Observational & Quasi-experimental Research Methods

8th Annual Kathleen Foley Palliative Care Retreat
Methods Workshop
Park City, Utah, October 20, 2014

Helene Starks, PhD MPH

Associate Professor
University of Washington
Dept of Bioethics & Humanities;
Director, Metrics, Quality & Evaluation Core
UW Palliative Care Center of Excellence

Melissa M. Garrido, PhD

Assistant Professor
GRECC, James J Peters VAMC, Bronx, NY;
Brookdale Department of Geriatrics & Palliative Medicine,
Icahn School of Medicine at Mount Sinai, New York, NY

Overview of Workshop

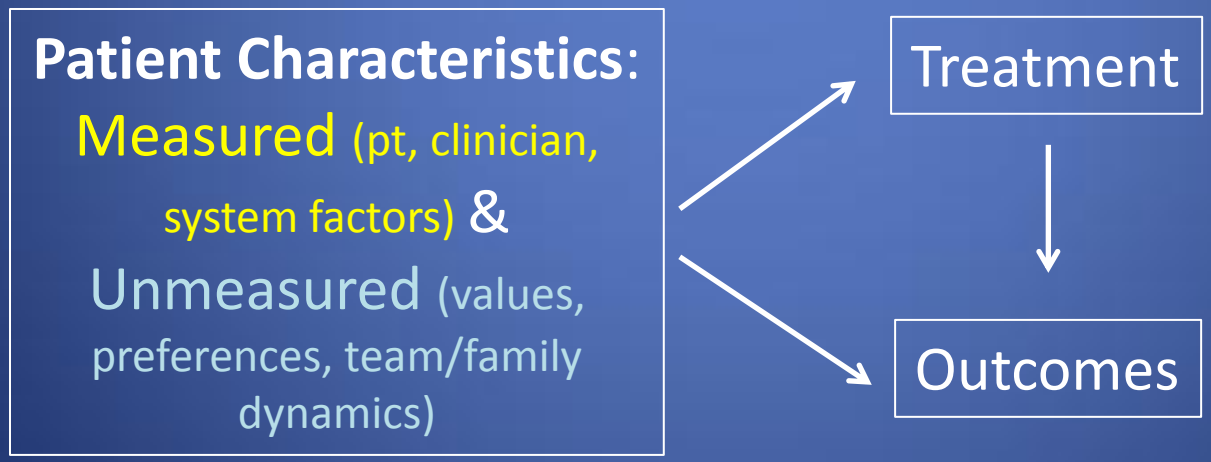
- Multivariable modeling vs. propensity scores to control for confounding
- Best practices in modeling
- What propensity scores can & cannot do
- Nuts & bolts of propensity score analysis
- Practice designing an analysis (variable selection, balancing/matching your sample)
- Q & A

Confounding due to Selection Bias in Observational Data

Patients not randomized to treatment

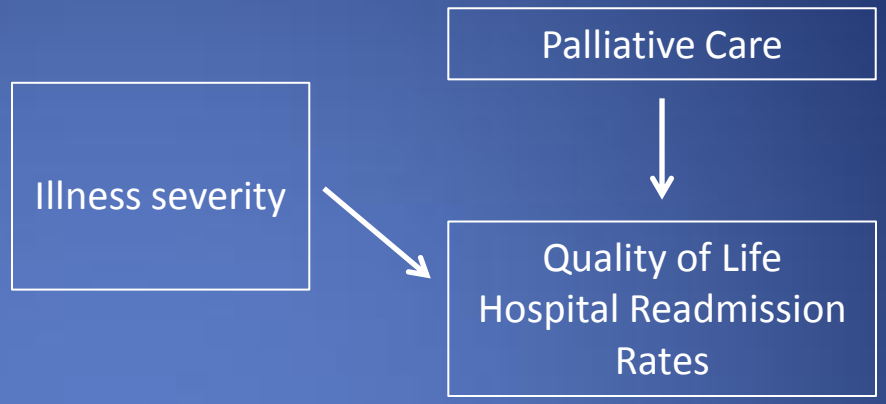
↓

Patient characteristics may be associated with both participation in treatment *and* outcome

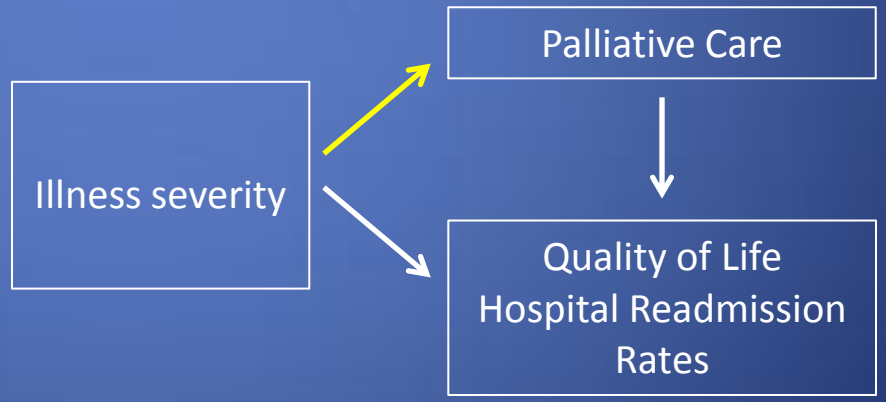


Tools to Address Confounding

- Multivariable models



- Matching
- Propensity scores
- Instrumental variables



Best Practices in Modeling: Variable Selection

- Do: Choose variables based on theory
- Do: Exclude variables that are highly correlated with each other

- Don't: Exclude variables via stepwise algorithms (unless building a parsimonious prediction model)
- Don't: Exclude variables based on p-values from bivariate tests

Step-Wise Modeling for Variable Selection

- NOT recommended for ***choosing*** variables
  - Choices should be theory-driven
- Good for ***evaluating*** which variables and interactions lead to the best predictive model

Tools to Address Confounding due to Selection Bias

- Matching
  - Compare treated and comparison individuals who are similar on one or two key covariates
- Propensity scores
  - Compare treated and comparison individuals who have similar "propensities" or likelihoods for receiving treatment, conditional on a set of several covariates
- Instrumental variables
  - Include an additional variable in your model (the "instrument") that is associated with treatment likelihood but not with outcome

Small Groups

Get in groups of 3 with at least one person who is designing an analysis using propensity scores

Take 5 minutes to:

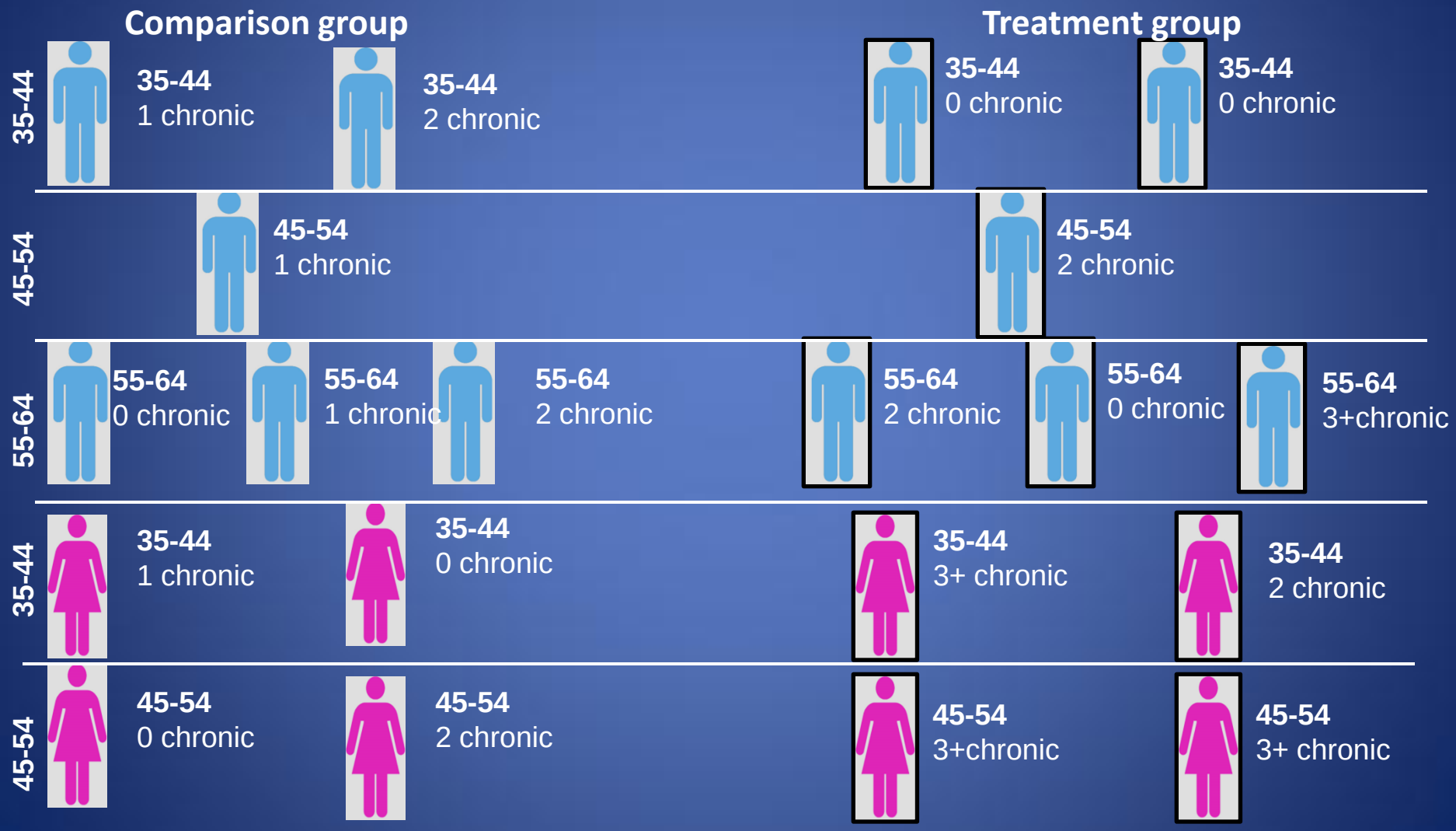
- Pitch a study to the group & articulate a research question
- Identify the treatment and outcome variables
- Report back to big group

Addressing Selection Bias with Matching

- Goal: Match patients so well that you could imagine that they were “randomly” assigned to each group
- For each patient in the treatment group, find at least one untreated patient from the comparison group who is identical or as similar as possible on all baseline characteristics
- By matching patients at the individual level, the treatment and comparison groups will be matched at the group level

