500 Class 6 Slides

github.com/THOMASELOVE/2019-500

2019-03-21

Agenda for Class 6

- Project Details / Tips
- Rosenbaum, Chapters 7 and 8
- Homework 3
- The Right Heart Catheterization Study

Project Proposals We've Yet to Discuss

Project Schedule

Time	April 18	April 25	May 2
8:35 AM to 8:50 AM	*Sarah*	Harry	Julijana
8:52 AM to 9:07 AM	*Gwen/Roberto*	Brian	Hyun Jo
9:09 AM to 9:24 AM	*Laura B.*	*Sneha*	Kelly
9:26 AM to 9:41 AM	*Andrew*	Sebastian	*Bob*
9:41 AM to 9:52 AM	BREAK	BREAK	BREAK
9:52 AM to 10:07 AM	Michael	Laura C.	*Jenny*
10:09 AM to 10:24 AM	Jordan	*Xin-Xin*	*Hasina/Gi-Ming*
10:26 AM to 10:41 AM	Zhanwen	*Kedar*	*Abhishek/Sandra*
10:43 AM to 10:58 AM	Dr. Love	Yousef	Dr. Love

Project Titles (people speaking May 2)

1. Kelly Steller "Fruit and Vegetable Intake and Body Mass Index in Adult U.S. Men and Women"
2. Hyun Jo Kim "The other side of beta blockers that doctors don't want you to know"
3. Julijana Conic "Length of survival in mitral valve repair for ischemic mitral regurgitation"

See http://bit.ly/500-2019-project-list for more details. Abstract 2 PM 04-17, Slides due 4 PM 05-01

Project Titles (people speaking April 25)

4. Yousef Mustafa "Diabetes and Blood Pressure Status in NHANES 2015-16 Subjects"
5. Laura Cremer "Is Marijuana a gateway drug?"
6. Sebastian Garcia-Medina "The influence of Marijuana consumption on sleeping habits"
7. Brian Richardson "Topology of the stomach: The effects of tumor location on gastric cancer progression"
8. Harry Persaud "Does fasting improve low-density lipoprotein and blood glucose? A re-analysis of de Toledo et al."

See http://bit.ly/500-2019-project-list for more details. Abstract 2 PM 04-17, Slides due 4 PM 04-24

Project Titles (people speaking April 18)

9. Zhanwen Du "Effect of baseline diabetes on postoperative in hospital complications"
10. Jordan Fiegl "Understanding Challenges of Primary Care for Medicaid Patients"
11. Michael King "Understanding the Impact of Physically Demanding Jobs on Sleep"

See http://bit.ly/500-2019-project-list for more details. Abstract 2 PM 04-17, Slides due 4 PM 04-17

Project Tips

1. If your project has roughly equal numbers of "exposed" and "control" subjects, you will run into a problem if you try to do 1:1 matching without replacement, in that your results will be virtually unchanged. You have the following alternatives available to you:
   - Match 1:1 with replacement (which may still have only a small impact, depending on how you do it)
   - Match 1:k with replacement, where k $>$ 1 (which may or may not useful)
   - Match 1:1 without replacement, but only on a random sample of (perhaps 1/3 of) your "exposed" subjects.
   - More on matching in today's case study.

Project Tips

2. If you have more than 4000 observations in total, I suggest sampling down (without replacement) to 3000 (at most) to develop your project. You should do this with code near the start of your R Markdown file, that can be tweaked as necessary, either to take a different sample, or not sample at all.
    - If you're going to sample down like this, I recommend you maintain the rate of exposure that you see in the data as a whole, unless that rate is close to 50%, in which case I suggest you sample down to a rate that is 25-35% exposed and the rest unexposed, so that 1:1 matching without replacement will be more likely to produce useful results.

Project Tips

3. The definition of which group is the "exposed" group and which is the "control" group matters analytically, although it's essentially an arbitrary selection. You will make your life easier for our purposes in developing the project by making your "exposed" group the smaller of the two groups, if possible.

Some Project Reminders

Your Project Update is due April 2

- How has your abstract changed from your approved project proposal?
- Describe the data - tell me what you have, and what you are still waiting for.
- Describe the biggest problem you're currently having with regard to completing the design and analysis of the study.

Final Materials

- Abstract is due at 2 PM April 17 for everyone
- Presentations are April 18, 25 and May 2.
  - Your Slides are due at 4 PM the day before your presentation.
- All Materials (including revised Abstract/Slides) plus data, R Markdown and HTML including discussion due at noon May 6.

