Advanced Statistical Methods for Observational Studies

LECTURE 04

class management

- Mike's office hours aren't happening this Thursday.
  - I'm out of town.
  - If you'd like, we can schedule a call during that time, need to email me.
- Questions?

a matched study

reminder

- We're using pair matching as our "go to" model.
- In lecture 03 we learned how to do
  - 1:k matching
  - 1:k matching with variable k
  - Full matching
- Matching with more than one control is often better because you're using more of the data than you would in a pair match.

efficiency

Design of Observational Studies: chapter 8.7

efficiency

- Our primary concern is bias.
- Bias is what the critiques are going to hit us on.
- Bias doesn't go away as we get more and more data.
- Efficiency is good to pay attention to though.
- If we assume our naïve model and constant variance, and we standardize to infinite number of controls then

number of controls	1	2	4	6	10	∞
variance multiplier	2.00	1.50	1.25	1.17	1.10	1.00

- In the real world, going from 1:2 to 1:10 may actually not be as beneficial as it looks… this table assumes perfect matches are available.

unobserved confounding

There are more things in heaven and earth, Horatio,
Than are dreamt of in your philosophy.
- *Hamlet* (1.5.167-8)

Design of Observational Studies: chapter 3.4-3.8

naïve model

- Model
- Assumptions
- Implementation

naïve model: "natural" experiments

- What if we design our study such that $Z_l + Z_k = 1$?

$$\Pr(Z_k = 1, Z_l = 0 | \dots, Z_l + Z_k = 1)$$

$$= \frac{\Pr(Z_k=1,Z_l=0|\dots)}{\Pr(Z_k=1,Z_l=0|\dots)+\Pr(Z_k=0,Z_l=1|\dots)}$$

$$= \frac{\pi_k^{1+0}(1-\pi_k)^{(1-1)+(1-0)}}{\Pr(Z_k=1,Z_l=0|\dots)+\Pr(Z_k=0,Z_l=1|\dots)}$$

$$= \frac{\pi_k^{1+0}(1-\pi_k)^{(1-1)+(1-0)}}{\pi_k^{1+0}(1-\pi_k)^{(1-1)+(1-0)} + \pi_k^{0+1}(1-\pi_k)^{(1-0)+(1-1)}} = \frac{1}{2}$$

IF we can do this then we get to use the same tools developed for RCTs!

naïve model: assumption one

- Strongly Ignorable Treatment Assignment: Those that look alike (in our data set) are alike

$$\pi_i = \Pr(Z_i = 1 | r_{Ti}, r_{Ci}, \boldsymbol{x}_i, u_i) = \Pr(Z_i = 1 | \boldsymbol{x}_i)$$

and

$$0 < \pi_i < 1 \text{ for all } i = 1, 2, \ldots, n$$

- If two subjects have the same propensity score, then their values of $\boldsymbol{x}$ may be different.
- By SITA, if these two subjects have the same $e(\boldsymbol{x})$ then the differences in their $\boldsymbol{x}$ are not predictive of treatment assignment (i.e., $\boldsymbol{x} \perp\!\!\!\perp Z | e(\boldsymbol{x})$).
- Therefore the mismatches in $\boldsymbol{x}$ will be due to chance and will tend to balance. (more details)

naïve model: assumption two

- No Interference Between Units (*part of SUTVA*): the observation on one unit should be unaffected by particular assignment of treatments to other units.
- Can be written as:

$$R_i(Z_i = z_i) = R_i(\boldsymbol{Z}^*)$$

where $Z_i = z_i$ indicates the treatment level for the i^{th} unit and $\boldsymbol{Z}^*$ is a particular randomization from the set of all randomizations that have $Z_i = z_i$.

