Observational & Quasi-experimental Research Methods

10th Annual Kathleen Foley Palliative Care Retreat
Old Québec, October 24, 2016

Melissa M. Garrido, PhD[1] and Jay Magaziner, PhD[2]

1. Department of Veterans Affairs, Icahn School of Medicine at Mount Sinai, New York, NY
2. University of Maryland School of Medicine, Baltimore, MD

Overview of Workshop

- Randomized trials and observational studies
- Selection bias – what is it and why do we care?
- Whirlwind tour of some methods to address selection bias
  - Propensity scores
  - Coarsened exact matching
  - Instrumental variables
- Q & A

When our objective is to understand the effect of a treatment or management strategy on an outcome:

Options: Experiment or Observational Study
Considerations: Trade offs between approaches

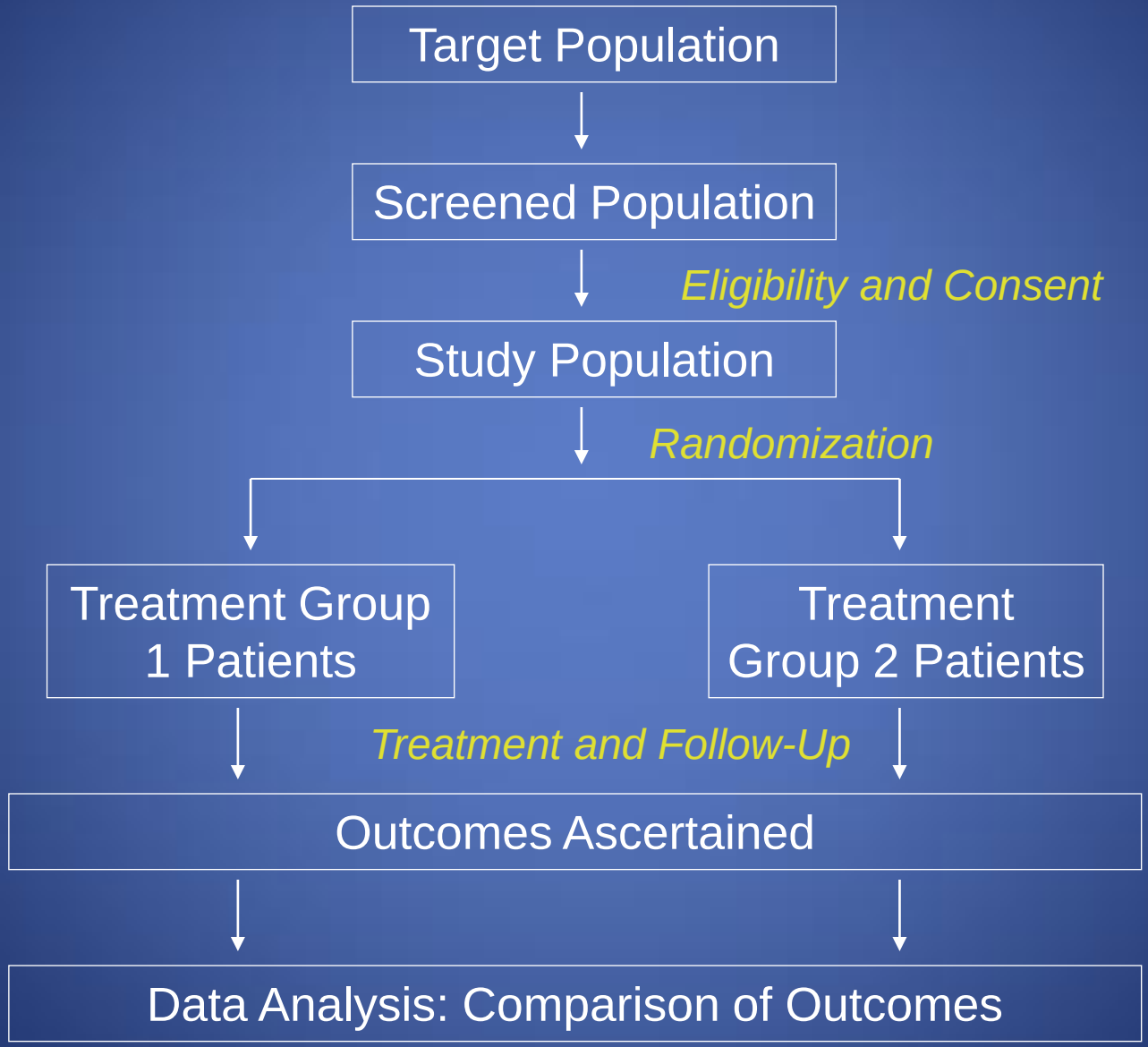
Experiments

- Best suited for evaluating efficacy
- Can be used to evaluate effectiveness

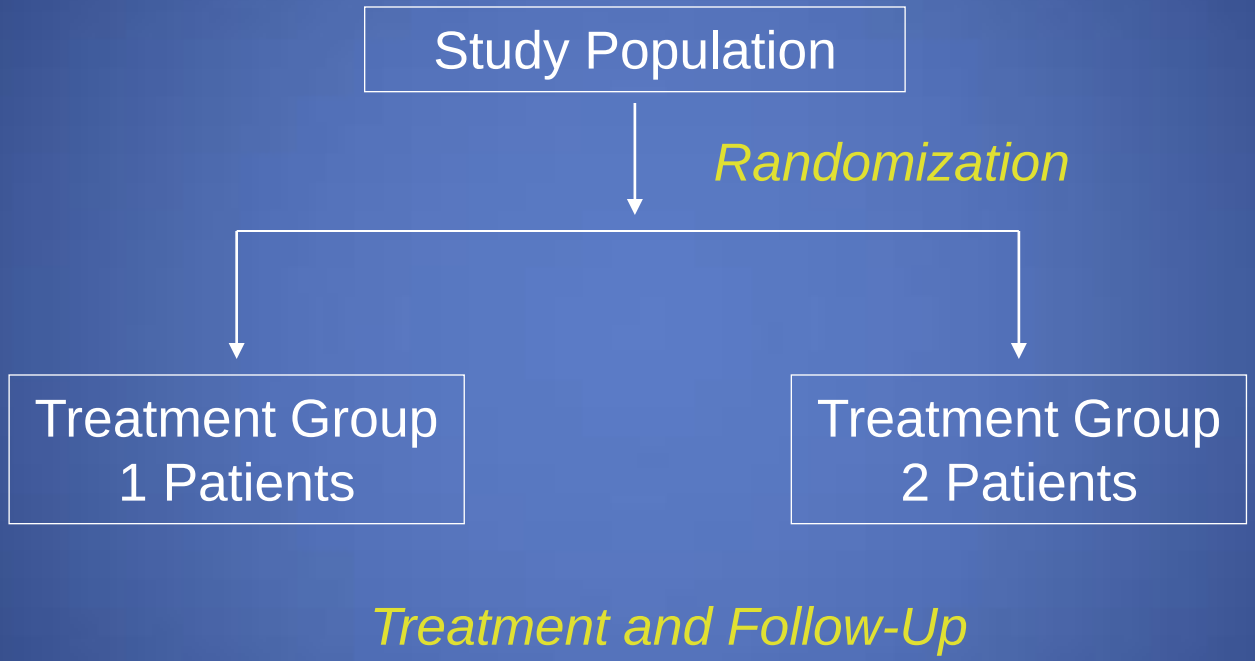
Randomization

Fundamental Procedure in an Experiment

The Clinical Trial Paradigm



Study Population



Randomization and Assignment

- Randomization is a scientific principle
- Assignment (Allocation) is by chance
- Assures that two groups are alike

Randomization Influence on Clinical Trial

- All aspects
- Definitive for Internal Validity
- Valid Statistical Tests without need to adjust for confounders because groups are alike in all ways (even those that are not easily measured or correlated with those that are easily measured)

Desired Features of Allocation

- Unpredictable
- Avoids selection effects (aka, confounding by indication)
- Secure (Not Switchable)

Intention to Treat Approach

Include anyone who was randomized in the group to which they were randomized.

- Advantages:
  - Provides the most fair comparison between groups because the groups should be equivalent with respect to prognostic factors due to randomization.
  - Straight-forward to implement, no subjective judgment or additional information is required.

Intention to Treat Approach (continued)

- Disadvantages:
  - Results in an estimate of the rate of events among those assigned to a treatment group, whereas, there might be greater interest in the rate of events among those who receive a particular treatment.
  - Includes individuals who did not get treatment.
  - Potential problem if a lot of people change treatment after randomization and before outcome is assessed.

Per-Protocol Analysis

Only include in the analysis those who received the treatment to which they were randomized.

- Advantages:
  - Provides an estimate of the rate of events among those who actually receive the treatment.
- Disadvantages:
  - Could result in a biased estimate of the impact of the treatment in the population because those who do not receive the treatment could be different than those who do.
  - Could result in a biased estimate of treatment differences because those excluded from one group may differ from those excluded from another group.

As-treated Analysis

Include all those in the study in the groups based on treatment received, not treatment randomized to.

- Advantages:
  - Similar to per-protocol but includes more patients.
- Disadvantages:
  - Similar to per-protocol, but is even more likely to result in differences between groups
  - Much like an observational study with a pre-selected group of participants.

