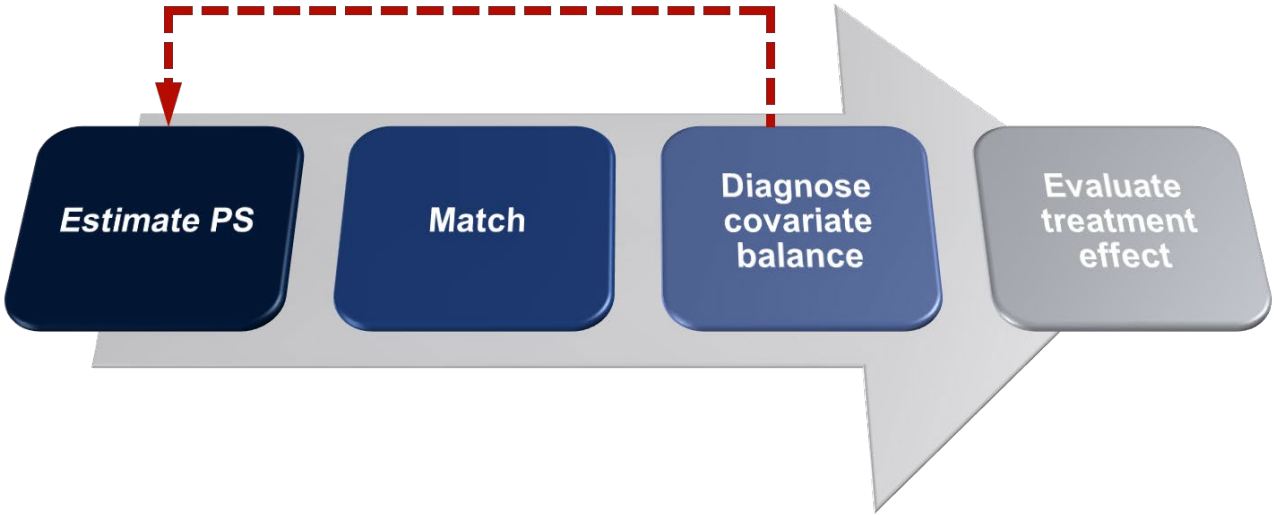
PSM

4 popular methods (Austin, 2009; Rosenbaum, 2002; Rosenbaum & Rubin, 1983)

- ***Matching*** treated *persons* with untreated *persons*
- ***Weighting*** *data*
- ***Stratifying*** *sample*
- ***Adjusting*** *parameter estimates*

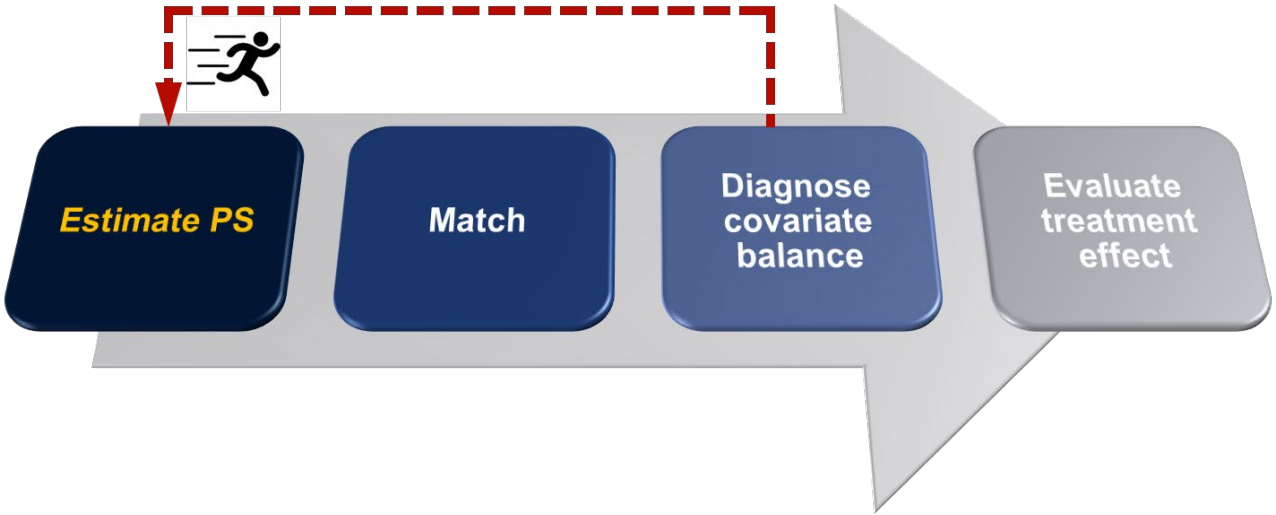
PROCESS

Propensity score matching



- Iteratively check *balance* on the *covariates* (*X*) between treated persons and untreated persons in the *"matched" sample*

ESTIMATING PS



ESTIMATING PS

2 popular methods of estimating PS

- *Parametric*: ***Logistic regression***
- *Non-parametric*: ***Generalized boosted modeling***