- Not true for most educational interventions and infectious disease applications.
- More details here and here.

naïve model: implementation

- Collect a bunch of covariates that are related to treatment level and to the outcome.
- Exact match if you can.
- You probably can't exact match so estimate propensity scores and match on a hybrid of pscores and Mahalanobis distance.
- Play around with the matching until you achieve acceptable comparison groups.
- Die a little bit inside when you read your critiques' reviews because they point out all of the confounding that could exist. Reevaluate life choices.

sensitivity analysis

- Sensitivity models are a means for moving past the “you didn’t do X which could lead to bias.”
- A useful sensitivity model addresses one assumption at a time, quantifying and making understandable the impact of departures from the assumption being assessed.
- We’re going to discuss the Γ sensitivity model which addresses the ignorable treatment assignment (SITA), not interference (SUTVA).

sensitivity analysis

- *A word of warning*: many people find the Γ sensitivity model confusing.
  - This lecture will only give you a sense of what's going on with this model; it isn't intended to be sufficient to fully understand Γ sensitivity.
  - Read section 3.4-3.8.
  - If you are so inclined then this might be a very nice place to produce your own framework for sensitivity.

model: sensitivity analysis

- Start with two observational units who have probability of treatment π_i and π_j (which may not be the same values).
  - Recall we defined this as $\pi_i = \Pr(Z_i = 1 | r_{Ti}, r_{Ci}, \boldsymbol{x}_i, u_i)$.

- We can talk about the odds of i receiving treatment:

$$\frac{\pi_i}{1 - \pi_i}$$

- And we can put the odds into a ratio:

$$\frac{\pi_i / (1 - \pi_i)}{\pi_j / (1 - \pi_j)}$$

model: sensitivity analysis

- Our sensitivity model asserts that we can bound the odds ratio like so:

$$\frac{1}{\Gamma} \leq \frac{\pi_i/(1-\pi_i)}{\pi_j/(1-\pi_j)} \leq \Gamma$$

  whenever $\boldsymbol{x}_i = \boldsymbol{x}_j$.
- We are making a particular statement about how "far off" the actual treatment probabilities are from the pscore (which only depends on the observed covariates).
- If $\Gamma = 1$ then this forces $\pi_i = \pi_j$.
- If $\Gamma = 2$ then π_i can depart from π_j
  - For example: if $\pi_i = 1/2$ and $\pi_j = 2/3$ then

$$\frac{\pi_i/(1-\pi_i)}{\pi_j/(1-\pi_j)} = \frac{0.5/(1-0.5)}{0.\overline{6}/(1-0.\overline{6})} = 2$$

model: sensitivity analysis

- With this model in place we can think about "worst case" scenarios regarding violations of SITA.
- If someone is willing to give you a particular framework for how the violation must occur (to the exclusion of all other possible ways it can fail) then use that parametric model.
- The Γ sensitivity model is non-parametric and we look at the extreme values that might occur when $\Gamma > 1$.
  - We'll get ranges of p-values and estimates
- Every study is sensitive to sufficiently large violations of the SITA assumption. Just let $\Gamma \to \infty$.
- If we're going to make progress then the question becomes what level of Γ is sufficiently large to proceed.

naïve model: "natural" experiments

- What if we design our study such that $Z_l + Z_k = 1$?

$$\Pr(Z_k = 1, Z_l = 0 | \dots, Z_l + Z_k = 1)$$

$$= \frac{\Pr(Z_k=1,Z_l=0|\dots)}{\Pr(Z_k=1,Z_l=0|\dots)+\Pr(Z_k=0,Z_l=1|\dots)}$$

$$= \frac{\pi_k^{1+0}(1-\pi_k)^{(1-1)+(1-0)}}{\Pr(Z_k=1,Z_l=0|\dots)+\Pr(Z_k=0,Z_l=1|\dots)}$$

$$= \frac{\pi_k^{1+0}(1-\pi_k)^{(1-1)+(1-0)}}{\pi_k^{1+0}(1-\pi_k)^{(1-1)+(1-0)} + \pi_k^{0+1}(1-\pi_k)^{(1-0)+(1-1)}} = \frac{1}{2}$$

IF we can do this then we get to use the same tools developed for RCTs!

model: sensitivity analysis

- If we design our study such that $Z_l + Z_k = 1$:

$$\Pr(Z_k = 1, Z_l = 0 | \ldots, Z_l + Z_k = 1) = \frac{\pi_i}{\pi_i + \pi_j}$$

- Combining this with the sensitivity model, and doing some vaguely enjoyable algebra, we get:

$$\frac{1}{1+\Gamma} \leq \frac{\pi_i}{\pi_i + \pi_j} \leq \frac{\Gamma}{1+\Gamma}$$

- We get ½ if $\Gamma = 1$.