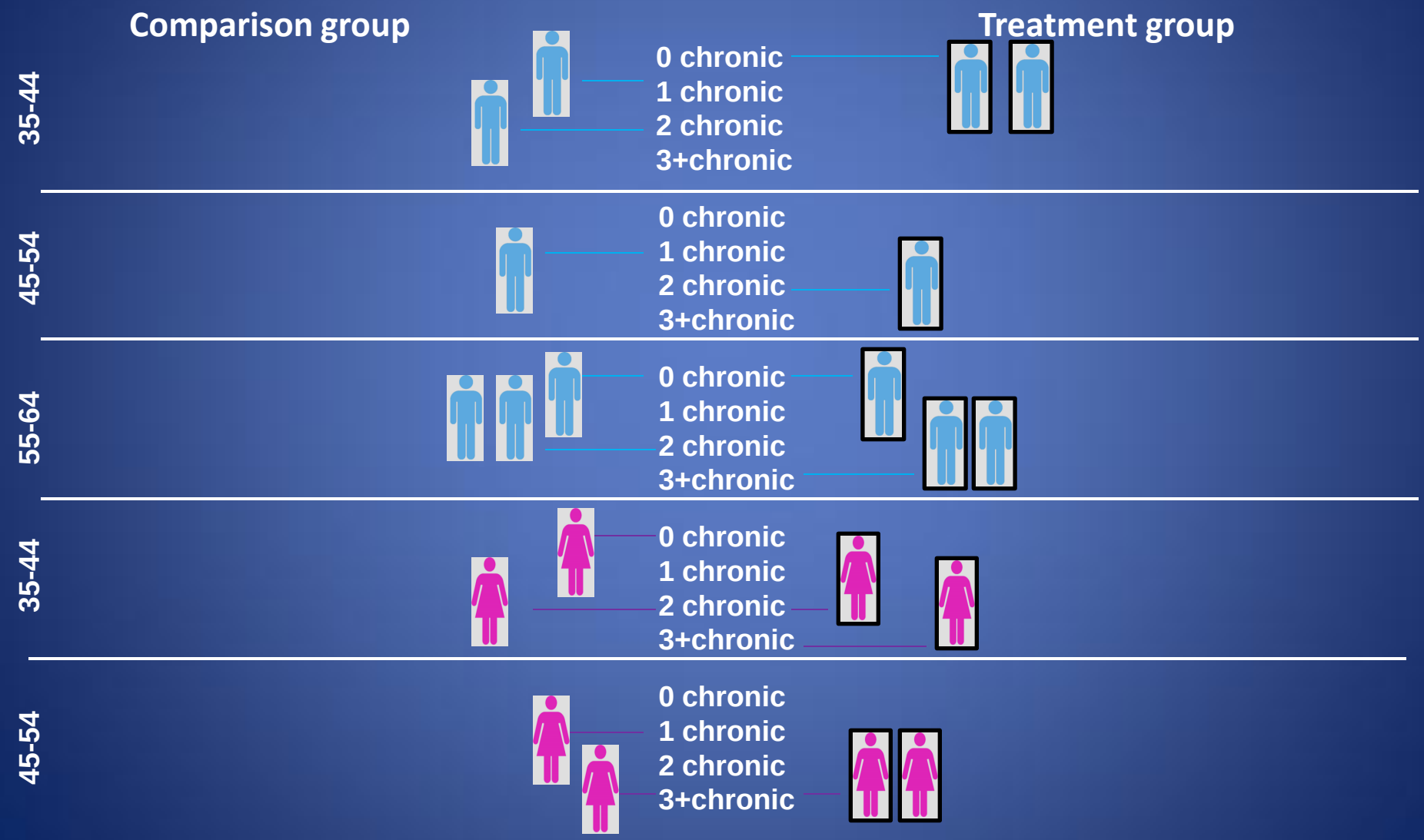
Matching on Specific Variables:

Match on gender and age



Matching on Specific Variables:

Gender, age, number of chronic conditions



Isn’t There an Easier Way?

Couldn’t we match on a
single composite score instead?

Propensity Score Matching

What ***IS*** a propensity score?

- Estimate of the likelihood that any given individual would be in the treatment group, given a set of measured characteristics
- Logistic regression with the treatment group (coded as 0/1) as the dependent variable
- Scores range from 0-1
- Cases matched on proximity of scores to each other

What propensity scores can & cannot do

- Help find matches from comparison group so that ***measured*** confounders can be equally distributed between treatment & comparison groups
- Helps improve precision of estimates of treatment effects
- Cannot account for ***unmeasured*** confounders
  - only control for observed variables and only to the extent that they are accurately measured
- Some residual confounding possible

Choosing Variables for Propensity Scores

- Include: Theoretically related to treatment & outcome
- Include: Available & easy/reliable to collect on everyone
- Include: Correlated with unmeasured confounders

- Do not include: Variables hypothesized to be associated with treatment but not with outcome
- Do not include: Variables that may be affected by the treatment
- Do not include: Variables that predict treatment status perfectly

Common Variables in PC

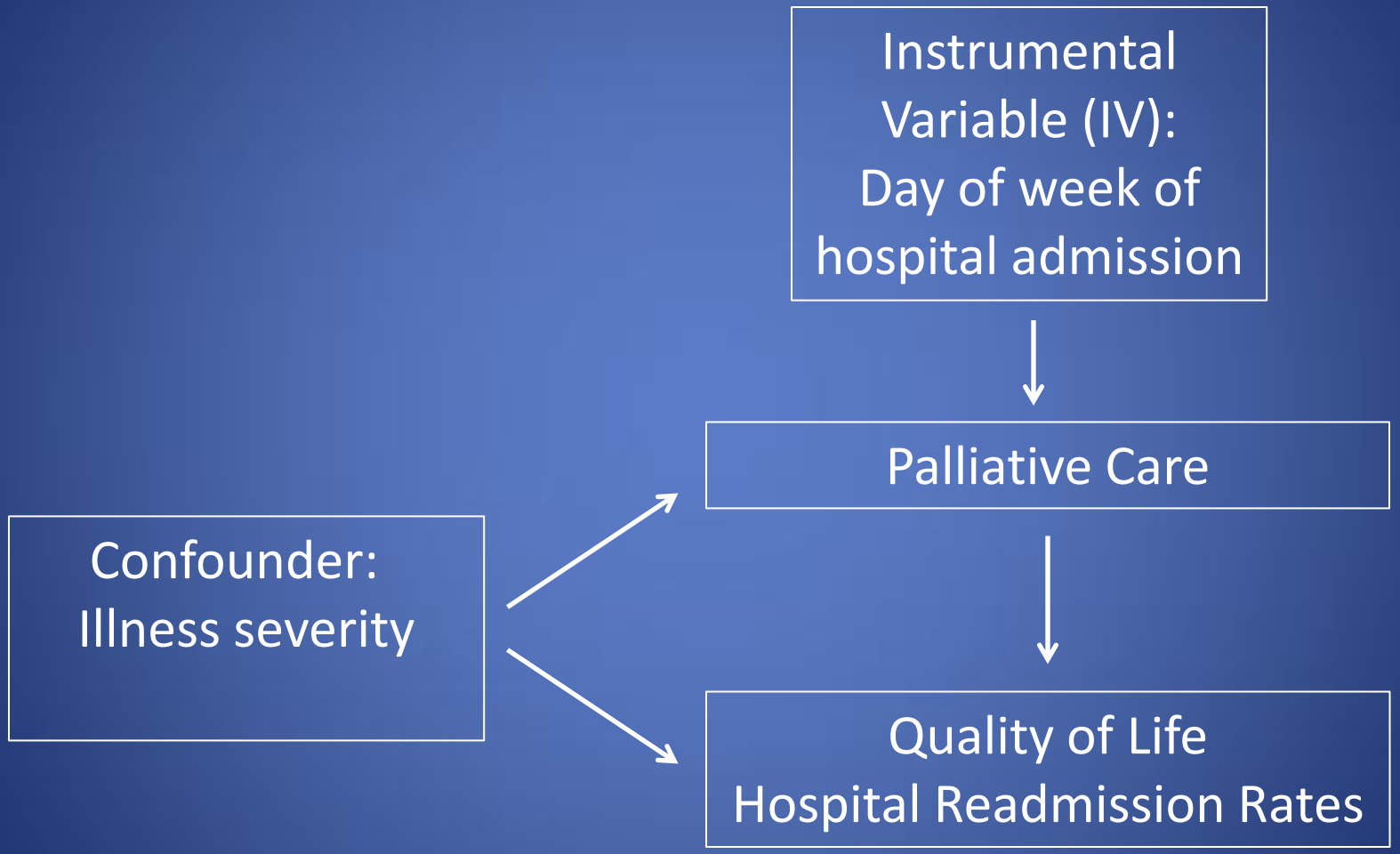
Patient variables

- Demographics (age, gender, ethnicity/race, marital status, insurance status, domicile [home v. LTC/institution])
- Illness-related factors (primary dx, comorbid conditions, severity of illness [APR-DRG])
- Prior utilization (ED visits, hospitalizations, outpt visits, home health/hospice enrollment, days in LTC)

Contextual variables

- Setting (urban/rural, hospice/SNF beds in community, for-profit status, geographic region/zip code, hospital site/type)
- Time (year of death, season of year)
- Clinician characteristics (yrs in practice, specialty, frequency of referral to PC/hospice)

Confounders vs Instrumental Variables



Small Groups

Take 5 minutes to:

- Identify potential confounders to include in your propensity score model
- Discuss theoretical justifications for your choices (i.e., specify how these are related to both treatment and outcome).

General Procedure

Step 1: Choose variables to include in propensity score

Step 2: Ensure that propensity score is balanced across treatment and comparison groups

Step 3: Ensure that covariates are balanced across treatment and comparison groups within blocks of the propensity score

Step 4: Choose a matching or weighting strategy

Step 5: Ensure that covariates are balanced across treatment and comparison groups in sample matched or weighted by propensity score

Calculating a propensity score is an iterative process. Steps 1-5 may be repeated several times.

