

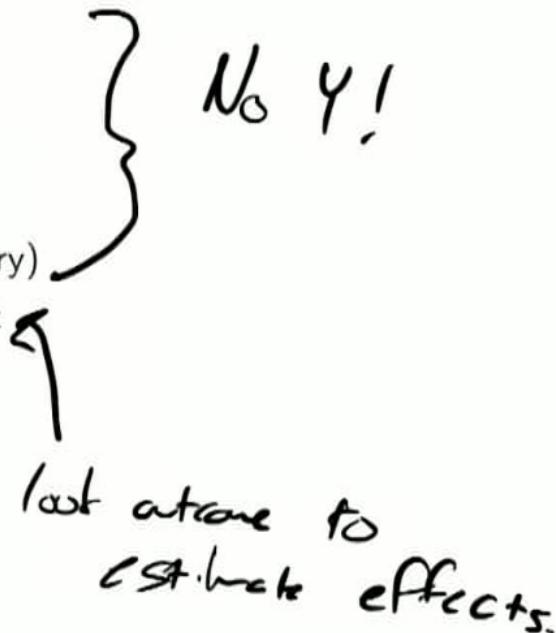
Segment 5: Analyzing Observational Studies

Section 01: Five Steps to Restructure the Data

5 Steps to Restructure the Data

According to Gelman, Hill, Vehtari

1. Estimands and Confounders
2. Calculate Distance Metrics
3. Restructuring the Data
4. Diagnostics
(Iterate Steps 2-4 as Necessary)
5. Estimate the treatment effect



What is the Question?

And to whom does it pertain?

No Data

Considerations:

- ▶ Defining the population of interest
 - ▶ Who are the units?
 - ▶ Is the "treatment" actually relevant to all of them?
 - ▶ Am I interested in all of them?
- ▶ Define the treatments being compared
 - ▶ What are the "treatment" and "control" conditions?
- ▶ Conceptualize the idealized version of the experiment!

Population + Treatment → Potential outcomes + Estimand
⇒ Precise Question to be Answered

"inclusion criteria"

Example Questions

- ▶ What is the causal effect of smoking on lung cancer?
 - ▶ Among male smokers over the age of 30 in the U.S. who started smoking by age 18 and were cancer-free at age 30, what is the causal effect of smoking at least 1 pack per day vs. smoking less than 1 pack per day on the incidence of lung cancer?
- ▶ What is the causal effect of air pollution on cardiovascular disease?
 - ▶ What is the causal effect of exposure to high vs. low levels of pollution during 2015-2020 on cardiovascular disease among people aged over 65 in the United States who live near an air pollution monitoring station?
- ▶ What is the causal effect of hormone replacement therapy on coronary heart disease (CHD) among post-menopausal women?
 - ▶ What is the effect of initiating hormone replacement therapy (vs. not) on CHD among post-menopausal women who do not have a history of previous heart disease?

What is the Population of Interest

A *causal estimand* is defined for a particular population of interest

- ▶ Average Treatment Effect (ATE)
 - ▶ Pertains to the entire sample/population

$$E[Y_i^{\circ} - Y_i^{\circ}]$$

- ▶ Average Treatment Effect on the Treated (ATT)
 - ▶ Pertains only to those observed to be *treated*

$$E[Y_i^{\circ} - Y_i^{\circ} | Z_i = 1]$$

- ▶ Average Treatment effect on the Control (ATC)
 - ▶ Pertains only to those observed to be *untreated*

$$E[Y_i^{\circ} - Y_i^{\circ} | Z_i = 0]$$

- ▶ Above quantities might not be equivalent
- ▶ Different analysis methods might target different estimands

Example Questions/Estimands

- ▶ Among male smokers over the age of 30 in the U.S. who started smoking by age 18 and were cancer-free at age 30, what is the causal effect of smoking at least 1 pack per day vs. smoking less than 1 pack per day on the incidence of lung cancer?
 - ▶ ATE
 - ▶ What would lung cancer have been among those smoking ≥ 1 pack per day if they had smoked < 1 pack
 - ▶ What would lung cancer have been among those smoking < 1 pack per day if they had smoked ≥ 1 pack?
- ▶ Among male smokers over the age of 30 in the U.S. who started smoking by age 18 and were cancer-free at age 30 **and smoked ≥ 1 pack per day**, what is the causal effect of smoking at least 1 pack per day vs. smoking less than 1 pack per day on the incidence of lung cancer?
 - ▶ ATT
 - ▶ What would lung cancer have been among those smoking ≥ 1 pack per day if they had smoked < 1 pack

How Were Treatments “Assigned”?