model: sensitivity analysis

- The randomization tests we have can be reworked under the understanding that we can vary the odds ratio within

$$\frac{1}{1+\Gamma} \leq \frac{\pi_i}{\pi_i + \pi_j} \leq \frac{\Gamma}{1+\Gamma}$$

- Setting $\frac{\pi_i}{\pi_i+\pi_j} = \frac{\Gamma}{1+\Gamma}$ will get you one extreme.
- Setting $\frac{1}{1+\Gamma} = \frac{\pi_i}{\pi_i+\pi_j}$ will get you the other.
- For notational purposes, let's say that the usual Wilcoxon signed rank test (when $\Gamma = 1$) is written as T.
- Then we'll write the test statistic under our sensitivity model as $\bar{T}$.

model: sensitivity analysis

- We can calculate the exact distribution of $\bar{T}$ under either extreme, but for large matched sets it'll be easier (and not far off) to use an approximation.
- The $\bar{T}$ has known expected value and variance

$$E[\bar{T}] = \frac{\Gamma}{1+\Gamma}\frac{\mathrm{I}(\mathrm{I}+1)}{2}$$

$$var(\bar{T}) = \frac{\Gamma}{(1+\Gamma)^2}\frac{\mathrm{I}(\mathrm{I}+1)(2\mathrm{I}+1)}{6}$$

where I is the number of matched pairs.

model: sensitivity analysis

- We can calculate the exact distribution of $\bar{T}$ under either extreme, but for large matched sets it'll be easier (and not far off) to use an approximation.
- The $\bar{T}$ has known expected value and variance

$$E[\bar{T}] = \frac{1}{1+\Gamma}\frac{\mathrm{I}(\mathrm{I}+1)}{2}$$

$$var(\bar{T}) = \frac{\Gamma}{(1+\Gamma)^2}\frac{\mathrm{I}(\mathrm{I}+1)(2\mathrm{I}+1)}{6}$$

where I is the number of matched pairs.

model: sensitivity analysis

- The standardized deviate of T (the Wilcoxon signed rank statistic) can be approximated using:

$$\frac{T - E[\bar{T}]}{\sqrt{var(\bar{T})}} \sim N(0,1)$$

example: sensitivity analysis

obs	b_weight	gest_age	dose	hearing
1	2412	36	1	0.12
2	2205	29	1	0.24
3	2569	36	1	0.02
4	2443	34	1	-0.16
5	2569	36	0	0.58
6	2436	35	0	-0.22
7	2461	34	0	-0.07
8	2759	32	0	-0.55
9	2324	27	0	-0.36
10	2667	34	0	0.28
…	…	…	…	…
500	2349	33	1	-0.55

Similar to data set from lecture 03, but different number of observations and outcome of interest.

example: sensitivity analysis

obs	b_weight	gest_age	dose	hearing
1	2412	36	1	0.12
2	2205	29	1	0.24
3	2569	36	1	0.02
4	2443	34	1	-0.16
5	2569	36	0	0.58
6	2436	35	0	-0.22
7	2461	34	0	-0.07
8	2759	32	0	-0.55
9	2324	27	0	-0.36
10	2667	34	0	0.28
…	…	…	…	…
500	2349	33	1	-0.55

Outcome of interest: Hearing is some standardized metric with population mean=0 and sd=1.

example: sensitivity analysis

- Create 250 pair matches.
- Using T, the usual Wilcoxon signed rank statistic:
- We know that E[T]=15,688 and sd(T)=812
- Get T=13,250
- Using the approximation:

$$\frac{T - E[T]}{\sqrt{var(T)}} \sim N(0,1)$$

- $\frac{13{,}250-15{,}688}{812} = -3.00$, which has a small p-value, under the naïve model.

example: sensitivity analysis

- Create 250 pair matches.
- Using $\bar{T}$, the usual Wilcoxon signed rank statistic:
- We know that E[$\bar{T}$]=16,540 and sd($\bar{T}$)=810
- Get T=13,250
- Using the approximation:

$$E[\bar{T}] = \frac{\Gamma}{1+\Gamma}\frac{I(I+1)}{2}$$

Set Γ=1.11

$$\frac{T - E[\bar{T}]}{\sqrt{var(\bar{T})}} \sim N(0,1)$$

- $\frac{13{,}250-16{,}540}{810} = -4.00$, which has a small p-value.