Step 6: Proceed with analyses based on sample matched or weighted by propensity score

General Procedure

Step 1: Choose variables to include in propensity score

Step 2: Ensure that propensity score is balanced across treatment and comparison groups

Step 3: Ensure that covariates are balanced across treatment and comparison groups within blocks of the propensity score

Step 4: Choose a matching or weighting strategy

Step 5: Ensure that covariates are balanced across treatment and comparison groups in sample matched or weighted by propensity score

↓

Step 6: Proceed with analyses based on sample matched or weighted by propensity score

List potential confounders

Evaluate feasibility of including these confounders

Calculate propensity score with logit or probit regression

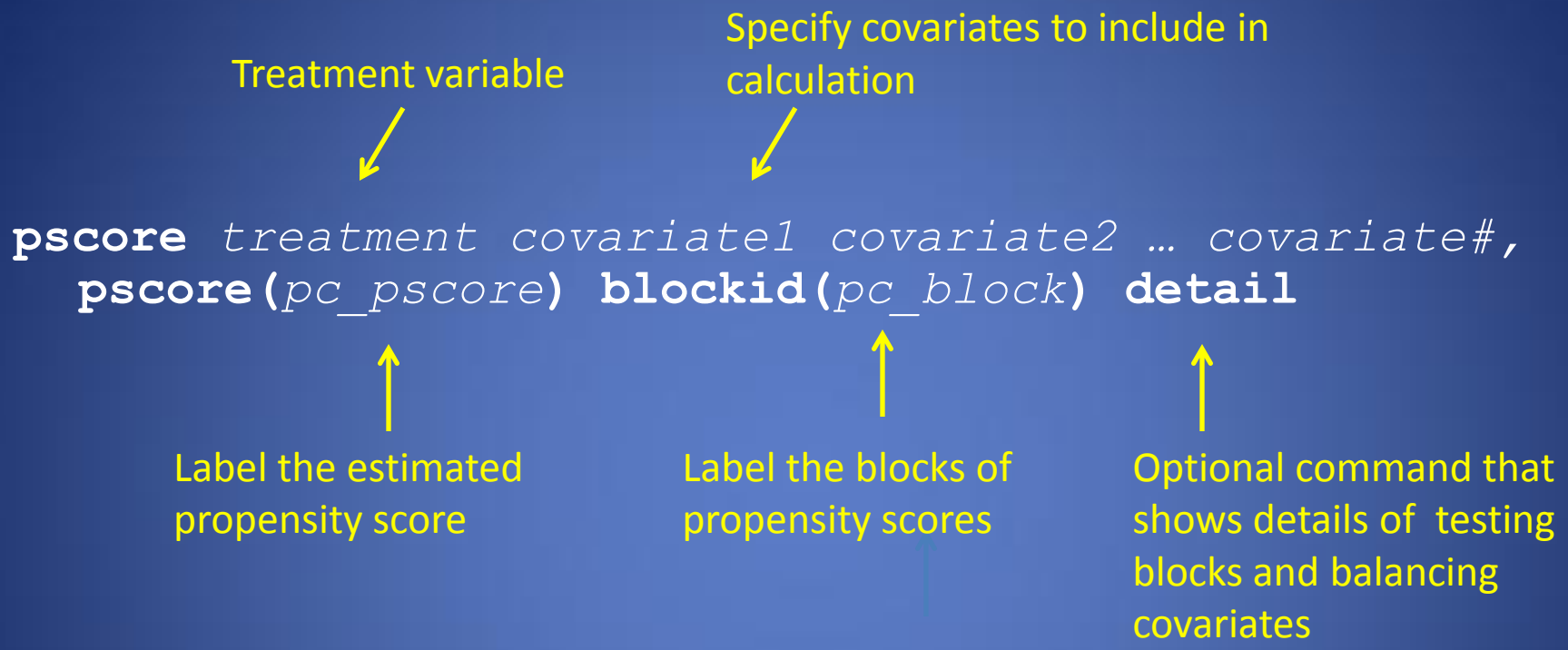
Working Example

- 2008 Healthcare Cost and Utilization Project (H-CUP) Nationwide Inpatient Sample
  - Discharge data for hospitalizations throughout the US
- 12,686 patients with metastatic cancer who died during the hospitalization
- Treatment: Palliative Care Consultation
- Outcome: Average total charges per day
- Contrived example – Please do not draw any conclusions from data presented here!

How Many Variables to Include in Propensity Score?

- Tradeoff between
  - Bias: Distance of estimated treatment effect from true effect
  - Efficiency/Variance: Precision of estimated treatment effect

Stata Code to Calculate Propensity Score



*pscore is not part of Stata's built-in commands. Type "findit pscore" in Stata's command line and follow link in pop-up window to install (st0026, Becker & Ichino)

Working Example: Propensity Score

Treatment variable

Specify covariates to include in calculation

```
pscore palliative ///
age35to44 age45to54 age55to64 age65to74 age75andup female race_Black ///
race_Hispanic race_other race_missing pay_Medicare pay_Medicaid  pay_outofpocket ///
pay_otherormiss NCHS2 NCHS3 NCHS4 NCHS5 NCHS6 ///
lung_ca liver_ca pancreas_ca leukemia stom_ca ///
septicemia pneumonia respfailure renalfailure, ///
pscore(pc_pscore) blockid(pc_block) detail
```

Label the estimated propensity score

Label the blocks of propensity scores

Beginning of output from pscore command

```
**************************************************
Algorithm to estimate the propensity score
**************************************************

The treatment is palliative
```

palliative	Freq.	Percent	Cum.
0	10,861	86.12	86.12
1	1,751	13.88	100.00
Total	12,612	100.00	

Frequency of treatment in sample

Probit regression to calculate probability of treatment given the covariates

```
Estimation of the propensity score

Iteration 0:   log likelihood = -5080.6799
Iteration 1:   log likelihood = -4839.3575
Iteration 2:   log likelihood = -4836.4592
Iteration 3:   log likelihood = -4836.4577

Probit regression                                 Number of obs   =      12612
                                                  LR chi2(28)     =     488.44
                                                  Prob > chi2     =     0.0000
Log likelihood = -4836.4577                       Pseudo R2       =     0.0481
```

palliative	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
age35to44	-.1966959	.1062481	-1.85	0.064	-.4049385	.0115466
age45to54	-.1114731	.0834791	-1.34	0.182	-.2750892	.0521429
age55to64	-.1496492	.0780981	-1.92	0.055	-.3027187	.0034203
age65to74	-.0636428	.077005	-0.83	0.409	-.2145698	.0872842
age75andup	-.078232	.076199	-1.03	0.305	-.2275793	.0711153
female	.055816	.0288977	1.93	0.053	-.0008225	.1124545
race_Black	-.1218066	.0495814	-2.46	0.014	-.2189843	-.0246288
race_Hispa~c	-.0708756	.0667917	-1.06	0.289	-.201785	.0600337
race_other	.1128539	.0595336	1.90	0.058	-.0038297	.2295376
race_missing	-.2662183	.0414828	-6.42	0.000	-.3475232	-.1849135
pay_Medicare	-.193896	.0413946	-4.68	0.000	-.2750279	-.1127641
pay_Medicaid	-.0507248	.056659	-0.90	0.371	-.1617744	.0603247
pay_outofp~t	.0283215	.0911258	0.31	0.756	-.1502818	.2069248
pay_othero~s	.4867789	.0666438	7.30	0.000	.3561594	.6173984
NCHS2	-.072554	.0407098	-1.78	0.075	-.1523436	.0072357
NCHS3	.1675837	.0417866	4.01	0.000	.0856836	.2494839
NCHS4	.1728791	.0562403	3.07	0.002	.0626502	.283108
NCHS5	.0499702	.0500913	1.00	0.318	-.048207	.1481474
NCHS6	.0169394	.0596206	0.28	0.776	-.0999148	.1337936

Working Example: Propensity Score

The -pscore- command provides you with a single score on which to match your treatment and comparison groups

Description of the estimated propensity score

Estimated propensity score

	Percentiles	Smallest		
1%	.0330385	.0158288		
5%	.0484464	.01833		
10%	.0607721	.0188756	Obs	12612
25%	.0876162	.0194706	Sum of Wgt.	12612
50%	.1289967		Mean	.1388897
		Largest	Std. Dev.	.0695261
75%	.1775761	.5015948		
90%	.2239398	.50889	Variance	.0048339
95%	.2577822	.5122628	Skewness	1.220613
99%	.3864837	.5437794	Kurtosis	5.721294

General Procedure

Step 1: Choose variables to include in propensity score

Step 2: Ensure that propensity score is balanced across treatment and comparison groups

Step 3: Ensure that covariates are balanced across treatment and comparison groups within blocks of the propensity score

Step 4: Choose a matching or weighting strategy

Step 5: Ensure that covariates are balanced across treatment and comparison groups in sample matched or weighted by propensity score

↓

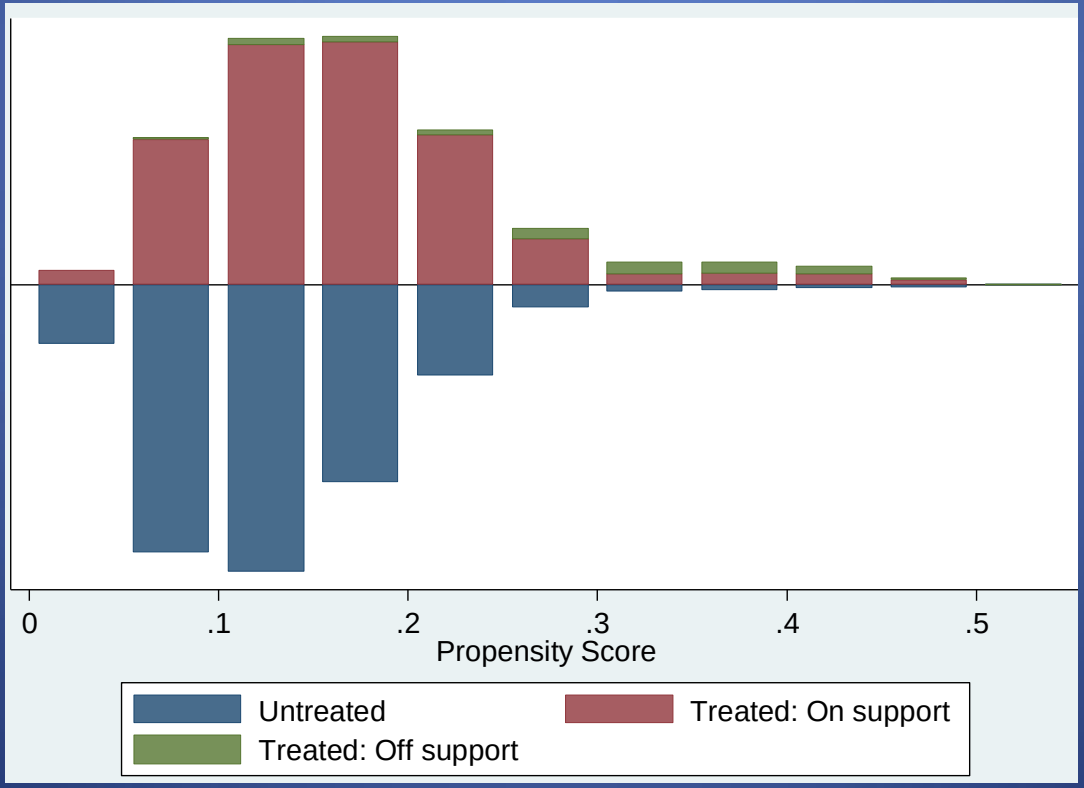
Step 6: Proceed with analyses based on sample matched or weighted by propensity score

Check range of common support
Check balance of propensity score

Check Range of Common Support

Extent to which distributions of propensity scores in treatment and comparison groups overlap

psgraph, treated(*treatment*)**pscore**(*pc_pscore*)



Check Balance of Propensity Score Across Groups

- Does the propensity score have a similar distribution across treatment and comparison groups?
- Estimate distribution by splitting sample by quintiles or other strata of propensity score
- Test whether mean of propensity score is equal in treatment and comparison groups within each quintile
- If not equal, split one or more quintiles into smaller blocks and compare means

Stata Output for Propensity Score Balance

(Continuation of -pscore- output, with "detail" option specified)

Distribution of treated and controls across blocks

Blocks of the pscore for treatment palliative	palliative 0	1	Total
1	9,387	1,222	10,609
2	1,425	479	1,904
3	49	50	99
Total	10,861	1,751	12,612

Stata stratifies your data based on the propensity score

Tests whether mean propensity score is equal for treated and controls within each block