No extensions.

Rosenbaum, Chapters 7-8

Rosenbaum, Chapter 7 (Elaborate Theories)

See some of my comments (and some essays for us to discuss) at this link

Rosenbaum, Chapter 8 (Quasi-Experimental Devices)

Provide an example of how the method of multiple control groups might be used to improve the design of an observational study with which you are familiar, either one you have been directly involved in, or one you've read about, or one you'd like to be able to do. Provide a basic description of your study's population, outcome, treatment and control conditions and covariates of interest so we can follow your logic. Be sure to specify the approach you would take in implementing multiple control groups, and the ways in which you expect the use of multiple control groups to be helpful in improving your understanding of the phenomenon you're studying.

Next Week: Chapter 9 (Sensitivity to Bias)

Homework 3

Tasks in Homework 3

1. Unadjusted analyses for binary outcomes
2. Fit propensity model.
3. Evaluate Rubin's Rules 1 and 2 without propensity adjustment.
4. Use direct PS adjustment to obtain a causal effect estimate (OR and CI) for the impact of treatment on hospice entry.
5. Use subclassification by PS quintile to obtain a pooled OR estimate across all strata, and assess if pooling is appropriate.
6. 1:1 PS match without replacement
7. Compare estimates from steps 1, 4, 5, and 6.

Answer Sketch will be available this afternoon.

Remember that Homework 4 is due next week

Also uses the `canc3` data. Start either with your version or the one produced by my sketch.

1. Do weighting using the ATT approach.
2. Assess balance after weighting.
3. Build a Love plot after weighting.
4. Evaluate Rubin's Rules 1 and 2 after weighting.
5. Try weighting with twang instead.
6. Pick your favorite weighting approach and use it to estimate the effects on survival and on hospice.
7. Perform a double-robust analysis.
8. Compare the results from 6 and 7 to the ones we got in Homework 3.

The Right Heart Catheterization Study

Complete Case Study is online

It's in the `rhc_2019` directory in our Data and Code page.

https://github.com/THOMASELOVE/2019-500/tree/master/data-and-code/rhc_2019

- The R Markdown code
- A Github Markdown result
- The HTML version is most easily viewed via RPubs at http://rpubs.com/TELOVE/rhc-500-2019
- The clean Rds data set, and the original .csv data (and those data are also housed at Vanderbilt, with this description).

What does this Case Study entail?

The example demonstrates:

1. Working with real data (and pulling it from a csv file on the web)
2. Using various approaches to matching with the propensity score (from the Matching package)
3. Assessing covariate balance after matching
4. Working with a binary outcome, a quantitative outcome and a survival (time-to-event) outcome
5. Sensitivity analyses after matching

To run this yourself, create a new directory for an R studio project, then place the R Markdown code in a subdirectory of that directory called `R` and also create an (empty) subdirectory called `data`. That should do it.