ESTIMATING PS: LR

Logistic regression

$$\hat{P}_i\left(T_i \middle| X_i\right)$$

$$\Leftrightarrow \ln\left(\frac{P_i(T_i=1)}{1-P_i(T_i=1)}\right)=\beta_0+\beta_1\left(X_{1i}\right)+\beta_2\left(X_{2i}\right)+\ldots+\beta_n\left(X_{ni}\right)+e_i$$

T_i = "actual" *treatment status* (1 = treatment, 0 = no treatment)

X_i, ... , X_{ni} = values on the *covariates* measured at baseline

ESTIMATING PS: LR

Summer Bridge Program

Variable	***Value***
Program participation	*Yes* (1), *No* (0)
Age	in year
Gender	Female / Male
Race / Ethnicity	White / Black / Hispanic / Asian
First generation	Yes / No
GPA	A=4, B=3, C=2, D=1, F=0
Dropout, STEM persistence, Degree completion, Transfer	Yes / No

ESTIMATING PS: LR

Logistic regression

```
> param <- matchit(treated ~ age + gender + race_ethnicity +
            first_gen_status
            family=binomial, data=dat)
> param$distance
```

	unique_id	age	gender	race_ethnicity	first_gen_status	treated	distance
1	23	17	Male	Unknown or not reported	Not First Generation	1	1.820141e-01
2	2	18	Male	Hispanic/Latino	First Generation	1	7.701892e-02
3	21	17	Female	Hispanic/Latino	First Generation	1	2.625690e-01
4	4	18	Female	Asian	First Generation	1	2.737584e-01
5	3	19	Female	Asian	First Generation	1	1.142132e-01

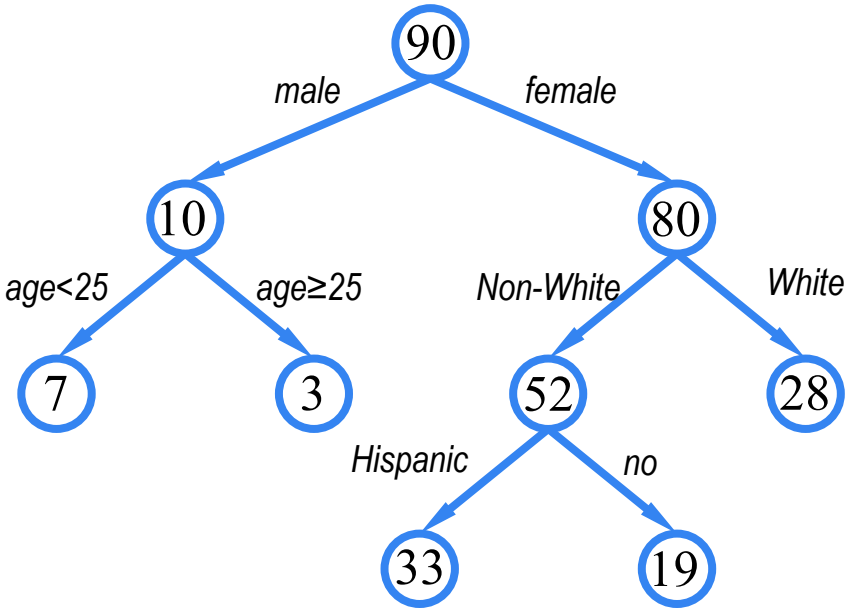
ESTIMATING PS: GBM

Generalized boosted modeling allows for multiway *product terms* modeled "naturally" as a result of sample splitting. (Friedman, 2001; McCaffrey, Ridgeway, & Morral, 2004)

- ***Step 1***: Randomly select 50% of the sample — "training data".
- ***Step 2***: Predict *treatment status* using ***Classification and Regression Trees (CART)***.

ESTIMATING PS: GBM

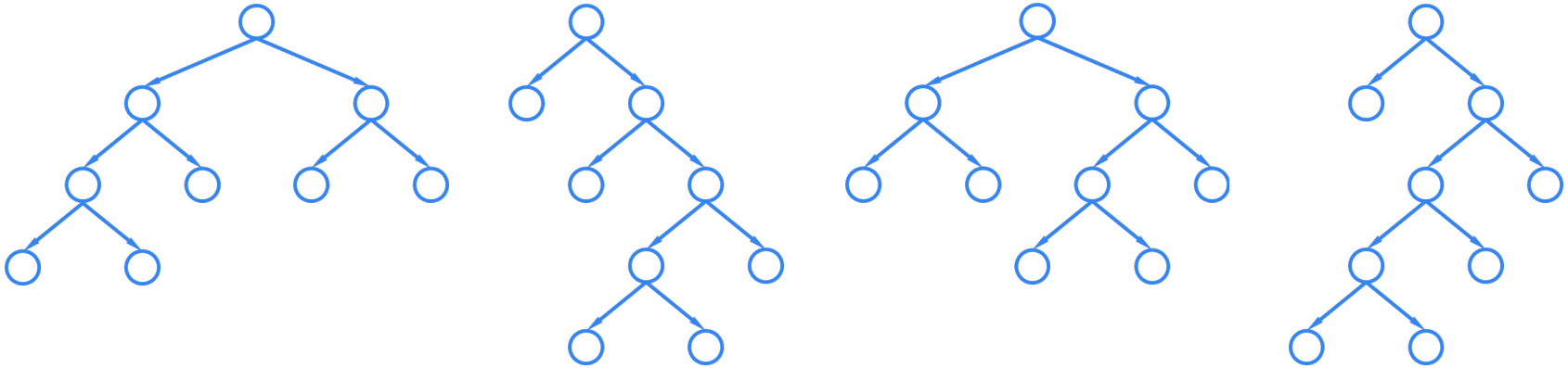
- ***Step 2***: Predict *treatment status* using ***CART***.