Test that the mean propensity score is not different for treated and controls

Test in block 1

Observations in block 1
obs: 10609, control: 9387, treated: 1222

Test for block 1

Two-sample t test with equal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
0	9387	.114782	.0004566	.0442347	.113887	.1156769
1	1222	.131742	.0011855	.0414404	.1294162	.1340677
combined	10609	.1167355	.0004296	.0442526	.1158933	.1175777
diff		-.01696	.0013357		-.0195783	-.0143417

diff = mean(0) - mean(1) t = -12.6971
Ho: diff = 0 degrees of freedom = 10607

Ha: diff < 0 Ha: diff != 0 Ha: diff > 0
Pr(T < t) = 0.0000 Pr(|T| > |t|) = 0.0000 Pr(T > t) = 1.0000

Groups are significantly different

Stata Output for Propensity Score Balance

(Continuation of -pscore- output, with "detail" option specified)

```
The mean propensity score is different for
treated and controls in block 1
Split the block 1 and retest

Check that blocks have shifted
```

Blocks of the pscore for treatment palliative	palliative 0	1	Total
1	9,387	1,222	10,609
3	1,425	479	1,904
4	49	50	99
Total	10,861	1,751	12,612

```
Test in block 1

Observations in block 1
 obs: 4041,  control: 3740,  treated: 301

Test for block 1

Two-sample t test with equal variances
```

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
0	3740	.069735	.000312	.0190816	.0691233	.0703467
1	301	.0757012	.0009512	.0165026	.0738293	.0775731
combined	4041	.0701794	.0002983	.0189646	.0695945	.0707643
diff		-.0059662	.0011325		-.0081865	-.0037459

```
    diff = mean(0) - mean(1)                                  t =  -5.2682
Ho: diff = 0                     degrees of freedom =     4039

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 1.0000
```

Stata splits Block 1 into two blocks and tests whether the propensity score is different for treated and controls in the new Block 1

Groups are still significantly different

Stata will automatically continue to split blocks and perform t-tests until it calculates the smallest # of blocks where the propensity score is equivalent across treated and controls in each block

General Procedure

Step 1: Choose variables to include in propensity score

Step 2: Ensure that propensity score is balanced across treatment and comparison groups

Step 3: Ensure that covariates are balanced across treatment and comparison groups within blocks of the propensity score

Step 4: Choose a matching or weighting strategy

Step 5: Ensure that covariates are balanced across treatment and comparison groups in sample matched or weighted by propensity score

Step 6: Proceed with analyses based on sample matched or weighted by propensity score

Check Balance of Covariates within Blocks of the Propensity Score

- Ideally, for each unique value of the propensity score, the distribution of X (composite of all covariates) is the same for the treatment and comparison groups

- This is practically impossible, so we check the balance of each observed covariate within blocks of the propensity score

Stata Output for Propensity Score Balance

(Continuation of -pscore- output, without "detail" option)

```
************************************************************
Step 1: Identification of the optimal number of blocks
Use option detail if you want more detailed output
************************************************************

The final number of blocks is 9

This number of blocks ensures that the mean propensity score
is not different for treated and controls in each blocks


************************************************************
Step 2: Test of balancing property of the propensity score
Use option detail if you want more detailed output
************************************************************

Variable NCHS6 is not balanced in block 1

Variable respfailure is not balanced in block 2

Variable NCHS6 is not balanced in block 3

Variable NCHS4 is not balanced in block 8

Variable age55to64 is not balanced in block 9

The balancing property is not satisfied

Try a different specification of the propensity score
```

Inferior of block of pscore	palliative 0	1	Total
0	670	27	697
.05	1,437	104	1,541
.075	1,633	170	1,803
.1	3,306	460	3,766
.15	1,283	222	1,505
.175	1,058	239	1,297
.2	1,285	389	1,674
.3	140	90	230
.4	49	50	99
Total	10,861	1,751	12,612

← Step 2 is completed (propensity score balanced across groups)

← Stata uses t-tests to determine whether each covariate is balanced within each block

← You will usually get an error message

Stata Output for Propensity Score Balance

(Continuation of -pscore- output, with "detail" option)

Testing the balancing property for variable respfailure in block 2

Two-sample t test with equal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
0	1437	.7974948	.0106049	.4020066	.7766921	.8182975
1	104	.6826923	.04586	.4676822	.5917398	.7736448
combined	1541	.7897469	.0103838	.4076206	.7693791	.8101147
diff		.1148025	.0413015		.0337893	.1958157

diff = mean(0) - mean(1) t = 2.7796
Ho: diff = 0 degrees of freedom = 1539

Ha: diff < 0 Ha: diff != 0 Ha: diff > 0
Pr(T < t) = 0.9972 Pr(|T| > |t|) = 0.0055 Pr(T > t) = 0.0028

Variable respfailure is not balanced in block 2

Groups are significantly different on "respfailure" in Block 2

Testing the balancing property for variable renalfailure in block 2

Two-sample t test with equal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
0	1437	.6256089	.0127714	.4841338	.6005564	.6506614
1	104	.6634615	.0465593	.4748137	.5711221	.755801
combined	1541	.6281635	.0123155	.4834519	.6040066	.6523204
diff		-.0378526	.0490983		-.1341593	.0584541

diff = mean(0) - mean(1) t = -0.7710
Ho: diff = 0 degrees of freedom = 1539

Ha: diff < 0 Ha: diff != 0 Ha: diff > 0
Pr(T < t) = 0.2204 Pr(|T| > |t|) = 0.4409 Pr(T > t) = 0.7796

Variable renalfailure is balanced in block 2

Groups are not significantly different on "renalfailure" in Block 2

Improving the Balance of the Propensity Score

- Some imbalance between the groups is usually expected
- Focus on balance of covariates that are more theoretically important

- Consider interactions/correlations between covariates
- Drop 1 or 2 covariates that are less important
- Re-categorize variables
- Include higher order terms or splines of variables

Small Groups

Take 5 minutes to:

- Discuss procedures for deciding which variables you might drop if your pscore doesn’t balance

Improving the Balance of the Propensity Score

1. Drop variables created by Stata for initial run of **pscore** command

```
drop pc_pscore pc_block
```

2. Change covariates
3. Re-run –**pscore**– command

```
pscore palliative ///
age35to44 age45to54 age55to64 age65to74 age75andup female race_Black ///
race_Hispanic race_other race_missing pay_Medicare pay_Medicaid  pay_outofpocket ///
pay_otherormiss NCHS2 NCHS3 NCHS4 NCHS5_6 ///
lung_ca liver_ca pancreas_ca leukemia stom_ca ///
septicemia pneumonia respfailure renalfailure, ///
pscore(pc_pscore) blockid(pc_block)
```

```
***********************************************************
Step 1: Identification of the optimal number of blocks
Use option detail if you want more detailed output
***********************************************************

The final number of blocks is 9

This number of blocks ensures that the mean propensity score
is not different for treated and controls in each blocks


***********************************************************
Step 2: Test of balancing property of the propensity score
Use option detail if you want more detailed output
***********************************************************

Variable respfailure is not balanced in block 2

Variable race_Hispanic is not balanced in block 7

Variable NCHS4 is not balanced in block 8

Variable age55to64 is not balanced in block 9

Variable age65to74 is not balanced in block 9

The balancing property is not satisfied

Try a different specification of the propensity score
```

Inferior of block of pscore	palliative 0	1	Total
0	673	26	699
.05	1,430	105	1,535
.075	1,638	168	1,806
.1	3,321	463	3,784
.15	1,265	222	1,487
.175	1,042	235	1,277
.2	1,303	393	1,696
.3	138	90	228
.4	51	49	100
Total	10,861	1,751	12,612

1st try: 4 variables unbalanced in 5 blocks

2nd try: 5 variables unbalanced in 4 blocks

← You will usually get an error message

```
**********************************************************
Step 1: Identification of the optimal number of blocks
Use option detail if you want more detailed output
**********************************************************

The final number of blocks is 9

This number of blocks ensures that the mean propensity score
is not different for treated and controls in each blocks


**********************************************************
Step 2: Test of balancing property of the propensity score
Use option detail if you want more detailed output
**********************************************************

Variable NCHS2 is not balanced in block 7

Variable NCHS4 is not balanced in block 8

The balancing property is not satisfied

Try a different specification of the propensity score
```

Inferior of block of pscore	palliative 0	1	Total
0	659	27	686
.05	1,417	97	1,514
.075	1,652	179	1,831
.1	3,326	460	3,786
.15	2,298	464	2,762
.2	1,061	291	1,352
.25	272	103	375
.3	124	83	207
.4	52	47	99
Total	10,861	1,751	12,612

```
**********************************************
End of the algorithm to estimate the pscore
**********************************************
```

1st try: 4 variables unbalanced in 5 block

2nd try: 5 variables unbalanced in 4 blocks

3rd try: 2 variables unbalanced in 2 blocks

← You will usually get an error message

Some imbalance between groups is usually expected

Assess Balance with Standardized Differences

- Account for means and variances
- Not sensitive to sample size

$$d = \frac{(\bar{x}_{\text{treatment}} - \bar{x}_{\text{control}})}{\sqrt{\frac{s^2_{\text{treatment}} + s^2_{\text{control}}}{2}}}$$