Effectiveness vs efficacy

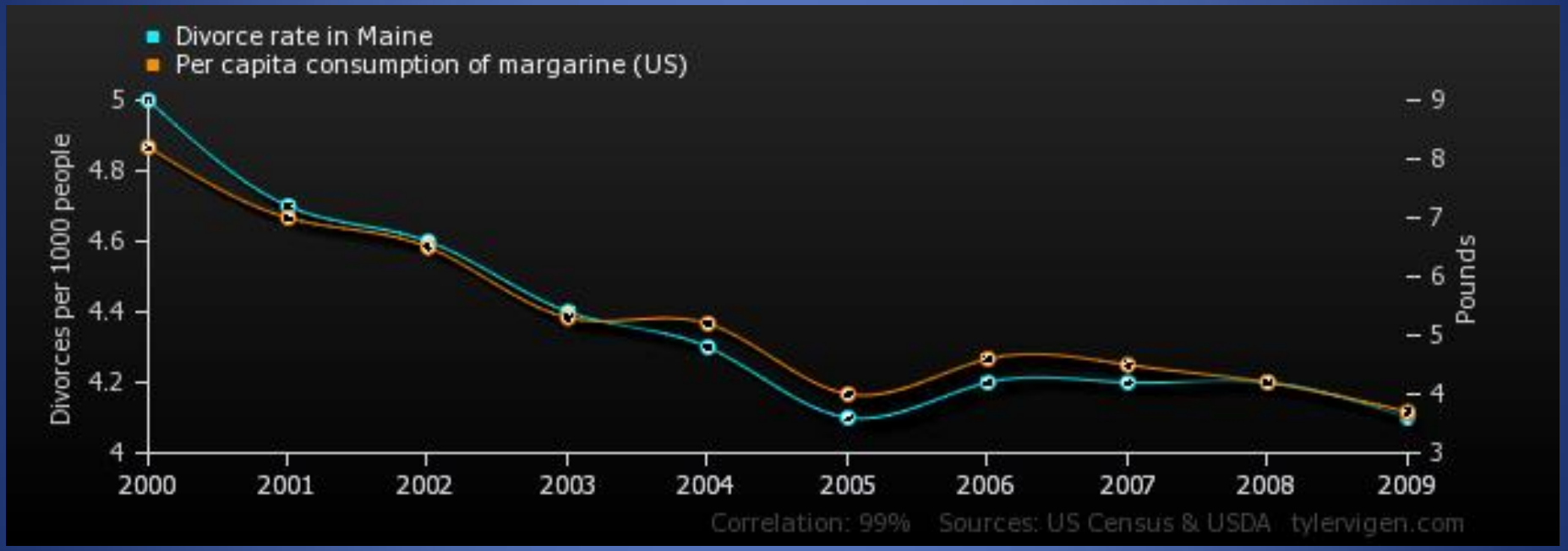
Effectiveness

- Generalizable to wider population
- External validity
- Generally less costly
- Can be done when experiment is not an option
- Not practical when treatment or outcome occur rarely in population being studied
- Other

Efficacy

- Generalizable only to those meeting entry criteria
- Internal validity
- Generally more costly
- Not always feasible
- Can be done as long as eligible patient group can be identified and enrolled
- Other

Making Sense of Observational Data



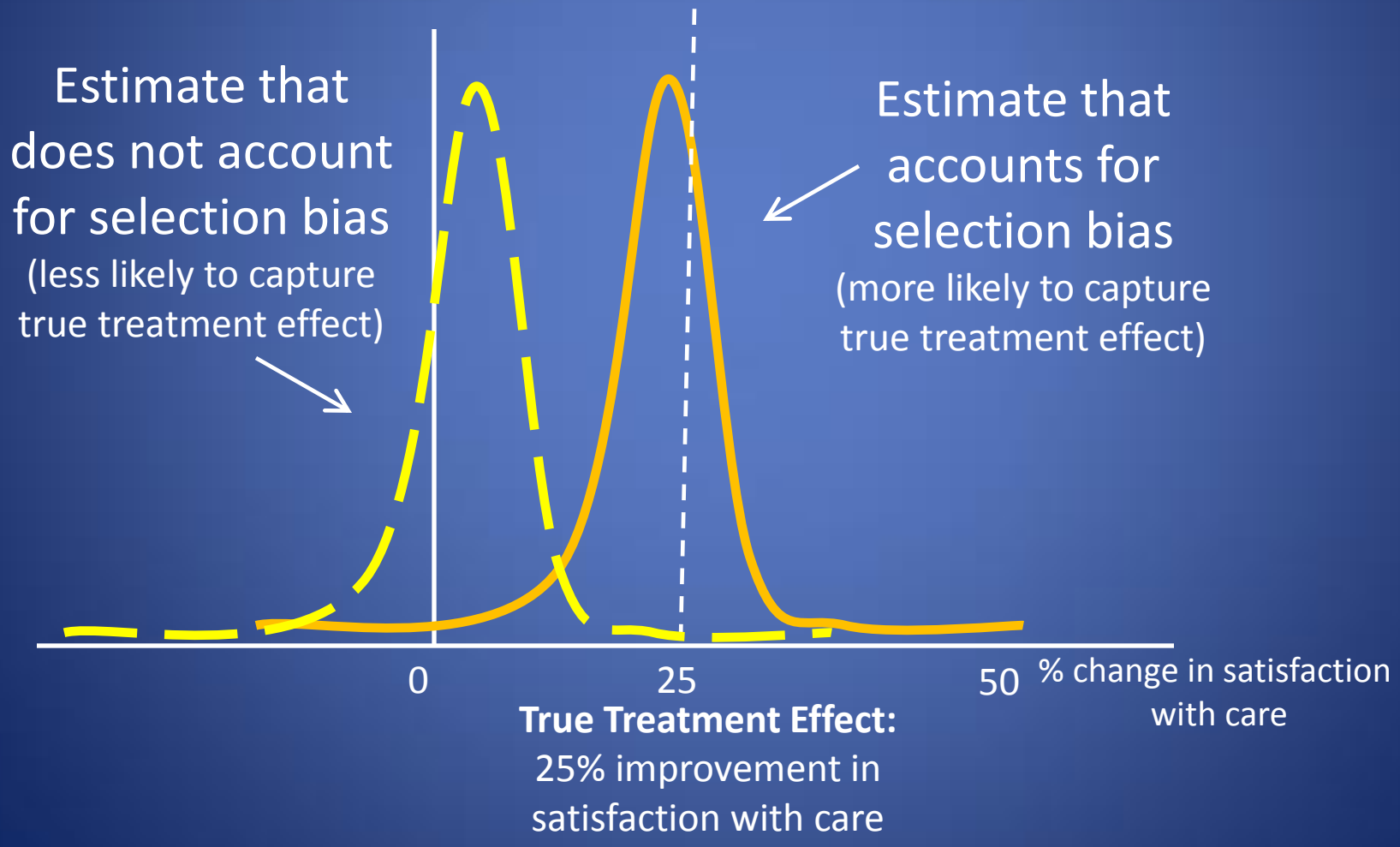
Confounding due to Selection Bias in Observational Data

Patients not randomized to treatment

↓

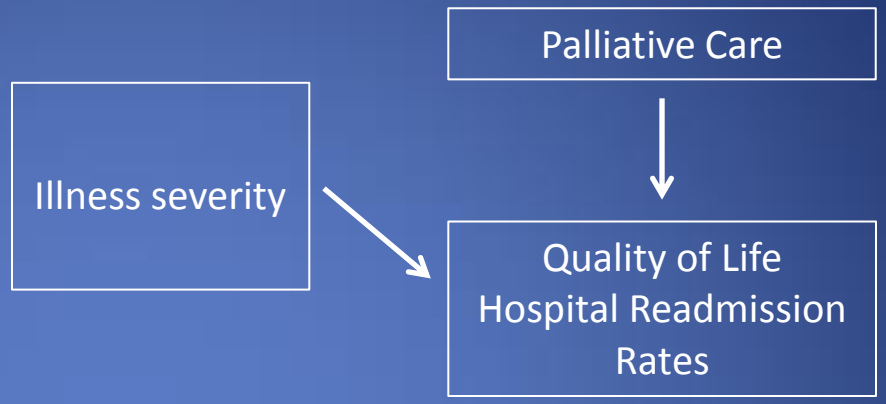
Patient characteristics may be associated with both participation in treatment *and* outcome

Impact of Selection Bias on Analytic Inferences

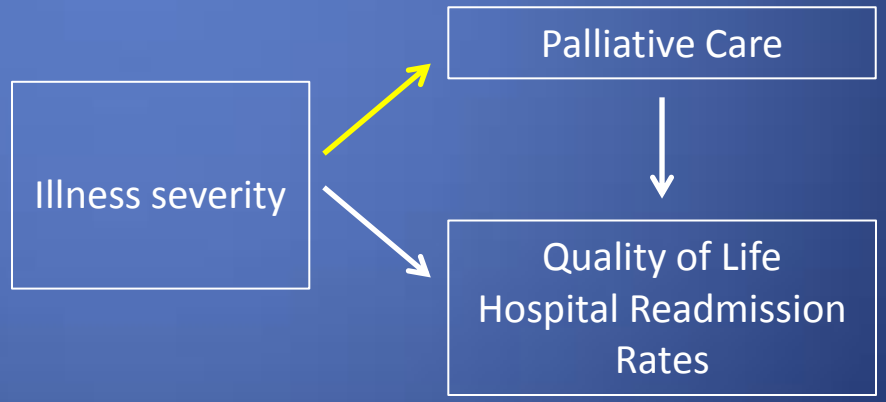


Tools to Address Confounding

- Multivariable models



- Matching
- Propensity scores
- Instrumental variables
- Regression discontinuity
- Difference-in-differences