- The selected sample is split by the covariate that, among all covariates, best predicts treatment status
- The difference between "estimated" PS and "actual" treatment status — *residual* — is computed within each split subset
- Additional splits are made by predicting the *residual* with the remaining covariates

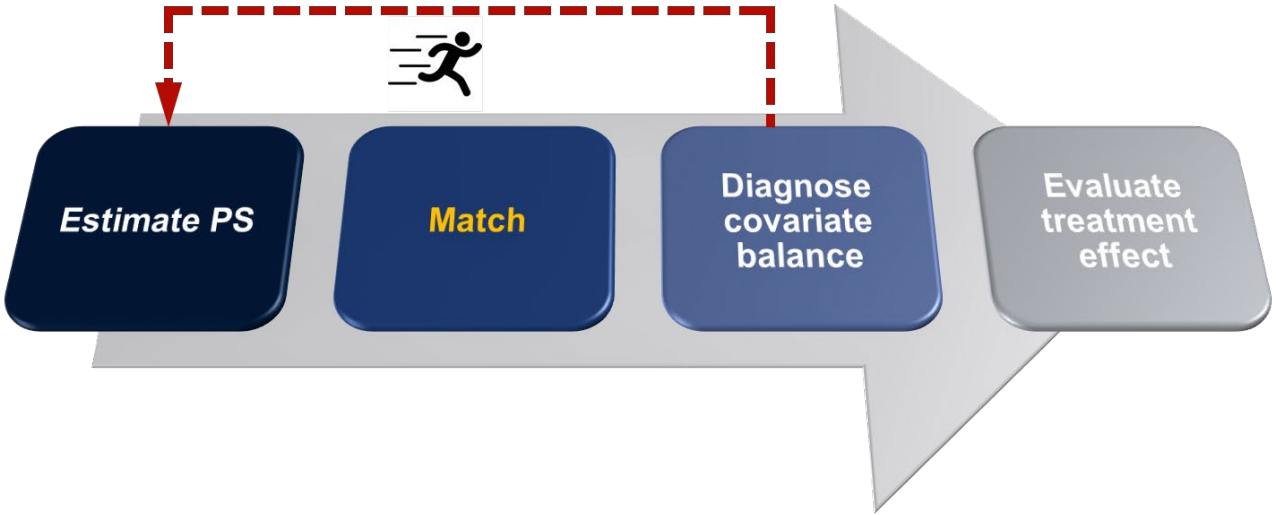
ESTIMATING PS: GBM

- ***Step 3***: Many trees are formed by repeating ***Steps 1 & 2***.



- ***Step 4***: The trees are combined together to calculate a final PS estimate for each person in the sample.

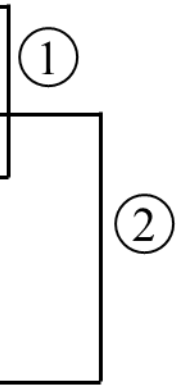
MATCHING


Estimate PS
Match
Diagnose covariate balance
Evaluate treatment effect

NEAREST NEIGHBOR MATCHING

In ***nearest neighbor (a.k.a. greedy) matching***, a treated person is matched to an untreated person if their PS are *most similar* — in the "smallest" distance.

ID	Treated	PS
1	1	.57
2	1	.36
3	0	.54
4	0	.60
5	0	.17
⋮	⋮	⋮

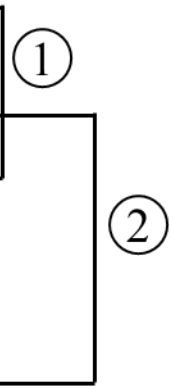


- Find a match for treated persons, one by one

NEAREST NEIGHBOR MATCHING

In ***nearest neighbor (a.k.a. greedy) matching***, a treated person is matched to an untreated person if their PS are *most similar* — in the "smallest" distance.