- ▶ What led units to be “assigned” certain “treatments”?
 - ▶ Or “exposed”
- ▶ What might the assignment mechanism have been?
 - ▶ Who are the decision makers?
 - ▶ What factors led to those decisions
 - ▶ Are they *confounders*?
- ▶ Are the “key covariates” for the assignment mechanism observed in the data?

Which Covariates Do I Include?

Can generally think of 5 types of covariates that may be available

1. *Confounders*: associated with treatment and outcome
 - ▶ Include these!
 - ▶ Required for unbiasedness
2. *Predictors*: associated only with outcome
 - ▶ Include these!
 - ▶ Not required for unbiasedness, but can increase precision
3. *Instruments*: associated only with treatment
 - ▶ Don't include these!
4. *Noise*: not associated with treatment or outcome
 - ▶ Don't include these!
5. *Posttreatment*: potentially affected by treatment
 - ▶ Do not include these!
 - ▶ Can induce *posttreatment selection bias*

Step 2: Calculating Distance Metrics

Goal: Find units with different treatments that are similar in their pre-treatment characteristics

- ▶ Easiest to think about as *matching* or *subclassifying*
 $Z = 0/Z = 1$ units that have the *exact same* covariate values, \mathbf{X}
 - ▶ Two units that look *exactly the same on all confounders* but received different treatments \Rightarrow "randomization"
 - ▶ Can be impractical for many \mathbf{X} with many levels
- ▶ Can more generally define a *distance metric* as a function between $\mathbf{X}_i, \mathbf{X}_j$
 - ▶ Euclidean distance
 - ▶ Mahalanobis distance
 - ▶ Propensity score distance
 - ▶ ...
- ▶ Use distance metric (usually a scalar) to identify matches or subclasses that are "close" in \mathbf{X}
 - ▶ But not necessarily exactly the same in \mathbf{X}

*inexact
match.,,*

Restructuring the Data to Approximate Randomization

- ▶ Subclassification
 - ▶ How exactly to define subclasses (or "blocks")
 - ▶ To approximate randomization *within* subclass
- ▶ Matching
 - ▶ Which units get matched to which units?
 - ▶ To approximate randomization *within* matches
- ▶ Weighting
 - ▶ How to re-weight units to approximate randomization in the *weighted* sample?
- ▶ Preprocessing/pruning/trimming
 - ▶ Should some units be excluded from the analysis because it is *impossible* to approximate randomization for these units?
 - ▶ I.e., lack of overlap

Interpreting Matching as Weighting

In general, a matching algorithm might match k untreated units to m treated units ($k : m$)

- ▶ Simplest case 1:1
- ▶ Subclassification $k : m$ with $k, m > 1$
- ▶ Many-to-one matching $k : 1$

Can interpret the $m : k$ ratio as a *weight* that can be used in a (regression) model strategy:

```
library(survey)
d_wts <- svydesign(ids~1, weights=~matching_output, ...)
reg <- svyglm(y ~ x1 + x2, design=d_wts, ...)
```

Inverse Probability Weighting

Goal: Weight the observed sample to create a *pseudo-population* that is balanced with respect to covariates (i.e., as though "randomized")

- ▶ The less likely a unit is to have received treatment $Z_i = z$, the larger representation it should get in the pseudopopulation
- ▶ Up-weight "weirdos": units whose treatment assignment is very different from what would be suggested by the propensity score
 - ▶ A unit with $Z_i = 1$ and $e_i = 0.05$ gets a weight of $\frac{1}{0.05} = 20$
 - ▶ A unit with $Z_i = 1$ and $e_i = 0.9$ gets weight of $\frac{1}{0.9} = 1.11$
- ▶ Weights adjust for confounding by creating a pseudopopulation where all individuals have the same probability of receiving $Z = 1$ and $Z = 0$ (like in a randomized study)
- ▶ Rooted in ideas of survey sampling

Restructuring \Leftrightarrow “Design”

- ▶ Whatever process we follow
 - ▶ Matching, subclassification, weighting, etc.
 - ▶ Propensity score or otherwise
- ▶ We should arrive at an instance of a “design”
 - ▶ Where the data have been restructured to (hopefully) approximate the features of a randomized study
 - ▶ E.g., balance on confounders
- ▶ But need to check whether the design was successful....

