# Some Techniques to Deal with Confounding in Observational Studies

| Technique | Strengths | Weaknesses |
|---|---|---|
| Matching | Simple | Difficult to find matches; requires good understanding of confounders involved; Inability to examine effect of confounders used for matching |
| Stratification | Simple; Ability to examine effect modifiers | Difficult to interpret study results when many subgroups; Requires good understanding of confounders involved |
| Multivariable regression | Can include many confounders; Can examine effects of individual confounders; ability to examine multilevel effects | More complicated analysis because lots of variables; Possibility of missing effect modification |
| Propensity score | Single number generated | Potential to compare very different patients with similar scores, which may be misleading |
| Instrumental variable | Only a single variable is needed | Difficult to ensure variable is not associated with the outcome |

# Propensity Score: Definition

- Subject's probability of receiving a specific treatment conditional on the **observed** covariates
- Reduces baseline information to a single composite summary of the covariates

$$\ln\left(\frac{PS_i}{1-PS_i}\right)=\beta_0+\beta_1X_{1i}+\beta_2X_{2i}+\ldots+\beta_pX_{pi}$$

$$PS_i=\frac{\exp\left(\beta_0+\beta_1X_{1i}+\beta_2X_{2i}+\ldots+\beta_pX_{pi}\right)}{1+\exp\left(\beta_0+\beta_1X_{1i}+\beta_2X_{2i}+\ldots+\beta_pX_{pi}\right)}$$

# Stukel et al. (2007)

- **Research question: Does invasive cardiac treatment improve long-term survival?**
- US cohort of 120K+ patients who were elderly (65-84 years), receiving Medicare, and hospitalized with acute myocardial infarction (AMI) in 1994-1995, and who were eligible for cardiac catheterization
- Patients were followed for up to 7 years to assess long-term survival
- The treatment is cardiac catheterization within 30 days of hospital admission

# Stukel et al. (2007)

- Analysis: Cox proportional hazards model
- Confounding variables
  - 65 patient, hospital, and ZIP code characteristics associated with post-AMI mortality
- Methods to address confounding:
  - Multivariable Cox model
  - Propensity score matching and conditional Cox model
  - Propensity score Cox model
  - Instrumental variable Cox model

# An Example: Stukel et al. (2007)

- PS matching and conditional Cox model
  - Patients receiving cardiac catheterization were matched to the closest control whose propensity score differed by less than 0.05, 0.10, or 0.15 among those patients within 5 years of age
  - Conditional Cox proportional hazards model was then applied to the matched data
- PS Cox model
  - Propensity scores were categorized into deciles
  - PS decile was included as a covariate in the Cox proportional hazards model

# An Example: Stukel et al. (2008)

**Table 2.** Distribution of Select Covariates by Propensity Score Deciles, According to Receipt of Cardiac Catheterization

| | Decile (Range) of Propensity Score* | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1 (0.00-0.26) | 2 (0.26-0.40) | 3 (0.40-0.50) | 4 (0.50-0.58) | 5 (0.58-0.65) | 6 (0.65-0.70) | 7 (0.70-0.75) | 8 (0.75-0.80) | 9 (0.80-0.85) | 10 (0.85-0.98) |
| No. of patients | | | | | | | | | | |
| No cardiac catheterization | 10021 | 8219 | 6873 | 5763 | 4834 | 3997 | 3283 | 2628 | 2060 | 1208 |
| Cardiac catheterization | 2191 | 3993 | 5340 | 6449 | 7378 | 8215 | 8930 | 9585 | 10151 | 11006 |
| Predicted 1-year mortality, %† | | | | | | | | | | |
| No cardiac catheterization | 54.5 | 39.2 | 31.8 | 27.5 | 23.4 | 20.0 | 17.3 | 15.3 | 14.0 | 13.6 |
| Cardiac catheterization | 51.2 | 38.9 | 31.8 | 27.4 | 23.5 | 20.0 | 17.3 | 15.3 | 13.5 | 12.8 |
| Mean age, y‡ | | | | | | | | | | |
| No cardiac catheterization | 79.4 | 78.0 | 77.0 | 75.5 | 74.3 | 72.9 | 71.9 | 70.8 | 70.1 | 70.0 |
| Cardiac catheterization | 79.3 | 77.9 | 76.8 | 75.7 | 74.3 | 73.0 | 71.8 | 70.9 | 70.0 | 69.9 |
| History of congestive heart failure, % | | | | | | | | | | |
| No cardiac catheterization | 59.8 | 40.0 | 27.0 | 18.8 | 10.8 | 7.3 | 4.2 | 2.7 | 2.0 | 2.1 |
| Cardiac catheterization | 61.4 | 40.0 | 26.5 | 16.7 | 10.5 | 5.7 | 3.6 | 2.5 | 2.0 | 1.7 |

*Propensity scores were rounded to 2 decimal points. There was no overlap across deciles.
†Predicted 1-year mortality was computed using the Cox proportional hazards regression model, including all baseline patient characteristics of age, sex, race, socioeconomic status, comorbidities, and clinical presentation.
‡SD for age was 4.3 years.