Studying Right Heart Catheterization in SUPPORT

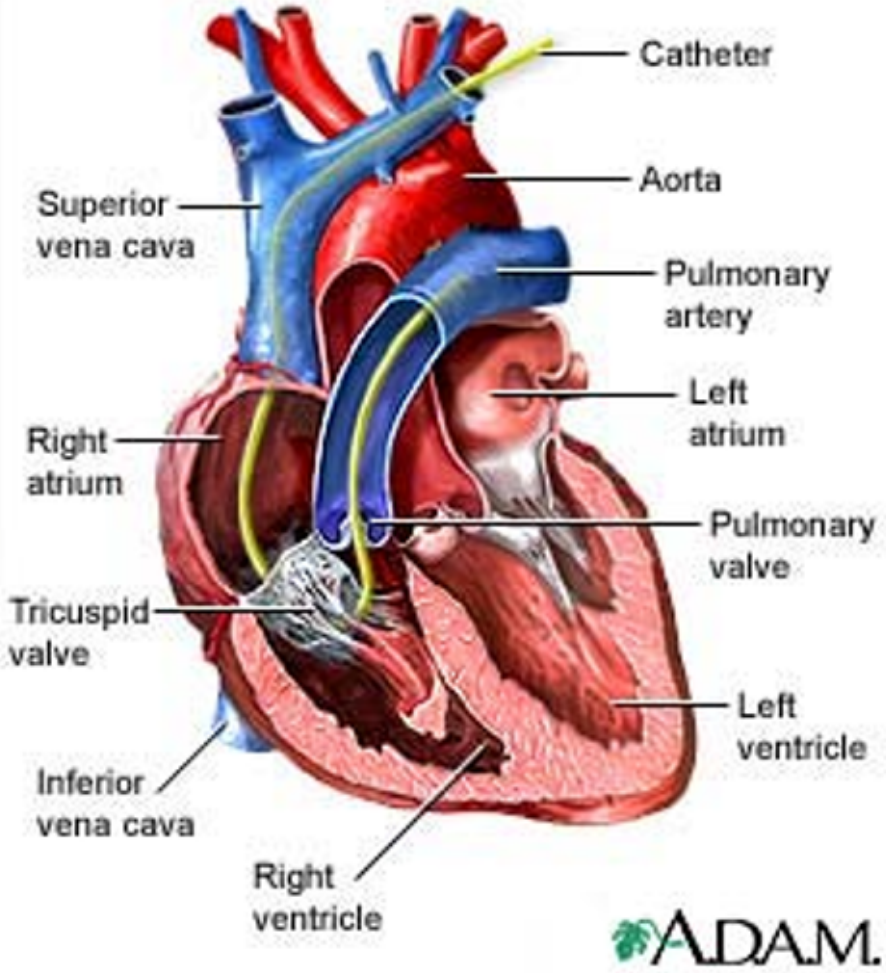
SUPPORT: Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatments[1]

- Goal: Examine the association between the use of RHC during the first 24 hours in the ICU and outcomes
- Outcomes: survival, length of stay, intensity and costs of care
- Sample: 5,735 critically ill adult ICU patients in nine disease categories

Study was prospective!

[1] Connors et al. 1996

Right Heart / Swan-Ganz / Pulmonary Artery Catheterization



Pass a thin flexible tube:

1. **to obtain diagnostic information about the heart**
2. **for monitoring of heart function in the critically ill.**

http://www.nlm.nih.gov/medlineplus/ency/imagepages/18087.htm

Does the RHC do more harm than good?

Prior (small) observational studies comparing RHC to non-RHC patients:

- RR of death higher in RHC elderly patients than non-RHC elderly
- RR of death higher in RHC patients with acute MI than non-RHC patients with MI
- Patients with higher than expected RHC use had higher mortality

Big Problem: Selection Bias. Physicians (mostly) decide who gets RHC and who doesn't.

Why not a RCT?

- RHC directly measures cardiac function
- Some MDs believe RHC is necessary to guide therapy for some critically ill patients
- Procedure is very popular - existing studies haven't created equipoise

81 Characteristics used to predict PS(RHC usage)

- Age, Sex, Race
- Education, Income, Insurance
- Primary and Secondary Disease category
- Admission diagnosis category (12 levels)
- ADL and DASI 2 weeks before admission
- DNR status on day 1
- Cancer (none, local, metastasized)
- 2 month survival model
- Weight, temperature, BP, heart rate, respiratory rate
- Comorbid illness (13 categories)
- Body chemistry (pH, WBC, PaCO_2_, etc.)

Panel (7 specialists in clinical care) specified important variables related to the decision to use or not use a RHC.