Step 4: Diagnostics

Balance and Overlap

- ▶ A successful *design* or *restructuring* is one that achieves good balance
 - ▶ And overlap
- ▶ This can (and should) be checked empirically
- ▶ Good balance \Rightarrow good approximation of a randomized study
 - ▶ (Assuming ignorability)
- ▶ Bad balance \rightarrow
 - ▶ Refine the restructuring
 - ▶ Fame expectations for causal inference with appropriate caution

Balance Checking

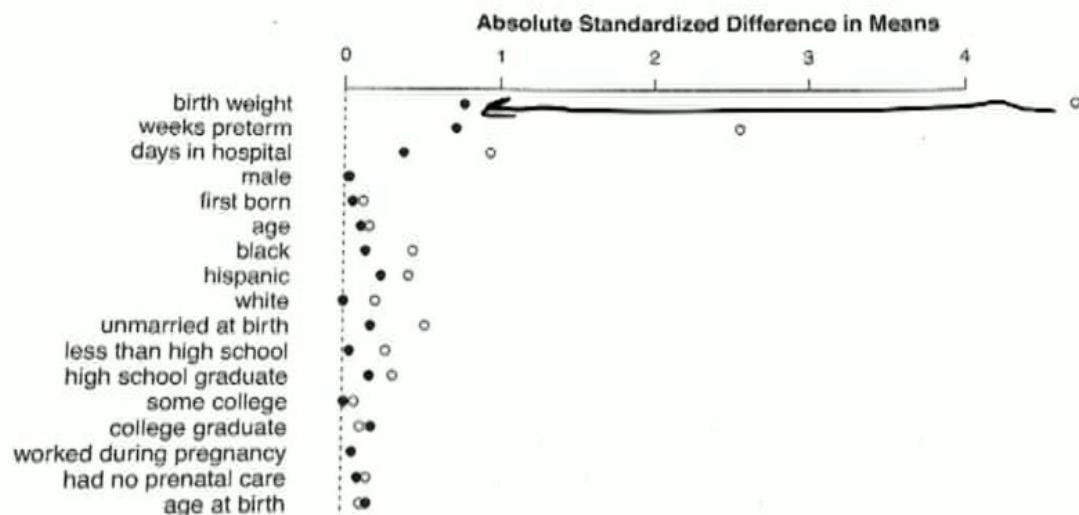
Balance can be checked by comparing the distribution of each potential confounder across treatment/control groups within (matches, subclasses, pseudopopulations) that result from the restructuring

- ▶ Compare the *distribution* of confounders within
 - ▶ Matched pairs, subclasses, weighted pseudopopulations, etc.
- ▶ Frequently done with *absolute standardized mean differences* (ASMD)
 - ▶
$$\frac{|\bar{X}_k^1 - \bar{X}_k^0|}{sd(\bar{X}_k^1)}$$
 for $k = 1, \dots, p$ covariates
 - ▶ Common to compare ASMD for every covariate *before* restructuring and *after* restructuring

Variations

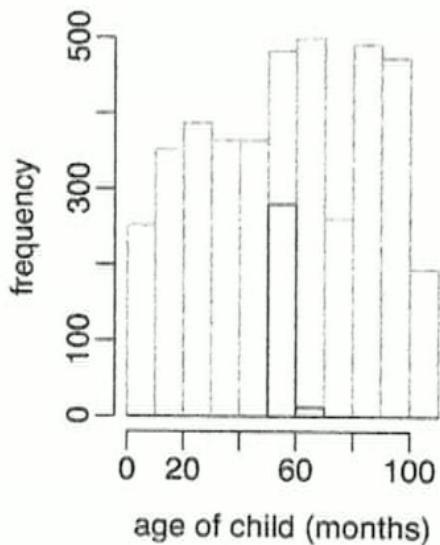
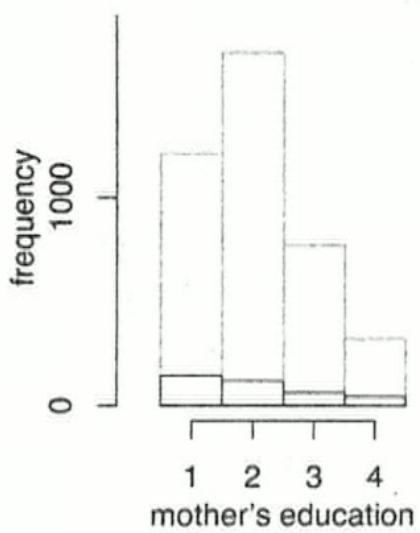
Balance Checks

