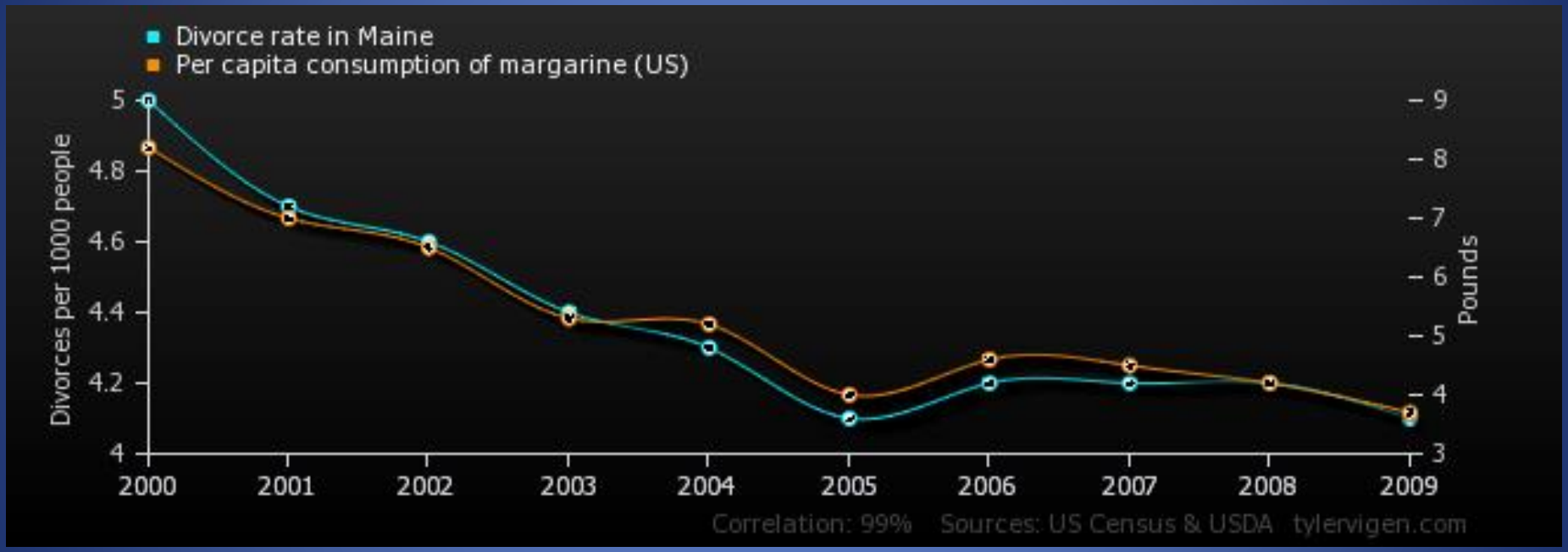
Making Sense of Observational Data



Melissa Garrido, PhD

Department of Veterans Affairs and Boston University School of Public Health

October 16, 2018

garrido@bu.edu
@GarridoMelissa

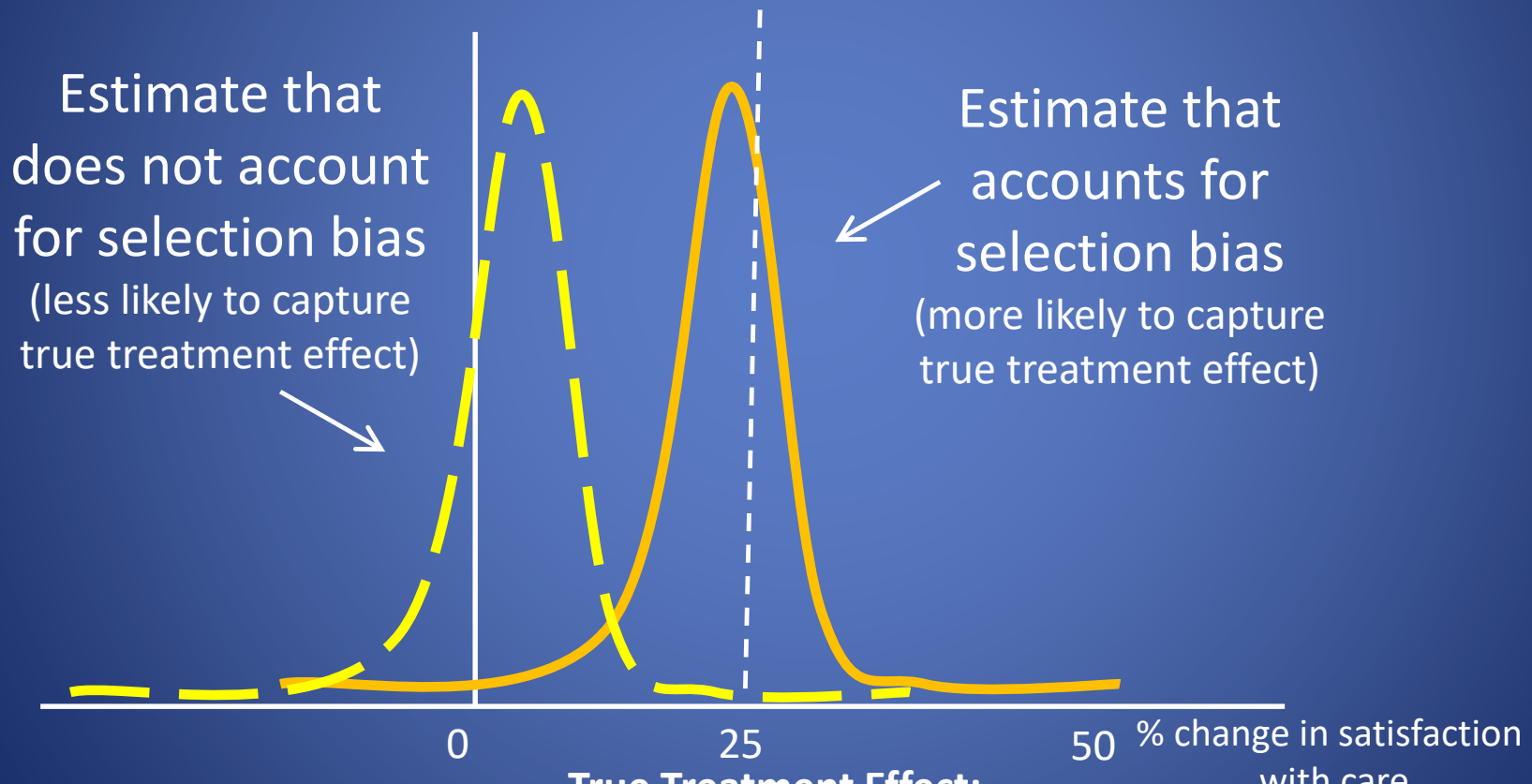
Confounding due to Selection Bias in Observational Data

Patients not randomized to treatment

↓

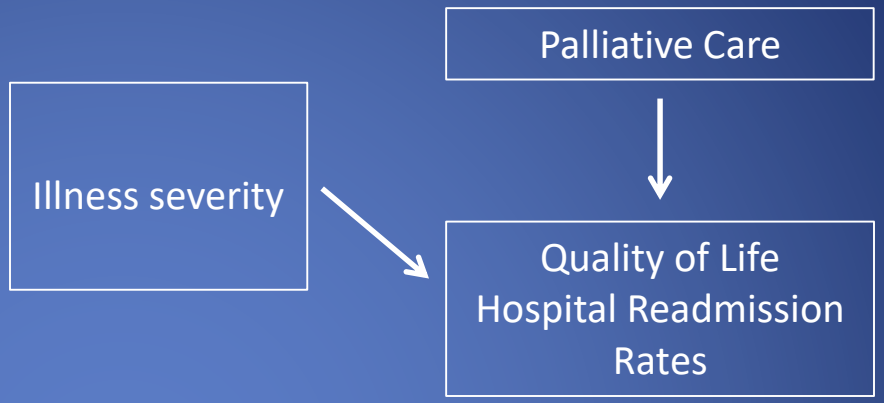
Patient characteristics may be associated with both participation in treatment *and* outcome

Impact of Selection Bias on Analytic Inferences

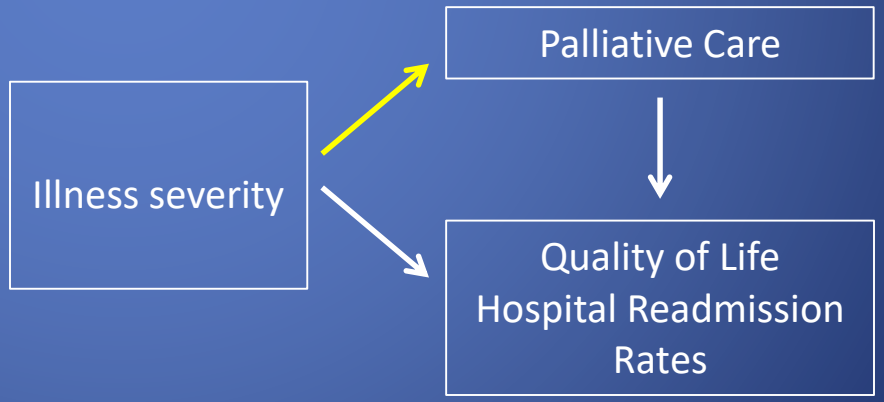


Tools to Address Confounding

- Multivariable models

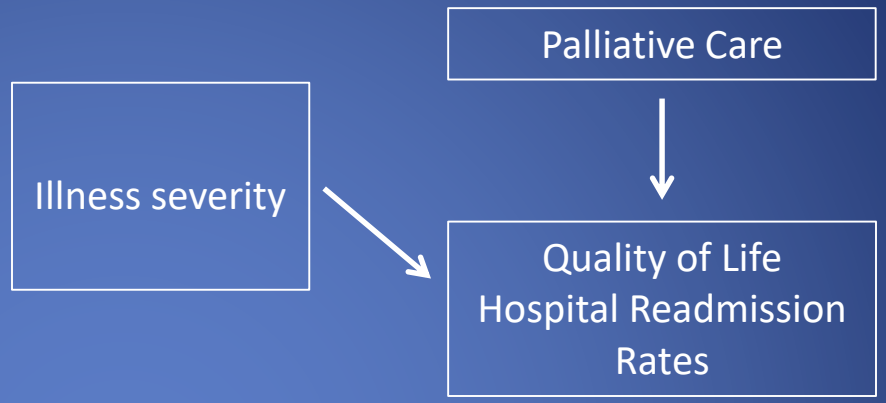


- Matching
- Propensity scores
- Coarsened exact matching
- Entropy balancing
- Instrumental variables
- Regression discontinuity
- Difference-in-differences