Tools to Address Confounding due to Selection Bias

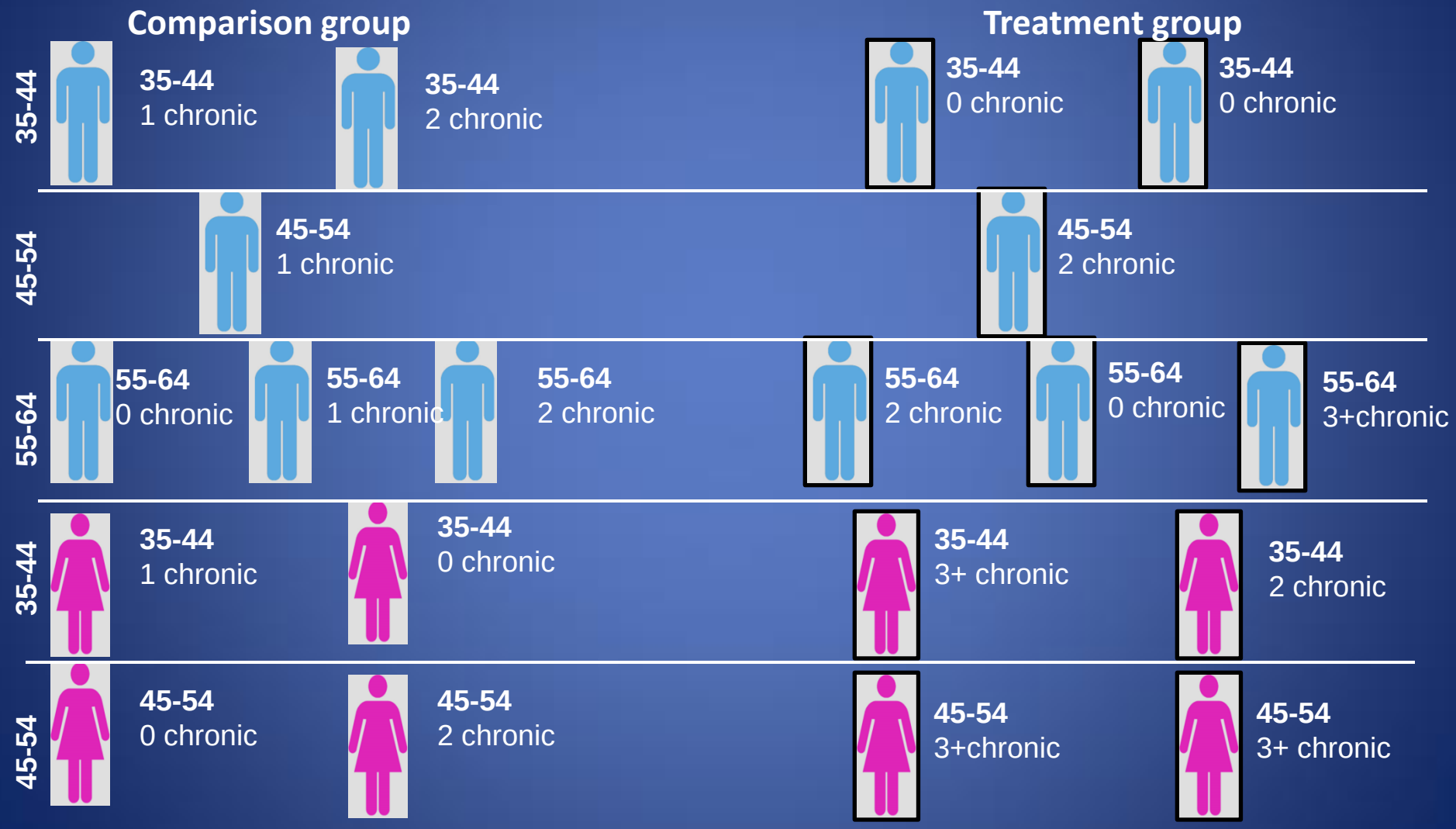
- Matching
  - Compare treated and comparison individuals who have the same values for a set of covariates
- Propensity scores
  - Compare treated and comparison individuals who have similar "propensities" or likelihoods for receiving treatment, conditional on a set of several covariates
- Instrumental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

Addressing Selection Bias with Exact 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

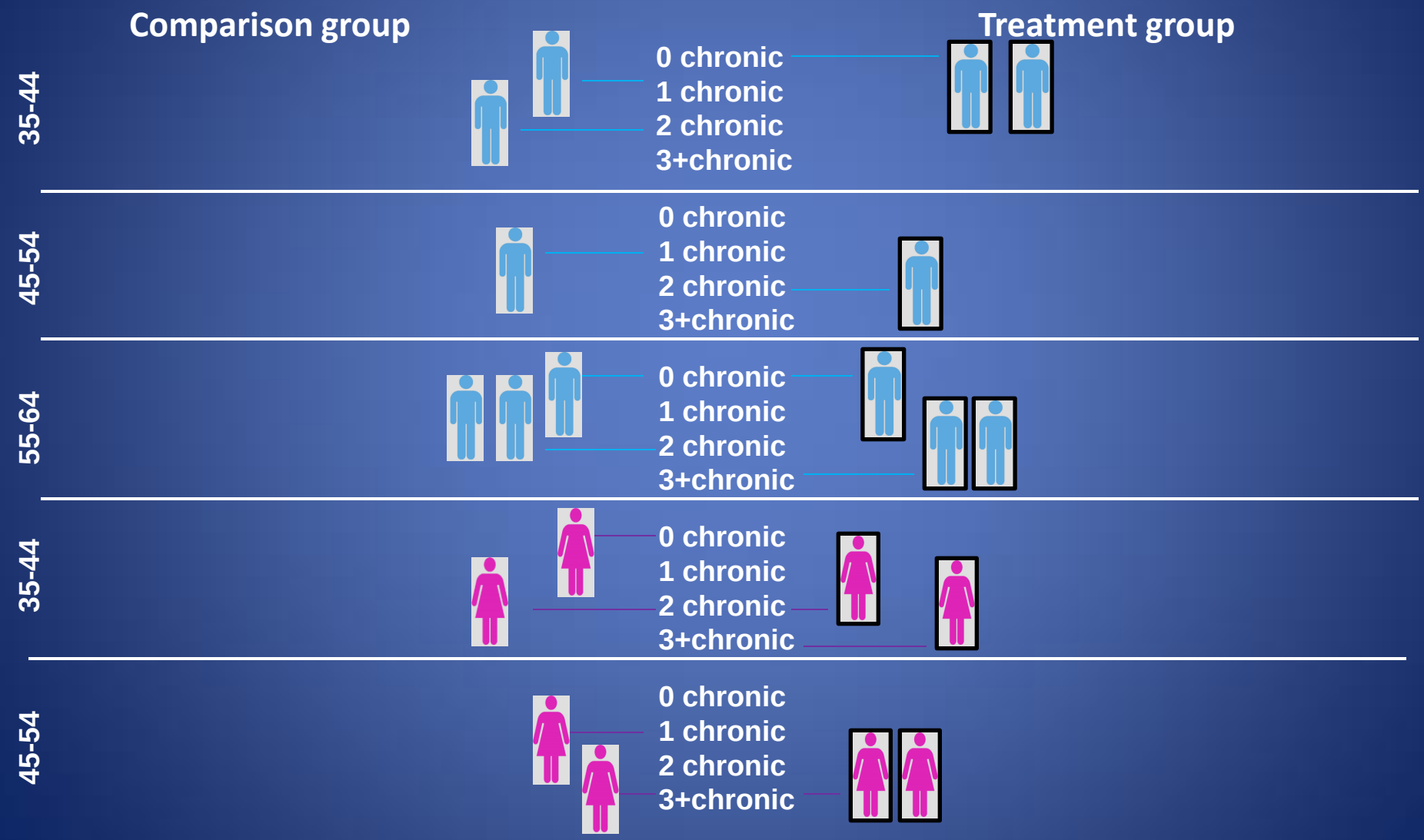
Matching on Specific Variables:

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

Propensity Scores: Big Picture

- Create a single composite score of all observed, measured potential confounders of the association between treatment and outcome
- Propensity score is the conditional probability of treatment given the observed covariates X

$$E(X) = P(D=1 \mid X)$$

- Match or weight on this one-dimensional score alone
- Do this without knowledge of the outcome variable

Propensity Score Assumption: Strongly Ignorable Treatment Assignment

- Given a set of covariates:
  - Treatment assignment and outcome are independent
  - Everyone has a nonzero chance of receiving the treatment

Rosenbaum & Rubin 1983. Biometrika 70: 41-45

What Propensity Scores Can & Cannot Do

- Propensity scores can:
  - Help find matches from comparison group so that ***measured*** confounders are equally distributed between treatment & comparison groups
  - Improve precision of treatment effect estimates

- Propensity scores cannot:
  - Account for ***unmeasured*** confounders

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

Garrido et al. 2014. HSR 49: 1701-1720

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

Choosing Variables for Propensity Scores

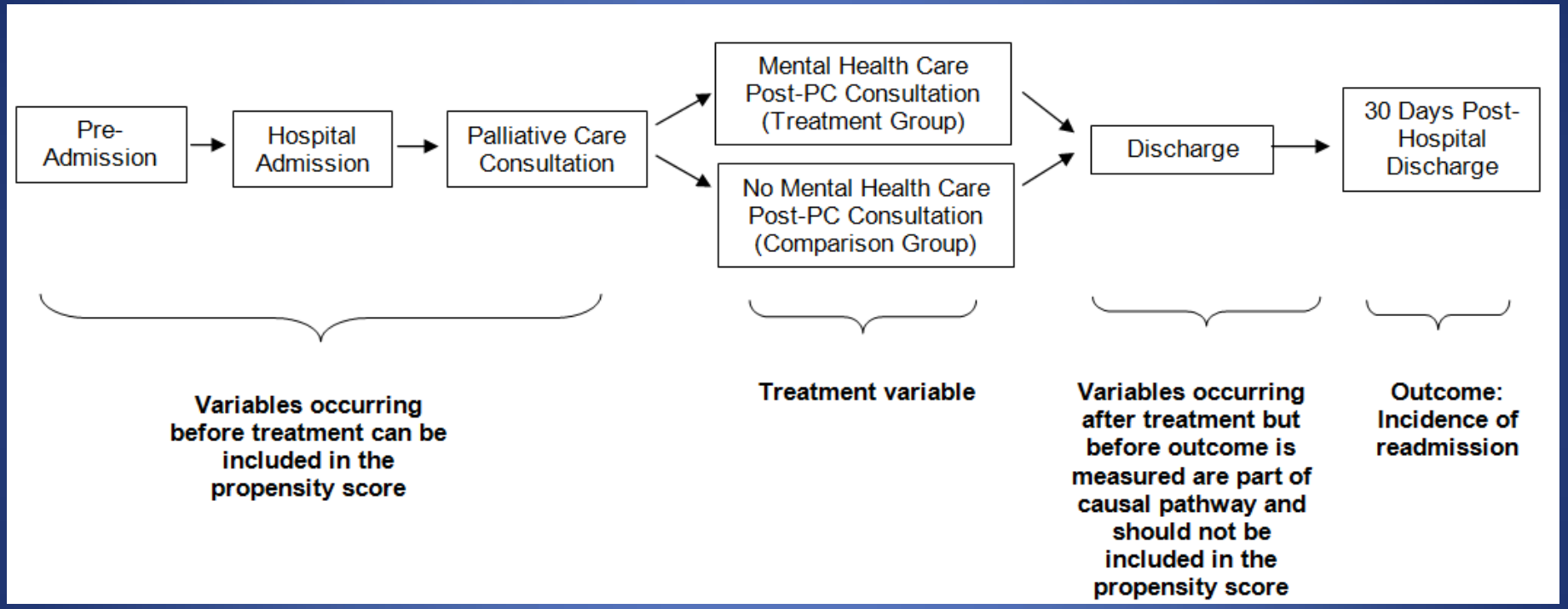
- Include:
  - Theoretically related to treatment and outcome
  - Available & easy/reliable to collect on everyone
  - Correlated with unmeasured confounders

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

Variable Selection Example

- Hospitalized veterans receiving a palliative care consultation in a VISN 3 acute care facility
- Treatment: Psychotherapy provided after a palliative care consultation
- Outcome: All-cause 30-day readmission

Choosing Variables for Propensity Score Models



Garrido 2014. JPSM 48:711-718

Working Example

- 2008 Healthcare Cost and Utilization Project (HCUP) 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!