From the GHV Child Care Example



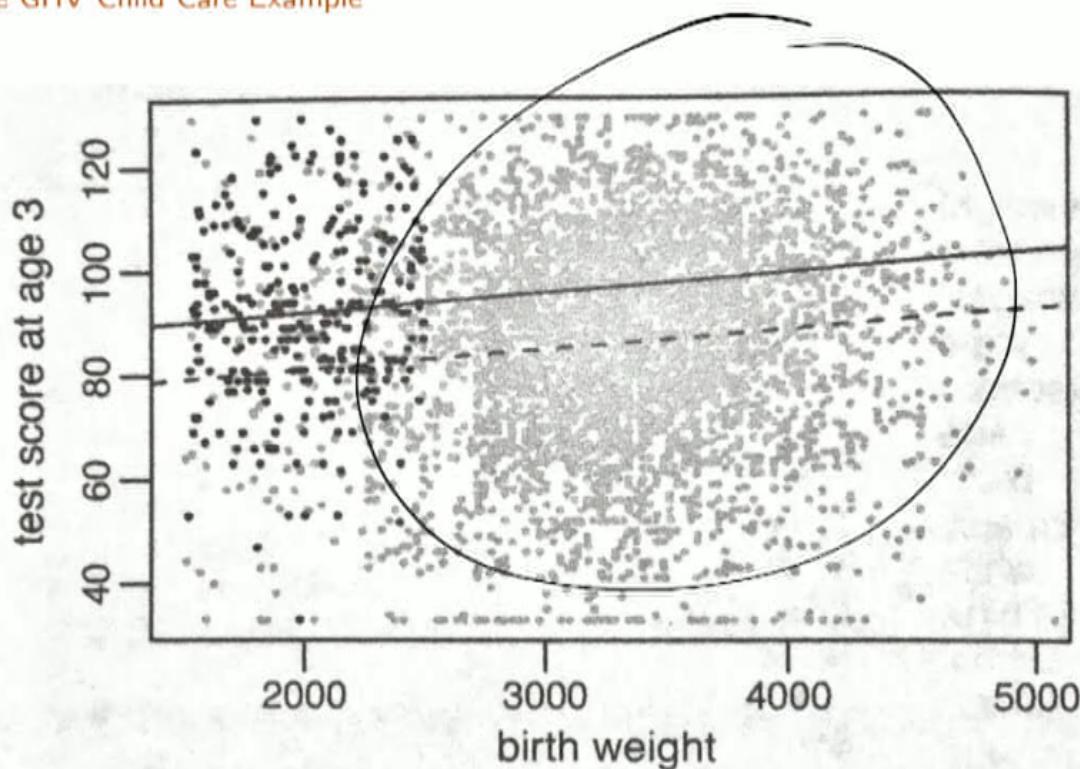
Overlap Checks

From the GHV Child Care Example



Overlap Checks

From the GHV Child Care Example



Balance Checks

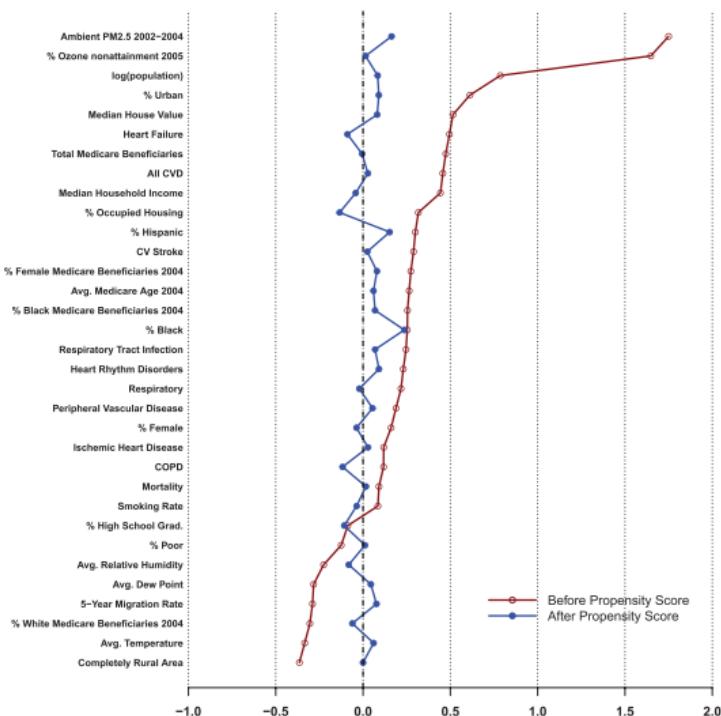
From Zigler et al. (2018)