RHC vs. Non-RHC patients

RHC patients were more likely to

- Be male, have private insurance, enter the study with ARF, MOSF or CHF

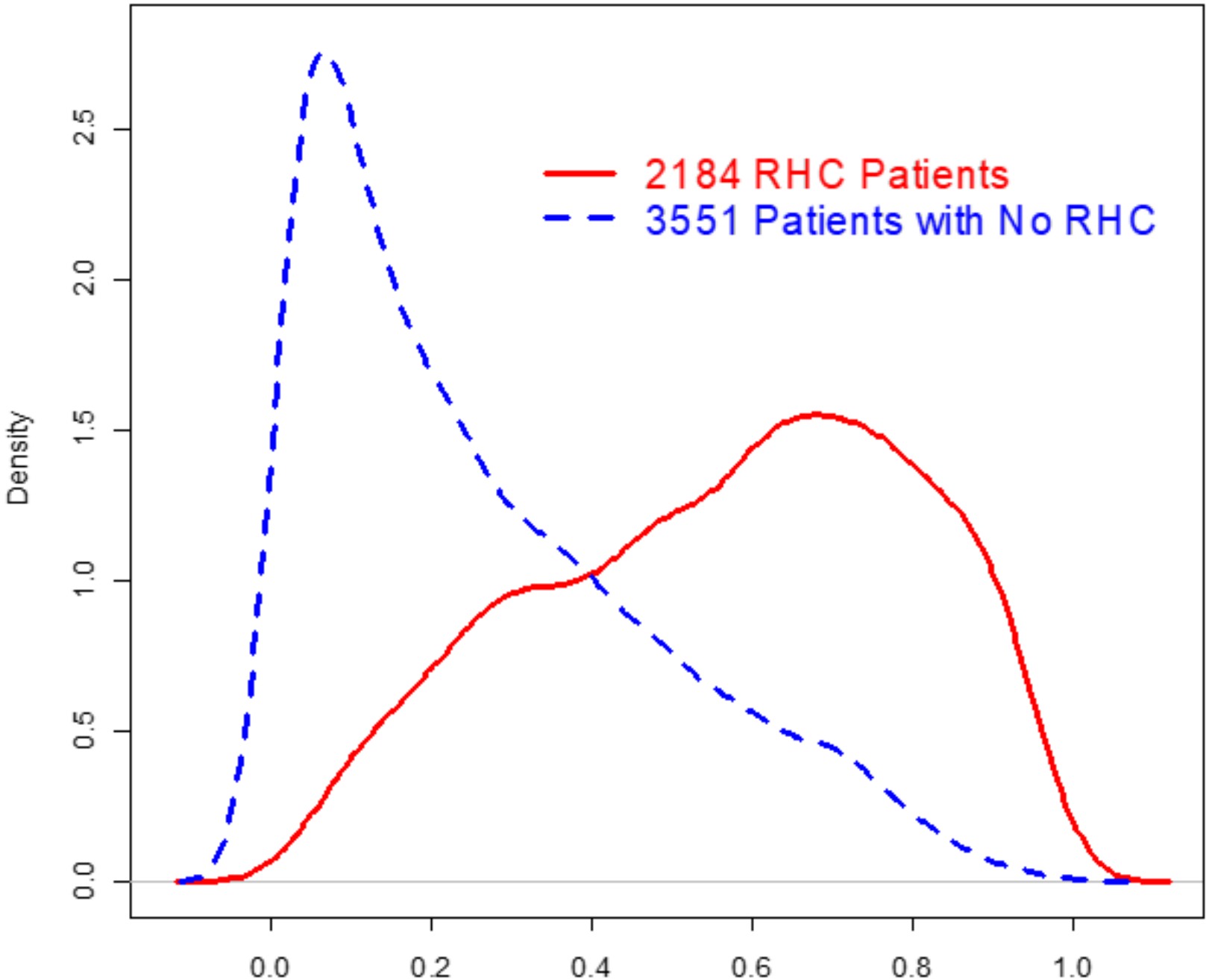
RHC patients were less likely to

- Be over 80 years old, have cancer, have a DNR order in the first 24 hours of hospitalization