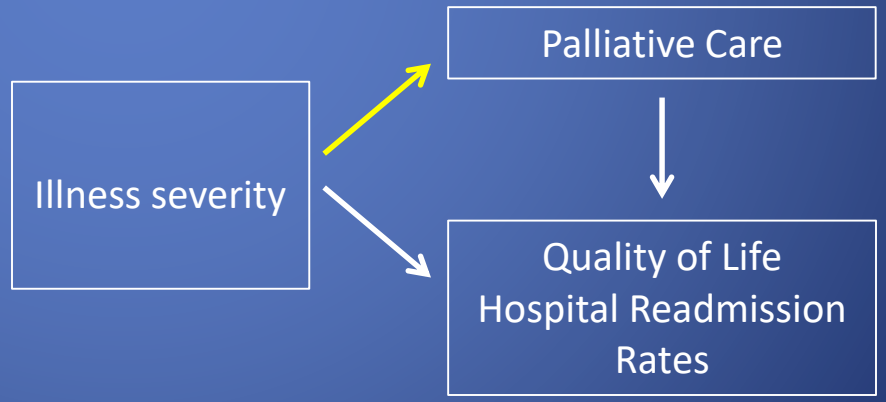


Tools to Address Confounding

- Multivariable models



- Matching
- Propensity scores
- Coarsened exact matching
- Entropy balancing
- Instrumental variables
- Regression discontinuity
- Difference-in-differences



Addressing Selection Bias by "Pre-Processing" Datasets

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

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

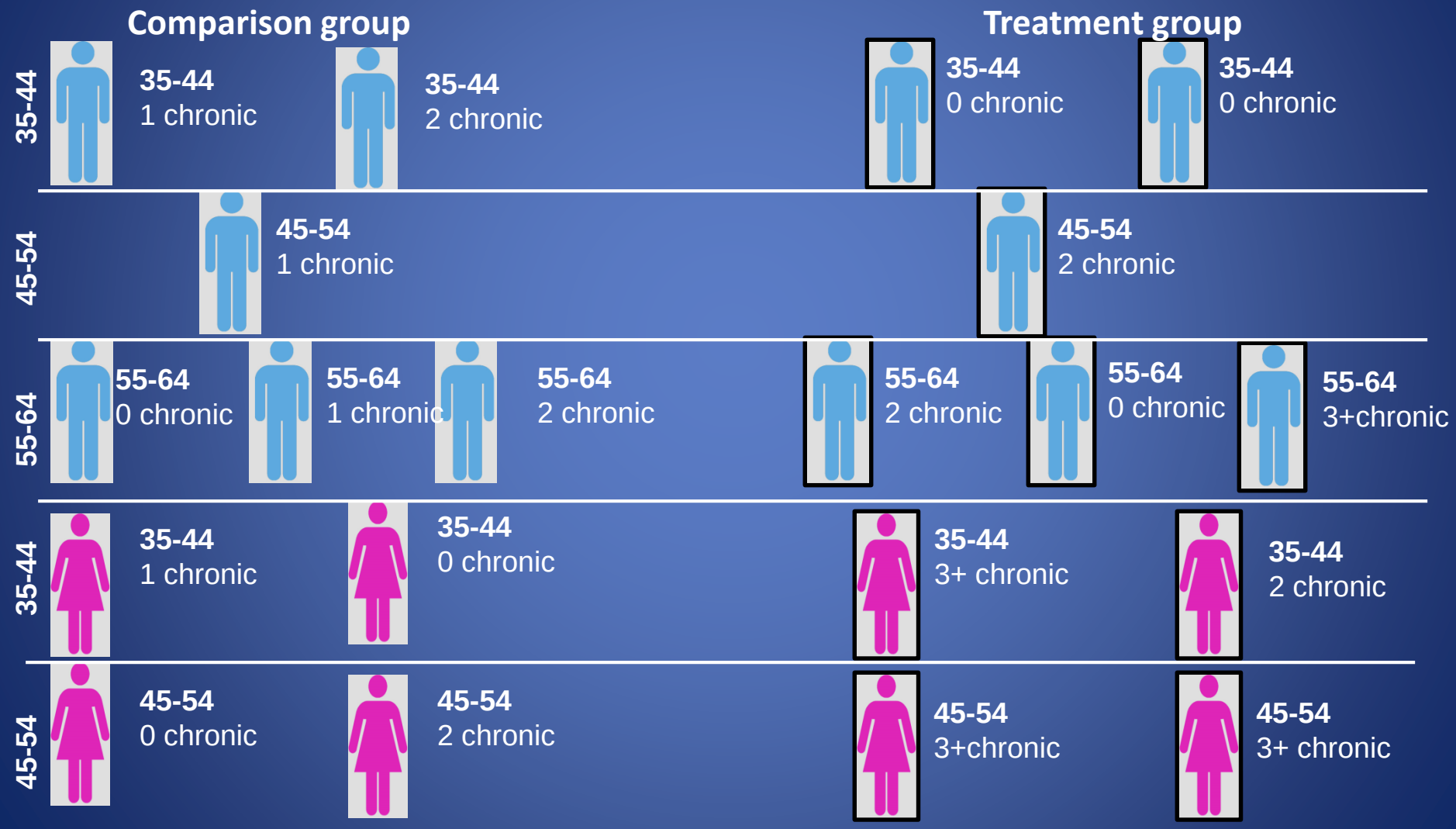
Ho et al. 2007. Political Analysis 15: 199-236
Stuart 2010. Statistical Science 25: 1-21.

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

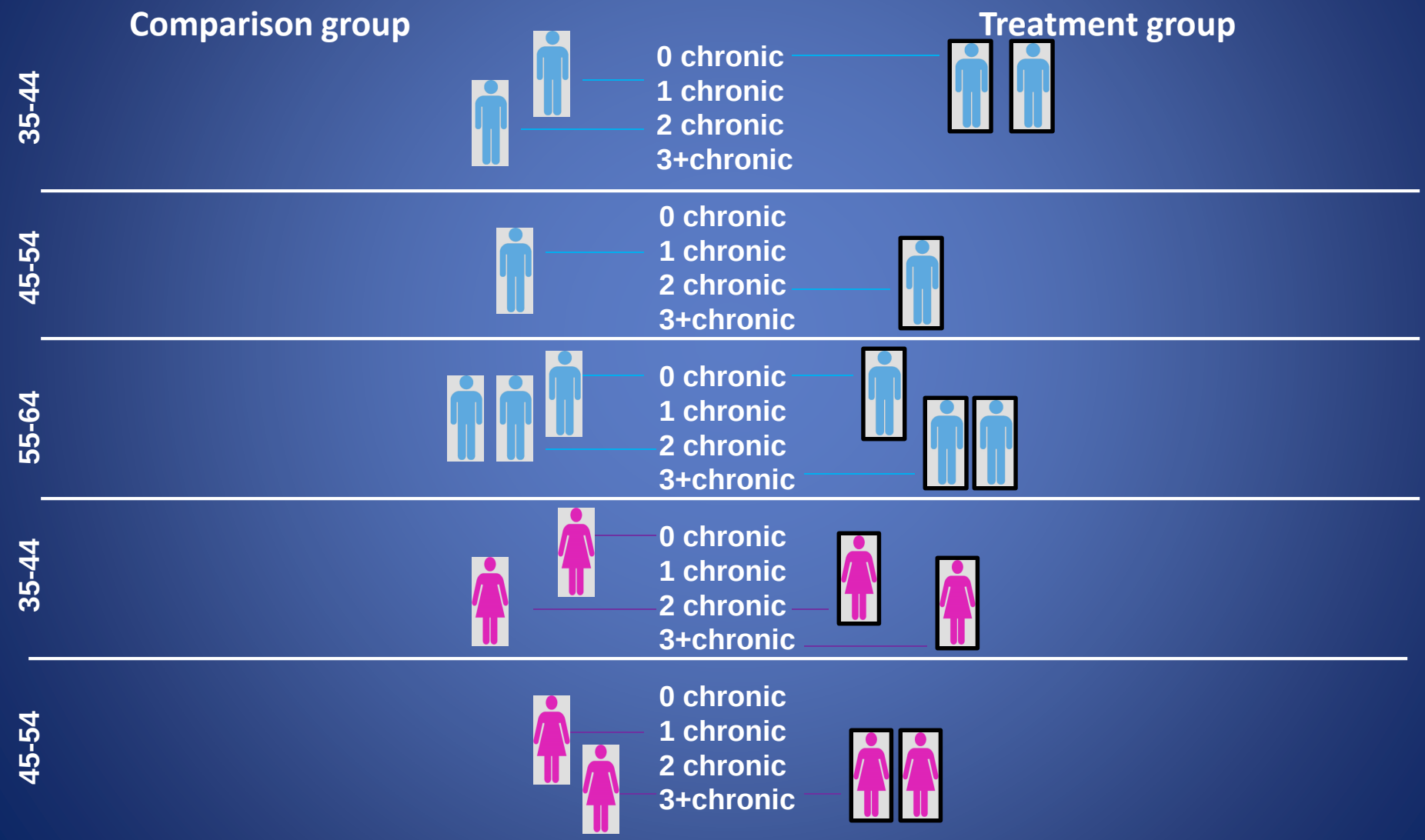
Matching on Specific Variables:

Match on gender and age



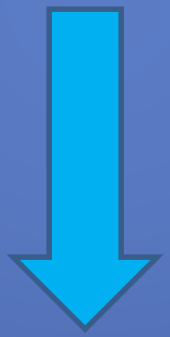
Matching on Specific Variables:

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

Estimate propensity score

Choosing Variables for Propensity Scores

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

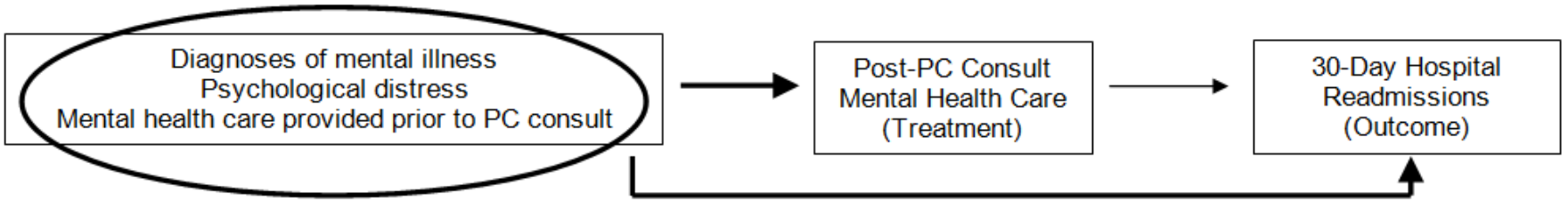
- Do not include:
  - Variables hypothesized to be associated with treatment but not with outcome
  - Variables that may be affected by the treatment
  - Variables that predict treatment status perfectly

Variable Selection Example

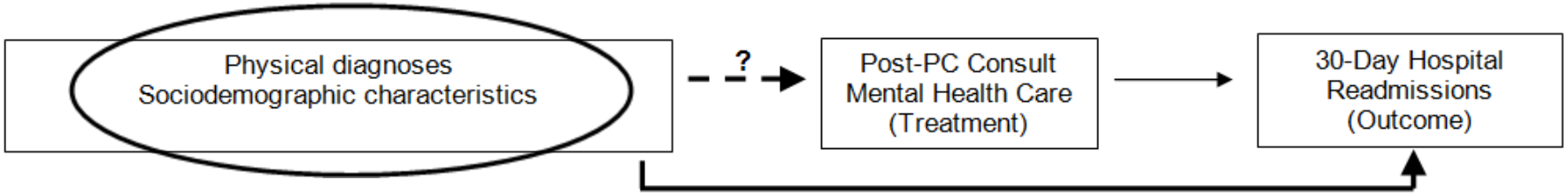
- Hospitalized veterans receiving a palliative care consultation in a VISN 3 acute care facility
- Treatment: Psychotherapy provided after a palliative care consultation
- Outcome: All-cause 30-day readmission

Choosing Variables for Propensity Score Models

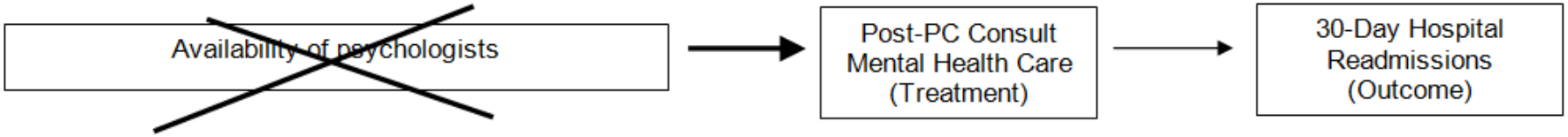
Guideline 1: INCLUDE variables hypothesized to be strongly associated with both treatment and outcome



Guideline 2: INCLUDE variables hypothesized to be associated with outcome but that may or may not be associated with treatment

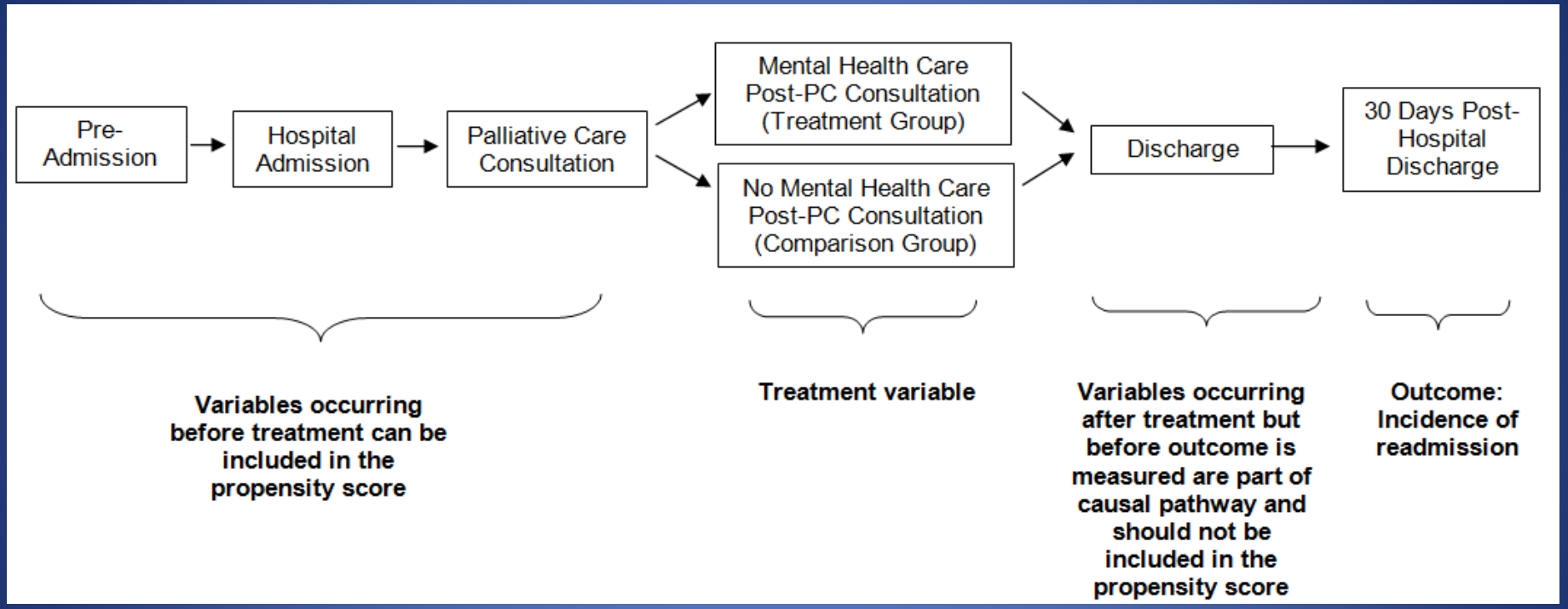


Guideline 3: DO NOT INCLUDE variables that are hypothesized to be only associated with treatment (instrumental variables)



Garrido 2014. JPSM 48:711-718

Choosing Variables for Propensity Score Models



Garrido 2014. JPSM 48:711-718

Calculate Propensity Score

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

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

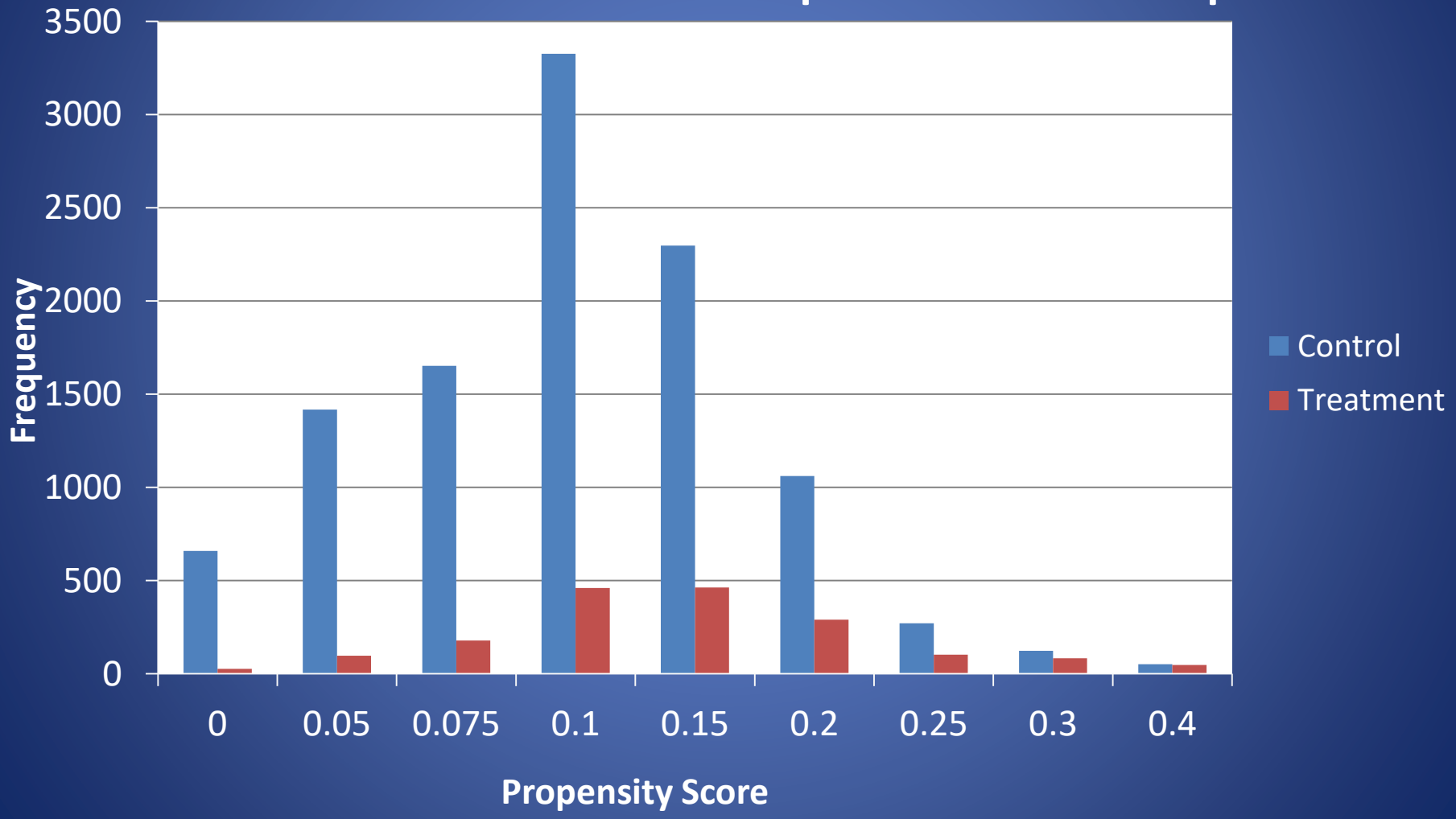
Step 5: Ensure that covariates are balanced across treatment and comparison groups in sample matched or weighted by propensity score

