Advanced Statistical Methods for Observational Studies

LECTURE 05

class management

class management

- Projects

class management

- Projects
  - Come talk to us.

class management

- Projects
  - Come talk to us.
  - The idea is you will do a 10 minute presentation (slide deck) for the two of us and we'll ask you questions.

class management

- Projects
  - Come talk to us.
  - The idea is you will do a 10 minute presentation (slide deck) for the two of us and we'll ask you questions. (Last year we held these in David's office.)

class management

- Projects
  - Come talk to us.
  - The idea is you will do a 10 minute presentation (slide deck) for the two of us and we'll ask you questions. (Last year we held these in David's office.)
- Take home exam will go up on the website this Friday.

class management

- Projects
  - Come talk to us.
  - The idea is you will do a 10 minute presentation (slide deck) for the two of us and we'll ask you questions. (Last year we held these in David's office.)
- Take home exam will go up on the website this Friday.
  - Will be due 10 days later (Monday class time).

class management

- Projects
  - Come talk to us.
  - The idea is you will do a 10 minute presentation (slide deck) for the two of us and we'll ask you questions. (Last year we held these in David's office.)
- Take home exam will go up on the website this Friday.
  - Will be due 10 days later (Monday class time).
  - Physical copies (not electronic).

class management

- Projects
  - Come talk to us.
  - The idea is you will do a 10 minute presentation (slide deck) for the two of us and we'll ask you questions. (Last year we held these in David's office.)
- Take home exam will go up on the website this Friday.
  - Will be due 10 days later (Monday class time).
  - Physical copies (not electronic).
- Questions?

unobserved confounding

Design of Observational Studies: chapter 3.4-3.8

unobserved confounding

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

naïve model

- Model
- Assumptions
- Implementation

naïve model: "natural" experiments

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

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

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

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

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

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

naïve model: assumption two

- No Interference Between Units:

naïve model: assumption two

- <u>No Interference Between Units</u>: the observation on one unit should be unaffected by particular assignment of treatments to other units.

naïve model: assumption two

- No Interference Between Units: 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 $Z_i = z_i$.

naïve model: assumption two

- No Interference Between Units: 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.

naïve model: assumption two

- No Interference Between Units: 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: 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

naïve model: implementation

- Collect a bunch of covariates that are related to treatment level and to the outcome.

naïve model: implementation

- Collect a bunch of covariates that are related to treatment level and to the outcome.
- Exact match if you can.

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.

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.

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 critics' reviews because they point out all of the confounding that could exist.

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 critics' reviews because they point out all of the confounding that could exist. Reevaluate life choices.

sensitivity analysis

sensitivity analysis

- Sensitivity models are a means for moving past the "you didn't do X which could lead to bias" argument.

sensitivity analysis

- Sensitivity models are a means for moving past the “you didn’t do X which could lead to bias” argument.
- A useful sensitivity model addresses one assumption at a time, quantifying and making understandable the impact of departures from the assumption being assessed.

sensitivity analysis

- Sensitivity models are a means for moving past the “you didn’t do X which could lead to bias” argument.
- 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)

sensitivity analysis

- Sensitivity models are a means for moving past the “you didn’t do X which could lead to bias” argument.
- 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:*

sensitivity analysis

- *A word of warning*: many people find the Γ sensitivity model confusing.

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.

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.

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

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).

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

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:

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

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

model: sensitivity analysis

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

model: sensitivity analysis

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

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

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 $\boldsymbol{x}_i = \boldsymbol{x}_j$.

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).

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 $\pi_i = \pi_j$.

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

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 $\pi_i = 2/3$ then

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 $\pi_i = 2/3$ then
$$\frac{\pi_i/(1-\pi_i)}{\pi_j/(1-\pi_j)}$$

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

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.

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…

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)…

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.

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 $\Gamma > 1$.

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

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.

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 $\Gamma \to \infty$.

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 $Z_l + Z_k = 1$:

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

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:

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:

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

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

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

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.

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.

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.

Wilcoxon rank-sign test

(SEE LECTURE 04 FOR MORE DETAIL)

the Wilcoxon signed rank

- So… what’s a “good” test statistic to use?

the Wilcoxon signed rank

- So… what's a "good" test statistic to use?
- It depends…

the Wilcoxon signed rank

- So… what’s a “good” test statistic to use?
- It depends…
- A reasonable one is the Wilcoxon signed rank statistic

the Wilcoxon signed rank

- So… what's a "good" test statistic to use?
- It depends…
- A reasonable one is the Wilcoxon signed rank statistic
- How it works:

1.
2.
3.

the Wilcoxon signed rank

- So… what's a "good" test statistic to use?
- It depends…
- A reasonable one is the Wilcoxon signed rank statistic
- How it works:

1. Take the absolute value of the difference between pairs, $|Y_i|$
2.
3.

the Wilcoxon signed rank

- So… what's a "good" test statistic to use?
- It depends…
- A reasonable one is the Wilcoxon signed rank statistic
- How it works:

1. Take the absolute value of the difference between pairs, $|Y_i|$
2. Assign ranks to these $|Y_i|$ based on magnitude, call these ranks q_i
3.

the Wilcoxon signed rank

- So… what's a "good" test statistic to use?
- It depends…
- A reasonable one is the Wilcoxon signed rank statistic
- How it works:

1. Take the absolute value of the difference between pairs, $|Y_i|$
2. Assign ranks to these $|Y_i|$ based on magnitude, call these ranks q_i
3. The Wilcoxon signed rank statistic is the sum of the q_i where $Y_i > 0$

the Wilcoxon signed rank

How it works:

1. Take the absolute value of the difference between pairs, $|Y_i|$
2. Assign ranks to these $|Y_i|$ based on magnitude from smallest to largest, call these ranks q_i
3. The Wilcoxon signed rank statistic is the sum of the q_i where $Y_i > 0$

pair	r_Ci1	r_Ti2	y_i
1	153	159	-6
2	163	148	15
3	159	150	9
4	142	159	-17
5	155	137	18
6	140	149	-9
7	148	154	-6
8	145	140	5
9	148	134	14
10	159	133	26

the Wilcoxon signed rank

How it works:

1. Take the absolute value of the difference between pairs, $|Y_i|$
2. Assign ranks to these $|Y_i|$ based on magnitude from smallest to largest, call these ranks q_i
3. The Wilcoxon signed rank statistic is the sum of the q_i where $Y_i > 0$

pair	r_Ci1	r_Ti2	y_i	abs(y_i)	q_i
1	153	159	-6		
2	163	148	15		
3	159	150	9		
4	142	159	-17		
5	155	137	18		
6	140	149	-9		
7	148	154	-6		
8	145	140	5		
9	148	134	14		
10	159	133	26		

the Wilcoxon signed rank

How it works:

1. Take the absolute value of the difference between pairs, $|Y_i|$
2. Assign ranks to these $|Y_i|$ based on magnitude from smallest to largest, call these ranks q_i
3. The Wilcoxon signed rank statistic is the sum of the q_i where $Y_i > 0$

pair	r_Ci1	r_Ti2	y_i	abs(y_i)	q_i
1	153	159	-6	6	
2	163	148	15	15	
3	159	150	9	9	
4	142	159	-17	17	
5	155	137	18	18	
6	140	149	-9	9	
7	148	154	-6	6	
8	145	140	5	5	
9	148	134	14	14	
10	159	133	26	26	

the Wilcoxon signed rank

How it works:

1. Take the absolute value of the difference between pairs, $|Y_i|$
2. Assign ranks to these $|Y_i|$ based on magnitude from smallest to largest, call these ranks q_i
3. The Wilcoxon signed rank statistic is the sum of the q_i where $Y_i > 0$

pair	r_Ci1	r_Ti2	y_i	abs(y_i)	q_i
1	153	159	-6	6	
2	163	148	15	15	
3	159	150	9	9	
4	142	159	-17	17	
5	155	137	18	18	
6	140	149	-9	9	
7	148	154	-6	6	
8	145	140	5	5	1
9	148	134	14	14	
10	159	133	26	26	

the Wilcoxon signed rank

How it works:

1. Take the absolute value of the difference between pairs, $|Y_i|$
2. Assign ranks to these $|Y_i|$ based on magnitude from smallest to largest, call these ranks q_i
3. The Wilcoxon signed rank statistic is the sum of the q_i where $Y_i > 0$

pair	r_Ci1	r_Ti2	y_i	abs(y_i)	q_i
1	153	159	-6	6	2.5
2	163	148	15	15	
3	159	150	9	9	
4	142	159	-17	17	
5	155	137	18	18	
6	140	149	-9	9	
7	148	154	-6	6	2.5
8	145	140	5	5	1
9	148	134	14	14	
10	159	133	26	26	

the Wilcoxon signed rank

How it works:

1. Take the absolute value of the difference between pairs, $|Y_i|$
2. Assign ranks to these $|Y_i|$ based on magnitude from smallest to largest, call these ranks q_i
3. The Wilcoxon signed rank statistic is the sum of the q_i where $Y_i > 0$

pair	r_Ci1	r_Ti2	y_i	abs(y_i)	q_i
1	153	159	-6	6	2.5
2	163	148	15	15	7
3	159	150	9	9	4.5
4	142	159	-17	17	8
5	155	137	18	18	9
6	140	149	-9	9	4.5
7	148	154	-6	6	2.5
8	145	140	5	5	1
9	148	134	14	14	6
10	159	133	26	26	10

the Wilcoxon signed rank

How it works:

1. Take the absolute value of the difference between pairs, $|Y_i|$
2. Assign ranks to these $|Y_i|$ based on magnitude from smallest to largest, call these ranks q_i
3. The Wilcoxon signed rank statistic is the sum of the q_i where $Y_i > 0$

pair	r_Ci1	r_Ti2	y_i	abs(y_i)	q_i
1	153	159	-6	6	2.5
2	163	148	15	15	7
3	159	150	9	9	4.5
4	142	159	-17	17	8
5	155	137	18	18	9
6	140	149	-9	9	4.5
7	148	154	-6	6	2.5
8	145	140	5	5	1
9	148	134	14	14	6
10	159	133	26	26	10

4. Average ties

the Wilcoxon signed rank

How it works:

1. Take the absolute value of the difference between pairs, $|Y_i|$
2. Assign ranks to these $|Y_i|$ based on magnitude from smallest to largest, call these ranks q_i
3. The Wilcoxon signed rank statistic is the sum of the q_i where $Y_i > 0$

pair	r_Ci1	r_Ti2	y_i	abs(y_i)	q_i
1	153	159	-6	6	2.5
2	163	148	15	15	7
3	159	150	9	9	4.5
4	142	159	-17	17	8
5	155	137	18	18	9
6	140	149	-9	9	4.5
7	148	154	-6	6	2.5
8	145	140	5	5	1
9	148	134	14	14	6
10	159	133	26	26	10

4. Average ties
5. You can remove zeros

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 $\bar{T}$.