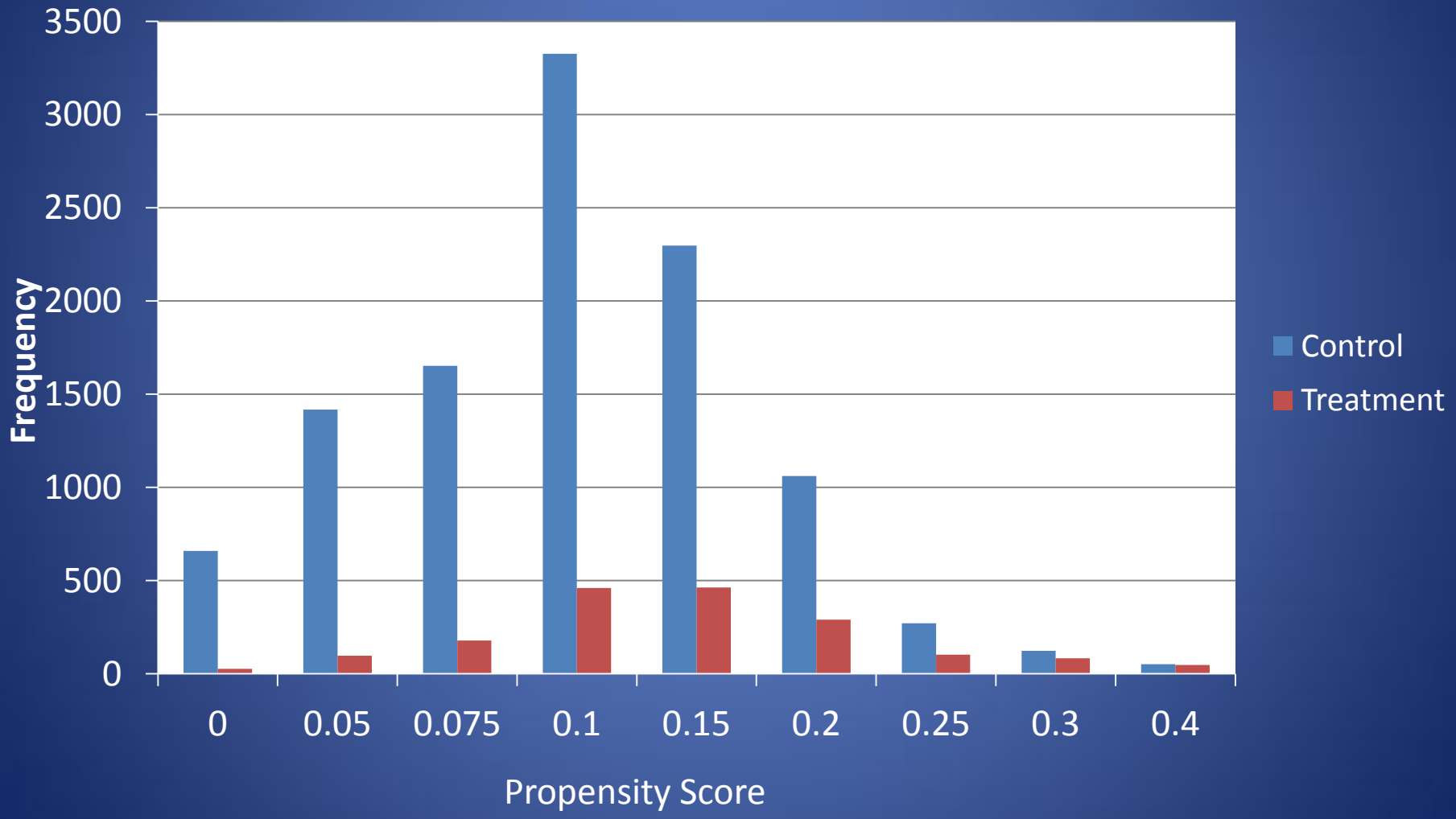
Continuous variables

$$d = \frac{(\hat{p}_{\text{treatment}} - \hat{p}_{\text{control}})}{\sqrt{\frac{\hat{p}_{\text{treatment}}(1 - \hat{p}_{\text{treatment}}) + \hat{p}_{\text{control}}(1 - \hat{p}_{\text{control}})}{2}}}$$

Dichotomous variables

Equations from Austin 2009. Statistics in Medicine 28: 3083-3107

Balanced Propensity Score



Balance of Covariates: Caution

- Propensity scores only balance measured confounders
- Balance in measured variables does not indicate balance in unmeasured variables
- Unmeasured confounders will bias treatment effect estimates

Balance of Covariates: Caution

- Do not use c-statistics, area under the curve, or any other model fit statistics to measure propensity score performance
  - They do not measure reduction in confounding

General Procedure

Step 1: Choose variables to include in propensity score

Step 2: Ensure that propensity score is balanced across treatment and comparison groups

Step 3: Ensure that covariates are balanced across treatment and comparison groups within blocks of the propensity score

Step 4: Choose a matching or weighting strategy

Step 5: Ensure that covariates are balanced across treatment and comparison groups in sample matched or weighted by propensity score

Step 6: Proceed with analyses based on sample matched or weighted by propensity score

Matching and Weighting Strategies

Quality  **Quantity**

Nearest Neighbor

Radius Matching

Kernel Weighting

Inverse Probability of Treatment Weighting

No universal “best” strategy

Choices When Matching Sample by Propensity Score

- How close of a match is acceptable?
- Should every treated individual have one or many matches in the comparison group?
- Should treated individuals be matched with or without replacement?
- Should matching be greedy or optimal?

Which Strategy to Choose?

- No best method
- Without examining outcome, evaluate covariate balance in several strategies (our next step – Step 5)
- Choose the method that has the best balance and still meets the analytic goal

Matching Strategy: Nearest Neighbor

- List the treated patients
- Identify comparison patient with closest propensity score
- Continue until all treated are matched; delete unused comparison patients

- Advantage
  - All treated patients are included in the new sample
- Disadvantage
  - Lose information from unmatched individuals in comparison group (more variance)
  - For some treated patients, the nearest match in the comparison group may have a very different propensity score (increases bias)

Matching Strategy: Radius Matching

- Define a “caliper” or maximum permissible difference that defines a match within a range of the propensity score
- .2 * standard deviation of logit of propensity score is often used

```
gen logitpscore = ln(mypscore/(1-mypscore))
sum logitpscore
```

- Individuals from both treatment and comparison groups are dropped from sample if no within-caliper match is found
- Advantage: Improving comparability of groups → less bias
- Disadvantage: Losing information from some observations → more variance

Matching vs Weighting

- Matching strategies reduce bias at the expense of sample size, increasing variance of treatment effect estimates

- Weighting allows you to keep the bulk of your sample while reducing bias by giving more weight to individuals with closer propensity scores

Weighting Strategy: Kernel Weighting

- Each treated individual is assigned a weight of 1
- For each treated individual, a composite of information from comparison individuals within a certain bandwidth is used
- Comparison individuals weighted by distance of propensity score from treated individual's propensity score (higher weights for better matches)
- Weight assigned by a nonparametric kernel function
- Leads to average treatment effect on *treated*

Weighting Strategy: Inverse Probability of Treatment Weighting (IPTW)

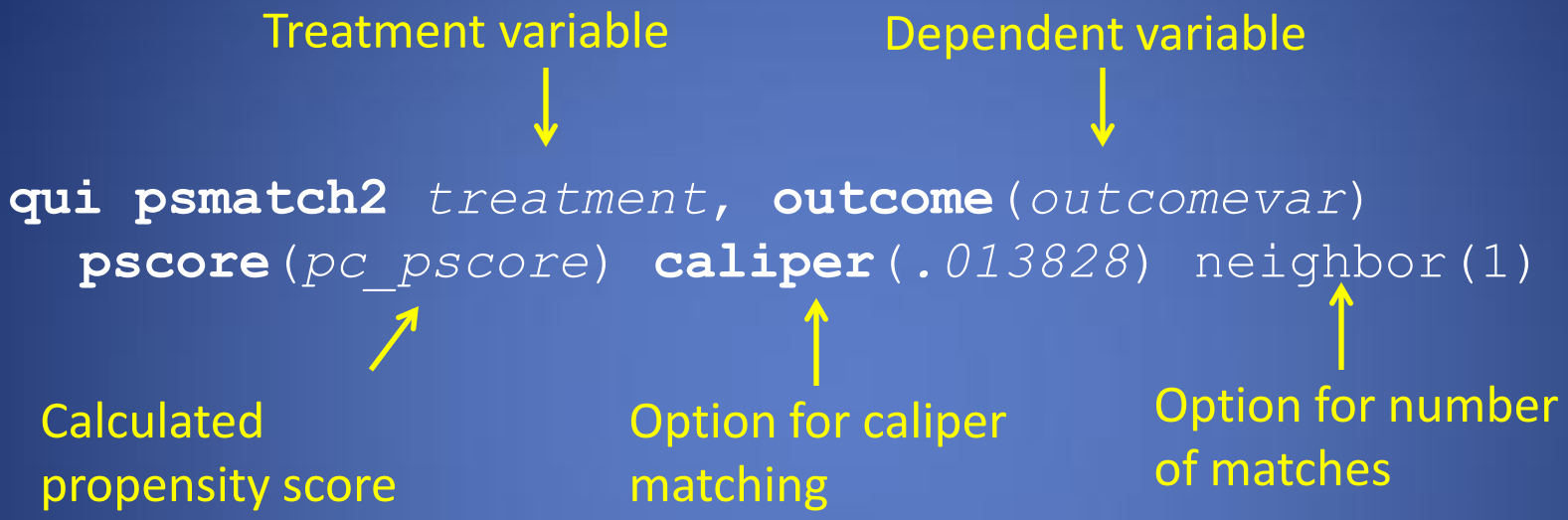
- Each treated individual receives a weight of 1/propensity score
- Each comparison individual receives a weight of 1/ (1-propensity score)
- Weights should be normalized to one
- Leads to average treatment effect for *sample*

Matching Example

- Specific Aim: Determine the impact of a palliative care consult on average hospital expenditures

- Matching Strategy:
  - 1:1 matching
  - With replacement
  - Caliper = 0.2* standard deviation of logit of propensity score

Stata Code to Match Sample on Propensity Score



Stata Code to Weight Sample on Propensity Score

Kernel Weight:

```
qui psmatch2 treatment, kernel outcome(outcomevar)
   pscore(pc_pscore)
```

IPTW:

```
qui dr outcomevar treatment covariate 1… covariate
   #, genvars
```

Creates variable "iptwt" that stores the weights calculated by this command

```
egen sumofweights = total(iptwt)
gen norm_weights  = iptwt/sumofweights
```

Normalize weights to sum to one

General Procedure

Step 1: Choose variables to include in propensity score

Step 2: Ensure that propensity score is balanced across treatment and comparison groups

Step 3: Ensure that covariates are balanced across treatment and comparison groups within blocks of the propensity score

Step 4: Choose a matching or weighting strategy

Step 5: Ensure that covariates are balanced across treatment and comparison groups in sample matched or weighted by propensity score

Perform multiple checks

↓

Step 6: Proceed with analyses based on sample matched or weighted by propensity score

Several Ways to Evaluate Balance in Sample Matched or Weighted by Propensity Score

- Standardized differences
- Graphs
  - Quantile-quantile plots
  - Plots of covariates in treated and comparison groups
- Ratios of variance

Evaluate standardized differences in matched sample

Immediately following –psmatch2–, run:

Treatment variable

```
pstest covariate1..covariate#, treated(treatment) both
```

show balance before and after match
(default is to only show after)

```
pstest age female race_Black ///
race_Hispanic race_other race_missing pay_Medicare pay_Medicaid  pay_outofpocket ///
pay_otherormiss NCHS2 NCHS3 NCHS4 NCHS5_6 ///
lung_ca liver_ca pancreas_ca leukemia stom_ca ///
septicemia pneumonia respfailure renalfailure, treated(palliative) both
```

Variable		Mean Treated	Mean Control	%bias	%reduct |bias|	t-test t	t-test p>|t|
age	Unmatched	67.007	67.572	-4.2		-1.66	0.097
	Matched	67.074	67.025	0.4	91.2	0.10	0.917
female	Unmatched	.49343	.46828	5.0		1.96	0.050
	Matched	.49365	.48602	1.5	69.7	0.43	0.669

Output from -pstest-

Summary of the distribution of the abs(bias)

BEFORE MATCHING

	Percentiles	Smallest		
1%	.4647914	.4647914		
5%	1.525013	1.525013		
10%	1.76495	1.76495	Obs	23
25%	2.446086	2.21497	Sum of Wgt.	23
50%	5.245142		Mean	10.00157
		Largest	Std. Dev.	9.429134
75%	17.56547	24.66997		
90%	25.51863	25.51863	Variance	88.90857
95%	25.5416	25.5416	Skewness	.9390208
99%	31.01482	31.01482	Kurtosis	2.455434