example: sensitivity analysis

- Create 250 pair matches.
- Using $\bar{T}$, the usual Wilcoxon signed rank statistic:
- We know that E[$\bar{T}$]=14,835 and sd($\bar{T}$)=810
- Get T=13,250
- Using the approximation:

$$E[\bar{T}] = \frac{1}{1+\Gamma}\frac{I(I+1)}{2}$$

Set Γ=1.11

$$\frac{T - E[\bar{T}]}{\sqrt{var(\bar{T})}} \sim N(0,1)$$

- $\frac{13{,}250-14{,}835}{810} = -1.95$, which has a p-value close to 0.05.
- Interpretation: If there was a small amount of bias $\Gamma = 1.12$ then this would nullify our qualitative claims.

implementation: sensitivity analysis

- In practice, software will do this for you and you will interpret.
- The key to keep in mind is that there are two different way things could go wrong: (i) units could be sorted into treatment or (ii) into control.
- This gives rise to three different distributions:
- Naïve model: $T \sim N\left(\frac{I(I+1)}{4}, \frac{I(I+1)(2I+1)}{24}\right)$
- Biased toward one way: $T \sim N\left(\frac{\Gamma}{1+\Gamma}\frac{I(I+1)}{2}, \frac{\Gamma}{(1+\Gamma)^2}\frac{I(I+1)(2I+1)}{6}\right)$
- Biased other way: $T \sim N\left(\frac{1}{1+\Gamma}\frac{I(I+1)}{2}, \frac{\Gamma}{(1+\Gamma)^2}\frac{I(I+1)(2I+1)}{6}\right)$

implementation: sensitivity analysis

- Use the new distributions to test your statistic to see where its critical values are.
- This will lead you to provide wider intervals for everything:
  - If you had a point estimate of (to pick a random number): 5 then, for a particular Γ, you may end up with a "point estimate" of (4, 6). This new interval is not due to randomness in assignment, it is due to the difference in treatment assignment probabilities.
  - If you had a p-value of 0.012 , for a particular Γ, you may end up with a p-value interval of (0.032, 0.0001).

implementation: sensitivity analysis

- In practice, it's common to just report the value of Γ which nullifies your qualitative conclusions (i.e., goes from significant to insignificant), and to help the reader in interpreting the meaning of Γ.
- For example, $\Gamma = 2$ means that within a given pair – even though the two matched individuals looked identical in the data set – the actual odds of assignment was up to twice as likely for one member in the pair than the other. Likely this difference is due to the unobserved covariates.
- The question then becomes: Is what's left lingering out there, outside of your data set, enough to cause that level of confounding?

practical issue

venturing out of the ivory tower.

assessing covariate balance

- Assessing covariate balance

unmatched

	High NICU	Low NICU	sd	Δ/sd
death	2.26%	1.25%	13.67%	0.07
birth weight (g)	2,454	2,693	739	-0.32
gestational age (months)	34.61	35.69	2.76	-0.39

matched

	High NICU	Low NICU	sd	Δ/sd
death	1.55%	1.94%	13.67%	-0.03
birth weight (g)	2,584	2,581	739	0.00
gestational age (months)	35.14	35.13	2.76	0.00

assessing covariate balance

- Standardized difference

(i) Create a weighted standard deviation using pre-match observations (i.e., use all observations).

$$s_{all,k} = \sqrt{\frac{s_{t,k}^2 + s_{c,k}^2}{2}}$$

where $s_{t,k}^2$ is the standard deviation of covariate $\boldsymbol{x}_k$ amongst the treated group prior to matching.

(i) Divide the difference of the observed means by the weighted standard deviation.