Sensitivity analysis

(END INTERLUDE)

model: sensitivity analysis

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.

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

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

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

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

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

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

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

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

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

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	death
1	2412	36	1	0
2	2205	29	1	1
3	2569	36	1	0
4	2443	34	1	0
5	2569	36	0	0
6	2436	35	0	0
7	2461	34	0	0
8	2759	32	0	0
9	2324	27	0	1
10	2667	34	0	0
...	...	...	...	...
500	2349	33	1	0

example: sensitivity analysis

obs	b_weight	gest_age	dose	death
1	2412	36	1	0
2	2205	29	1	1
3	2569	36	1	0
4	2443	34	1	0
5	2569	36	0	0
6	2436	35	0	0
7	2461	34	0	0
8	2759	32	0	0
9	2324	27	0	1
10	2667	34	0	0
…	…	…	…	…
500	2349	33	1	0

Similar to data set from lecture 03, but different number of observations…

example: sensitivity analysis

obs	b_weight	gest_age	dose	death
1	2412	36	1	0
2	2205	29	1	1
3	2569	36	1	0
4	2443	34	1	0
5	2569	36	0	0
6	2436	35	0	0
7	2461	34	0	0
8	2759	32	0	0
9	2324	27	0	1
10	2667	34	0	0
…	…	…	…	…
500	2349	33	1	0

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

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

example: sensitivity analysis

- Create 250 pair matches.

example: sensitivity analysis

- Create 250 pair matches.
- Using T, the usual Wilcoxon signed rank statistic:

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

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

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:

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

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:

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

- $\frac{13{,}250 - 15{,}688}{812} =$

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.

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 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:

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

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

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

example: sensitivity analysis

- Create 250 pair matches.
- Using 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:

upper bound

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

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

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

example: sensitivity analysis

- Create 250 pair matches.
- Using 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:

upper bound

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

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

Set Γ=1.11

- $\frac{13{,}250 - 15{,}688}{812} = -3.00$, 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:

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

upper bound

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

Set Γ=1.11

- $\frac{13{,}250 - 16{,}540}{810} = -4.00$, 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[$\bar{T}$]=16,540 and sd($\bar{T}$)=810
- Get T=13,250
- Using the approximation:

upper bound

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

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

Set Γ=1.11

- $\frac{13,250-16,540}{810} = -4.00$, 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[$\bar{T}$]=16,540 and sd($\bar{T}$)=810
- Get T=13,250
- Using the approximation:

upper bound

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

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

Set Γ=1.11

- $\frac{13{,}250 - 16{,}540}{810} = -4.00$, 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[$\bar{T}$]=16,540 and sd($\bar{T}$)=810
- Get T=13,250
- Using the approximation:

upper bound

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

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

Set Γ=1.11

- $\frac{13{,}250 - 16{,}540}{810} = -4.00$, 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:

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

lower bound

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

Set Γ=1.11

- $\frac{13{,}250-14{,}835}{810} = -1.95$, which has a p-value close to 0.05.

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:

lower bound

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

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

Set Γ=1.11

- $\frac{13{,}250 - 14{,}835}{810} = -1.95$, which has a p-value close to 0.05.

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:

lower bound

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

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

Set Γ=1.11

- $\frac{13,250-14,835}{810} = -1.95$, 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

implementation: sensitivity analysis

- In practice, software will do this for you and you will interpret.

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:

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.

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:

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: $T \sim N(\frac{I(I+1)}{4}, \frac{I(I+1)(2I+1)}{24})$

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: $T \sim N(\frac{I(I+1)}{4}, \frac{I(I+1)(2I+1)}{24})$
- Biased toward one way: $T \sim N(\frac{\Gamma}{1+\Gamma}\frac{I(I+1)}{2}, \frac{\Gamma}{(1+\Gamma)^2}\frac{I(I+1)(2I+1)}{6})$

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: $T \sim N(\frac{I(I+1)}{4}, \frac{I(I+1)(2I+1)}{24})$
- Biased toward one way: $T \sim N(\frac{\Gamma}{1+\Gamma}\frac{I(I+1)}{2}, \frac{\Gamma}{(1+\Gamma)^2}\frac{I(I+1)(2I+1)}{6})$
- Biased other way: $T \sim N(\frac{1}{1+\Gamma}\frac{I(I+1)}{2}, \frac{\Gamma}{(1+\Gamma)^2}\frac{I(I+1)(2I+1)}{6})$

implementation: sensitivity analysis

- Use the new distributions to test your statistic to see where its critical values are.

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

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).

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.

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

- 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). Generally (though not always), you'll report the largest p-value in the interval.

implementation: sensitivity analysis

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)…

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.

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…

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

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.

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.

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?

sensitivity analysis: understanding gamma

sensitivity analysis: understanding gamma

- When you report Gamma, you report the maximal Gamma that still returns a qualitatively similar conclusion (e.g., if you found the treatment positive and significant then you report the value of Gamma that just barely makes the test not-significant).

sensitivity analysis: understanding gamma

- When you report Gamma, you report the maximal Gamma that still returns a qualitatively similar conclusion (e.g., if you found the treatment positive and significant then you report the value of Gamma that just barely makes the test not-significant).
- Note: This says nothing about the case at hand.

sensitivity analysis: understanding gamma

- When you report Gamma, you report the maximal Gamma that still returns a qualitatively similar conclusion (e.g., if you found the treatment positive and significant then you report the value of Gamma that just barely makes the test not-significant).
- Note: This says nothing about the case at hand. You are looking at the strength of the argument (i.e., how many in treated need to be switched to control before null).

sensitivity analysis: understanding gamma

- When you report Gamma, you report the maximal Gamma that still returns a qualitatively similar conclusion (e.g., if you found the treatment positive and significant then you report the value of Gamma that just barely makes the test not-significant).
- Note: This says nothing about the case at hand. You are looking at the strength of the argument (i.e., how many in treated need to be switched to control before null). But we haven't measured the level of unobserved (it's still unobserved!!).

sensitivity analysis: understanding gamma

- When you report Gamma, you report the maximal Gamma that still returns a qualitatively similar conclusion (e.g., if you found the treatment positive and significant then you report the value of Gamma that just barely makes the test not-significant).
- Note: This says nothing about the case at hand. You are looking at the strength of the argument (i.e., how many in treated need to be switched to control before null). But we haven't measured the level of unobserved (it's still unobserved!!).
- This is like saying a building can withstand a magnitude 8 earthquake.

sensitivity analysis: understanding gamma

- When you report Gamma, you report the maximal Gamma that still returns a qualitatively similar conclusion (e.g., if you found the treatment positive and significant then you report the value of Gamma that just barely makes the test not-significant).
- Note: This says nothing about the case at hand. You are looking at the strength of the argument (i.e., how many in treated need to be switched to control before null). But we haven’t measured the level of unobserved (it’s still unobserved!!).
- This is like saying a building can withstand a magnitude 8 earthquake. There is no statement in there about what magnitude earthquake the building will experience.

sensitivity analysis: understanding gamma

sensitivity analysis: understanding gamma

- Explaining Gamma sensitivity is hard because the concept of confounding is tough.

sensitivity analysis: understanding gamma

- Explaining Gamma sensitivity is hard because the concept of confounding is tough.
- Confounding (usually) arises from sorting into treatment/control…

sensitivity analysis: understanding gamma

- Explaining Gamma sensitivity is hard because the concept of confounding is tough.
- Confounding (usually) arises from sorting into treatment/control on variables important in determining the outcome.

sensitivity analysis: understanding gamma

- Explaining Gamma sensitivity is hard because the concept of confounding is tough.
- Confounding (usually) arises from sorting into treatment/control (**propensity**) on variables important in determining the outcome.

sensitivity analysis: understanding gamma

- Explaining Gamma sensitivity is hard because the concept of confounding is tough.
- Confounding (usually) arises from sorting into treatment/control (**propensity**) on variables important in determining the outcome (**prognostic**).

sensitivity analysis: understanding gamma

- Explaining Gamma sensitivity is hard because the concept of confounding is tough.
- Confounding (usually) arises from sorting into treatment/control (**propensity - Fisher**) on variables important in determining the outcome (**prognostic**).

sensitivity analysis: understanding gamma

- Explaining Gamma sensitivity is hard because the concept of confounding is tough.
- Confounding (usually) arises from sorting into treatment/control (**propensity - Fisher**) on variables important in determining the outcome (**prognostic – Mill**).

sensitivity analysis: understanding gamma

- Explaining Gamma sensitivity is hard because the concept of confounding is tough.
- Confounding (usually) arises from sorting into treatment/control (propensity - Fisher) on variables important in determining the outcome (prognostic – Mill).
- Gamma sensitivity is kind of weird because it only focusses on the propensity, and then assumes the worst case for prognostic.

an example

matching on more than one metric

matching on more than one metric

- Intuition: matching on just propensity scores is like uniform randomization, whereas a Mahalanobis & pscores is more like a matched pairs randomization.

example

- House, MD:

example

- House, MD:

example

- Example: House examines the patient and wants to treat for sarcoidosis.

example

- Example: House examines the patient and wants to treat for sarcoidosis. He is always considering treating for sarcoidosis…

example

- Example: House examines the patient and wants to treat for sarcoidosis. He is always considering treating for sarcoidosis… but in a way that is unrelated to how sick the patient is.

example

- Example: House examines the patient and wants to treat for sarcoidosis. He is always considering treating for sarcoidosis… but in a way that is unrelated to how sick the patient is.
- To make this example easier to follow, let's consider two data generating functions

example

- Example: House examines the patient and wants to treat for sarcoidosis. He is always considering treating for sarcoidosis… but in a way that is unrelated to how sick the patient is.
- To make this example easier to follow, let's consider two data generating functions
  1. Treatment:
  2. Outcome:

example

- Example: House examines the patient and wants to treat for sarcoidosis. He is always considering treating for sarcoidosis… but in a way that is unrelated to how sick the patient is.
- To make this example easier to follow, let's consider two data generating functions
  1. Treatment:

     B

A		0	1
	0	0.5	0.1
	1	0.5	0.1

  2. Outcome:

example

- Example: House examines the patient and wants to treat for sarcoidosis. He is always considering treating for sarcoidosis… but in a way that is unrelated to how sick the patient is.
- To make this example easier to follow, let's consider two data generating functions

1. Treatment:

B

A

	0	1
0	0.5	0.1
1	0.5	0.1

2. Outcome:

B

A

	0	1
0	0.1	0.1
1	0.5	0.5

propensity score vs. prognostic score

propensity score vs. prognostic score

- This departure arises when the variables predictive of treatment differs from the prognostically relevant variables

propensity score vs. prognostic score

- This departure arises when the variables predictive of treatment differs from the prognostically relevant variables

- This insight led to an interesting paper:
  - Bhattacharya & Vogt "Do Instrumental Variables Belong in Propensity Scores?"

propensity score vs. prognostic score

- This departure arises when the variables predictive of treatment differs from the prognostically relevant variables

- This insight led to an interesting paper:
  - Bhattacharya & Vogt "Do Instrumental Variables Belong in Propensity Scores?"