↓

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

Check range of common support
Check balance of propensity score

Initial Checks: Common Support and Balance across Treatment and Comparison Groups



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

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

Assess Balance with Standardized Differences

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

- Do not use t-tests

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

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

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

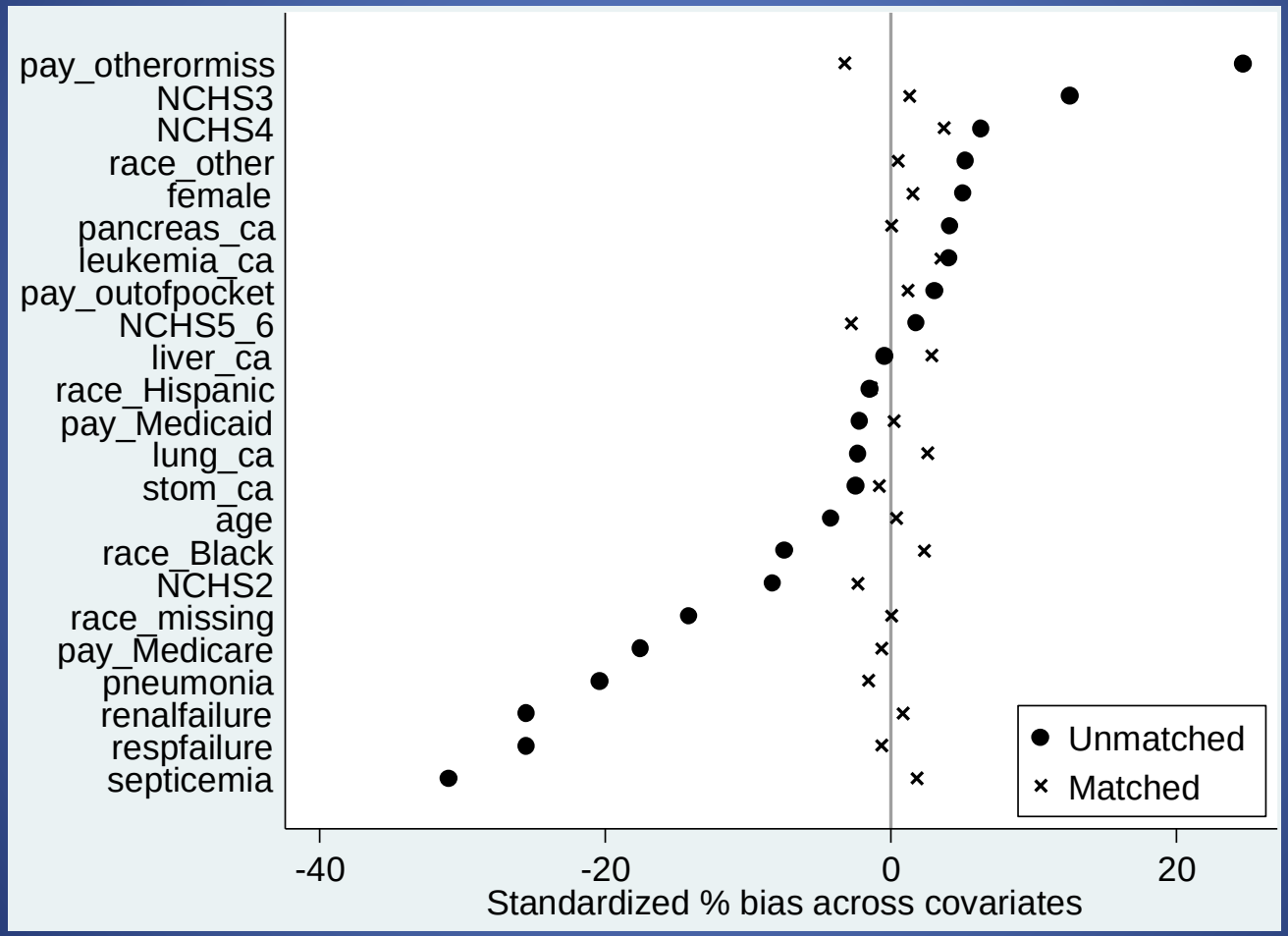
↓

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

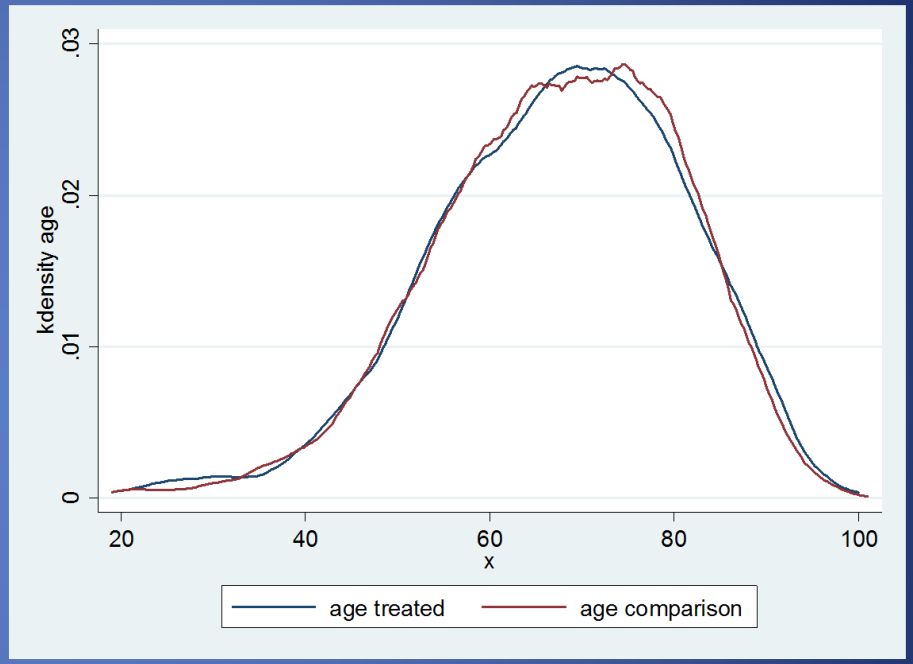
- Standardized differences
- Graphs
  - Quantile-quantile plots
  - Plots of covariates in treated and comparison groups
- Ratios of variance

Visual Inspection of Standardized Differences



Plots of Covariates in Treated and Comparison Groups

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



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

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
- Choice impacts how propensity score weights are constructed

Need to Correct Standard Errors for Treatment Effect Estimates

- Ignoring uncertainty
  - Makes standard errors for ATEs more conservative
  - Makes 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

Interpretation of Treatment Effect Estimates From Propensity Score Analyses

- Generalizability
- Meaning of other coefficients in the model
- Sensitivity to unobserved confounding

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

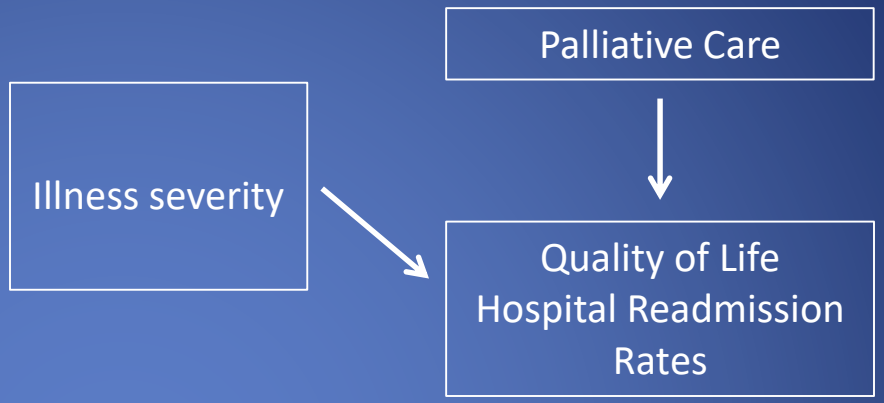
Liu et al 2013. Prevention Science 14: 570-580