FIGURE 2. Locations of monitoring locations in the study area. Locations of all (A) the 829 PM_{2.5} monitoring locations available for the initial analysis set and (B) the 404 locations retained after propensity score pruning.

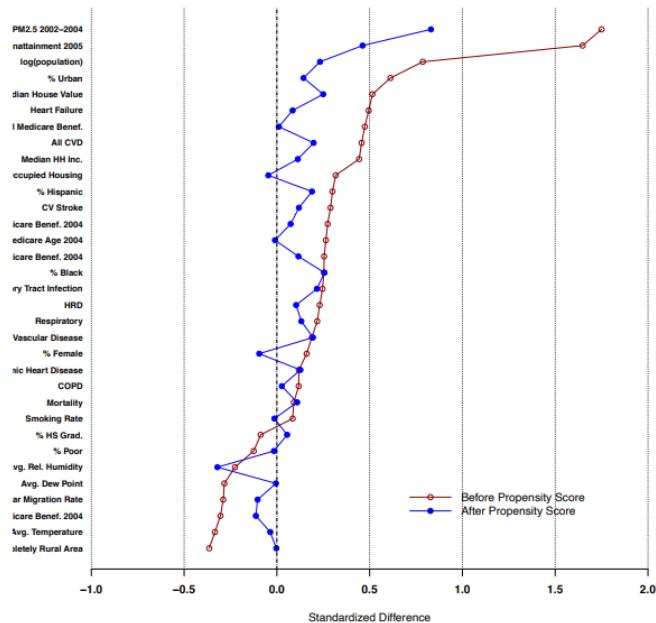
Balance Checks

From Zigler et al. (2018), after propensity score pruning



Balance Checks

From Zigler et al. (2018), without propensity score pruning



Other Diagnostic Considerations

- ▶ Don't use overlap to understand model fit
 - ▶ A model with no predictive power will lead to great overlap!
- ▶ Overlap and "changing the estimand"
 - ▶ Omit (prune, trim) some observations → changes the quantity being estimated
 - ▶ Changes the "inclusion criteria" of the hypothetical experiment
- ▶ Overlap in *confounders* not *covariates*
 - ▶ Only need overlap with respect to *confounders*
 - ▶ Including extraneous covariates could make overlap appear poor in ways that are actually inconsequential
- ▶ Beyond balance in just means
 - ▶ Should achieve balance in *other* distributional features too (e.g., standard deviations)
 - ▶ Can be checked similarly

Iterating Steps 2-4

Achieving adequate balance may require several iterations

- ▶ Estimate propensity score → subclasses → balance diagnostics
- ▶ Repeat

Note: This is ok as long as you don't choose your "design" based on which one gives the best treatment effect estimate

- ▶ Don't "peek" at the treatment effect estimate during the design phase
- ▶ Don't even look at outcomes at all (cf. Rubin 2008 comments on "objectivity")
- ▶ Evaluate different designs based on information on the covariates, treatments

There may be many different ways to achieve a "good design"

- ▶ And they may not all provide equivalent answers for the treatment effect

Step 5: Estimate the Treatment Effect

After Satisfactory Restructuring...

Analyze just as you would an experiment!

- ▶ Regression
- ▶ Comparison of means (across subclasses)
- ▶ etc.

One main point of the restructuring/design is that the specifics of the analysis strategy *should not make that big of a difference*

- ▶ Just as would be the case in a randomized trial
- ▶ Conclusions about causal effects should be *robust* with respect to specific analysis strategy **if the design was successful and assumptions like ignorability are satisfied**