- Prognostic score is one way to address this:
  - Ben Hansen "The prognostic analog of the propensity score"

The Buffalo

- There are more ways to use information that is commonly considered.

The Buffalo

- There are more ways to use information that is commonly considered.

Dylan Greaves

The Buffalo

- There are more ways to use information that is commonly considered.

Dylan Greaves

Rocky Aikens

The Buffalo

- There are more ways to use information that is commonly considered.
- What if, instead of using many controls to match to a treated, we use some of the controls to understand outcome variation.

Dylan Greaves

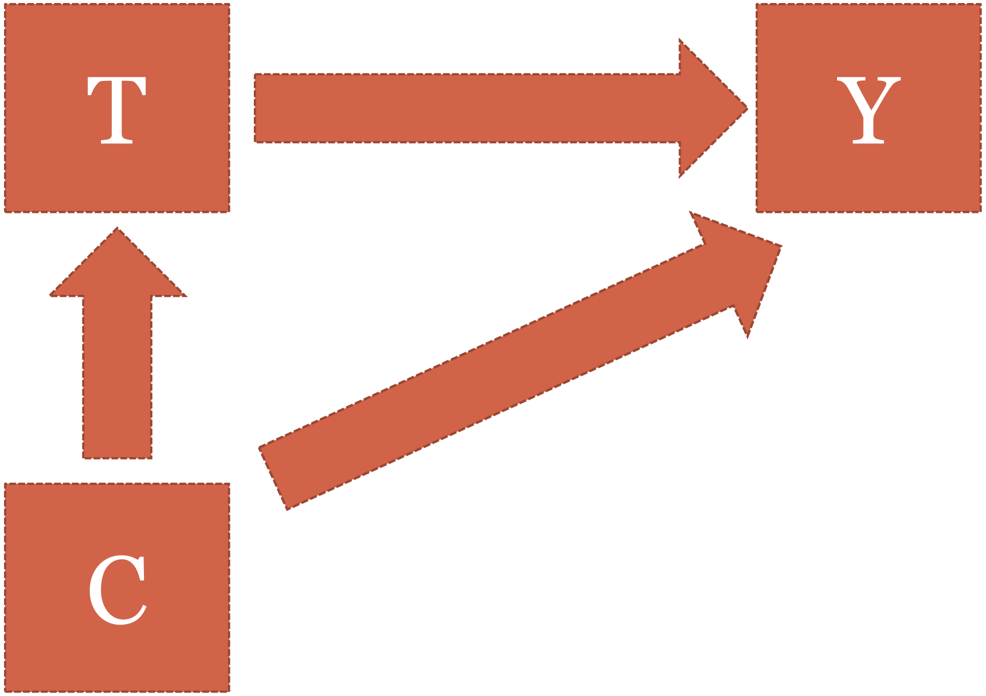
Rocky Aikens

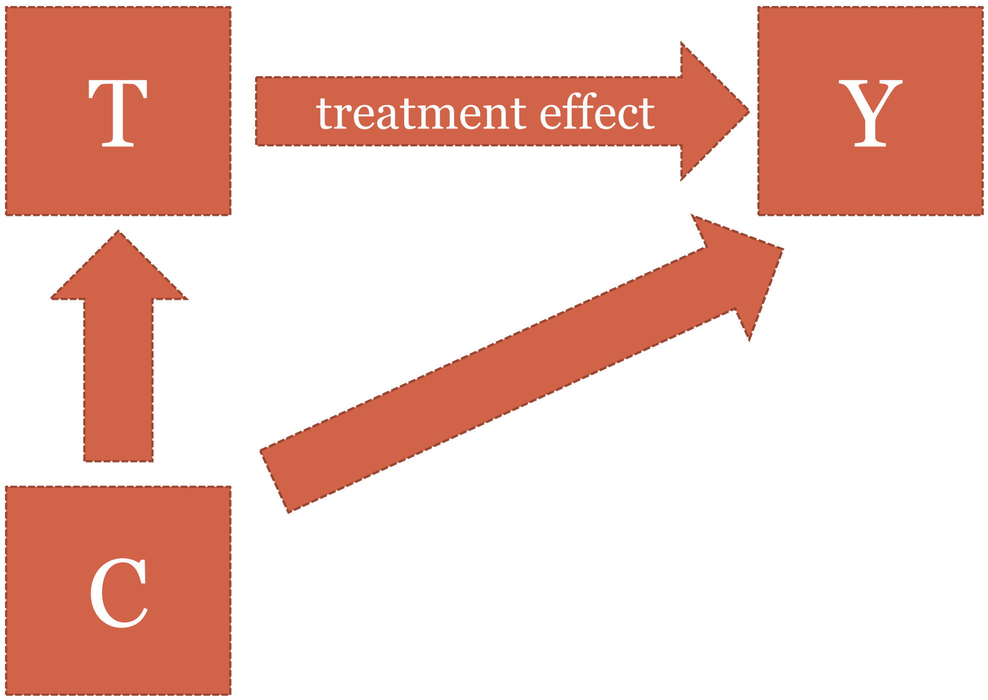
The Buffalo

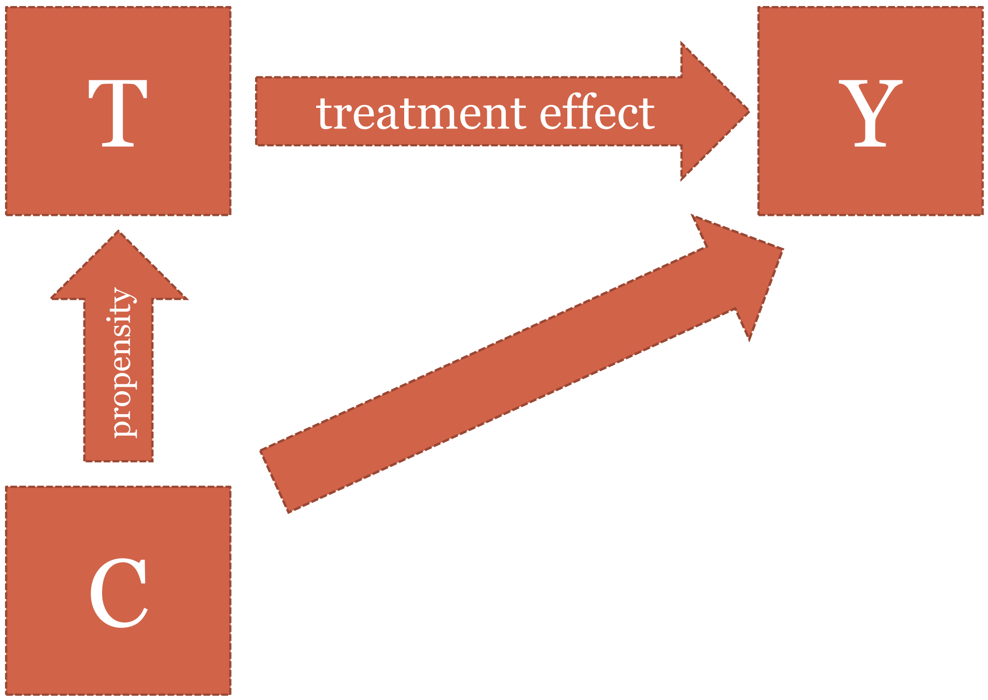
- There are more ways to use information that is commonly considered.
- What if, instead of using many controls to match to a treated, we use some of the controls to understand outcome variation.
- Problematic because we are looking at the outcome for some observations, but maybe we can "burn" those and still benefit.