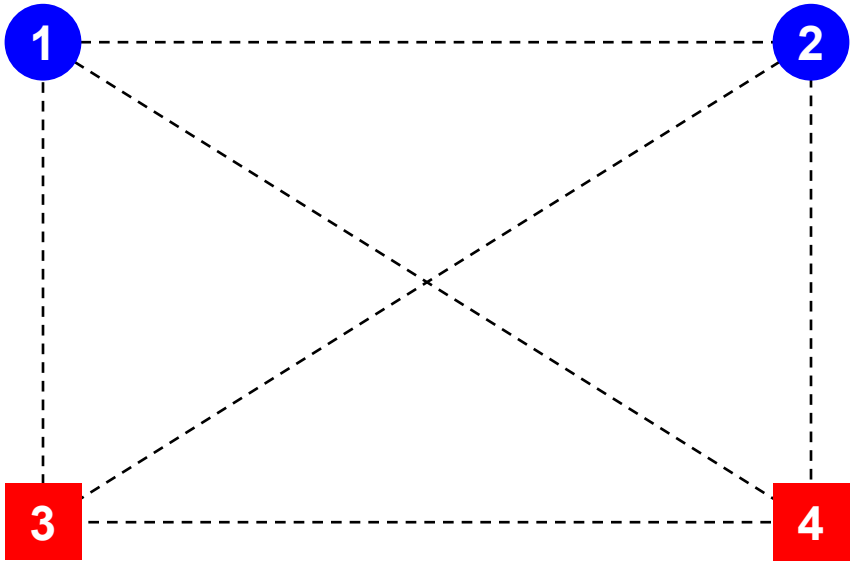
ID	Treated	PS
1	1	.57
2	1	.36
3	0	.54
4	0	.60
5	0	.17
⋮	⋮	⋮



- Often the PS is *not* close for possible pairs
- To avoid bad matches, define a "*caliper*" — the *maximum* distance in PS by which matches are allowed
- *0.25 x standard deviation of the logit of PS*
  (Rosenbaum & Rubin, 1985)

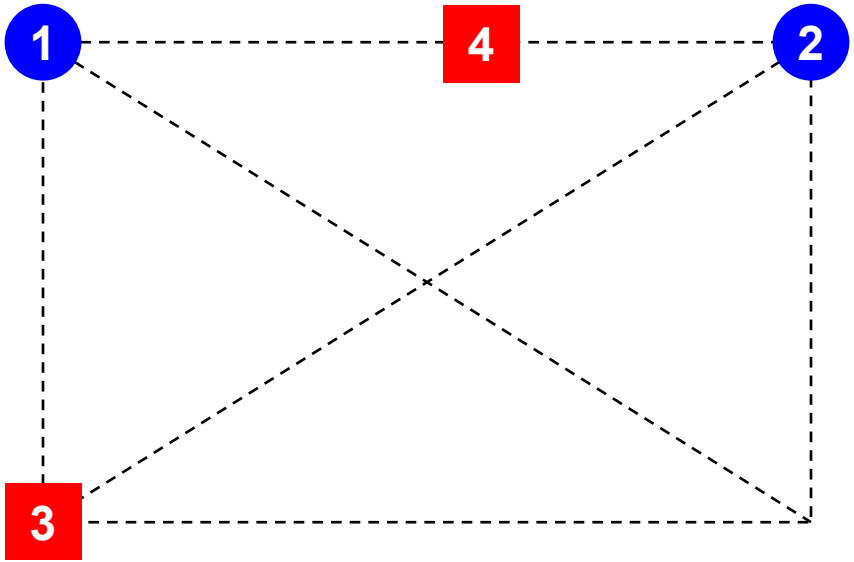
OPTIMAL MATCHING

In ***optimal matching***, matches are formed by *minimizing* the *global distance* in PS, defined as the sum of PS distances in the whole matched sample.



OPTIMAL MATCHING

In ***optimal matching***, matches are formed by *minimizing* the *global distance* in PS, defined as the sum of PS distances in the whole matched sample.