$$\frac{\overline{x_{t,k}} - \overline{x_{c,k}}}{s_{all,k}}$$

Links: 1 and 2

assessing covariate balance

- Assessing covariate balance

unmatched

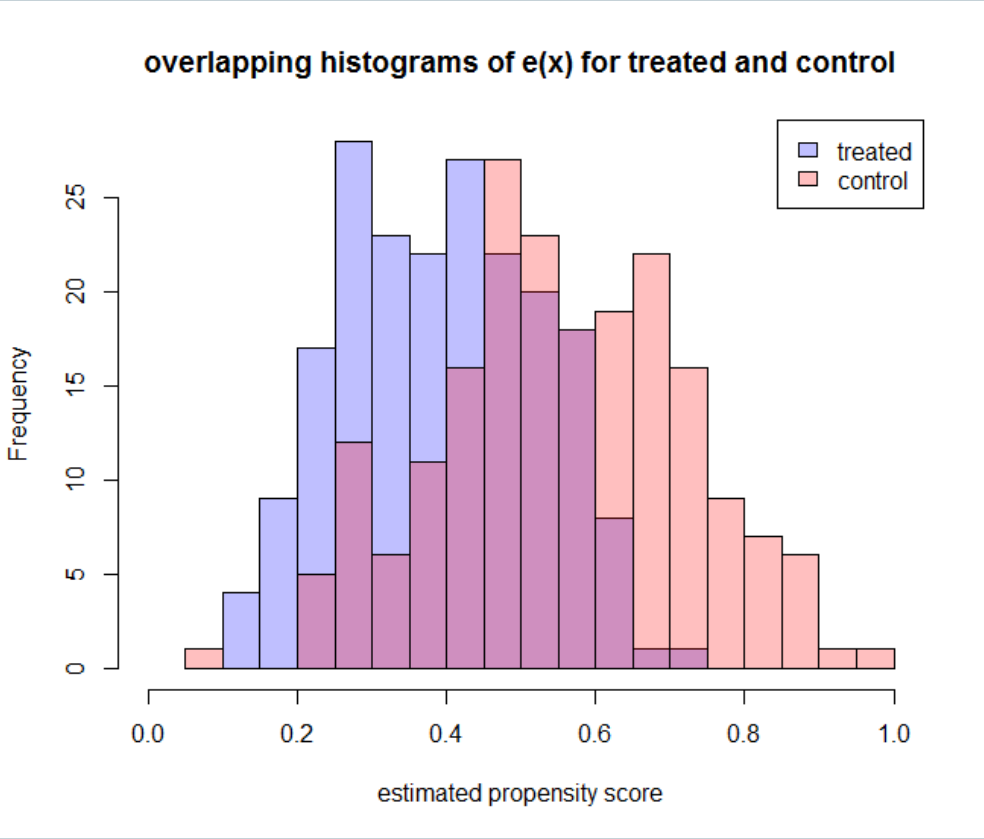
	High NICU	Low NICU	sd	Δ/sd
death	2.26%	1.25%	13.67%	0.07
birth weight (g)	2,454	2,693	739	-0.32
gestational age (months)	34.61	35.69	2.76	-0.39

matched

	High NICU	Low NICU	sd	Δ/sd
death	1.55%	1.94%	13.67%	-0.03
birth weight (g)	2,584	2,581	739	0.00
gestational age (months)	35.14	35.13	2.76	0.00

- The observed difference between the treated and control groups is judged by the typical variation in that covariate.

assessing covariate balance



dealing with lots of observations

- If you get lots of observations then you should be happy.
- If you try to put them all into a matching algorithm then you will be sad.
- The complexity of matching algorithms grows really fast so cutting down the problem into smaller chunks helps a lot.
- Look at your covariates:
  - Is there one or two that are binary or categorical?
  - Break your data set into separate data sets and match within a given level of a variable (or variables).
  - Choose variables that are prognostically important.
  - It's nice if these variables are close to uniformly distributed (e.g., p=0.5, or p=<1/3, 1/3, 1/3>).

dealing with lots of observations

- In the NICU example, we had millions of babies.
- I sub-setted the data on gestational age (i.e., 26 weeks only matched to 26 weeks).
- For larger gestational age groups, I further sub-setted on birth weight.
  - This was much less satisfactory because it's more continuous.
  - I picked arbitrary boundaries and didn't look back…
- You can fret about the matching method, but do not confuse that for the quality of the match which is assessed by looking at the covariates.

missing covariates

- Missing covariates

obs	b_weight	gest_age	dose	death	e^(x)
1	2412	36	1	0	
2	NA	29	1	1	
3	2569	36	1	0	
4	2443	34	1	0	
5	2569	36	0	0	
6	2436	NA	0	0	
7	2461	34	0	0	
8	2759	32	0	0	
9	2324	27	0	1	
10	2667	34	0	0	