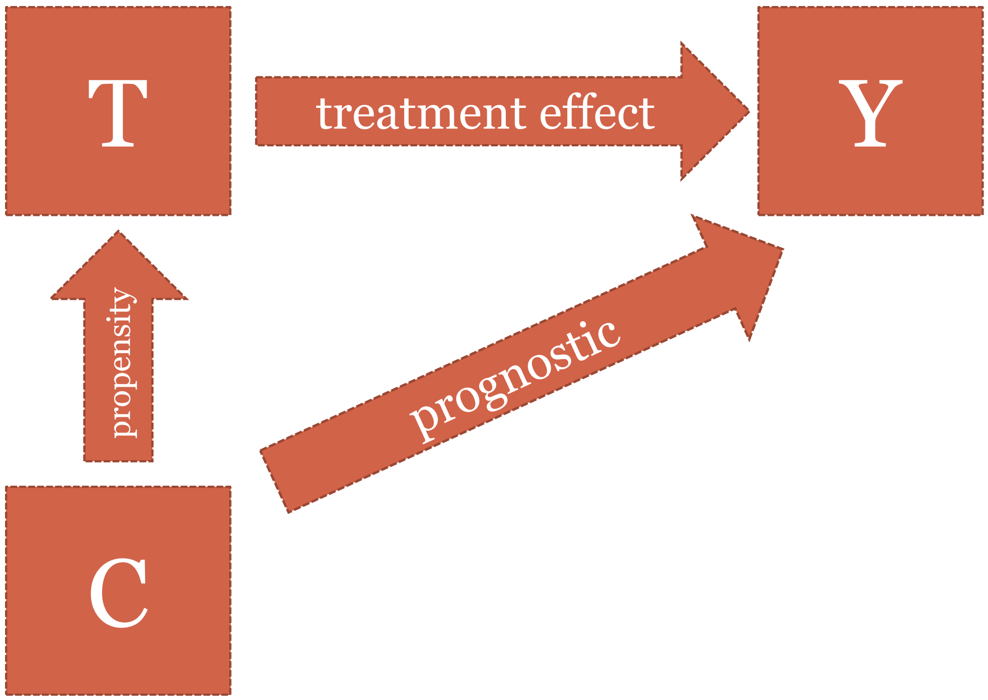
Dylan Greaves

Rocky Aikens

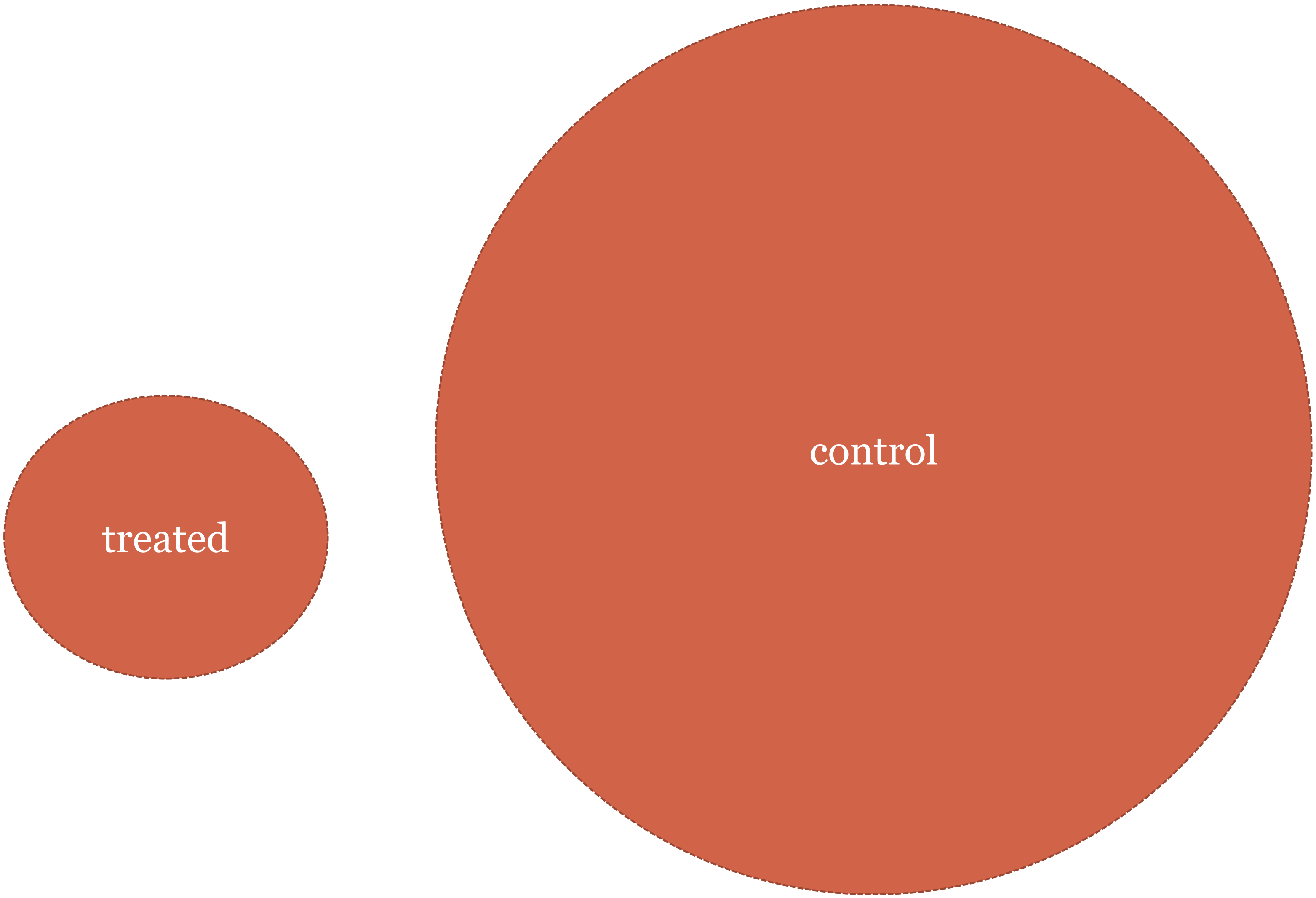

T
Y
C


T
treatment effect
Y
C

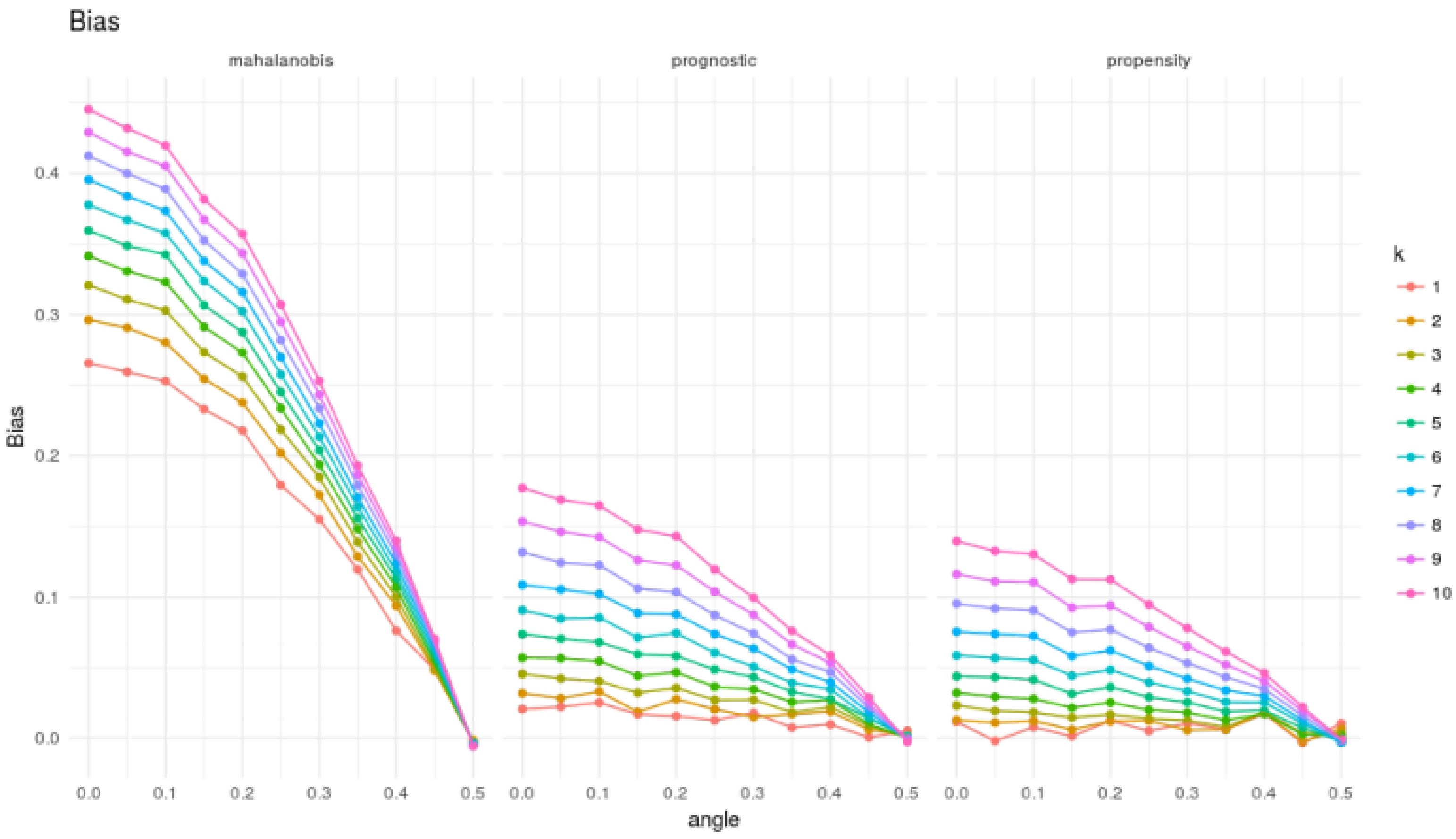

T
treatment effect
Y
propensity
C


T
treatment effect
Y
propensity
prognostic
C

treated

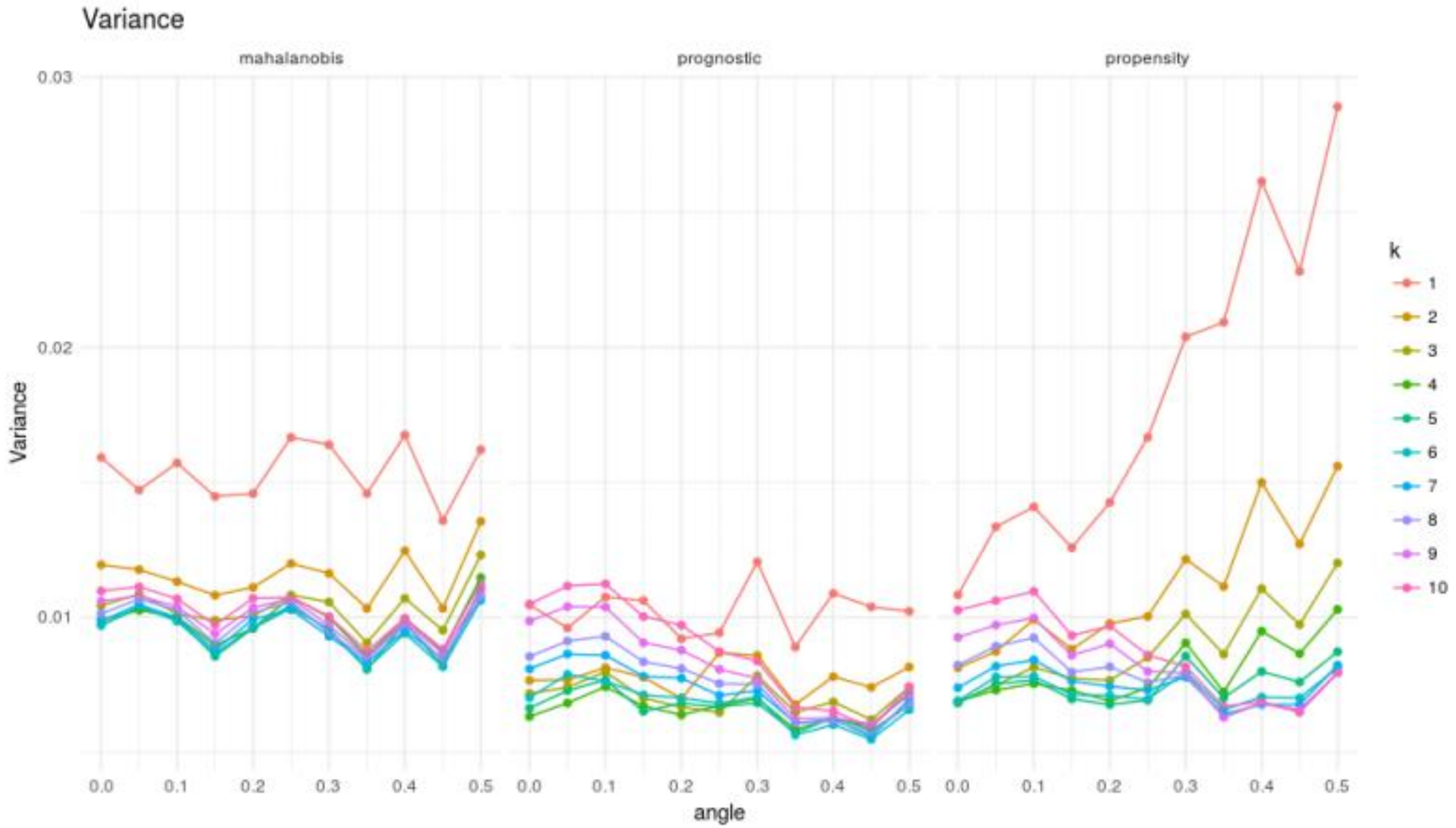

treated
control

The Buffalo

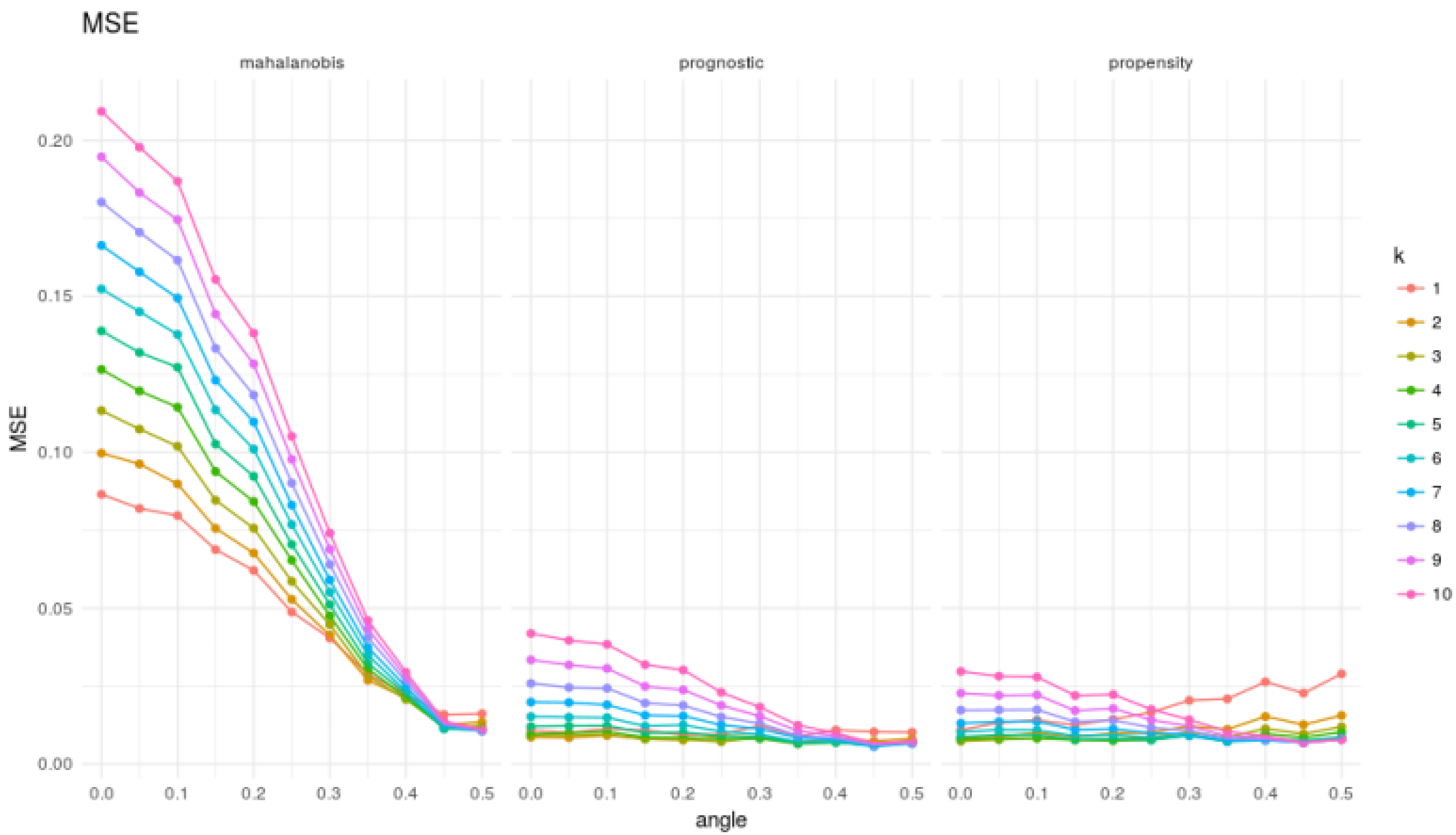


Bias
mahalanobis