MATCHING

Nearest neighbor matching

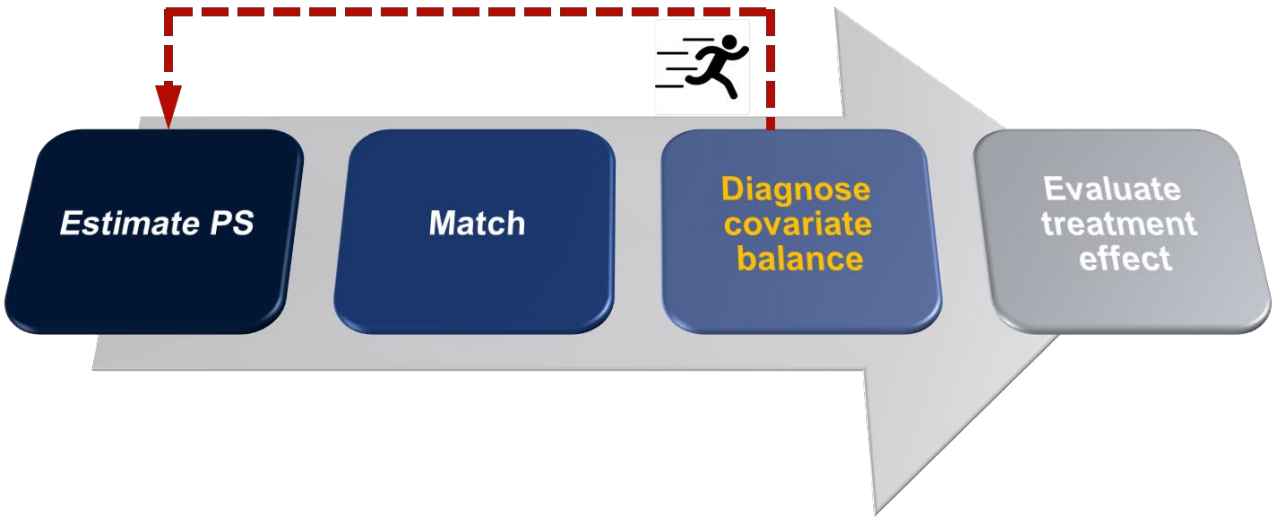
```
matchit(treated ~ age + gender + race_ethnicity +
        first_gen_status,
        family=binomial, data=dat,
        method="nearest", caliper=0.25)
```

```
Sample Sizes:
          Control Treated
All           650      24
Matched        23      23
Unmatched     627       1
Discarded       0       0
```

Optimal matching

```
matchit(treated ~ age + gender + race_ethnicity +
        first_gen_status,
        family=binomial, data=dat,
        method="optimal")
```

BALANCE DIAGNOSTICS



BALANCE DIAGNOSTICS

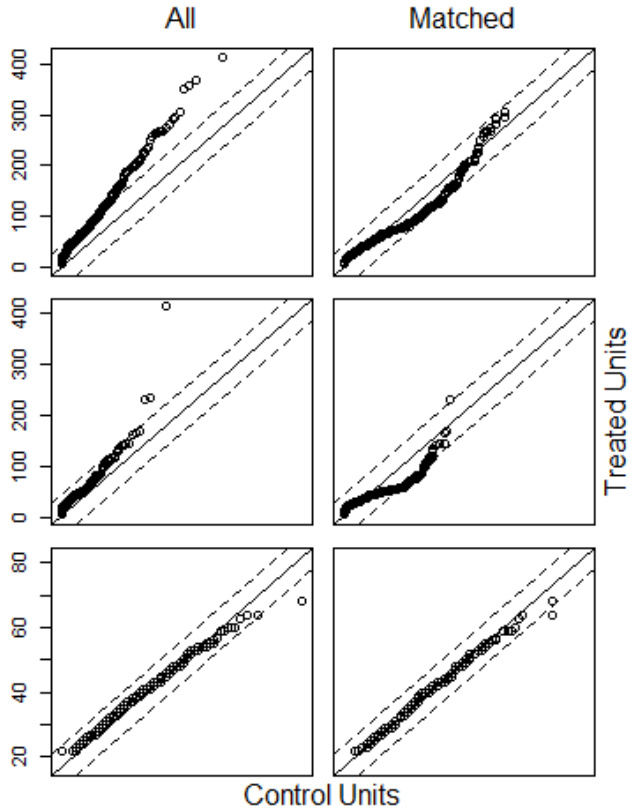
Once a matching is successfully implemented, the next step is to examine if *balance* is made on the covariates. (Austin, 2009; Flury & Riedwyl, 1986)

- Inspection of *distributions* — ***Q-Q plot***
- *Standardized difference* in *means* — ***Cohen's d***