Calculate Propensity Score

- Maximum Likelihood Estimation (logit, probit models)
- Generalized Boosting Methods
- Generalized Method of Moments (Covariate Balancing Propensity Score [CBPS])

Stata Code to Calculate Propensity Score

Treatment variable

Specify covariates to include in calculation

```
pscore treatment covariate1 covariate2 … covariate#,
   pscore(pc_pscore) blockid(pc_block) detail
```

Label the estimated propensity score

Label the blocks of propensity scores

Optional command that shows details of testing blocks and balancing covariates

*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 the propensity score

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

↓

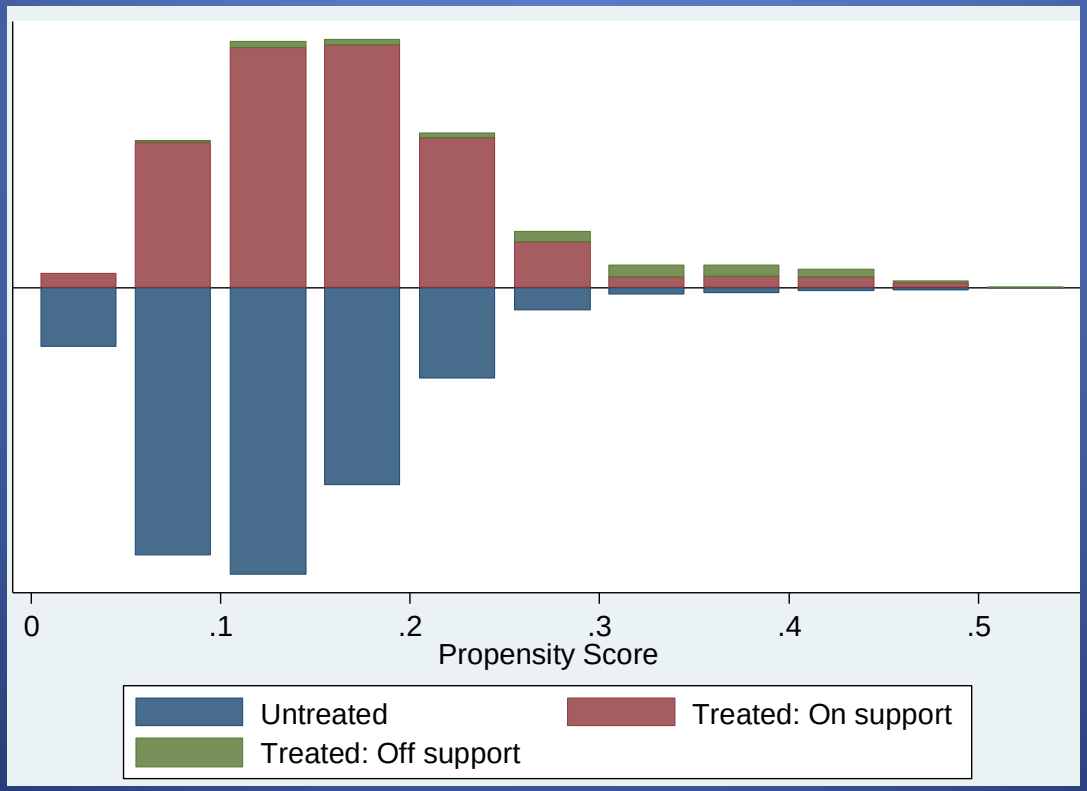
Step 6: Proceed with analyses based on sample matched or weighted by propensity score

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

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 blocks

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

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

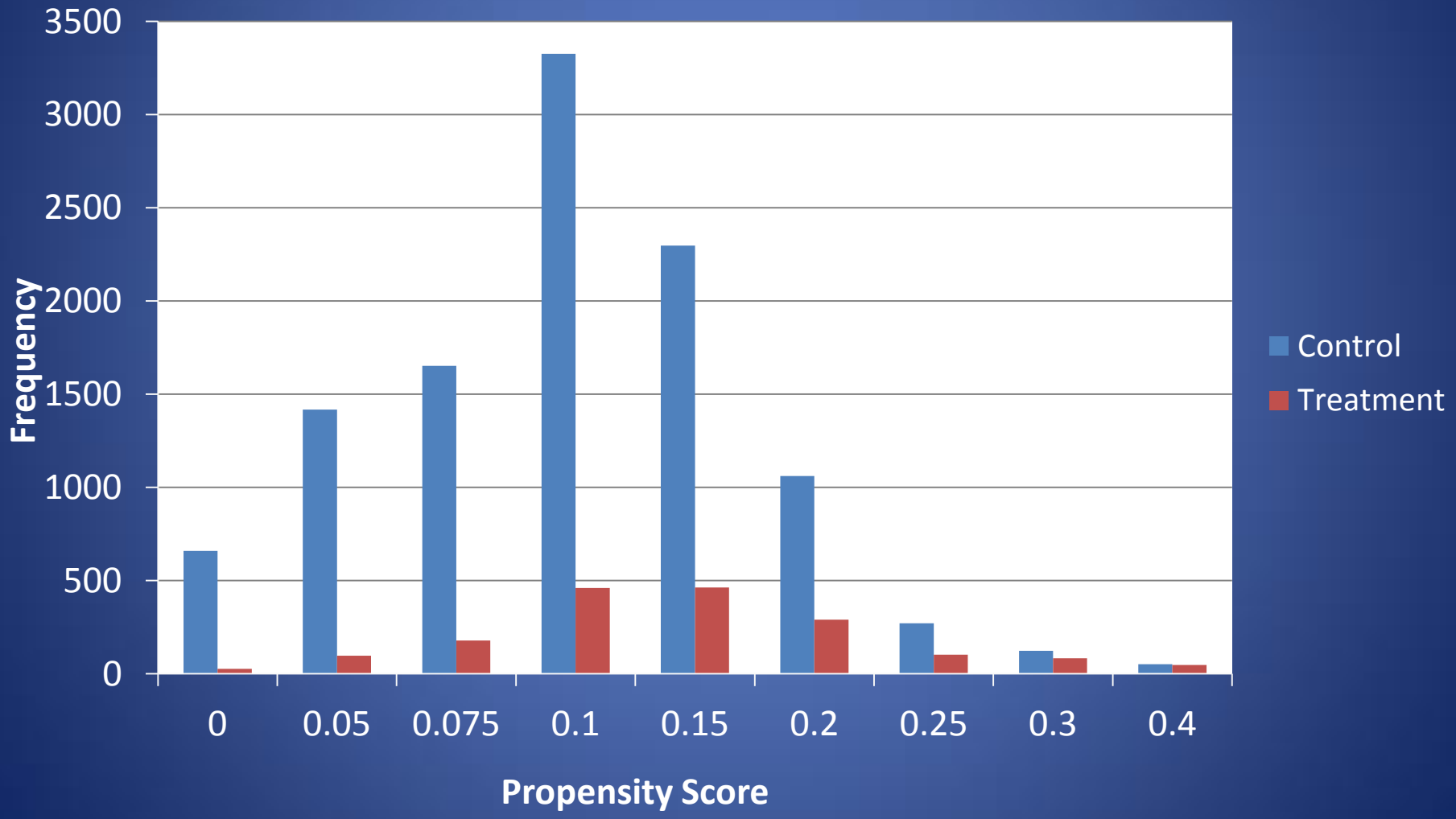
Continuous variables

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

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

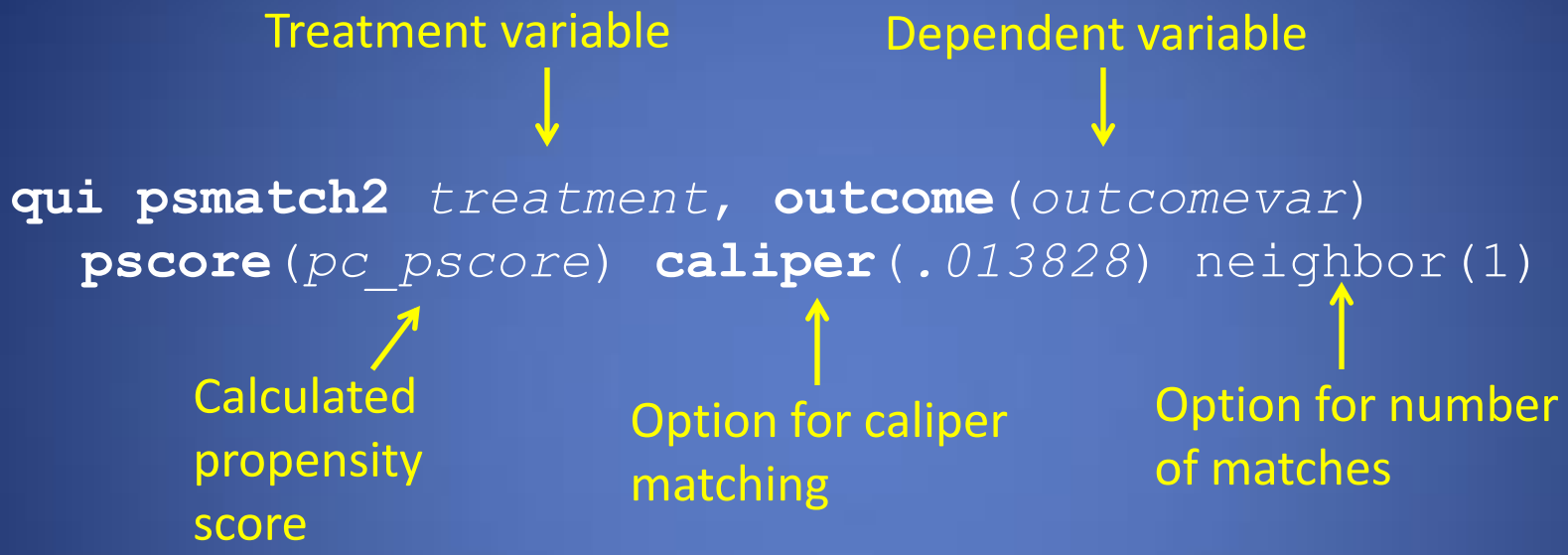
Choices When Matching Sample by Propensity Score