prognostic
propensity
Bias
0.0
0.1
0.2
0.3
0.4
0.0 0.1 0.2 0.3 0.4 0.5
angle
k
1
2
3
4
5
6
7
8
9
10

The Buffalo

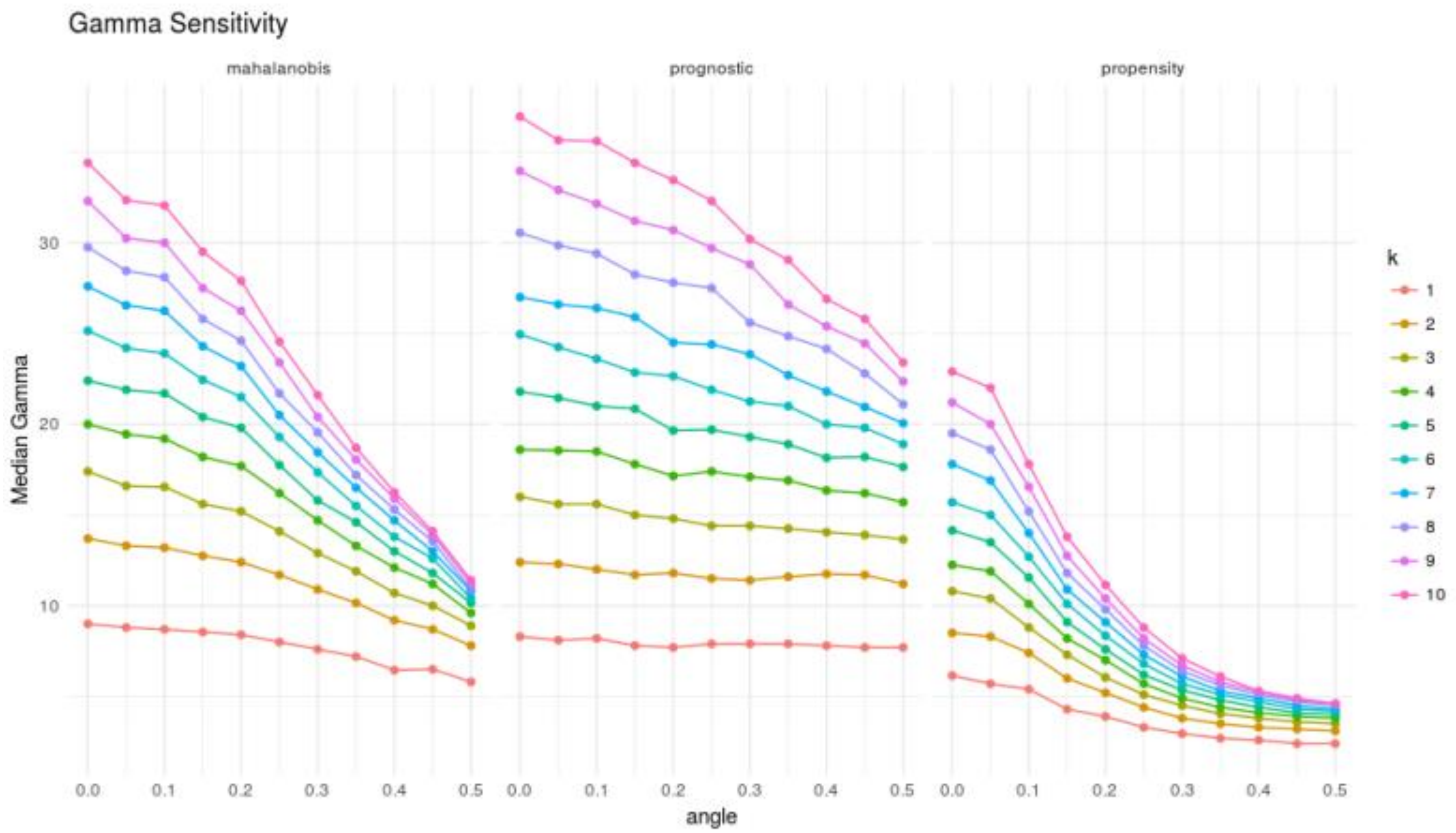

Variance
mahalanobis
prognostic
propensity
0.03
0.02
0.01
Variance
0.0
0.1
0.2
0.3
0.4
0.5
angle
k
1
2
3
4
5
6
7
8
9
10

The Buffalo



MSE
mahalanobis
prognostic
propensity
MSE
0.00
0.05
0.10
0.15
0.20
0.0
0.1
0.2
0.3
0.4
0.5
angle
k
1
2
3
4
5
6
7
8
9
10