AFTER MATCHING

	Percentiles	Smallest		
1%	0	0		
5%	0	0		
10%	.22809	.22809	Obs	23
25%	.6399435	.370463	Sum of Wgt.	23
50%	1.423716		Mean	1.575237
		Largest	Std. Dev.	1.141444
75%	2.5283	2.865676		
90%	3.245421	3.245421	Variance	1.302894
95%	3.478944	3.478944	Skewness	.3601008
99%	3.72565	3.72565	Kurtosis	1.953917

Sample	Pseudo R2	LR chi2	p>chi2	MeanBias	MedBias
Raw	0.048	482.73	0.000	10.0	5.2
Matched	0.002	7.12	0.999	1.6	1.4

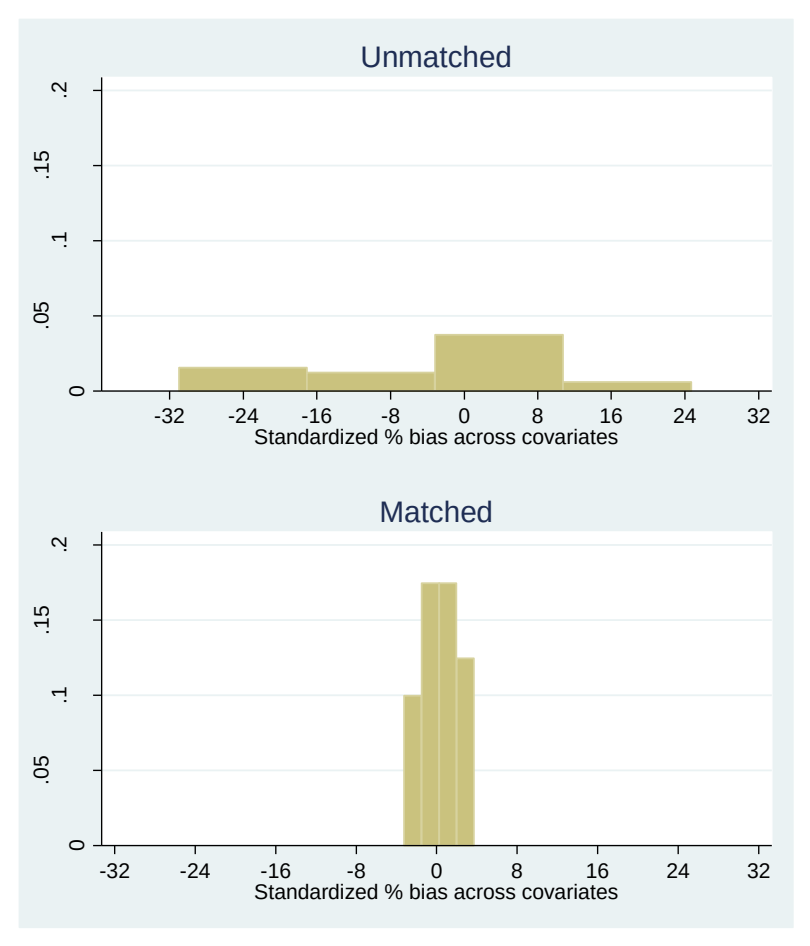
Summary of covariate imbalance

Summary of mean and median bias before and after matching

Visual inspection of standardized differences

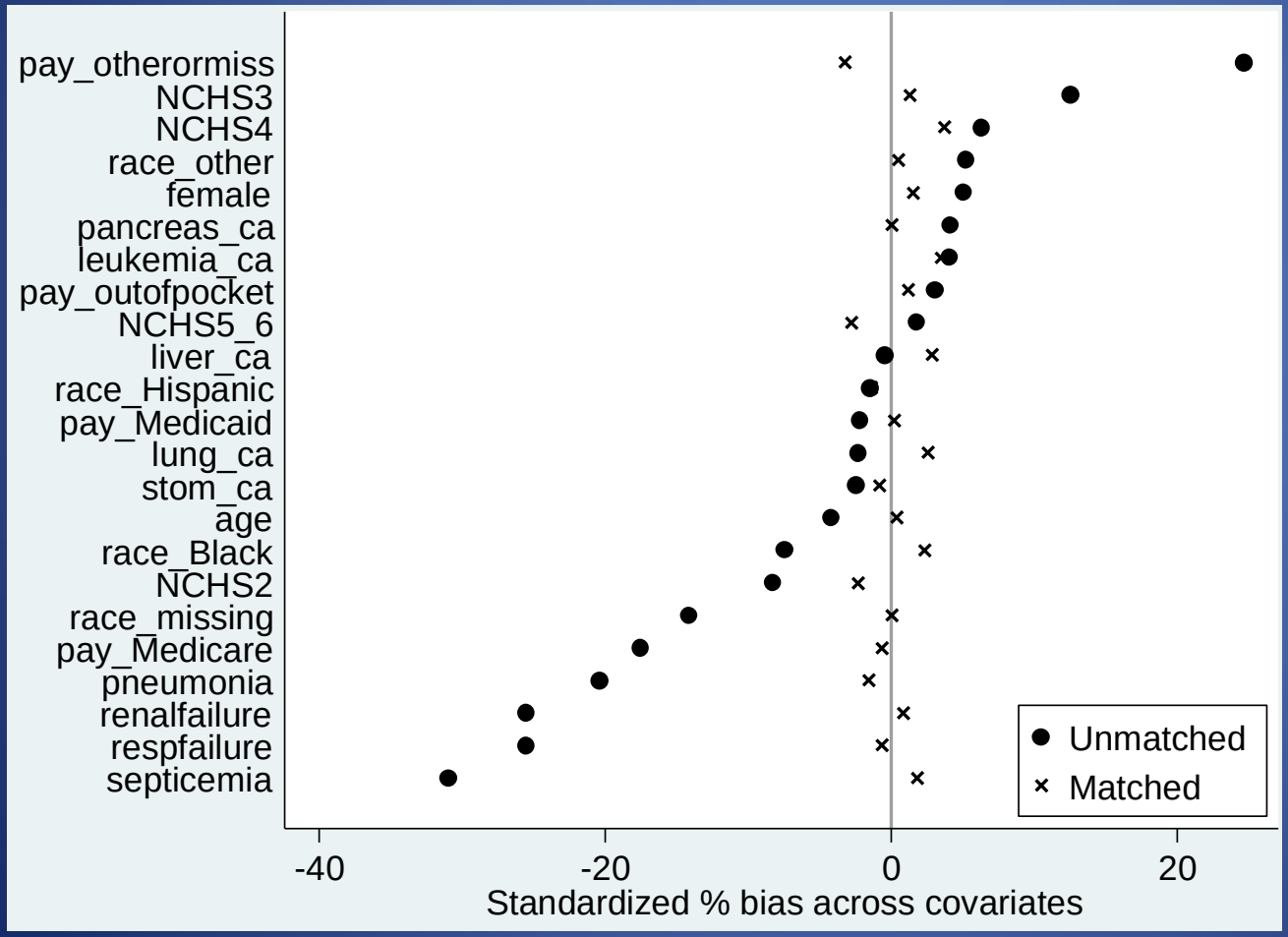
pstest *covariate1..covariate#*, **treated**(*treatment*) **both** **hist**

Optional command to get histogram of covariate balance



Visual inspection of standardized differences

pstest *covariate1..covariate#*, **treated**(*treatment*) **both** **graph**



Optional command to get dot graph of covariate balance

Evaluate standardized differences in weighted sample

- Kernel: Uses **`-psmatch2-`** so can use same procedure as for matched samples

- IPTW:

  After running **`-dr-`** and normalizing weight variable, run:

```
pbalchk treatment covariate1… covariate#,
wt(norm_weights)
```

Name of weight variable created earlier

Output from **-pbalchk-**

	Mean in treated	Mean in Untreated	Standardised diff.
age	67.47	67.52	-0.003
female	0.48	0.47	0.008
race_Black	0.11	0.11	0.011
race_Hispa~c	0.05	0.05	-0.006
race_other	0.06	0.06	-0.002
race_missing	0.17	0.18	-0.007
pay_Medicare	0.58	0.58	0.001
pay_Medicaid	0.09	0.09	0.005
pay_outofp~t	0.03	0.02	0.014
pay_othero~s	0.04	0.04	-0.009
NCHS2	0.21	0.23	-0.033
NCHS3	0.17	0.17	-0.006
NCHS4	0.08	0.08	0.021
NCHS5_6	0.20	0.18	0.047
lung_ca	0.30	0.29	0.016
liver_ca	0.02	0.02	0.020
pancreas_ca	0.05	0.05	0.001
leukemia_ca	0.01	0.01	-0.009
stom_ca	0.03	0.02	0.025
septicemia	0.34	0.34	-0.015
pneumonia	0.34	0.32	0.023
respfailure	0.45	0.46	-0.013
renalfailure	0.32	0.32	-0.009