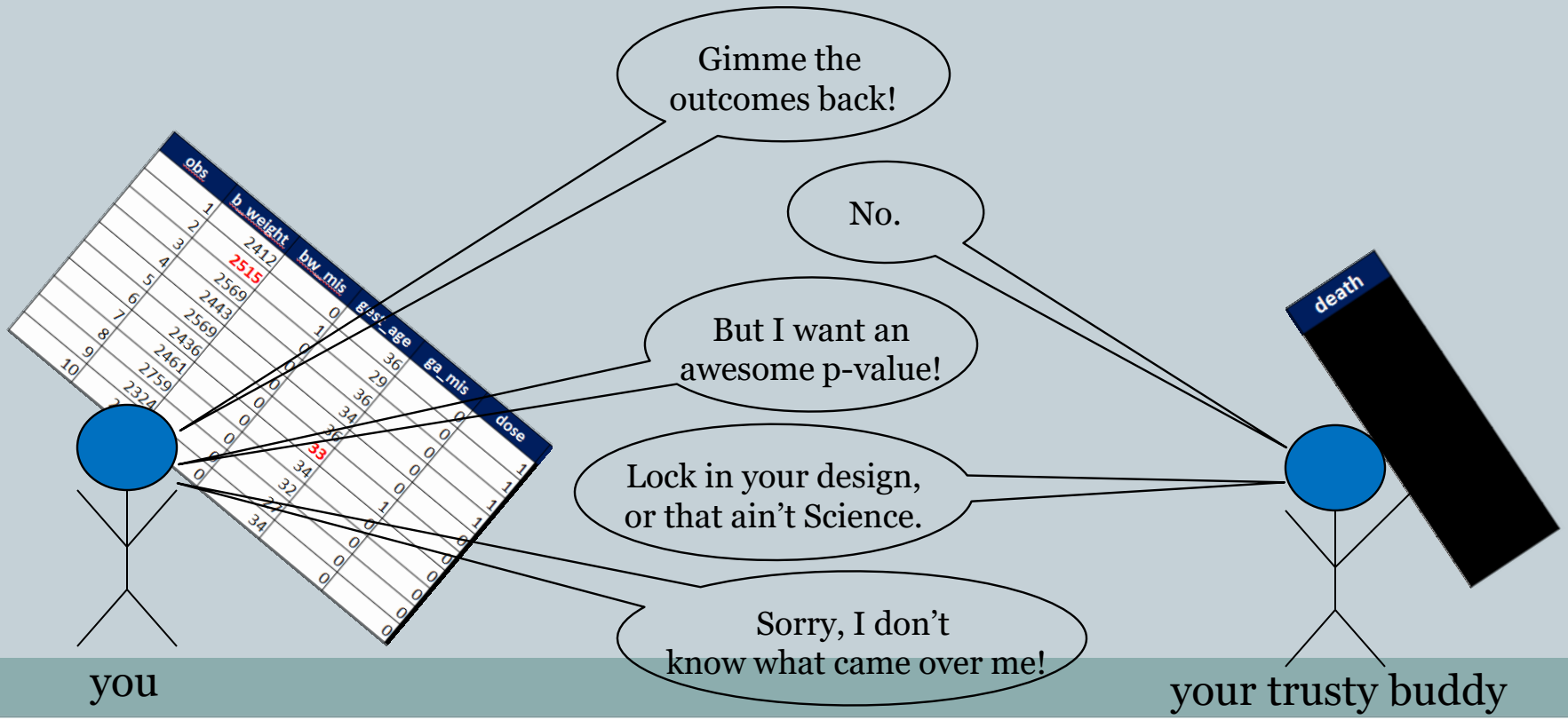
missing covariates

- Missing covariates

obs	b_weight	bw_mis	gest_age	ga_mis	dose	death
1	2412	0	36	0	1	0
2	**2515**	1	29	0	1	1
3	2569	0	36	0	1	0
4	2443	0	34	0	1	0
5	2569	0	36	0	0	0
6	2436	0	**33**	1	0	0
7	2461	0	34	0	0	0
8	2759	0	32	0	0	0
9	2324	0	27	0	0	1
10	2667	0	34	0	0	0

(i) Build pscores using the imputed value and the missing indicators.
(ii) Use imputed values and missing indicators in calculating the Mahalanobis distance.

a small but important point



fin.