- 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

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 covariate1… covariate#,
   genvars
```

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

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

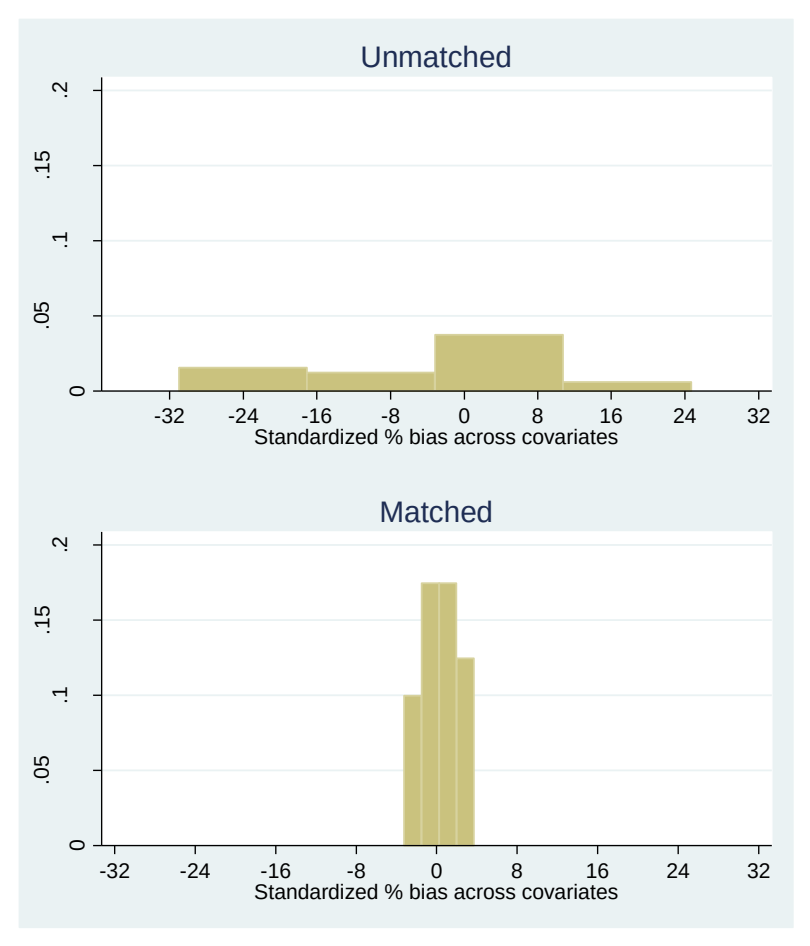
Summary of covariate imbalance

Summary of mean and median bias before and after matching

Visual inspection of standardized differences

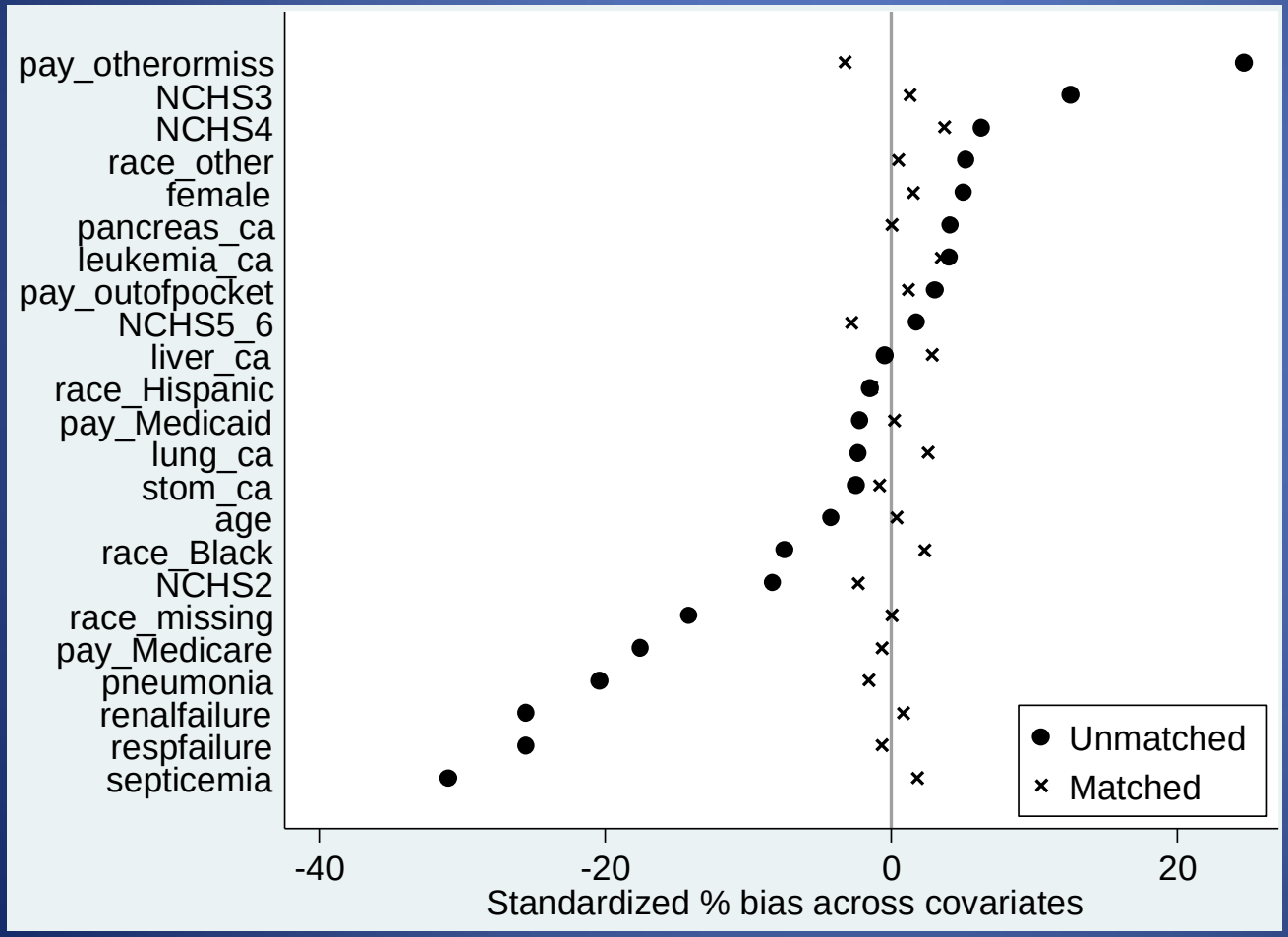
pstest *covariate1..covariate#*, **treated**(*treatment*) **both** **hist**

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

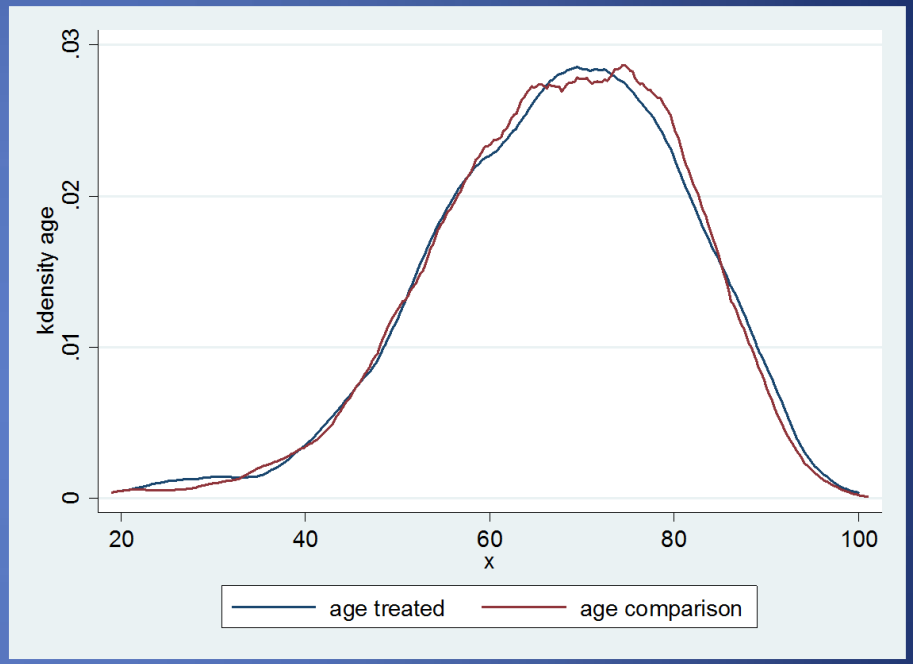
Output from **-pbalchk-**

	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

Plots of Covariates in Treated and Comparison Groups

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




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

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

Alternative:

Covariate Balancing Propensity Score Method

↓

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

Covariate Balancing Propensity Scores

- Generalized method of moments to estimate a propensity score model that optimizes covariate balance across treatment groups
- Typically used with IPTW

- Advantage: Less subject to investigator bias
- Disadvantage: No control over relative weight provided to confounders

- Software – R package ‘CBPS’

Imai & Ratkovic. J R Statist Soc B 2014; 76: 243-246.

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

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

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 (Stata's `aweight` and `pweight` commands)
  - As long as *either* the propensity score *or* the regression model is specified correctly, the treatment effect estimates will not be biased

Interpreting Propensity Score Analysis Results

- 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

Sensitivity Analyses for Residual (Unobserved) Confounding

- Identify smallest amount of unobserved confounding that would need to exist to change your inference from rejection to acceptance of H_0
- Test effect of treatment variable on a lagged outcome