BALANCE DIAGNOSTICS: Q-Q PLOT

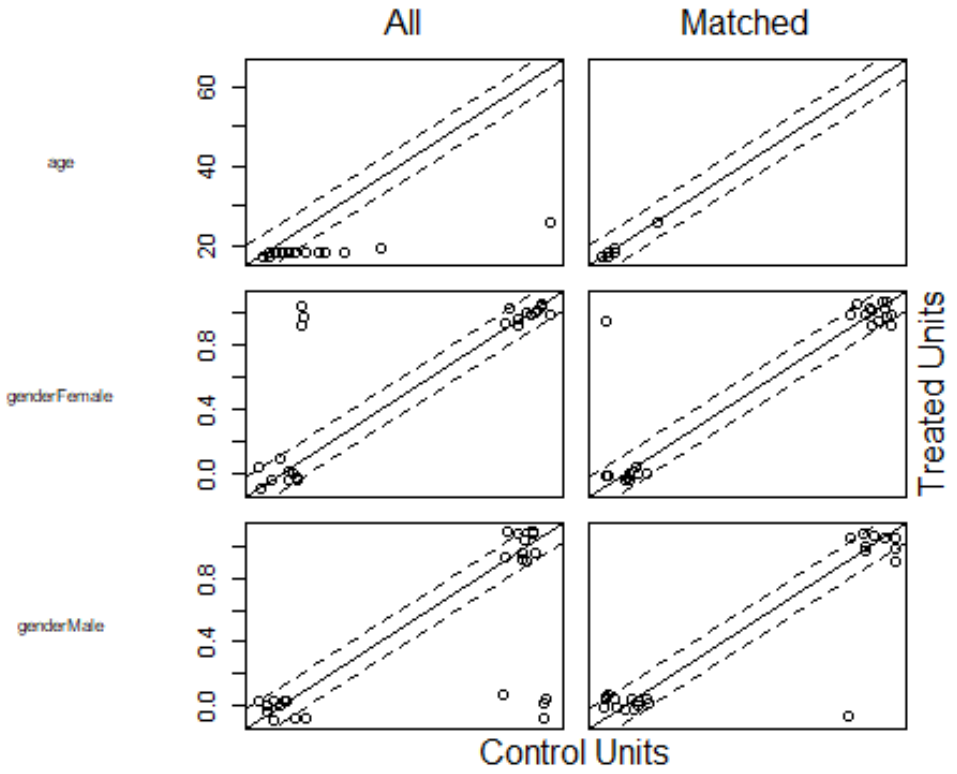
Q-Q plot

- The *distribution* of a covariate in the *Treatment* group is plotted against the distribution in the *Control* group
- *Deviations from a 45-degree line* indicate that the distributions are dissimilar.



BALANCE DIAGNOSTICS: Q-Q PLOT

Q-Q plot



BALANCE DIAGNOSTICS: Cohen's D

Cohen's d

- Continuous covariates:

$$\frac{\left(\bar{X}_T - \bar{X}_U\right)}{\sqrt{\frac{s_T^2 + s_U^2}{2}}}$$

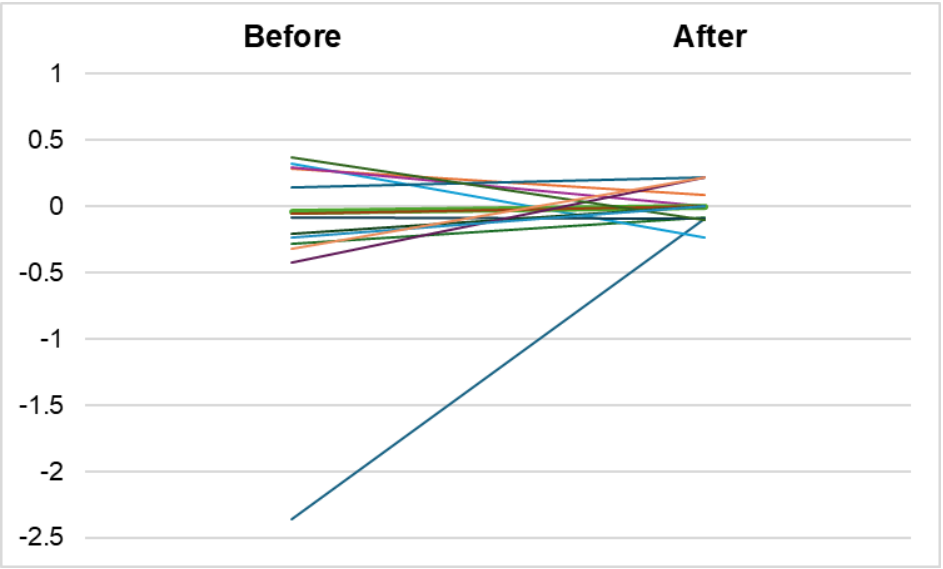
Binary covariates:

$$\frac{\left(\hat{P}_T - \hat{P}_U\right)}{\sqrt{\frac{\hat{P}_T\left(1-\hat{P}_T\right) + \hat{P}_U\left(1-\hat{P}_U\right)}{2}}}$$

  - A covariate with *n-categories* is *dichotomized* into *n* variables (by dummy-coding) and then examined for balance