The Buffalo


Gamma Sensitivity
mahalanobis
prognostic
propensity
Median Gamma
30
20
10
0.0
0.1
0.2
0.3
0.4
0.5
angle
k
1
2
3
4
5
6
7
8
9
10

person id	x1	x2	...	x_p	treated
1	109	119		71	0
2	106	48		91	1
3	106	81		35	1
4	102	70		79	0
5	65	68		38	0
6	110	118		74	0
7	73	101		69	1
8	50	90		81	0
9	94	71		54	0
10	38	120		42	1
11	54	116		56	0
12	56	106		50	0
13	112	88		75	1

person id	x1	x2	…	x_p	treated	propensity
1	109	119		71	0	0.32
2	106	48		91	1	0.54
3	106	81		35	1	0.32
4	102	70		79	0	0.32
5	65	68		38	0	0.32
6	110	118		74	0	0.54
7	73	101		69	1	0.54
8	50	90		81	0	0.54
9	94	71		54	0	0.54
10	38	120		42	1	0.32
11	54	116		56	0	0.32
12	56	106		50	0	0.54
13	112	88		75	1	0.32

person id	x1	x2	...	x_p	treated	propensity
1	109	119		71	0	0.32
3	106	81		35	1	0.32
4	102	70		79	0	0.32
5	65	68		38	0	0.32
10	38	120		42	1	0.32
11	54	116		56	0	0.32
13	112	88		75	1	0.32
2	106	48		91	1	0.54
6	110	118		74	0	0.54
7	73	101		69	1	0.54
8	50	90		81	0	0.54
9	94	71		54	0	0.54
12	56	106		50	0	0.54

person id	x1	x2	...	x_p	treated	propensity	prognostic
1	109	119		71	0	0.32	10
3	106	81		35	1	0.32	16
4	102	70		79	0	0.32	6
5	65	68		38	0	0.32	13
10	38	120		42	1	0.32	18
11	54	116		56	0	0.32	18
13	112	88		75	1	0.32	5
2	106	48		91	1	0.54	18
6	110	118		74	0	0.54	13
7	73	101		69	1	0.54	7
8	50	90		81	0	0.54	18
9	94	71		54	0	0.54	16
12	56	106		50	0	0.54	5

person id	x1	x2	...	x_p	treated	propensity	prognostic
13	112	88		75	1	0.32	5
12	56	106		50	0	0.54	5
4	102	70		79	0	0.32	6
7	73	101		69	1	0.54	7
1	109	119		71	0	0.32	10
5	65	68		38	0	0.32	13
6	110	118		74	0	0.54	13
3	106	81		35	1	0.32	16
9	94	71		54	0	0.54	16
10	38	120		42	1	0.32	18
11	54	116		56	0	0.32	18
2	106	48		91	1	0.54	18
8	50	90		81	0	0.54	18

takeaway

takeaway

- This toy example highlights that the propensity score focuses on treatment, which may be unrelated to outcomes.

takeaway

- This toy example highlights that the propensity score focuses on treatment, which may be unrelated to outcomes.
- This is OK – the theory of inference is predicated on randomization, not identical units going into the groups (Fisher)

takeaway

- This toy example highlights that the propensity score focuses on treatment, which may be unrelated to outcomes.
- This is OK – the theory of inference is predicated on randomization, not identical units going into the groups (Fisher)
- But it is better to start with similar groups (Mill)

a second outcome

Design of Observational Studies: chapter 5.2.3 and 5.2.4
Rosenbaum, "The Role of Known Effects in Observational Studies"

the structure of the argument: two outcomes

- If your theory is well developed then you might be able to locate multiple outcomes that will support your understanding of the mechanism of the intervention.

the structure of the argument: two outcomes

- If your theory is well developed then you might be able to locate multiple outcomes that will support your understanding of the mechanism of the intervention.
- Two ways this can happen:
  - The second outcome can be compatible (show violation)
  - The confirmation of a “null effect” can help rebuff claims of unobserved biases

the structure of the argument: two outcomes

- If your theory is well developed then you might be able to locate multiple outcomes that will support your understanding of the mechanism of the intervention.*
- Two ways this can happen:
  - The second outcome can be compatible (show violation)
  - The confirmation of a "null effect" can help rebuff claims of unobserved biases

*Keep this idea separate from "intermediate effects," not because there's a deep fundamental difference in these concepts but rather conflating them will tend to confuse discussions.

coherence

- (Rough) Definition:

coherence

- (Rough) Definition: A claim is made that an intervention must have a certain form (i.e., there’s a detailed hypothesis).

coherence

- (Rough) Definition: A claim is made that an intervention must have a certain form (i.e., there's a detailed hypothesis). In this situation, coherence means a pattern of observed associations compatible with this anticipated form, and incoherence means a pattern of observed associations incompatible with this form.

coherence

- (Rough) Definition: A claim is made that an intervention must have a certain form (i.e., there's a detailed hypothesis). In this situation, coherence means a pattern of observed associations compatible with this anticipated form, and incoherence means a pattern of observed associations incompatible with this form.
- Claims of coherence or incoherence are arguable to the extent that the anticipated form of treatment effect is arguable.

coherence

- (Rough) Definition: A claim is made that an intervention must have a certain form (i.e., there's a detailed hypothesis). In this situation, coherence means a pattern of observed associations compatible with this anticipated form, and incoherence means a pattern of observed associations incompatible with this form.
- Claims of coherence or incoherence are arguable to the extent that the anticipated form of treatment effect is arguable.
- If you want to see the technical details of how to build a statistical argument around this then check out *Observational Studies*, section 17.2 (coherent signed rank statistic).