- Estimate treatment effect in multiple comparison groups

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

“Pre-processing” Datasets

Make treatment and comparison group as similar as possible on observed confounders before proceeding with analysis

- Propensity Scores
- Exact Matching
- Coarsened Exact Matching
- Entropy Matching

Ho et al. 2007. Political Analysis 15: 199-236
Stuart 2010. Statistical Science 25: 1-21.

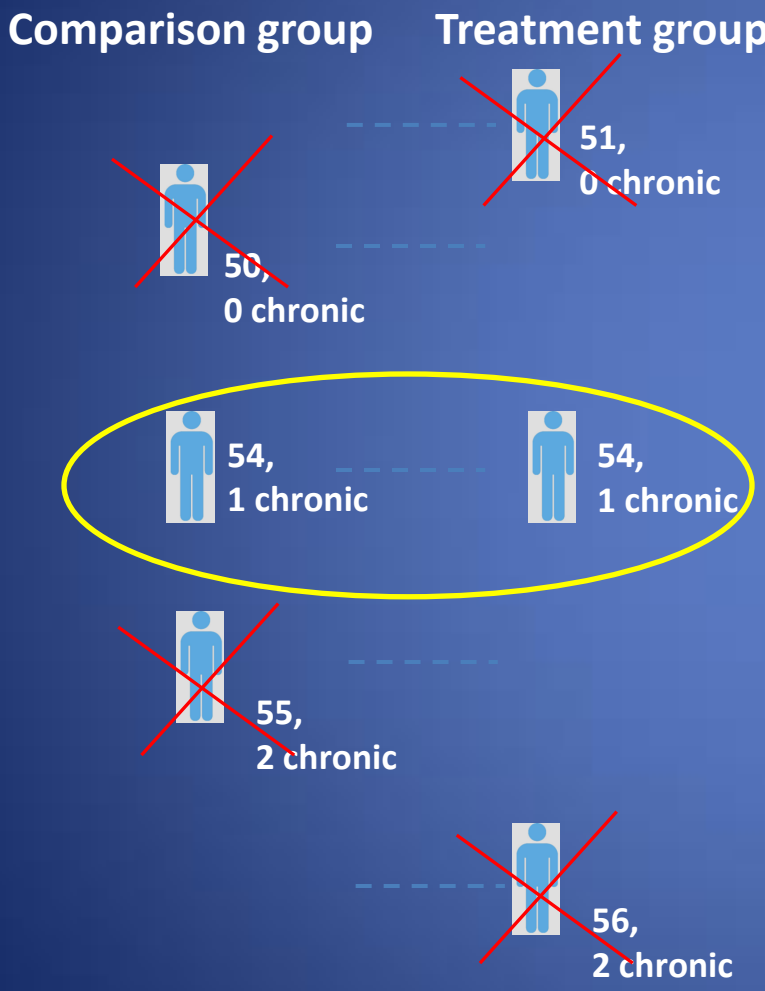
Coarsened Exact Matching

- Match on broad categories (coarsened values) of important variables
- More feasible than exact matching on large set of potential confounders
- Not susceptible to worsened balance due to model misspecification (a strong risk with propensity score matching when data on important confounders are not available)

King 2015. http://gking.harvard.edu/publications/why-propensity-scores-should-not-be-used-formatching

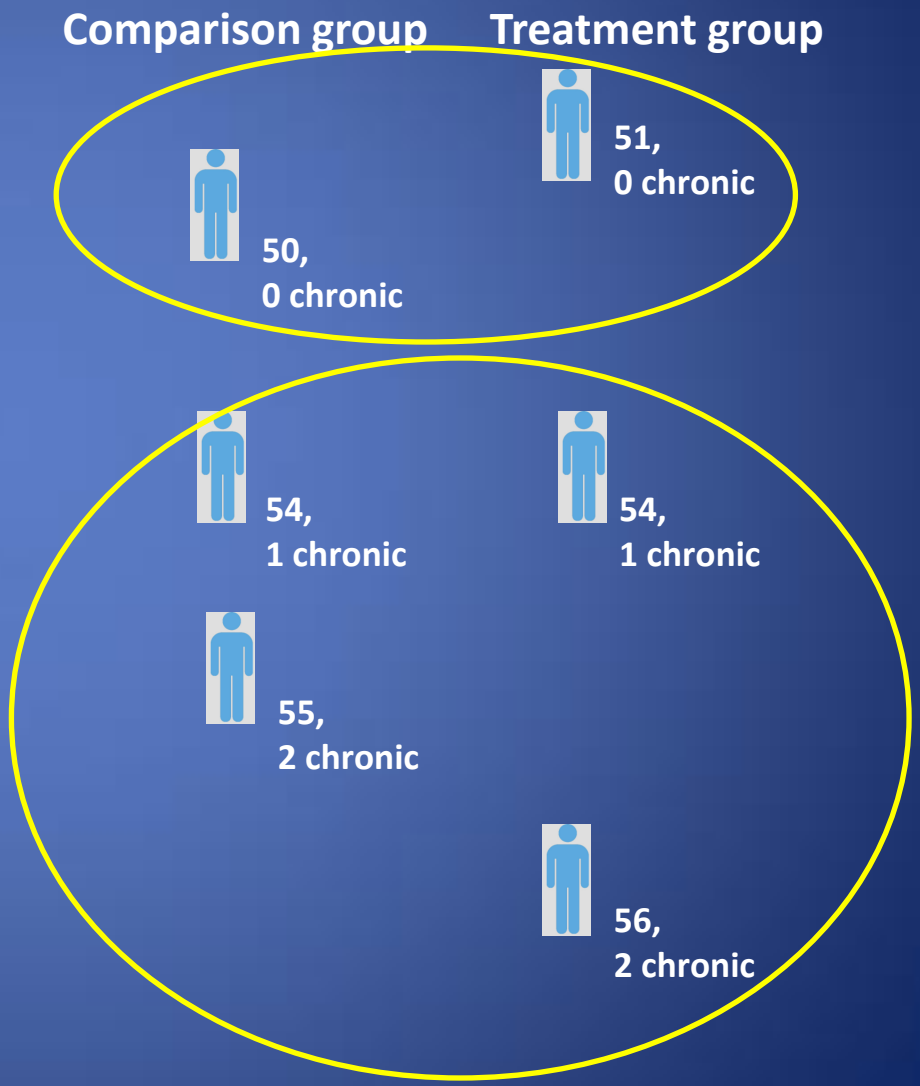
Exact Matching

Match on exact age and number of chronic conditions



Coarsened Exact Matching

Match on age category (50-59, 60-69) and presence of chronic conditions



Coarsened Exact Matching Procedure

- Divide sample into strata that have treated and comparison individuals with the same coarsened values of covariates
- Within strata,
  - Treated individuals assigned a weight of 1
  - Comparison individuals are assigned a weight that accounts for the number of: treated observations within the strata, comparison observations within the strata, matched treated observations within the dataset, and matched comparison observations within the dataset
- Strata without both treated and comparison individuals are assigned a weight of 0
- Traditional multivariable analyses are run on the weighted dataset

What CEM Can & Cannot Do

- CEM can:
  - Help find matches from comparison group so that ***measured*** confounders can be equally distributed between treatment & comparison groups
  - Improve precision of treatment effect estimates

- CEM cannot:
  - Account for ***unmeasured*** confounders

Stata Code to Perform CEM

Specify covariates to include in calculation

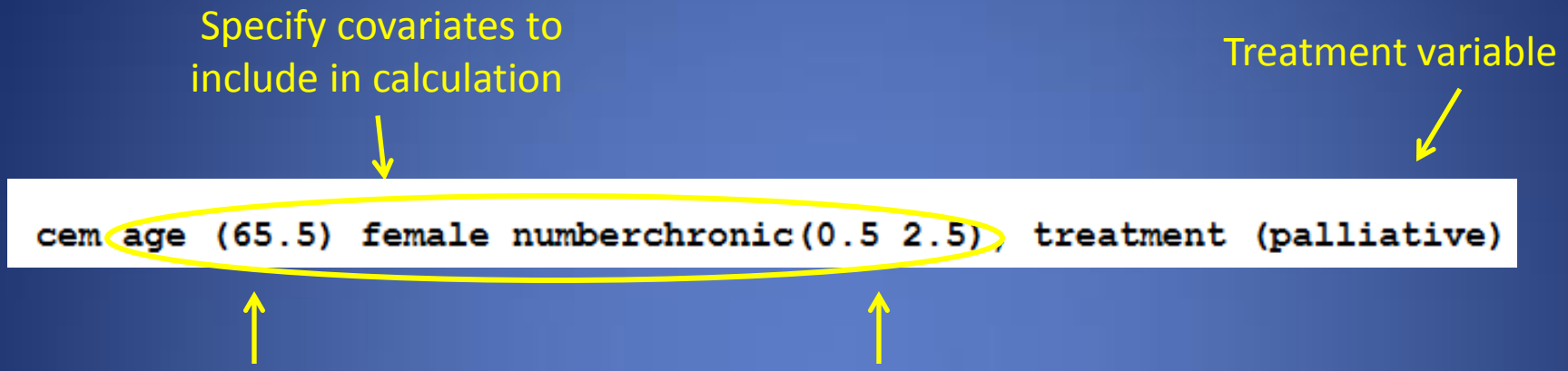
Specify coarsened values of continuous variables