BALANCE DIAGNOSTICS: Cohen's D

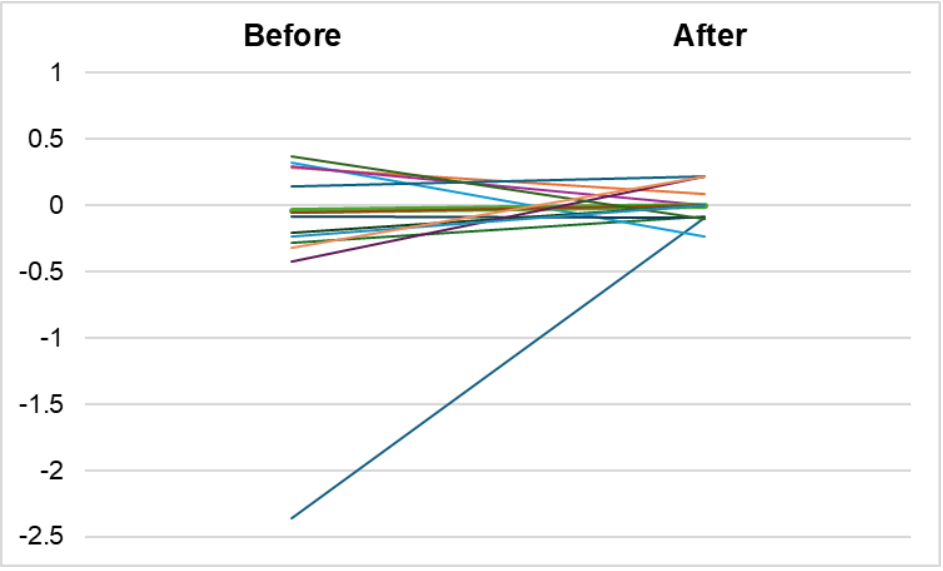
Cohen's d



	Std. Mean Diff.
distance	0.0025
age	-0.0976
genderFemale	0.0882
genderMale	-0.0882
race_ethnicityAsian	-0.2333
race_ethnicityBlack	0.0000
race_ethnicityHawaiian/Pacific Islander	0.0000
race_ethnicityHispanic/Latino	-0.0882
race_ethnicityNative American	0.0000
race_ethnicityTwo or more races	0.0000
race_ethnicityUnknown or not reported	0.2176
race_ethnicityWhite	0.2141
first_gen_statusFirst Generation	-0.1004
first_gen_statusNot First Generation	0.0000
first_gen_statusUnknown/Not Reported	0.2176

BALANCE DIAGNOSTICS: Cohen's D

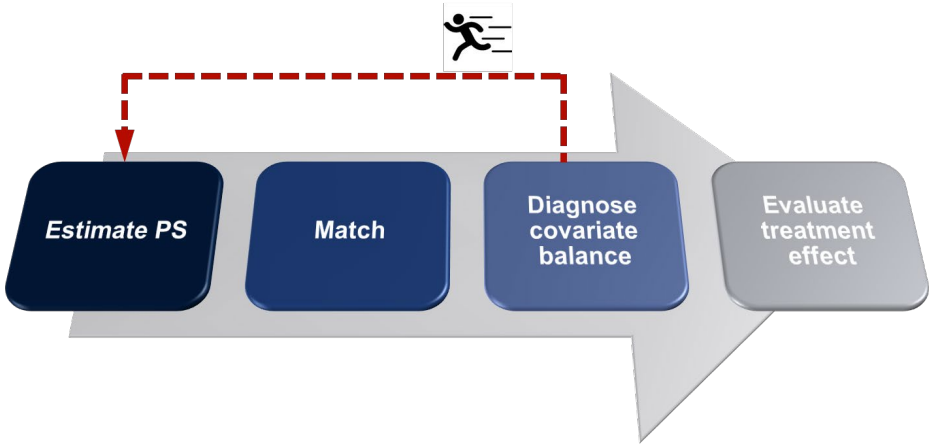
Cohen's d



	Std. Mean Diff.
distance	0.0025
age	-0.0976
genderFemale	0.0882
genderMale	-0.0882
race_ethnicityAsian	-0.2333
race_ethnicityBlack	0.0000
race_ethnicityHawaiian/Pacific Islander	0.0000
race_ethnicityHispanic/Latino	-0.0882
race_ethnicityNative American	0.0000
race_ethnicityTwo or more races	0.0000
race_ethnicityUnknown or not reported	0.2176
race_ethnicityWhite	0.2141
first_gen_statusFirst Generation	-0.1004
first_gen_statusNot First Generation	0.0000
first_gen_statusUnknown/Not Reported	0.2176

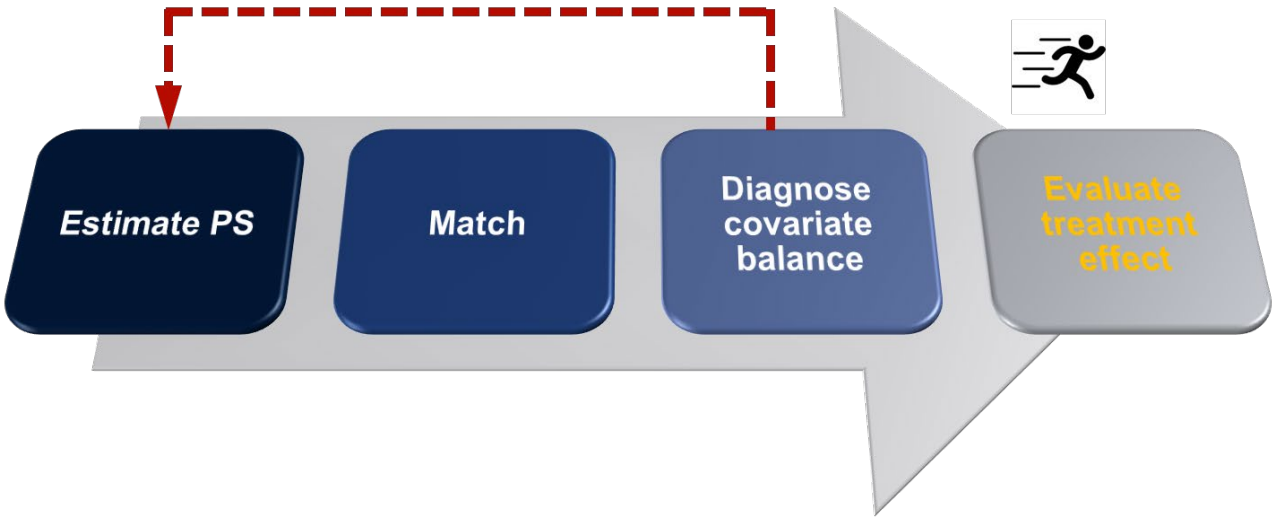
0.00 ≤ ES Difference ≤ 0.05	0.05 < ES Difference ≤ 0.25	ES Difference > 0.25
Satisfies baseline equivalence	*Statistical adjustment required to satisfy baseline equivalence*	Does not satisfy baseline equivalence