Same information as %bias in -pstest- output, but not expressed as a percentage

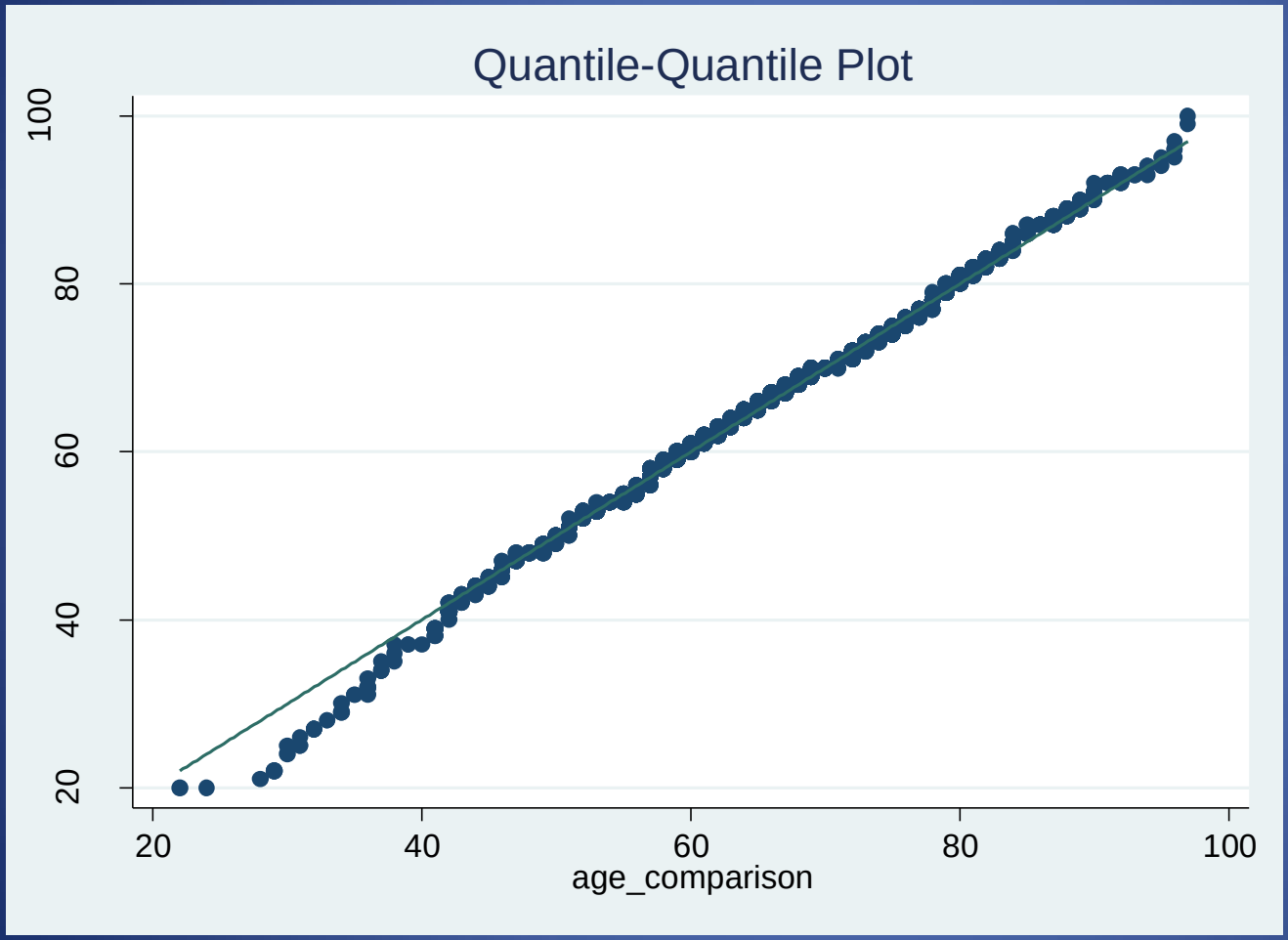
Quantile-Quantile Plots

- For unweighted continuous variables
- Plot covariate in treated group against covariate in comparison group (will need to create 2 new variables)

```
qqplot covariate_treated
  covariate_comparison if _weight==1
```

Restricts plot to those matched in a one-to-one nearest neighbor match

Output from `-qqplot-`



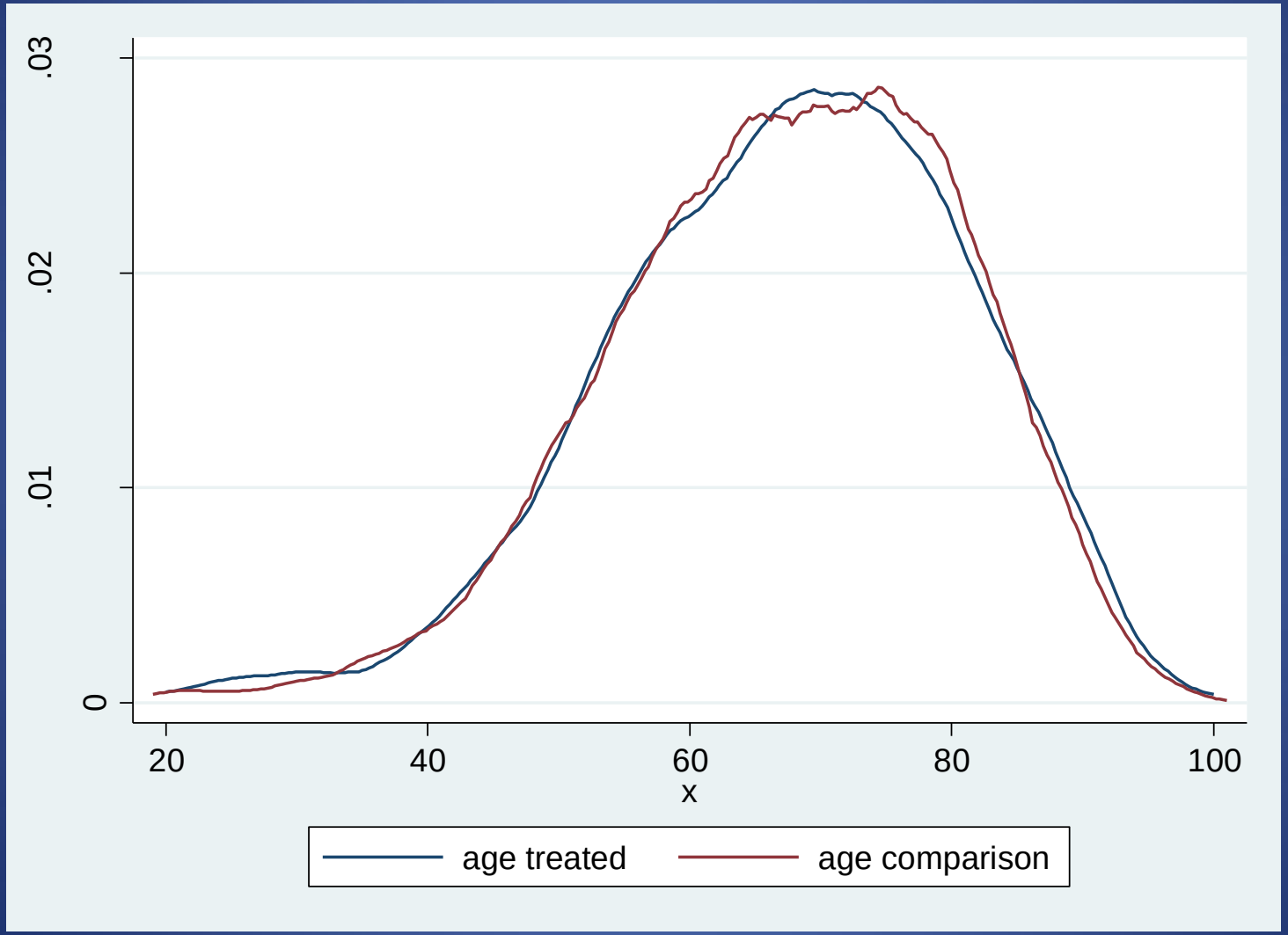
If points lie along 45 degree, covariate is balanced

Plots of Covariates in Treated and Comparison Groups

- For weighted continuous variables
- Plot density of covariate in treated group against density of covariate in comparison group
- Subjective comparison

```
twoway kdensity covariate  if treatment
  [aweight= norm_weights] || kdensity
  covariate if !treatment [aweight=
  norm_weights]
```

Example of Density Plot



Evaluate Ratio of Variances

- Ratio of variance of covariate in treated group to variance of covariate in comparison group should be near one if covariate is balanced
- Rubin: “1/2 or 2 are far too extreme”
- Compare ratio before and after matching or weighting sample by propensity score

Rubin 2001. Health Services & Outcomes Research Methodology 2: 169-188

Compare Results of Balance Tests

- If multiple tests indicate balance, there is a greater likelihood that covariates are balanced across treatment and comparison groups in the propensity score matched or weighted sample

General Procedure

Step 1: Choose variables to include in propensity score

Step 2: Ensure that propensity score is balanced across treatment and comparison groups

Step 3: Ensure that covariates are balanced across treatment and comparison groups within blocks of the propensity score

Step 4: Choose a matching or weighting strategy

Step 5: Ensure that covariates are balanced across treatment and comparison groups in sample matched or weighted by propensity score

Step 6: Proceed with analyses based on sample matched or weighted by propensity score

Analysis of Data Matched or Weighted by Propensity Score

- Delete observations from individuals not within the range of common support
- Choose the treatment effect of interest
- Calculate correct standard error for propensity score matched or weighted sample
- Guard against misspecification of the propensity score

Delete Individuals Outside of the Range of Common Support

psmatch2: Treatment assignment	psmatch2: Common support On suppor	Total
Untreated	10,070	10,070
Treated	1,574	1,574
Total	11,644	11,644

Nearest neighbor

No unmatched individuals

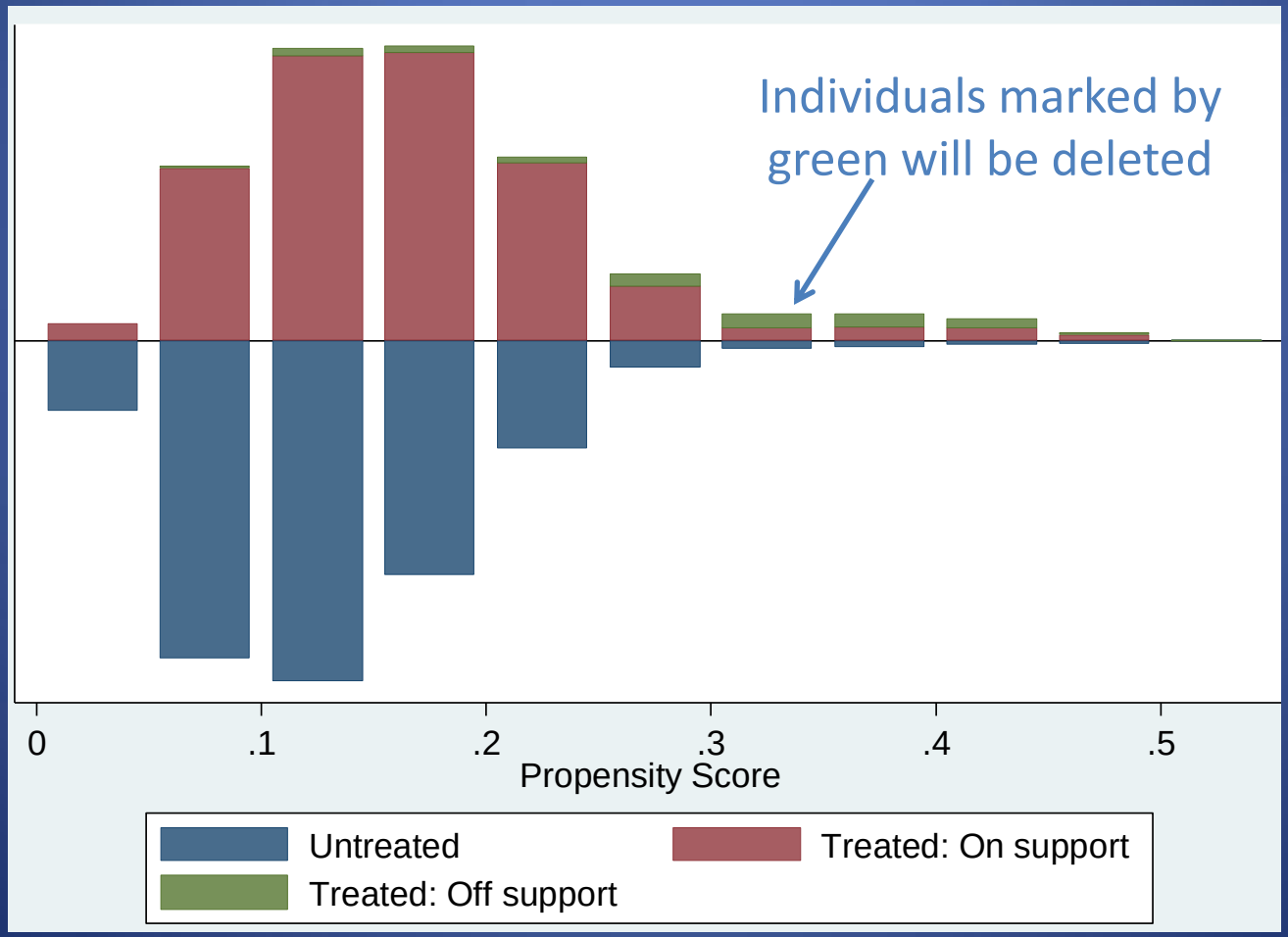
psmatch2: Treatment assignment	psmatch2: Common support Off suppo	On suppor	Total
Untreated	0	10,070	10,070
Treated	106	1,468	1,574
Total	106	11,538	11,644