```
cem covariate1 (cutpoint1 cutpoint2..cutpoint3)
   covariate2 … covariate#, treatment(treatmentvar)
```

Treatment variable

*-cem- is not part of Stata's built-in commands. Type "findit cem" in Stata's command line and follow link in pop-up window to install

Working Example: CEM



Working Example: Stata Output for CEM

```
.
. cem age (65.5) female numberchronic(0.5 2.5), treatment (palliative)
(using the scott break method for imbalance)

Matching Summary:
-----------------
Number of strata: 9
Number of matched strata: 8

                0      1
      All   10897   1789
  Matched   10895   1789
Unmatched       2      0

Multivariate L1 distance: .26375911

Univariate imbalance:

                  L1      mean       min       25%       50%       75%       max
         age    .05185   -.05637         1         0         0         0        -1
      female   2.3e-14   3.9e-14         0         0         0         0         0
numberchronic    .0817   -.45616         0         0        -1        -1        -3
```

2 unmatched observations (weight = 0)

0 = perfect balance, 1 = complete imbalance
**Interpret relative to output from other matches

```
. cem age (65.5) female hospsize_large hosp_private_nfp ///
> numberchronic(0.5 2.5), treatment (palliative)
(using the scott break method for imbalance)

Matching Summary:
-----------------
Number of strata: 34
Number of matched strata: 32

                0      1
      All   10897   1789
  Matched   10895   1789
Unmatched       2      0


Multivariate L1 distance: .43637651
```

```
. cem age (45.5 65.5) female hospsize_large hosp_private_fp hosp_private_nfp ///
> numberchronic(0.5 1.5 2.5), treatment (palliative)
(using the scott break method for imbalance)

Matching Summary:
-----------------
Number of strata: 94
Number of matched strata: 73

                0      1
      All   10897   1789
  Matched   10799   1788
Unmatched      98      1


Multivariate L1 distance: .45118723
```

As variables become less coarsened, finding matches for every observation becomes more difficult

CEM: Run Planned Analyses on Weighted Sample

- `-cem-` produces variable *`cem_weights`*
- Use un-coarsened values of variables used for matching

Continuous variables can be returned to original form

```
.
. glm TOTCHG i.palliative age female i.hospsize_small i.hospsize_med i.hosp_private_nfp ///
> i.hosp_private_fp numberchronic [pweight = cem_weights], family(gamma) link(log)
```

Weight command

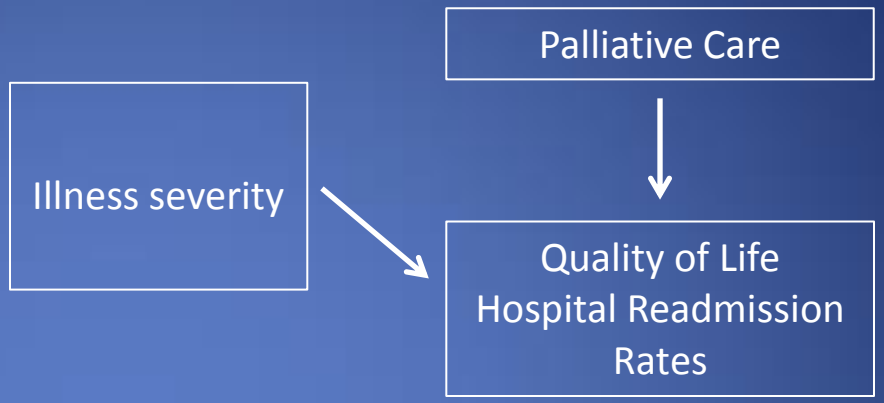
Interpreting Results of Analyses Using CEM

- Generalize to individuals similar to those included in the matched sample

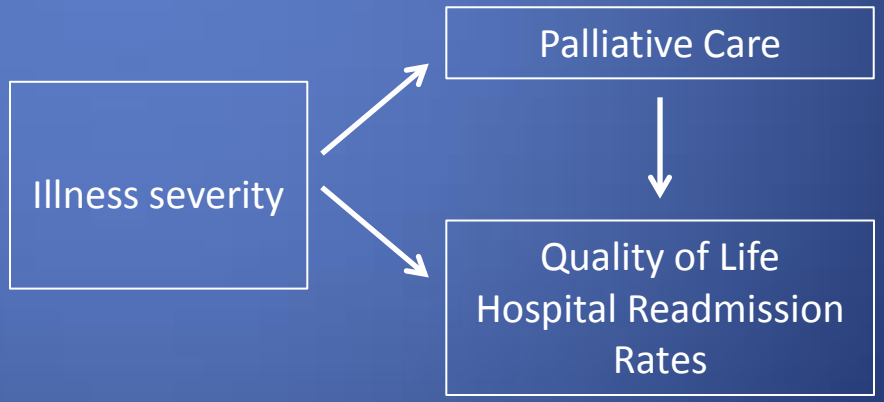
- ATT

Tools to Address Confounding

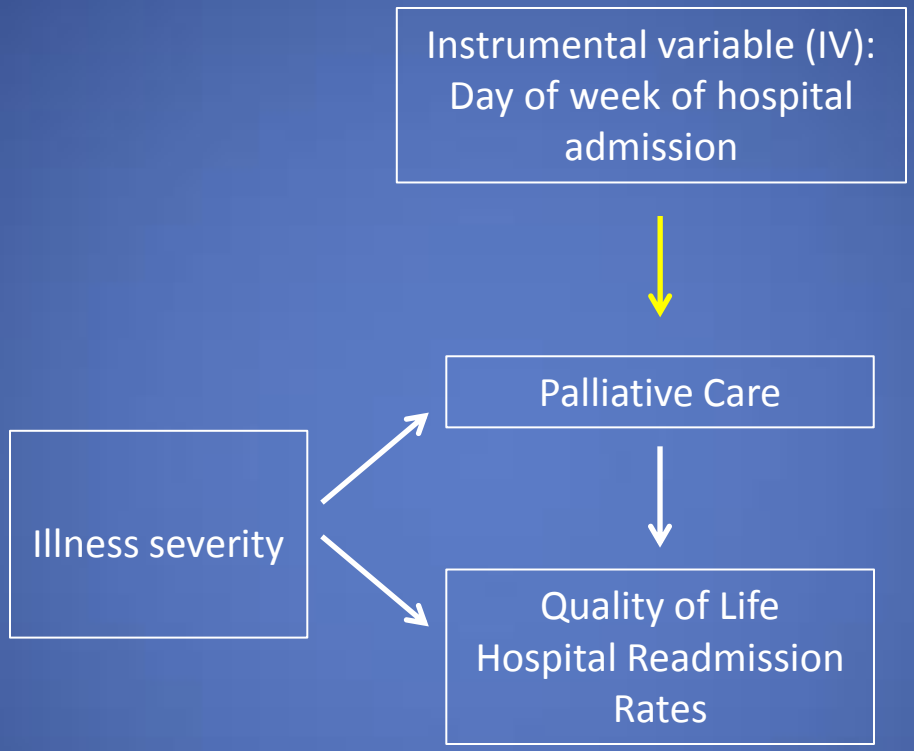
- Multivariable models



- Matching
- Propensity scores
- Instrumental variables
- Regression discontinuity
- Difference-in-differences



Instrumental Variable Analyses



"…Finding a little RCT inside a lot of observational data"

Pizer 2016. HSR. 51: 790-811

What Makes a Good Instrument?

- Related to treatment likelihood
  - *F-statistic and partial r^2*
- Not independently related to outcome (exclusion restriction)
  - *Falsification tests*
- Unrelated to other patient characteristics
  - *Standardized differences*

Brookhart et al. 2010. Pharmacoepidemiology and Drug Safety 19: 537-554.

What IV Analysis Can & Cannot Do

- IV analysis can:
  - Reduce selection bias due to both ***measured and unmeasured*** confounders
  - Estimate treatment effect for individuals who may or may not get treatment, depending on the value of the IV

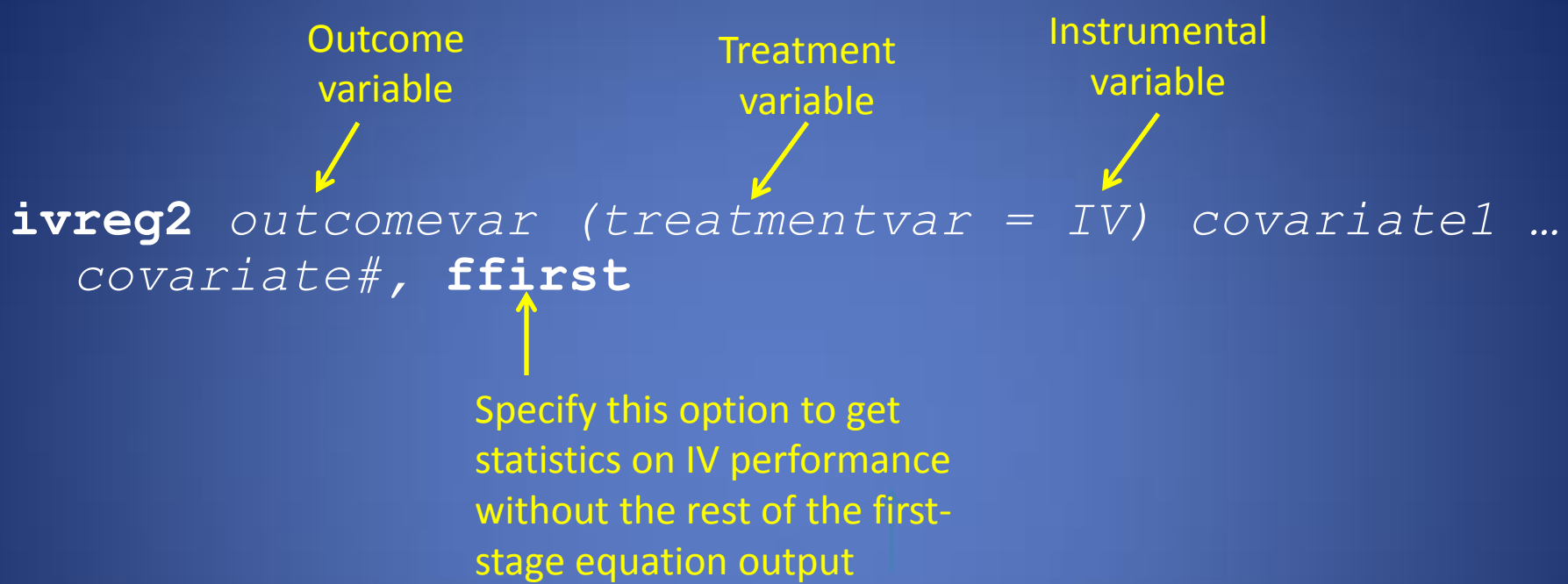
- IV analysis cannot:
  - Generalize to individuals who would not be sensitive to the value of the instrumental variable

Instrumental Variable Methods: Two-Stage Least Squares (2SLS) and Control Functions

- Step 1: Model treatment likelihood, include instrumental variable

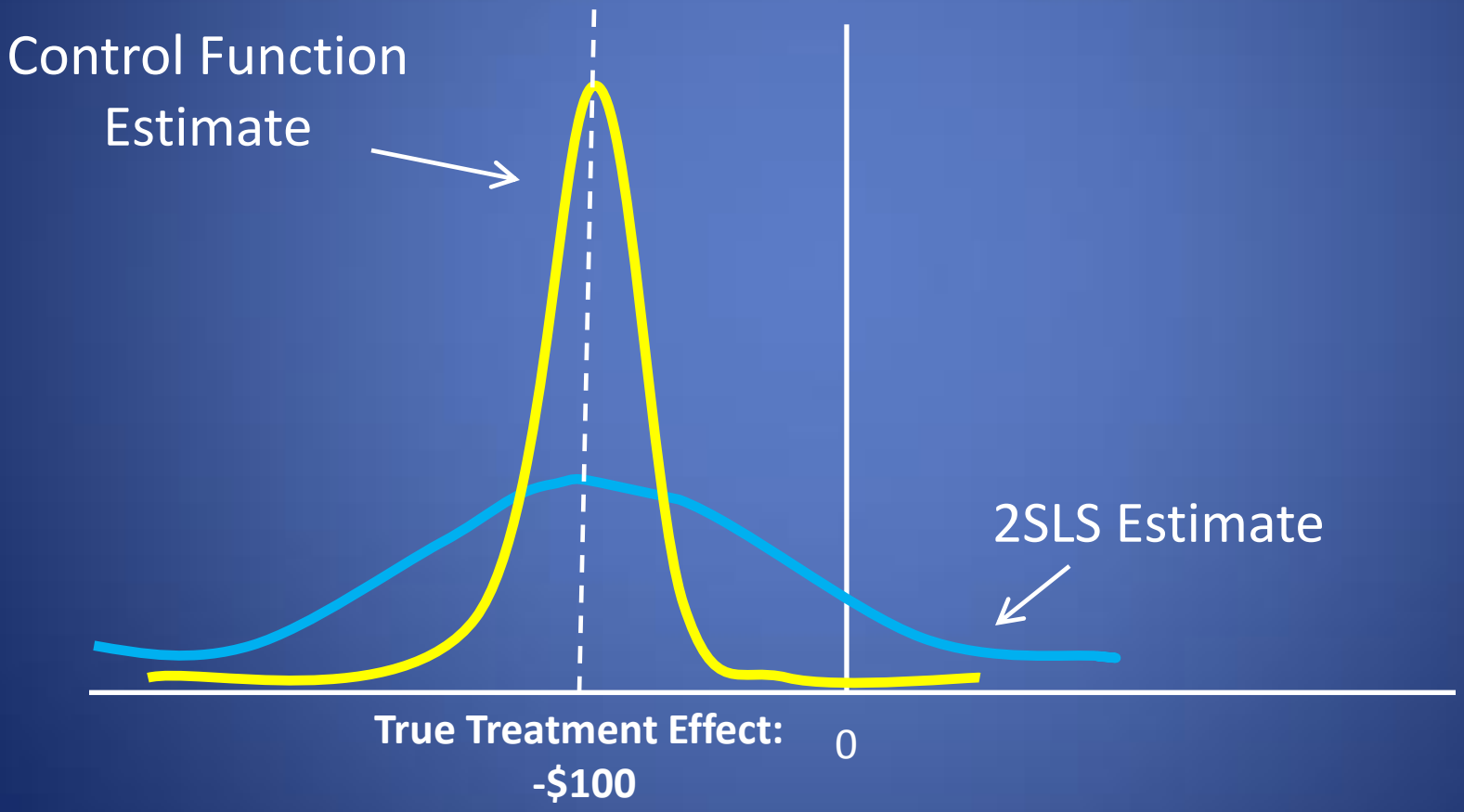
- Step 2: Model outcome
  - 2SLS: Include treatment likelihood from Step 1
  - Control Function: Include a *function of the residuals* from Step 1

Stata Code to Perform 2SLS



*-`ivreg2`- is not part of Stata's built-in commands. Type "`ssc install ivreg2, replace`" in Stata's command line

Estimates from Control Functions More Efficient than Estimates from Two-Stage Least Squares Models



Stata Code for Control Functions: Two-stage residual inclusion

Model treatment likelihood, include IV