# An Example: Stukel et al. (2008)

**Table 3.** Adjusted Relative Mortality Rate Associated With Receipt of Cardiac Catheterization Among Patients With AMI Using Standard Risk-Adjustment Methods

| Risk-Adjustment Method | Relative Mortality Rate (95% CI) |
|---|---|
| Unadjusted survival model | 0.364 (0.358-0.370) |
| Multivariable survival model (65 covariates) | 0.510 (0.502-0.519) |
| Survival models using simple propensity score* | |
| Propensity deciles alone | 0.538 (0.529-0.547) |
| Propensity deciles plus all covariates | 0.520 (0.511-0.529) |
| Survival models using complex propensity score† | |
| Propensity deciles alone | 0.540 (0.531-0.549) |
| Propensity deciles plus all covariates | 0.522 (0.513-0.531) |
| Survival models using propensity-based matching cohort | |
| Match within ±0.05 of propensity score and 5 y of age | 0.538 (0.518-0.558) |
| Match within ±0.10 of propensity score and 5 y of age | 0.528 (0.514-0.542) |
| Match within ±0.15 of propensity score and 5 y of age | 0.511 (0.499-0.523) |

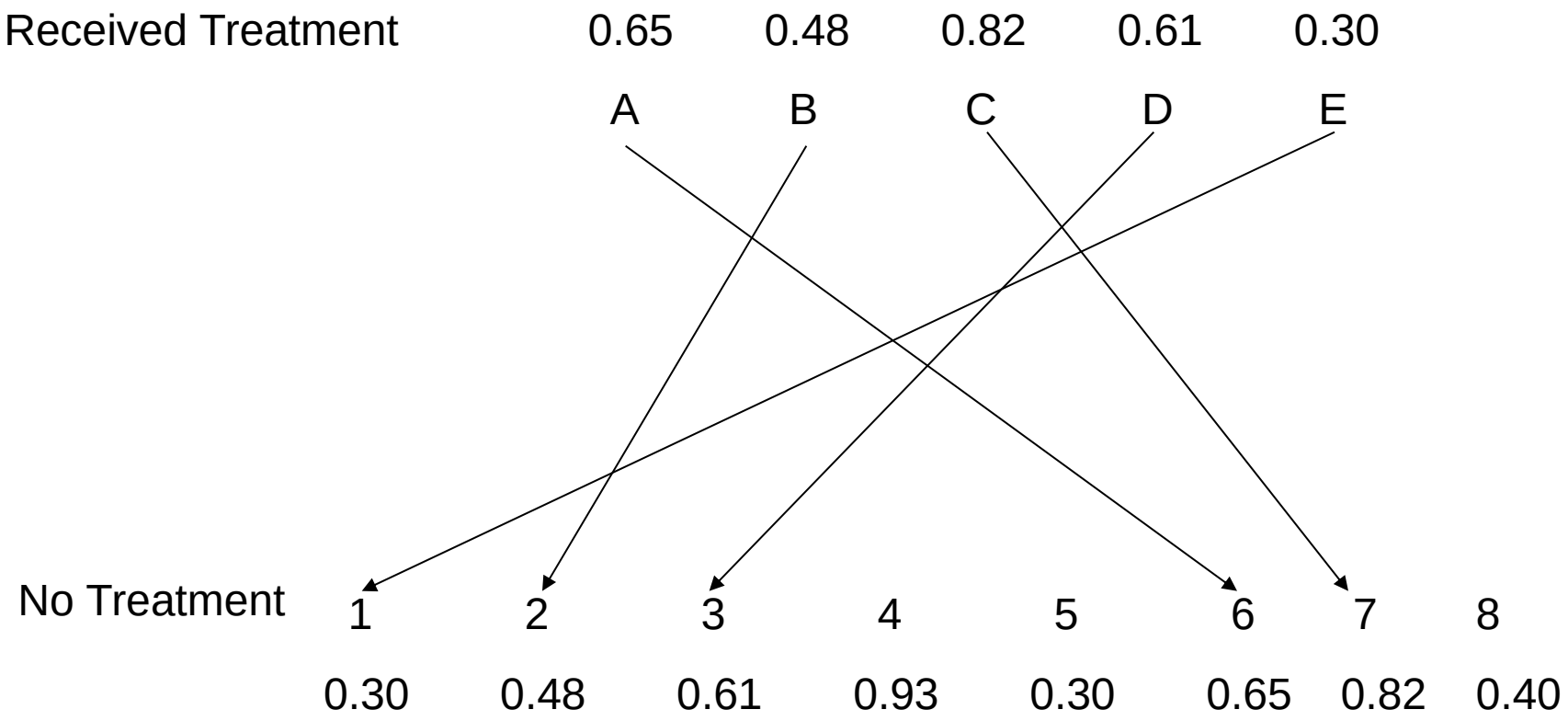
Abbreviations: AMI, acute myocardial infarction; CI, confidence interval.
*Simple propensity score included all 65 patient, hospital, and ZIP code characteristics.
†Complex propensity score included all patient, hospital, and ZIP code characteristics and all interactions of age, sex, and race with the other characteristics (750 variables).