106 treated individuals will be deleted

Delete Individuals Outside of the Range of Common Support

After -**psmatch2**-, can run: **psgraph, pscore(***pc_pscore***)**

to visualize distribution of individuals who will be deleted



Treatment Effects

- ATT: Average Treatment Effect on the Treated

- ATE: Average Treatment Effect for sample within range of common support
  - Incorporates ATT and average treatment effect on untreated

Need to Correct Standard Errors for Treatment Effect Estimates

- Uncertainty from propensity score estimate influences uncertainty of treatment effect estimate

- Ignoring uncertainty
  - Makes standard errors for ATEs more conservative (might conclude that there was no evidence of a significant treatment effect when there was)
  - Can make standard errors for ATTs more conservative *or* more generous

How to Correct Standard Errors

- Do nothing
  - If propensity score and treatment effect are estimated simultaneously, no need for further correction
- Bootstrap
  - When propensity score created in a separate step from treatment effect estimate and sample is *weighted* by propensity score
- Abadie-Imbens method
  - When propensity score created in a separate step from treatment effect estimate and sample is *matched* by propensity score

Guarding Against Misspecification of the Propensity Score

- "Doubly-robust" estimation
- Perform multivariable regression analysis on a sample matched or weighted by the propensity score
- As long as *either* the propensity score *or* the regression model is specified correctly, the treatment effect estimates will not be biased

Be careful when interpreting results of a propensity score analysis

- Generalizability
  - Excluded individuals differ from those within the range of common support
  - Treated and comparison individuals with missing values for any variables used in the propensity score are usually deleted
- Meaning of other coefficients in the model
  - Would need to create a new propensity score to test other interventions in the dataset

Treatment Effects with and without Propensity Scores

Analytic Approach	Adjustment for Selection Bias	Adjustment for Other Covariates	Sample Size	ATT
Regression of costs on outcome in original sample	No	No	1751	- $2,014
Regression of costs on outcome and control variables in original sample	No	Yes	1751	- $1,230
Propensity score matched sample, single method (ATT from -psmatch2- or -teffects-)	Yes	Yes	1468	-$937
Propensity score matched sample, doubly robust method (regression of costs on outcome and covariates within sample matched by propensity score)	Yes	Yes	1468	-$861

Controlling for Confounding

	Strengths	Limitations
Multivariable modeling	• Most commonly used approach • People are familiar with techniques & assumptions • Produces specific ß-coefficients for each individual confounder • Allows examination of these specific contributions to the outcome	• Requires parsimony to conserve degrees of freedom • Cannot detect group differences in distributions of measured confounders • Model assumptions may not fit the data • Only adjusts for measured confounders
Propensity scores	• Good for small data sets – summarizes set of confounders into a single measure; parsimony less of an issue • Distributions of confounders are similar between groups • Shows group differences • Allows for closer examination across strata	• Confounders balanced at group level – 2 people with the same score may not share the same characteristics • Balancing all the variables across strata can be hard to achieve • Only adjusts for measured confounders

Other Issues in Propensity Score Analysis

- Multi-valued and continuous treatments
- Residual confounding
- Power analysis

Multi-Valued Categorical Treatment

- Example: Effect of discharge status (home with no services, home with home health aide, post-acute care facility) on hospital readmissions
- Matching is not practical
- In Stata 13, can calculate the effect of a multi-valued categorical treatment on an outcome with the -teffects- package through IPTW

Treatment Effects for Multi-Valued Categorical Treatments

Average Treatment Effect on the Treated

$$E\{(y_{treatmentA} - y_{treatment0}) \mid t=treatmentB\}$$

- Mean difference in treatment effects between treatment of interest (treatment A) and comparison/baseline treatment (treatment 0), given that individual received a certain level of treatment (treatment B)

- Treatment A and B can refer to the same treatment group

Treatment Effects for Multi-Valued Categorical Treatments

Average Treatment Effect on the Treated

$$E\{(y_{\text{post-acute care}} - y_{\text{home with no services}}) \mid t=\text{post-acute care}\}$$

Mean difference in treatment effects between treatment of interest (post-acute care) and comparison/baseline treatment (home with no services), given that individual received a certain level of treatment (post-acute care)

Stata Code for Multi-Valued Categorical Treatment

```
teffects ipwra (outcome covariate1…

  covariate#)(treatment covariate1…

  covariate#),

atet control(treatmentlevel0)

tlevel (treatmentlevelB)
```

Default is ATE

Baseline treatment level

"Treated" group for whom you want ATT

Continuous Treatment Generalized Propensity Score (GPS)

```
doseresponse  covariate1… covariate#,

    outcome(outcomevar) t(treatmentvar)

    gpscore(newpscorevar) predict(hat_treat)

    sigma(hat_sd) cutpoints(cut) index(mean)

    nq_gps(#) dose_response(newdoseresponsevar)

    test(Bayes_factor) detail
```

Creates propensity score

Variable that splits treatment into intervals

Number of quantiles of GPS

Alternative to t-test for balance diagnostics

Sensitivity Analyses for Residual (Unobserved) Confounding

	Rosenbaum	VanderWeele and Arah
Study Design	For 1:1 matched samples only	Any
Sensitive to sample size?	Yes	No
Information supplied by researcher	Relationship between outcome (**y**) and confounder (**u**) Relationship between treatment (**d**) and **u**	Relationship between **y** and **u** Prevalence of **u** when **d** = 0 Prevalence of **u** when **d**= 1
Result	Strength of relationships among **y**, **u**, and **d** needed to make treatment effect estimate no longer significant	Treatment effect estimate is adjusted for **u**

Adapted from Liu et al 2013. Prevention Science 14: 570-580

Power Analyses for Propensity Scores

Traditional Power Analysis:

Estimate power to detect an increase in % of patients with a goals of care conversation after a PC intervention

- Hypothesized change from 50% to 70% of patients
- n (Treatment group) fixed at 75 patients

Usually, power increases with sample size

- Power when n(control group) is 75 = .71
- Power when n(control group) is 150 = .83

Power Analyses for Propensity Scores

Traditional power analysis does not account for:

- Precision of matches when multiple comparison group individuals are matched to one treated individual
- Dependence of observations
- Propensity score's reduction of variation from observed confounders
- Unequal contribution of observations to analysis if using propensity score weights

One solution: Power calculation via simulation

References

Austin PC. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples Statistics in Medicine 2009;28: 3083-3107.

Garrido MM. Propensity scores: A practical method for assessing treatment effects in pain and symptom management research. JPSM. 2014; 48(4): 711-718.

Garrido MM, Kelley AS, Paris J, Roza K, Meier DE, Morrison RS, Aldridge MD. Methods for constructing and assessing propensity scores. Health Services Research. 2014; 49(5): 1701-1720.

Liu W, Kuramoto SJ, Stuart EA. An introduction to sensitivity analysis for unobserved confounding in nonexperimental prevention research. Prev Sci. 2013 Dec;14(6):570-80.

Rubin DB. Using propensity scores to help design observational studies: Application to the tobacco litigation. Health Services & Outcomes Research Methodology 2001; 2(3-4): 169-188

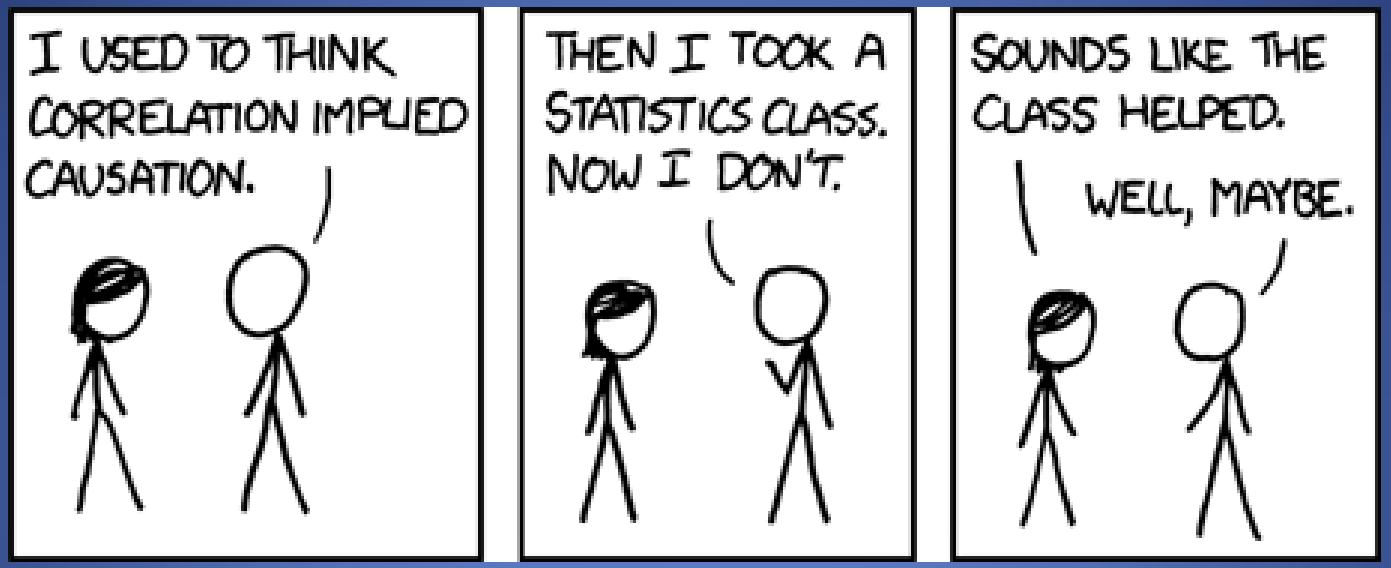
Starks H, Diehr P, Curtis JR. The challenge of selection bias and confounding in palliative care research. J Palliat Med. 2009 Feb;12(2):181-7.

Acknowledgements

- NPCRC Career Development Award, 2007-2010 (Starks)
- VA HSR&D (CDA 11-201, CDP 12-255) (Garrido)
- PC4C Study Group (NCI/NINR 5R01CA116227, National Palliative Care Research Center)
- Diane Meier, R. Sean Morrison, Melissa Aldridge, Amy Kelley, Julia Paris, Katherine Roza
- Icahn School of Medicine at Mount Sinai Claude D. Pepper Older Americans Independence Center (NIH/NIA P30 AG028741-01A2)

The views expressed in this presentation are those of the authors and do not necessarily reflect the position or policy of the Department of Veterans Affairs or the United States government.

Questions?



xkcd.com

melissa.garrido@mssm.edu
tigiba@u.washington.edu