```
qui glm ivreg2 treatmentvar IV covariate1 …
covariate#, f(family) link(link)
```

Predict the residual from the treatment likelihood equation

```
predict treatment_res, response
```

Model outcome, include residual from treatment likelihood equation

```
glm outcomevar treatmentvar covariate1..covariate#
treatment_res, f(family) link(link)
```

Generate marginal effects and calculate bootstrapped standard errors

Interpreting Results of IV Analyses

- Generalize to individuals similar to those included in the matched sample

- Local ATE or local ATT

Falsification tests

- Cannot prove the exclusion restriction
- Falsification tests can strengthen argument that exclusion restriction is valid
- Rerun analyses in situations where treatment should not have an effect, but potential confounders might have an effect
  - Alternate outcome
  - Alternate population
- If no evidence of an effect from confounders, strengthens confidence in IV results

Pizer 2016. HSR. 51: 790-811

Summary

- Observational data can be rich source of information for improving patient outcomes
- Many tools to improve treatment effect estimation from observational data
- Important to understand assumptions, generalizability, and limitations of each tool

Questions?

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.

Brookhart MA et al. Instrumental variable methods in comparative safety and effectiveness research. Pharmacoepidemiology and Drug Safety 2010; 19: 537-554.

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 et al. Methods for constructing and assessing propensity scores. HSR 2014; 49:1701-1720.

Ho DE et al. Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. Political Analysis 2007; 15: 199–236.

Imai K, Ratkovic M. Covariate balancing propensity score. J R Statist. Soc. B. 2014; 76(1): 243-246.

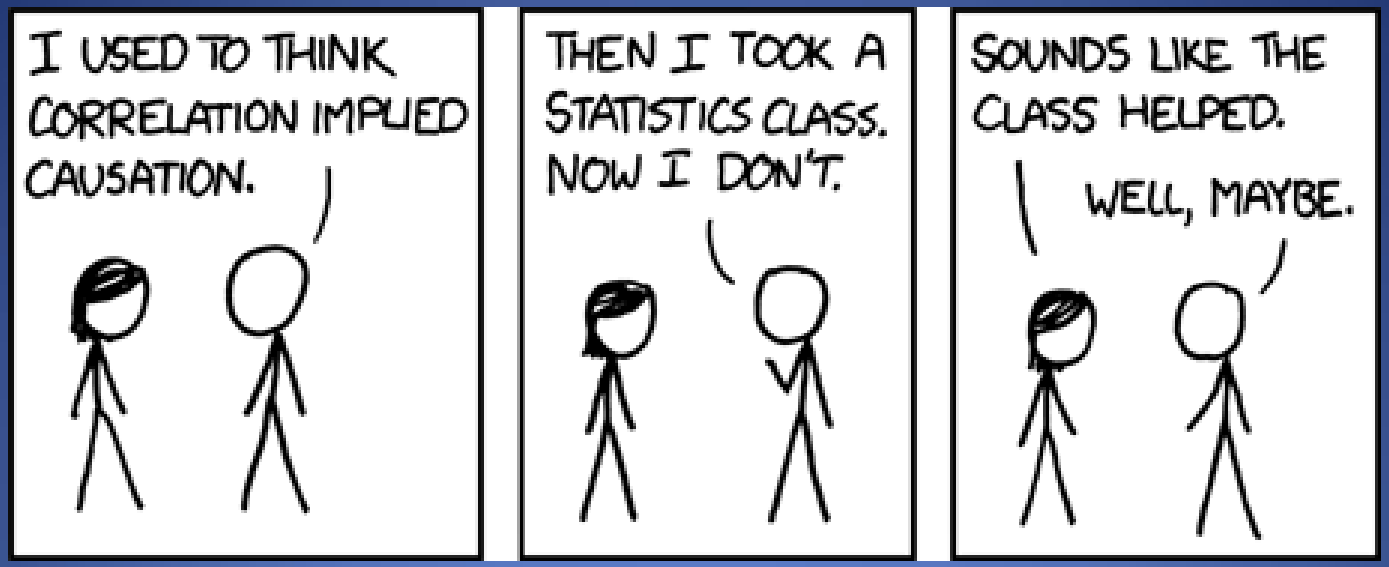
King G. 2015. http://gking.harvard.edu/publications/why-propensity-scores-should-not-be-used-formatching

Liu W et al. An introduction to sensitivity analysis for unobserved confounding in nonexperimental prevention research. Prev Sci 2013;14(6):570-80.

Pizer S. Falsification testing of instrumental variables methods. HSR 2016; 51(2): 790-811.

Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects**.** Biometrika. 1983; 70: 41-45.

Stuart EA. Matching methods for causal inference: A review and look forward. Statistical Science 2010; 25 (1): 1–21.



xkcd.com

melissa.garrido@va.gov
@GarridoMelissa

Acknowledgements

- VA HSR&D (CDA 11-201, CDP 12-255)
- PC4C Study Group (NCI/NINR 5R01CA116227, National Palliative Care Research Center)
- Melissa Aldridge, PhD MBA; Helene Starks, PhD
- 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.