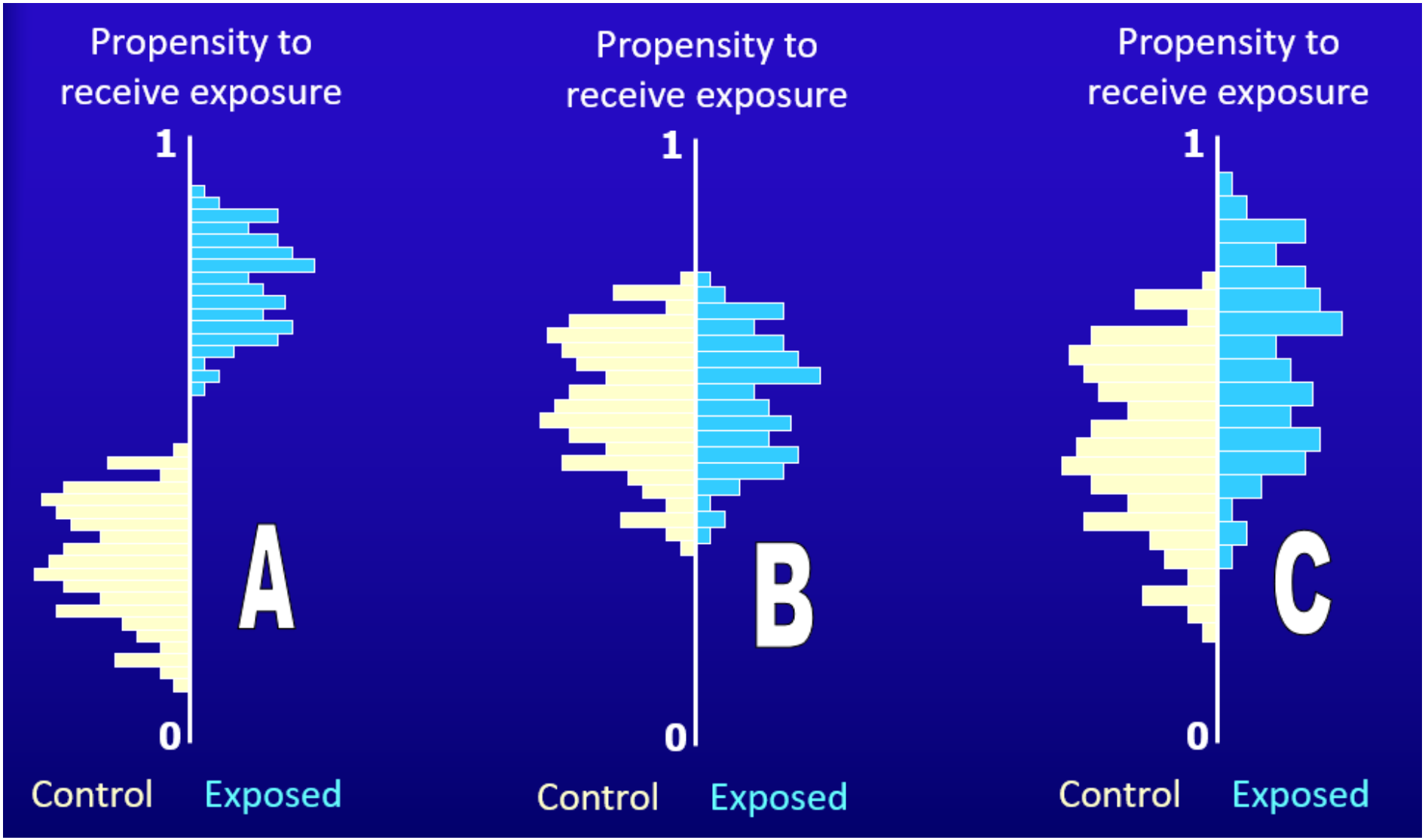
RHC patients had significantly

- Fewer comorbid conditions,
- More abnormal results of vital signs, WBC count, albumin, creatinine, etc.
- Lower model probability of 2-month survival

How Much Overlap do we see in the RHC data?



How Much Overlap do we want?



To The Case Study

It's in the `rhc_2019` directory in our Data and Code page.

https://github.com/THOMASELOVE/2019-500/tree/master/data-and-code/rhc_2019

- The R Markdown code
- A Github Markdown result
- The HTML version is most easily viewed via RPubs at http://rpubs.com/TELOVE/rhc-500-2019
- The clean Rds data set, and the original .csv data (and those data are also housed at Vanderbilt, with this description).