Checklist: Crucial Information on Propensity Score Analyses to include in Grants or Papers

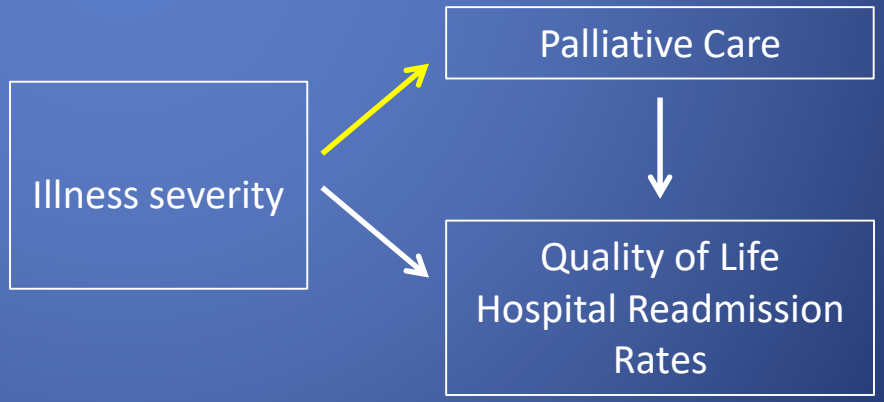
- ✓ Rationale for choosing propensity scores
- ✓ Rationale for variable choice
- ✓ Lists method of propensity score creation and matching/weighting strategy
- ✓ Assessed covariate balance with standardized differences
- ✓ No c-statistics or other model fit statistics for the propensity score model
- ✓ Multivariable regression run on sample matched or weighted by propensity score
- ✓ Standard error calculation applied appropriately
- ✓ Treatment effect (ATT or ATE) specified
- ✓ Generalizes results to appropriate population

Tools to Address Confounding

- Multivariable models



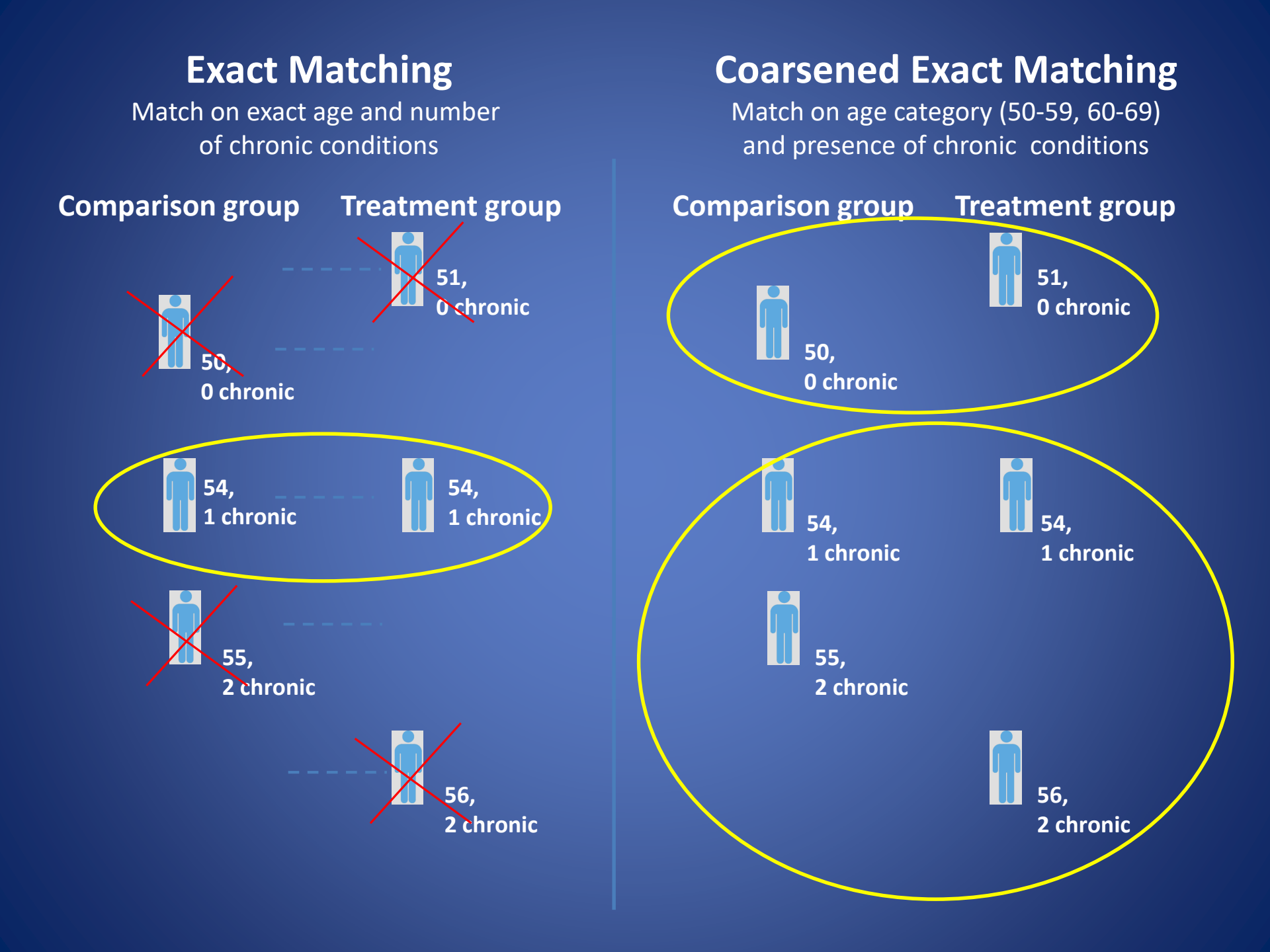
- Matching
- Propensity scores
- Coarsened exact matching
- Entropy balancing
- Instrumental variables
- Regression discontinuity
- Difference-in-differences



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
Comparison group
Treatment group
51, 0 chronic
50, 0 chronic
54, 1 chronic
54, 1 chronic
55, 2 chronic
56, 2 chronic
Coarsened Exact Matching
Match on age category (50-59, 60-69) and presence of chronic conditions
Comparison group
Treatment group
51, 0 chronic
50, 0 chronic
54, 1 chronic
54, 1 chronic
55, 2 chronic
56, 2 chronic

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

Interpreting Results of Analyses Using CEM

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

- ATT

CEM Example

- Question: Is participation in a mental health self-direction program associated with an increase in days worked with pay?
- Dataset: All adults in Florida with a documented serious and persistent mental illness
- Potential for confounding: What factors might be associated with voluntary enrollment in this program and with increased employment?

Croft et al. Psychiatric Services 2018; 69:819–825

CEM Example

- Variables included in matching:
  - Age
  - High school completion
  - Gender
  - Race/ethnicity
  - Schizophrenia diagnosis
  - Substance use disorder diagnosis
  - Marital status
  - County of residence
  - Veteran status
  - Limited English proficiency
  - Ever arrested during study period
  - Ever assessed as having an ADL limitation during study period
  - Ever spent one or more days outside of community during study period
  - Days between first and last assessments
  - Disability income receipt
- Identified matches for 67% of treatment group

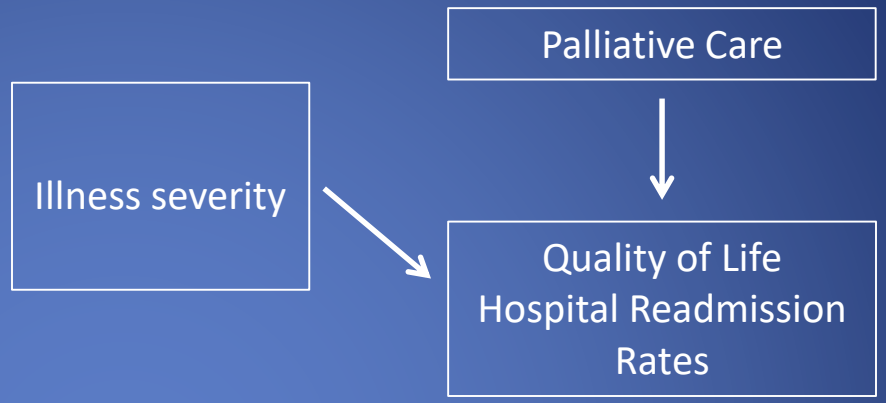
Croft et al. Psychiatric Services 2018; 69:819–825

Checklist: Crucial Information on CEM to include in Grants or Papers

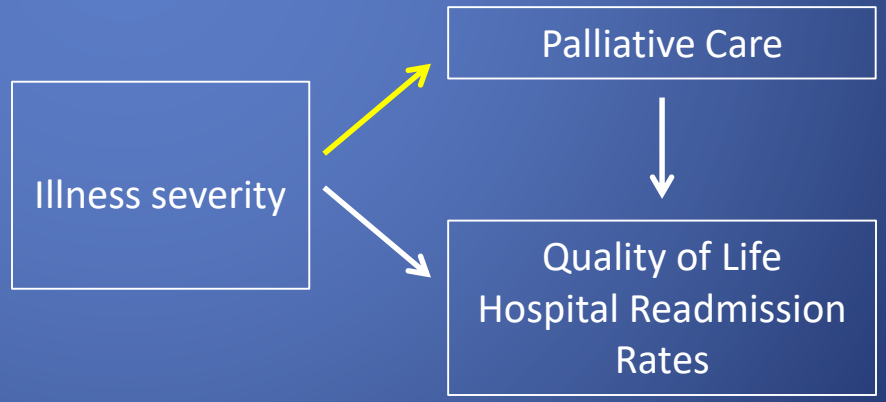
- ✓ Rationale for choosing CEM
- ✓ Rationale for variable choice
- ✓ Description of categorization of variables
- ✓ Assessed imbalance before matching with standardized differences
- ✓ Lists number of observations dropped from treatment and comparison groups
- ✓ Multivariable regression run on matched sample
- ✓ Treatment effect (ATT or ATE) specified
- ✓ Generalizes results to appropriate population