# Multivariable Model vs. Propensity Score Model

- Multivariable model: fit one model containing all covariates + treatment variable – this is also known as a one stage model
- PS model: fit PS model, then use the PS as a covariate in the outcome model – this is a two-stage model

- Limitations of the multivariable model:
  - When the number of covariates is large, it may be difficult to fit the model
  - If the distributions of some confounding variables do not overlap substantially in the treated and control groups, the regression relationship is determined primarily by treated subjects in one region of the covariate space and by control subjects in another
    - Estimates of causal effects are then based on extrapolation

# Propensity Score Matching



# Propensity Score Matching

- Do you match on the exact value of the PS?
- Do you match on the nearest value? How close is close enough?
- Do you match with replacement?
- How many control subjects should be matched to each treated subject ?

# Propensity Score Matching

- Nearest-neighbour matching
  - Randomly order the individuals in the treatment and control groups
  - Select the first treated individual and find the *m* control individuals with the closest PS values
- Caliper matching
  - Uses all of the control individuals within a predetermined PS radius (or “caliper”) of the PS for a treated individual
  - This methods uses only as many control individuals as are available within the calipers

# Propensity Score Matching

- $N \rightarrow 1$ matching
- Example: $5 \rightarrow 1$ match
  - Begin by attempting a match using first five digits of the PS
  - If an appropriate control subject can not be selected for a given treated subject, then try a four-digit match on the propensity score
  - If a match can not be formed on the first four digits of the propensity score, then a three-digit match is attempted.
  - This process is repeated, if no suitable match can be found, until matches are attempted on the first digit of the propensity score.
  - If a treated subject can not be matched to any control subject on the first digit of the propensity score, then the treated subject is discarded from the matched analysis.

# Austin (2008): Recommendations for Studies involving Propensity Score Matching

- Explicitly state the strategy for matching
- Specify whether sampling was done with or without replacement
- Describe the distribution of the characteristics of the treated and control groups

# Austin (2008): cont'd

- Assess differences in the distribution of propensity scores for population groups
- Analytic methods for the estimation of the treatment effect should be appropriate for matched data

# Some Issues in Using PS Methods: How Not to Select Covariates

| Covariate | Treatment/ Exposure | Control/ No Exposure | Significant? | Include? |
|---|---|---|---|---|
| Age | 54 | 60 | Yes | Yes |
| BMI | 23 | 22 | No | No |
| Physical Health | 58 | 62 | Yes | Yes |
| Mental Health | 63 | 65 | No | No |
| Illness Severity | 72 | 83 | Yes | Yes |

# Some Issues in Using PS Methods: Rich vs. Poor Covariate Sets

- With a rich set of covariates, adjustments for unobserved covariates may be less critical

- With less rich covariate sets, you may need to fit a more complex model
  - E.g., include an instrumental variable in the outcome model as well as propensity scores

# Some Issues in Using PS Methods: Model Diagnostics

- Rubin describes “confusion between two kinds of statistical diagnostics…”

(1) Diagnostics for the successful prediction of probabilities and parameter estimates underlying those probabilities

(2) Diagnostics for the successful design of observational studies based on estimated propensity scores

- Rubin’s advice: worry about #2 but not about #1 in PS models

# Some Issues in Using PS Methods: Model Goodness of Fit

- Are tests used to evaluate logistic regression model fit and discrimination helpful in detecting the omission of an important confounder?
  - Simulation studies show that the Hosmer-Lemeshow GOF test and *c*-statistic are of no value in detecting residual confounding

# Some Issues in Using PS Methods: Limitations

- If the two groups do not have substantial overlap in propensity score values, then bias may be introduced into the model
  - If only the worst cases from the control group are compared to only the best cases from the treatment group, the result may be regression toward the mean

# Some Issues in Using PS Methods: Limitations

- Hidden bias: Beware of unmeasured covariates which affect the outcome and/or assignment
  - Conducting a sensitivity analysis helps quantify the problem
- The choice of variables narrows as an investigation proceeds
  - Observational studies require careful attention to the choice of variables – think about potential confounding variables early on in the design of a study

# Final Thoughts: Advantages of Using PS Methods

- Results can be persuasive even to audiences with limited statistical training
- The comparability of the groups can be verified
- PS methods may be combined with other sorts of adjustments for confounding effects