coherence

X_1

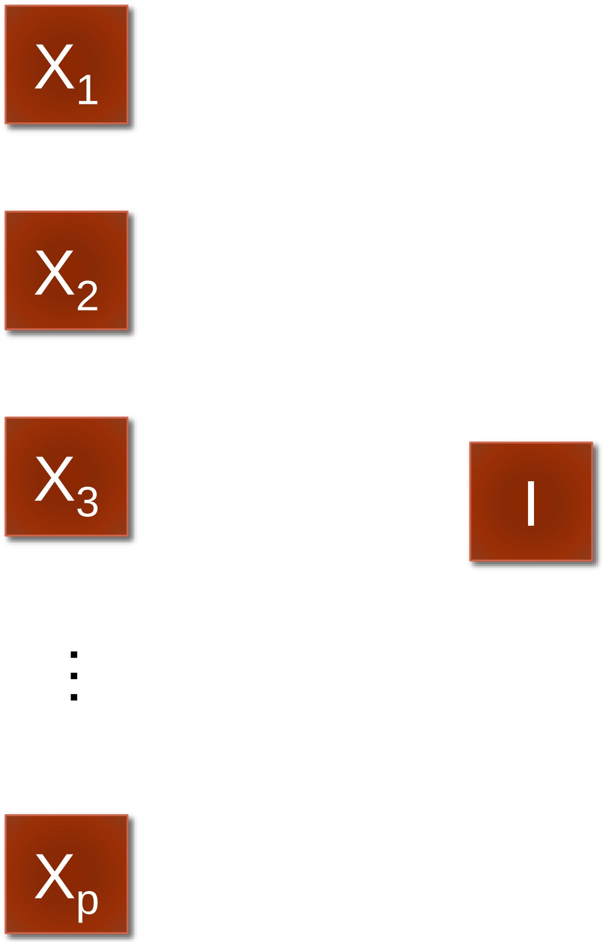
X_2

X_3

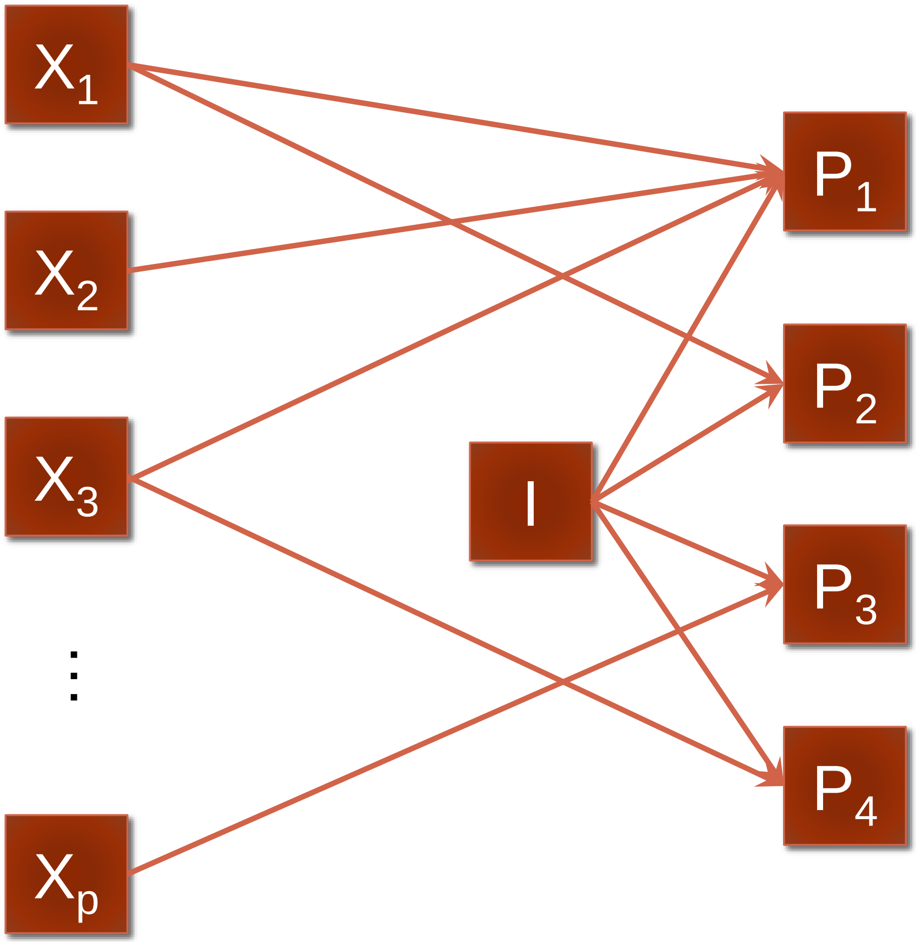
$\vdots$

X_p

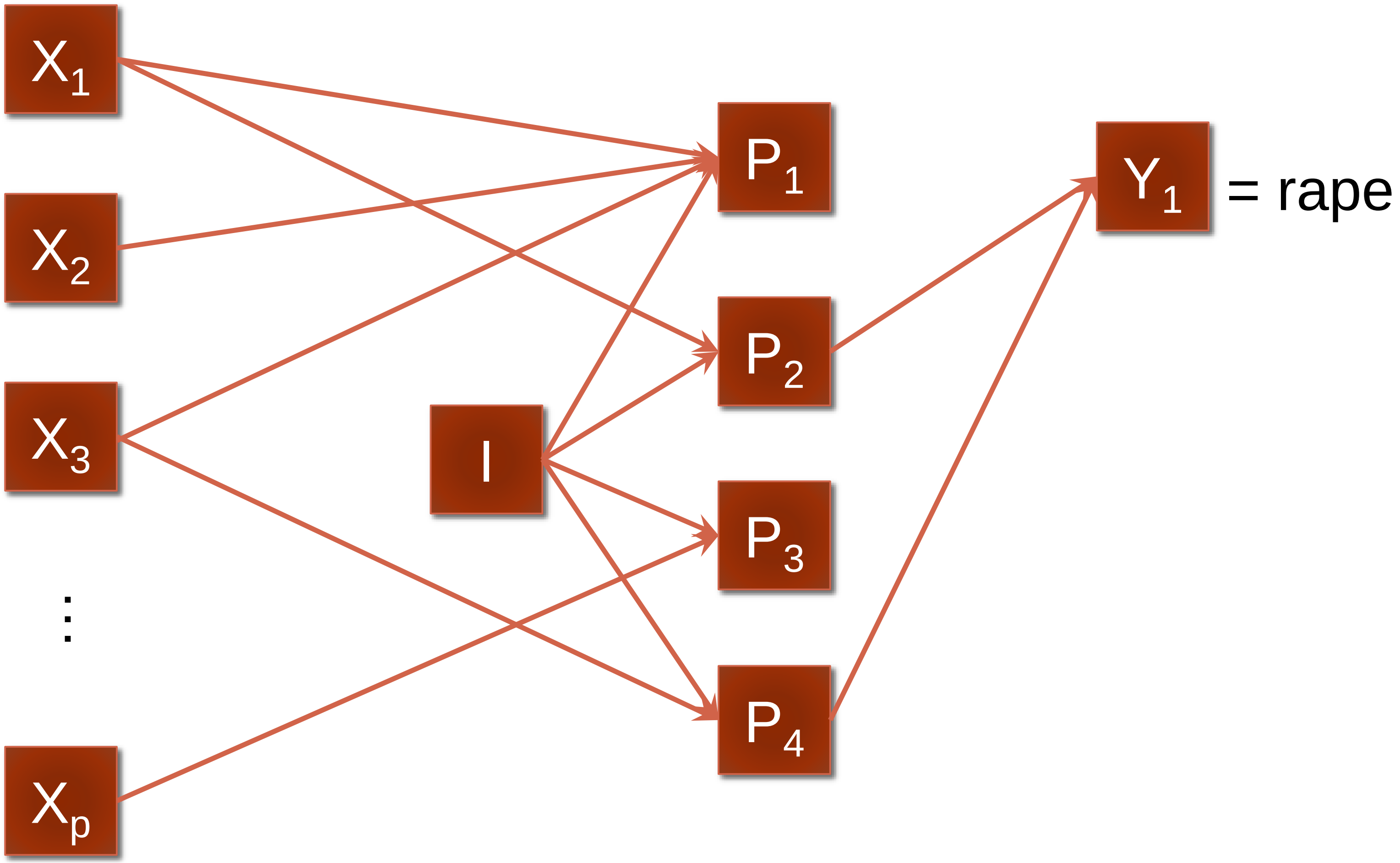
coherence



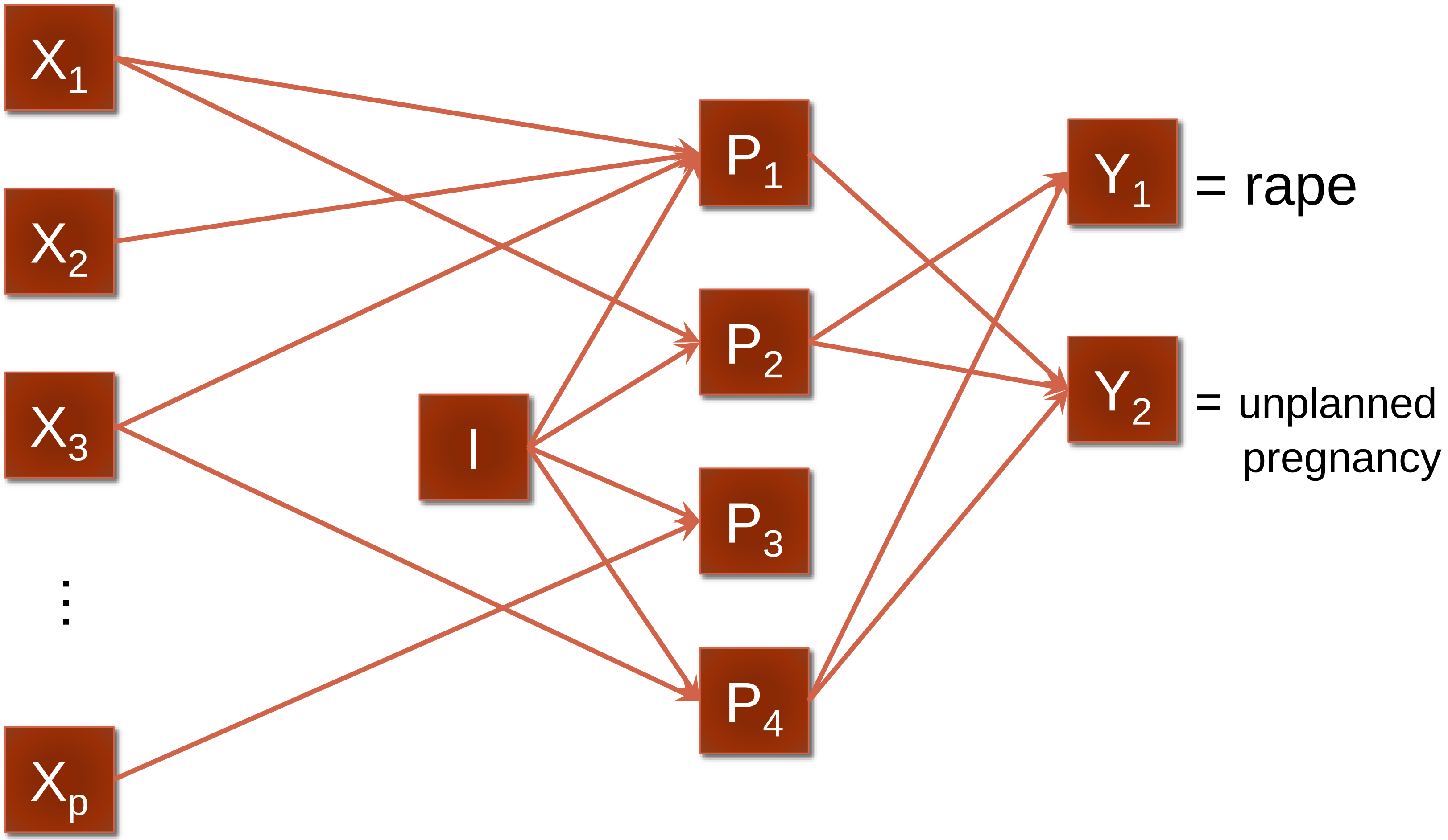
coherence



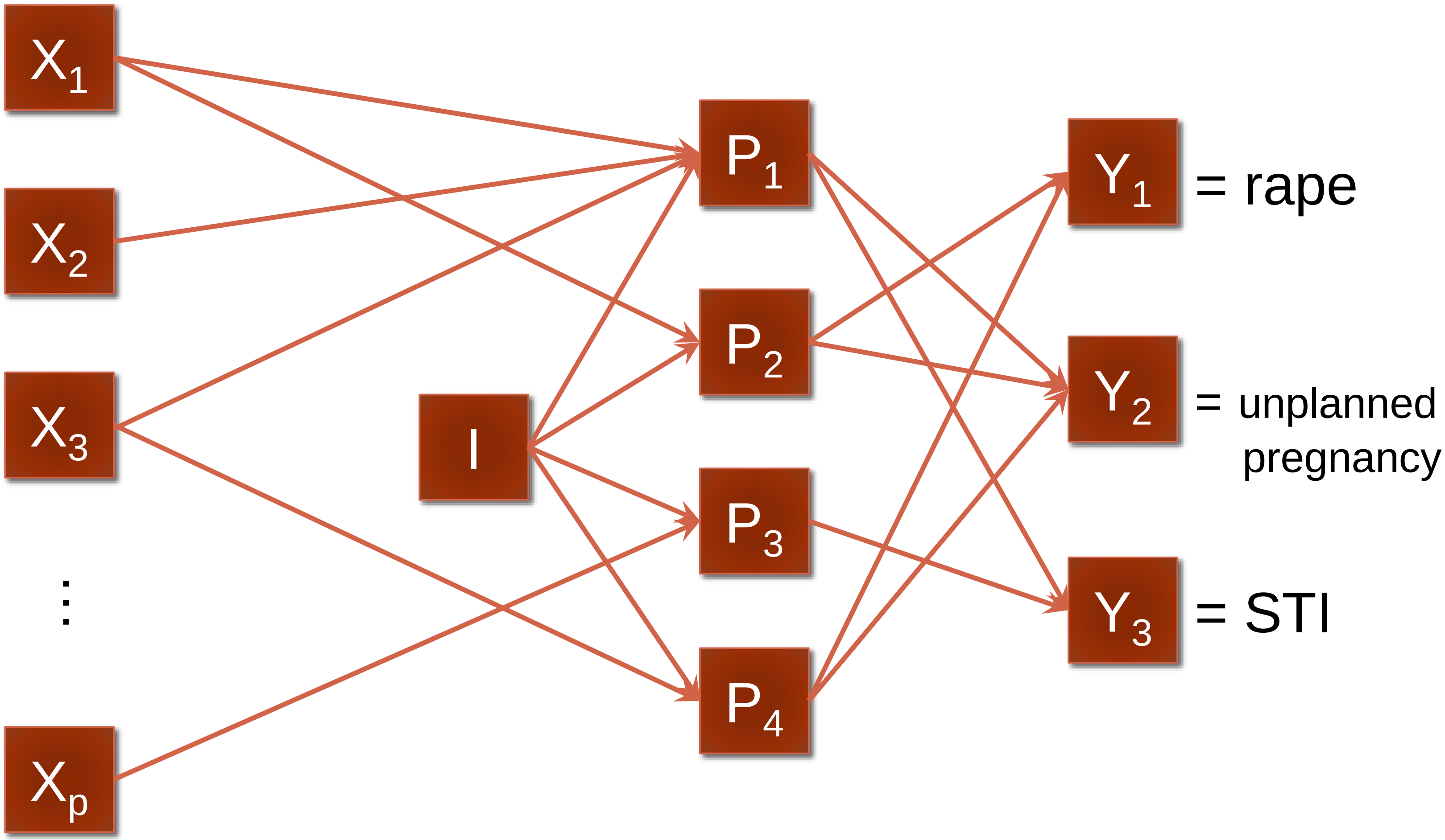
coherence



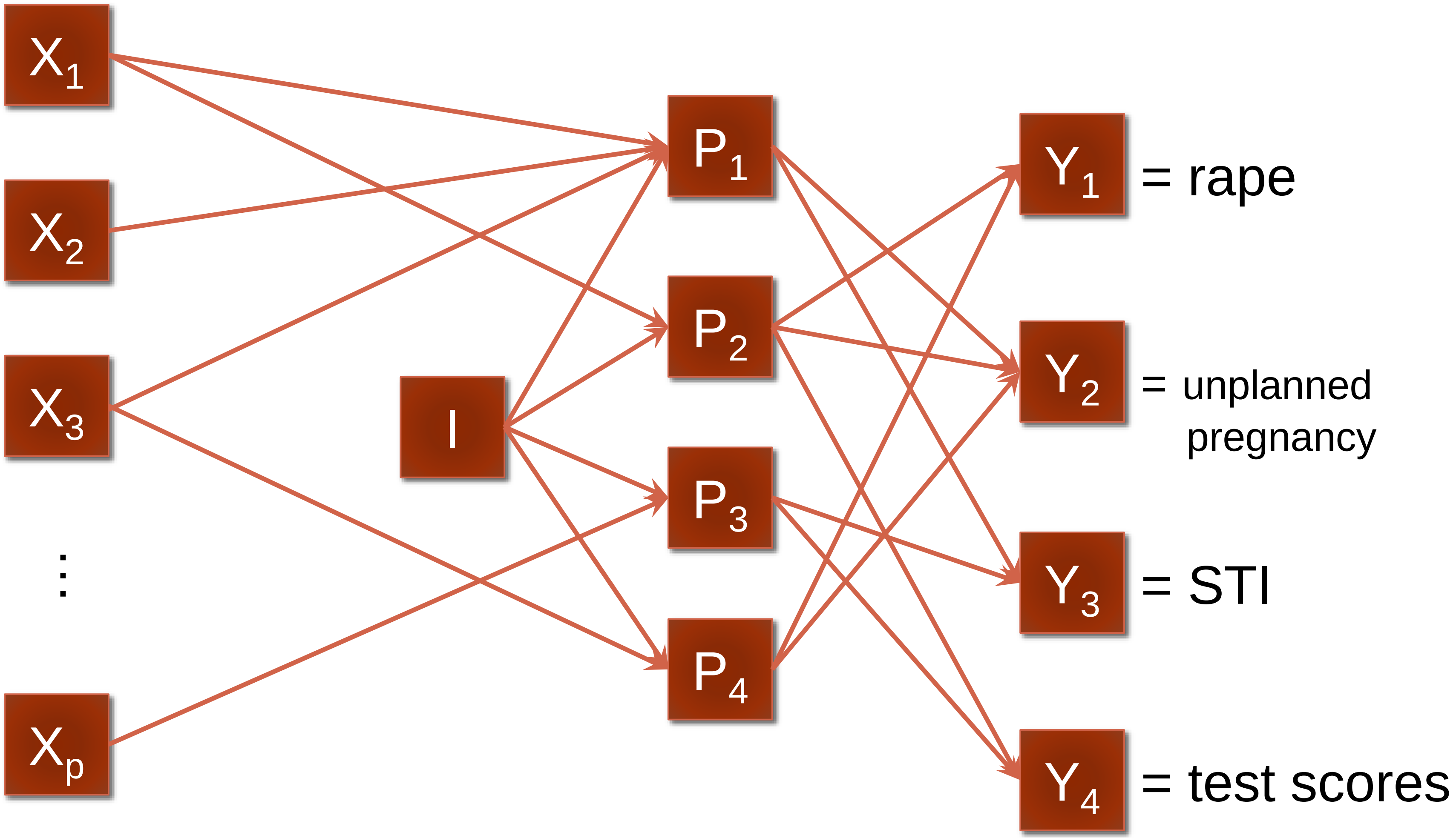
coherence



coherence



coherence


X1
X2
X3
⋮
Xp
I
P1
P2
P3
P4
Y1
Y2
Y3
Y4
= rape
= unplanned pregnancy
= STI
= test scores

the structure of the argument: null effect outcomes

- Basic idea:

the structure of the argument: null effect outcomes

- <u>Basic idea</u>: Suppose that a treatment is known to not change a particular outcome.

the structure of the argument: null effect outcomes

- Basic idea: Suppose that a treatment is known to not change a particular outcome. Then if we see differences between the treatment and control groups on this particular outcome…

the structure of the argument: null effect outcomes

- Basic idea: Suppose that a treatment is known to not change a particular outcome. Then if we see differences between the treatment and control groups on this particular outcome, this must mean that there are differences between the treatment and control group on unmeasured covariates and thus there is hidden bias.

example: methylmercury fish

- Example:

example: methylmercury fish

- Example: [Skerfving (1974)] studied whether eating fish contaminated with methylmercury causes chromosome damage.

example: methylmercury fish

- Example: [Skerfving (1974)]() studied whether eating fish contaminated with methylmercury causes chromosome damage. The outcomes of interest was the percentage of cells exhibiting chromosome damage. Pairs were matched for age and sex.

example: methylmercury fish

- Example: Skerfving (1974) studied whether eating fish contaminated with methylmercury causes chromosome damage. The outcomes of interest was the percentage of cells exhibiting chromosome damage. Pairs were matched for age and sex.

```
control.cu.cells <- c(2.7,.5,0,0,5,0,0,1.3,0,1.8,0,0,1,1.8,0,3.1)
exposed.cu.cells <- c(.7,1.7,0,4.6,0,9.5,5,2,2,2,1,3,2,3.5,0,4);
library(exactRankTests)
```

example: methylmercury fish

- Example: Skerfving (1974) studied whether eating fish contaminated with methylmercury causes chromosome damage. The outcomes of interest was the percentage of cells exhibiting chromosome damage. Pairs were matched for age and sex.

```
control.cu.cells <- c(2.7,.5,0,0,5,0,0,1.3,0,1.8,0,0,1,1.8,0,3.1)
exposed.cu.cells <- c(.7,1.7,0,4.6,0,9.5,5,2,2,2,1,3,2,3.5,0,4);
library(exactRankTests)

wilcox.exact(exposed.cu.cells,control.cu.cells,paired=TRUE)

        Exact Wilcoxon signed rank test

data:  exposed.cu.cells and control.cu.cells
V = 84, p-value = 0.04712

alternative hypothesis: true mu is not equal to 0
```

example: methylmercury fish

- In the absence of hidden bias, there’s evidence that eating large quantities of fish containing methylmercury causes chromosome damage.

example: methylmercury fish

- In the absence of hidden bias, there's evidence that eating large quantities of fish containing methylmercury causes chromosome damage.
- Going further, *Skerfving* described other health conditions of these subjects including other diseases such as…

example: methylmercury fish

- In the absence of hidden bias, there's evidence that eating large quantities of fish containing methylmercury causes chromosome damage.
- Going further, *Skerfving* described other health conditions of these subjects including other diseases such as (i) hypertension, (ii) asthma, (iii) drugs taken regularly, (iv) diagnostic X-rays over the previous three years, (v) and viral diseases such as influenza.

example: methylmercury fish

- In the absence of hidden bias, there's evidence that eating large quantities of fish containing methylmercury causes chromosome damage.