Tools to Address Confounding

- Multivariable models



- Matching
- Propensity scores
- Coarsened exact matching
- Entropy balancing
- Instrumental variables
- Regression discontinuity
- Difference-in-differences

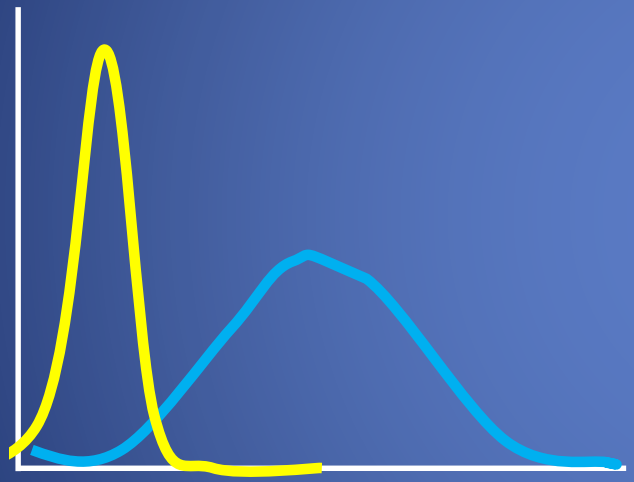


Entropy Balancing

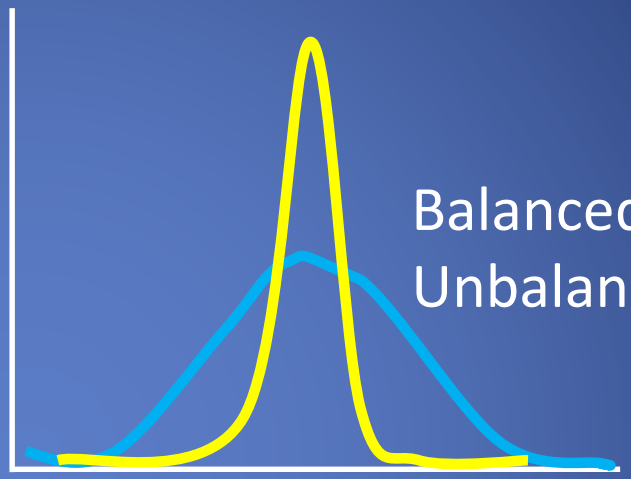
- Create treatment and comparison groups with similar moments (mean, variance, skew) of covariate distributions
- Eliminates step to verify covariate balance
- Not susceptible to worsened balance due to model misspecification (a strong risk with propensity score matching when data on important confounders are not available)
- Uses weights (fewer dropped observations than in methods based on matching)

Hainmueller. Political Analysis 2012: 20:25-46.

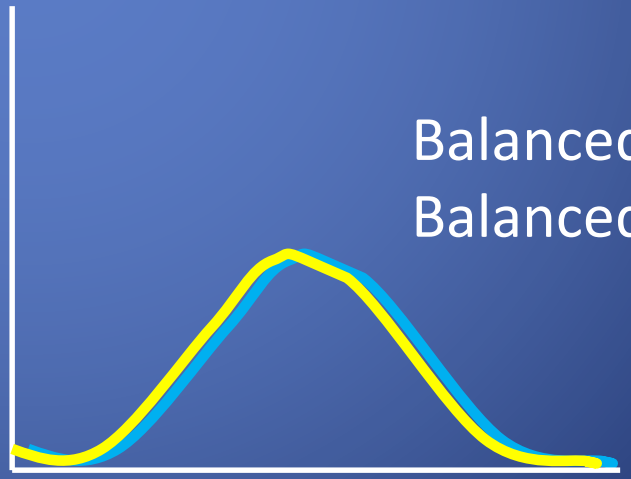
Entropy Balancing



Pre-balancing



Balanced mean
Unbalanced variance



Balanced mean
Balanced variance

What Entropy Balancing Can & Cannot Do

- Entropy balancing can:
  - Help create weights so that distributions of ***measured*** confounders are equal across treatment & comparison groups
  - Improve precision of treatment effect estimates

- Entropy balancing cannot:
  - Account for ***unmeasured*** confounders

Interpreting Results of Analyses Using Entropy Balancing

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

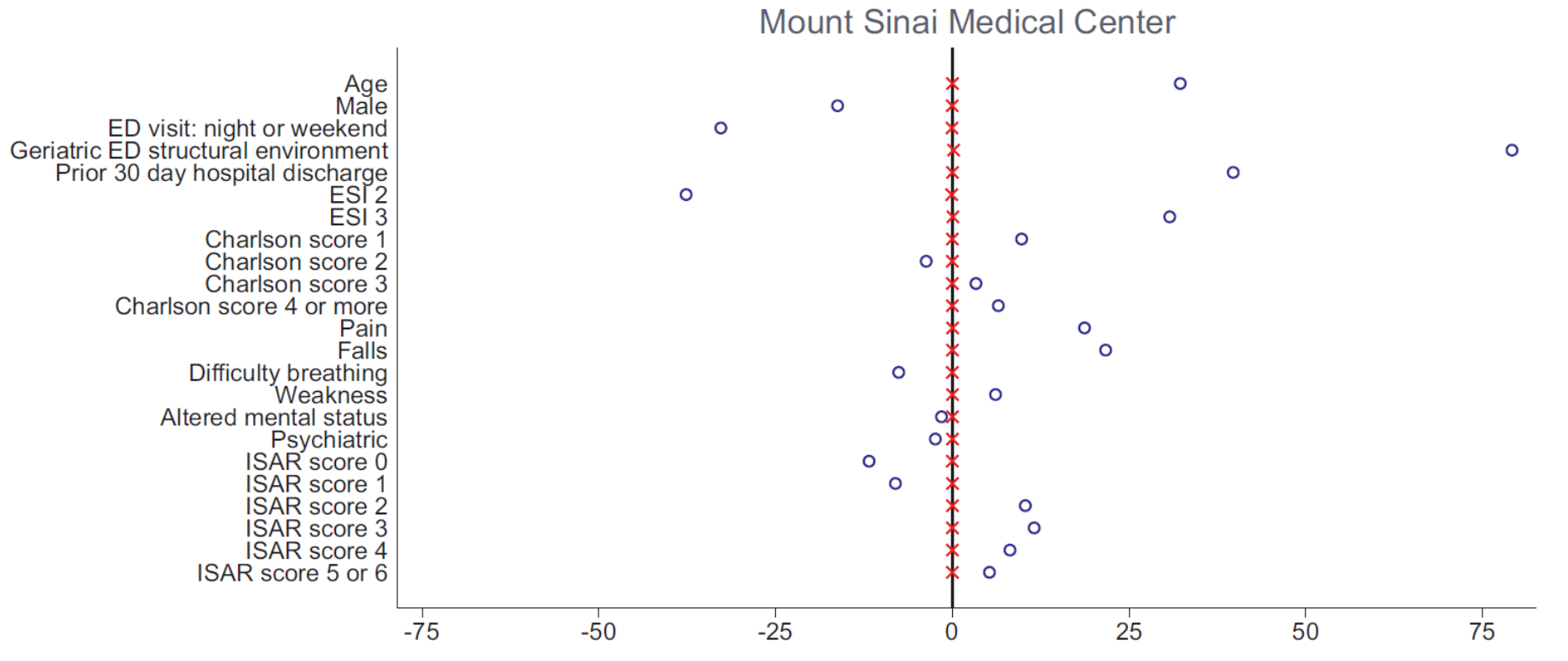
- ATT

Entropy Balancing Example

- Question: Is exposure to a transitional care nurse in the ED associated with reduced likelihood of inpatient admissions?
- Dataset: All patients 65 and older who visited a Mount Sinai ED from 1/1/2013-7/30/2015
- Potential for confounding: What factors might be associated with exposure to a transitional care nurse and inpatient admission?

Hwang et al. JAGS 2018; 66:459-466

Entropy Balancing Example



Graph of standardized differences before entropy balancing (blue circles) and after entropy balancing (X)

Hwang et al. JAGS 2018; 66:459-466

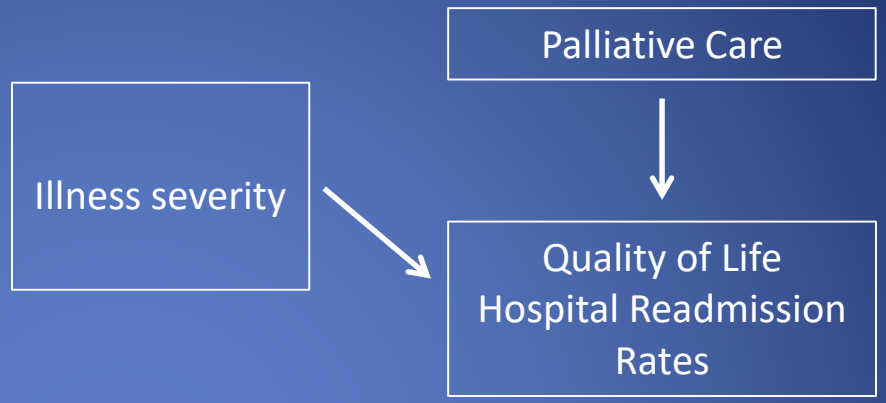
Checklist: Crucial Information on Entropy Balancing to include in Grants or Papers