BALANCE DIAGNOSTICS



- If an *imbalance* is indicated by dissimilar distribution and/or nontrivial *d*,
  - *Transform* or *re-categorize* the unbalanced covariates
  - Add *polynomial terms* of the unbalanced covariates
  - Add *product terms* of the unbalanced covariates and other covariates
  - Use a *smaller caliper*

ESTIMATING TREATMENT EFFECT



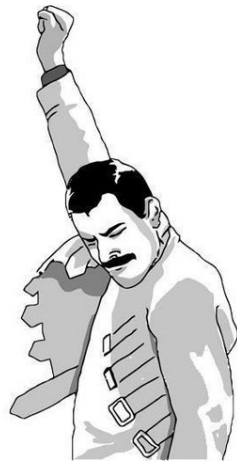
ESTIMATING TREATMENT EFFECT

Bivariate tests

- Estimate the *treatment effect* by comparing outcomes between treated persons and untreated persons in the *matched sample*

  - *Continuous* → difference in *means* → *t*-test
  - *Categorical* → difference in *proportions* → chi-square/Fisher test
  - *Binary* → difference in *probabilities* → chi-square test

$p<0.05$



ESTIMATING TREATMENT EFFECT

Bivariate tests

Spring 23 (after the summer program in Summer 22)

Outcome	*Treatment*	*Control*	*p*	*d / V*
GPA	2.8 ± 1.4	2.7 ± 0.8	0.90	0.04
Dropout	8%	25%	0.11	0.22
STEM persistence	86%	83%	0.79	0.04
Degree completion	8.3%	4.2%	0.55	0.09

- Positive outcomes are emerging, supporting the impacts of the ***Summer Bridge Program***

ESTIMATING TREATMENT EFFECT

Multivariate tests

- *Multivariate* analysis is also applicable, of course
    - Linear, logistic, or Poisson regression
    - Survival analysis (e.g., time to transfer to 4-year university)
    - Structural equation modeling
    - Hierarchical linear modeling
- The outcome models can include the covariates used for estimating PS, so as to further *eliminate* *residual imbalance* in "prognostically" important covariates (Harder et al., 2010; Ho et al., 2007)

ESTIMATING TREATMENT EFFECT

Concluding remarks

- *Selection bias* is a major threat to the *validity* of any observational study

- ***PS methodology*** offers researchers an integrative framework where…

  - Not only *"overt" bias* from the *measured covariates* can be corrected in the estimates of the treatment effect,

  - But also *"hidden" bias* from *unmeasured covariates* can be evaluated in terms of robustness of the effect estimate — ***sensitivity analysis***

Behaviour Research and Therapy 98 (2017) 76–90

Contents lists available at ScienceDirect

Behaviour Research and Therapy

journal homepage: www.elsevier.com/locate/brat

A practical guide to propensity score analysis for applied clinical research

Jaehoon Lee*, Todd D. Little

Department of Educational Psychology and Leadership, College of Education, Texas Tech University, United States

ARTICLE INFO

Article history:
Received 22 September 2016
Received in revised form
10 January 2017
Accepted 12 January 2017
Available online 19 January 2017

Keywords:
Propensity score
Matching
Subclassification
Weighting
R

ABSTRACT

Observational studies are often the only viable options in many clinical settings, especially when it is unethical or infeasible to randomly assign participants to different treatment régimes. In such case propensity score (PS) analysis can be applied to accounting for possible selection bias and thereby addressing questions of causal inference. Many PS methods exist, yet few guidelines are available to aid applied researchers in their conduct and evaluation of a PS analysis. In this article we give an overview of available techniques for PS estimation and application, balance diagnostic, treatment effect estimation, and sensitivity assessment, as well as recent advances. We also offer a tutorial that can be used to emulate the steps of PS analysis. Our goal is to provide information that will bring PS analysis within the reach of applied clinical researchers and practitioners.

QUESTIONS?

jaehoon.lee@ttu.edu
kwanghee.jung@ttu.edu