Another Key Point in the RHC data

Right Heart Catheterization and the Perils of Selective Weighting

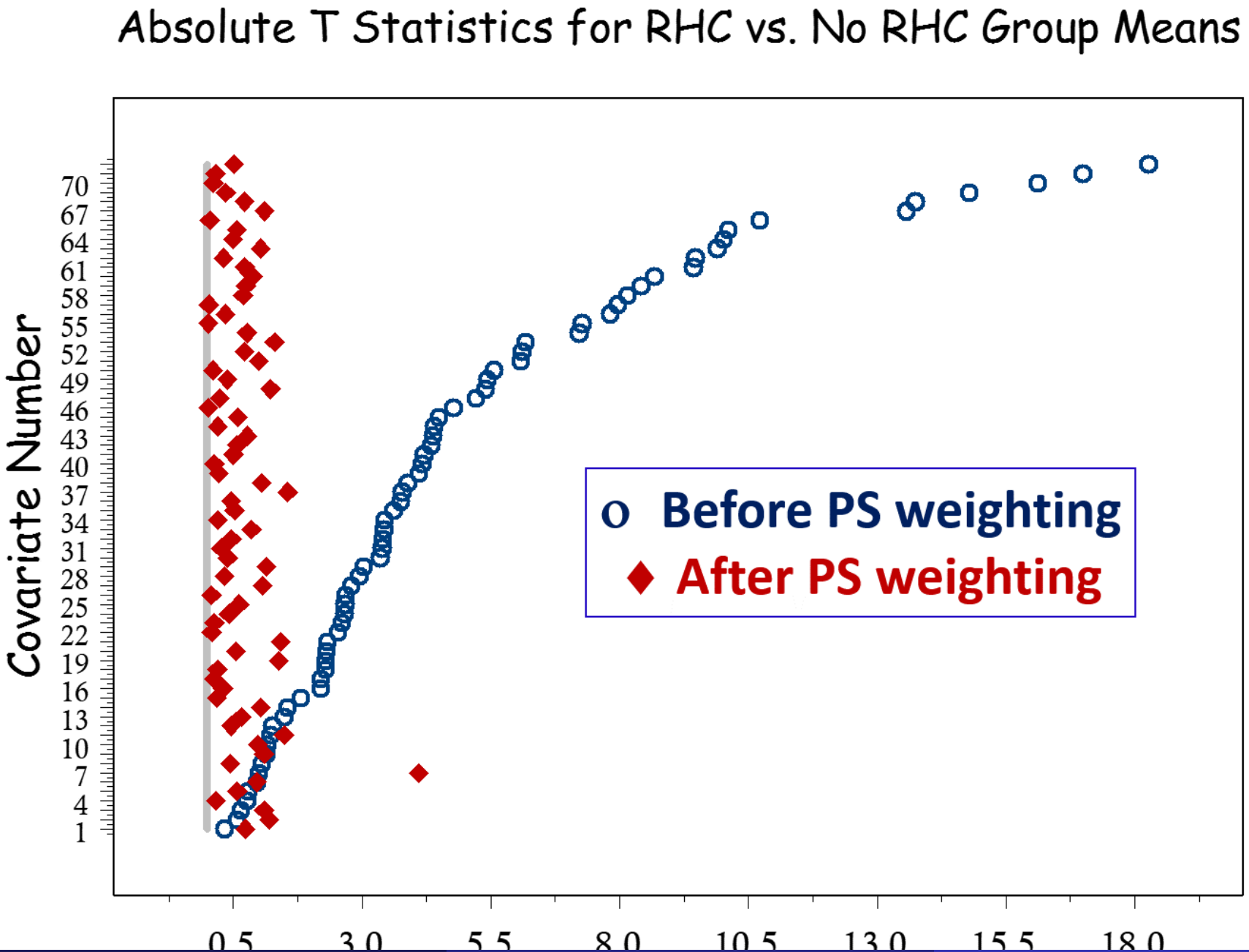
- 5,735 hospitalized patients in SUPPORT study
  - 2,184 treated (RHC) and 3,551 controls (no RHC).

Reweight each treated patient by 1/PS, and each control patient by 1/(1-PS).

- PS model estimated by Hirano and Imbens[2] using 57 of 72 available covariates
  - Selected only those with $|t| > 2.0$
  - Serum potassium, for instance, prior to weighting showed a mean of 4.04 in the RHC group and 4.07 in the "No RHC" group, for a t = -0.99, so it was not included in the propensity model.

[2] Hirano and Imbens 2001, Connors 1996, Hirano, Imbens & Ridder 2003

Results of this Weighting Approach



Effectiveness of RHC Propensity Score Weighting

- The weighting is based on a propensity model including 57 of the 72 covariates.
- Serum potassium not included in this PS.
- Most means are much closer, although six variables become less balanced (larger absolute t statistic) after weighting. None of these six were in the 57-variable PS model.
- Weighting by the propensity score appears to balance control and treatment groups well.

A "Double Robust" Estimator

1. Fit propensity score model
2. Weight the individual subjects (ATT, commonly) by the propensity score.
3. Directly adjust (via regression) for the propensity score in estimating the treatment effect.

- Forces you to think hard about selection.
- You don't care about parsimony in the PS, so you can maximize predictive value.
- Can fit a very complex PS model, and a smaller outcome model.
- Some hope that if PS model or weighting is helpful, the combination will be helpful.

For Next Time

- Read (or at least skim) the Rubin 2001 Tobacco Litigation Paper
- Homework 4
- Essay about Rosenbaum Chapter 9
- Bring questions about your projects, as they arise