- ✓ Rationale for choosing entropy balance
- ✓ Rationale for variable choice
- ✓ Assessed imbalance before matching with standardized differences
- ✓ Describes whether covariates were balanced on means only or also on other moments
- ✓ Balance constraints are listed and are reasonable
- ✓ There are more control observations than balance constraints
- ✓ Multivariable regression run on balanced sample
- ✓ Treatment effect (ATT or ATE) specified
- ✓ Generalizes results to appropriate population

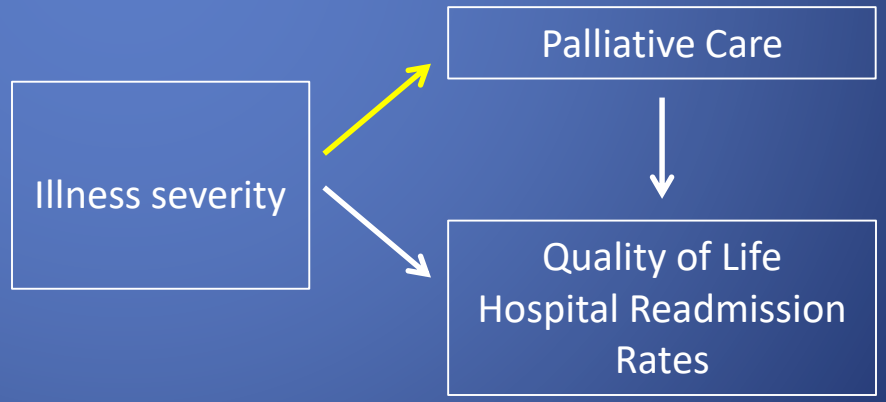
Hainmueller & Xu. Journal of Statistical Software. 2013; 54(7)

Tools to Address Confounding

- Multivariable models



- Matching
- Propensity scores
- Coarsened exact matching
- Entropy balancing
- Instrumental variables
- Regression discontinuity
- Difference-in-differences

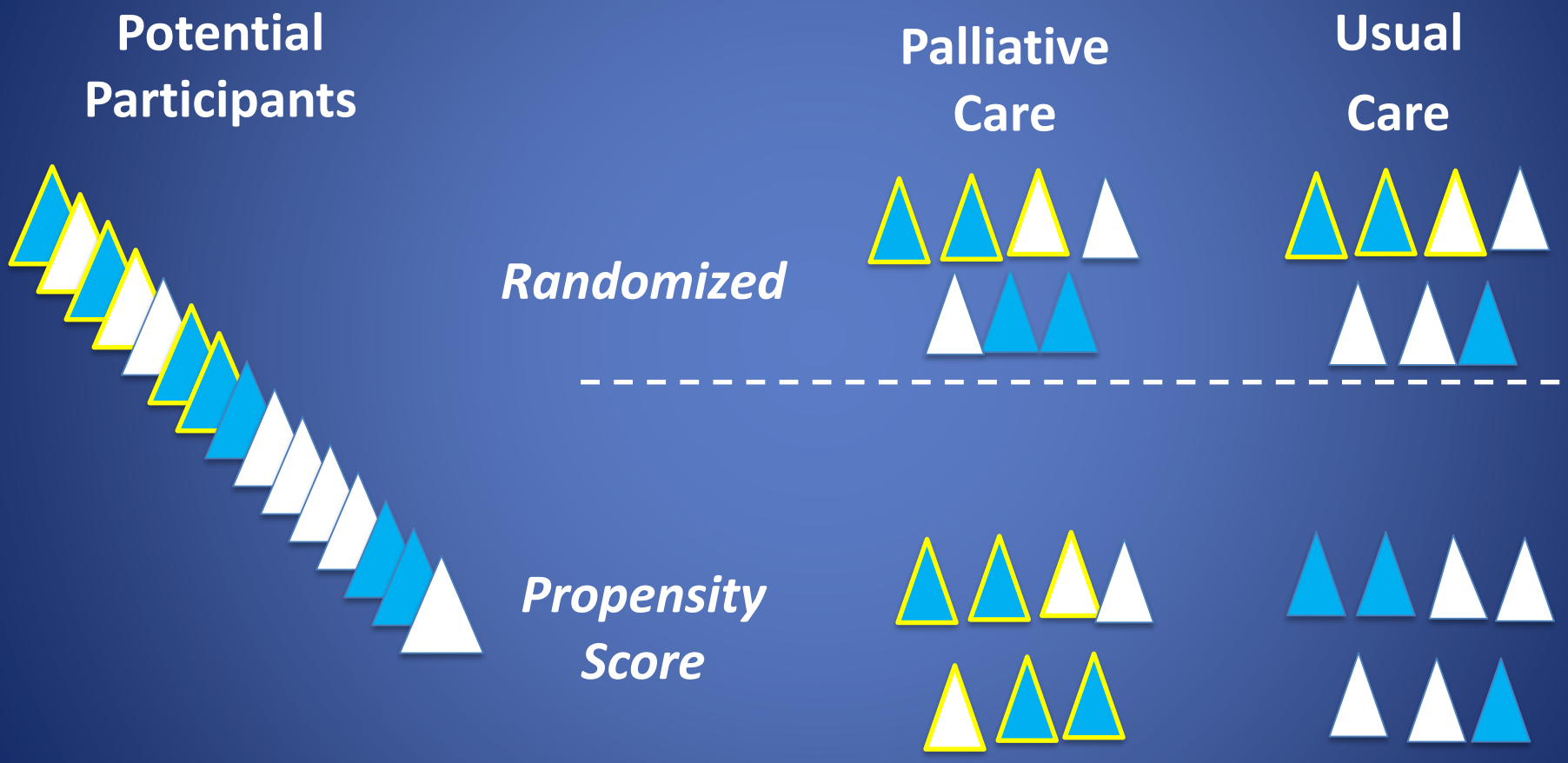


Why Might Pre-Processing and RCT Results Differ?

- Unobserved variables
- Analytic sample choice
- Treatment effect choice

Why Might Pre-Processing and RCT Results Differ?

Unobserved Variables



Why Might Pre-Processing and RCT Results Differ?
Analytic Sample Choice

Question: What is the impact of in-hospital mental health care on risk of readmission?

- After data collection, observational analysis with pre-processing would exclude:
  - Patients who would always receive the treatment
  - Patients who would never receive the treatment

- Before data collection, RCT cohort would exclude:
  - Patients who do not meet homogeneous diagnostic criteria
  - Patients who could not be ethically randomized to control group

Why Might Pre-Processing and RCT Results Differ?
Treatment Effect Choice

Average Treatment Effect (ATE)

vs

Average Treatment Effect on the Treated (ATT or ATET)

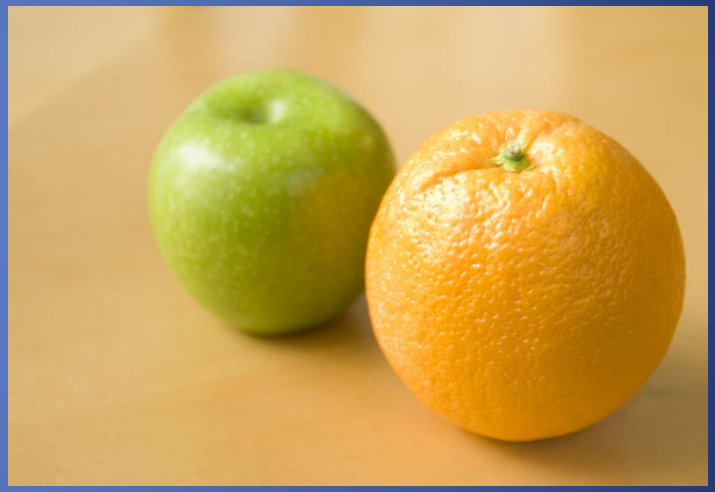
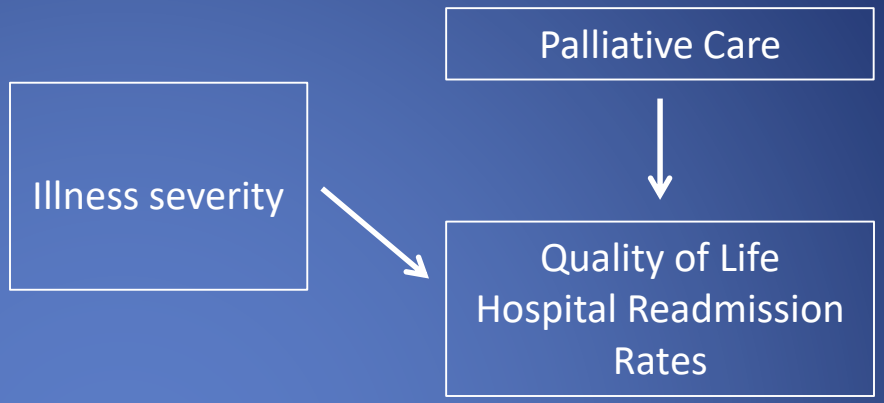


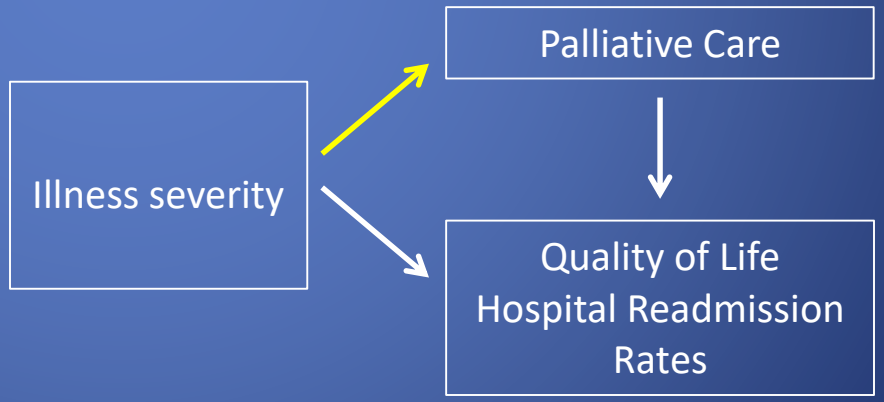
Image from http://flickr.com/photo/26176646@N04/2492945625

Tools to Address Confounding

- Multivariable models



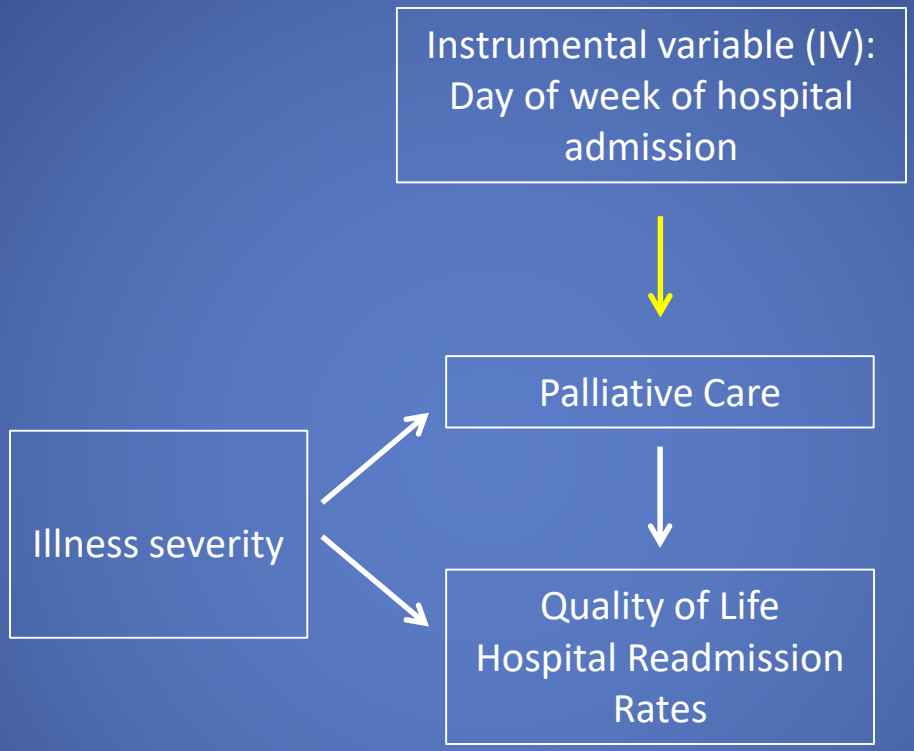
- Matching
- Propensity scores
- Coarsened exact matching
- Entropy balancing
- Instrumental variables
- Regression discontinuity
- Difference-in-differences



Instrumental Variable Analyses

- Requires identification of a variable (the instrument) that is associated with treatment but not the outcome
- Allows for estimation of treatment effect among individuals whose treatment receipt depends on the value of the instrument
- Accounts for both observed and unobserved confounders

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.

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

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

Interpreting Results of IV Analyses

- Generalize to individuals similar to those whose treatment receipt is sensitive to the value of the instrumental variable

- Local ATE or local ATT

Instrumental Variable Example

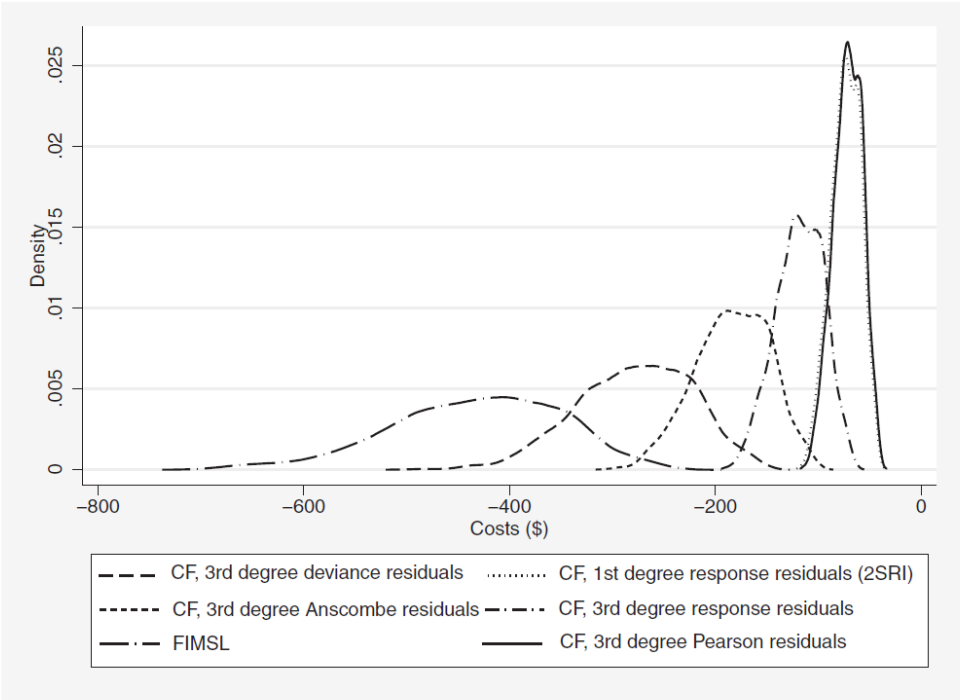
- Question: Is an inpatient palliative care consultation associated with reduced hospitalization costs?
- Dataset: Veterans with life-limiting illnesses admitted to a NY or NJ VA hospital in 2005-2006
- Potential for confounding: What factors might be associated with receipt of a PC consultation and hospitalization costs?
- Potential for instrumental variable: What factors might be associated with receipt of a PC consultation but not hospitalization costs?

Garrido et al. 2012. HSR 47(6): 2377-2397

Instrumental Variable Example

- Instrument: Physician likelihood of requesting a PC consultation
- How certain are we that this instrument is not independently associated with hospitalization costs?

Figure 3: LATE of a Palliative Care Consultation on Direct Costs per Day (CF and FIMSL Models)



CF, control function; FIMSL, full information maximum simulated likelihood; PC, palliative care; 2SRI, two-stage residual inclusion.

Garrido et al. 2012. HSR 47(6): 2377-2397

Falsification Tests

- Cannot prove the exclusion restriction (instrument not independently related to outcome)
- 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

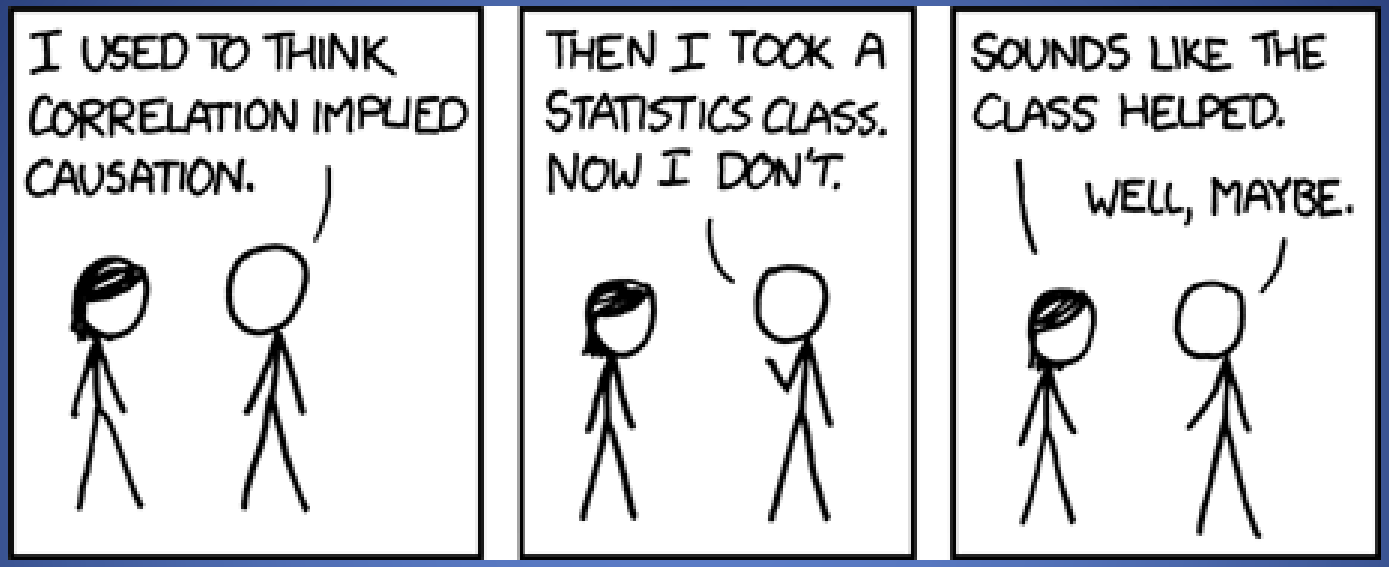
Checklist: Crucial Information on Instrumental Variables to include in Grants or Papers

- ✓ Rationale for choosing IV
- ✓ Theoretical rationale for choice of instrument
- ✓ Tests of instrument strength (how closely are the instrument and treatment probability related?)
- ✓ Tests of instrument's independence from other patient characteristics
- ✓ Falsification tests
- ✓ Treatment effect (local ATT or local ATE) specified
- ✓ Generalizes results to appropriate population

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?



xkcd.com

garrido@bu.edu
@GarridoMelissa

Funded in part by VA HSR&D IIR 16-140

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.

Resources

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.

Hainmueller J. Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. Political Analysis 2012: 20:25-46.

Hainmueller J & Xu Y. ebalance: A Stata package for entropy balancing. Journal of Statistical Software. 2013; 54(7)

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.