- Going further, *Skerfving* described other health conditions of these subjects including other diseases such as (i) hypertension, (ii) asthma, (iii) drugs taken regularly, (iv) diagnostic X-rays over the previous three years, (v) and viral diseases such as influenza.
- These can be considered outcomes since they describe the period when the exposed subjects were consuming contaminated fish.

example: methylmercury fish

- In the absence of hidden bias, there's evidence that eating large quantities of fish containing methylmercury causes chromosome damage.
- Going further, *Skerfving* described other health conditions of these subjects including other diseases such as (i) hypertension, (ii) asthma, (iii) drugs taken regularly, (iv) diagnostic X-rays over the previous three years, (v) and viral diseases such as influenza.
- These can be considered outcomes since they describe the period when the exposed subjects were consuming contaminated fish.
- However, it is difficult to imagine that eating fish contaminated with methylmercury causes influenza or asthma, or prompts X-rays of the hip or lumbar spine.

example: methylmercury fish

- The data

example: methylmercury fish

- The data

```
control.other.health.conditions <- c(rep(0,8),2,rep(0,3),2,1,4,1)
exposed.other.health.conditions <- c(0,0,2,0,2,0,0,1,1,2,0,9,0,0,1,0)
```

example: methylmercury fish

- The data

```
control.other.health.conditions <- c(rep(0,8),2,rep(0,3),2,1,4,1)
exposed.other.health.conditions <- c(0,0,2,0,2,0,0,1,1,2,0,9,0,0,1,0)
```

```
> wilcox.exact(control.other.health.conditions,exposed.other.health.conditions)

        Exact Wilcoxon rank sum test

data:  control.other.health.conditions and exposed.other.health.conditions
W = 112.5, p-value = 0.5257
alternative hypothesis: true mu is not equal to 0
```

example: methylmercury fish

- The data

```
control.other.health.conditions <- c(rep(0,8),2,rep(0,3),2,1,4,1)
exposed.other.health.conditions <- c(0,0,2,0,2,0,0,1,1,2,0,9,0,0,1,0)
```

```
> wilcox.exact(control.other.health.conditions,exposed.other.health.conditions)

        Exact Wilcoxon rank sum test

data:  control.other.health.conditions and exposed.other.health.conditions
W = 112.5, p-value = 0.5257
alternative hypothesis: true mu is not equal to 0
```

- There is no evidence of hidden bias.

example: methylmercury fish

- The data

```
control.other.health.conditions <- c(rep(0,8),2,rep(0,3),2,1,4,1)
exposed.other.health.conditions <- c(0,0,2,0,2,0,0,1,1,2,0,9,0,0,1,0)
```

```
> wilcox.exact(control.other.health.conditions,exposed.other.health.conditions)

        Exact Wilcoxon rank sum test

data:  control.other.health.conditions and exposed.other.health.conditions
W = 112.5, p-value = 0.5257
alternative hypothesis: true mu is not equal to 0
```

- There is no evidence of hidden bias.
- But absence of evidence is not evidence of absence.

example: methylmercury fish

- Questions:

  (1) When does such a test have a reasonable prospect of detecting hidden bias?

  (2) If no evidence of hidden bias is found, does this imply reduced sensitivity to bias in the comparisons involving the outcomes of primary interest?

  (3) If evidence of bias is found, what can be said about its magnitude and its impact on the primary comparisons?

null effect outcomes

- Power of the test of hidden bias:

null effect outcomes

- <u>Power of the test of hidden bias</u>: Let $\boldsymbol{y}$ denote the outcome for which there is a known effect of zero. For a particular unobserved covariate $\boldsymbol{u}$, what unaffected outcome $\boldsymbol{y}$ would be useful in detecting hidden bias from $\boldsymbol{u}$?

null effect outcomes

- Power of the test of hidden bias: Let $\boldsymbol{y}$ denote the outcome for which there is a known effect of zero. For a particular unobserved covariate $\boldsymbol{u}$, what unaffected outcome $\boldsymbol{y}$ would be useful in detecting hidden bias from $\boldsymbol{u}$?
- Precise statement of results in:
  - "The Role of Known Effects in Observational Studies"

null effect outcomes

- Power of the test of hidden bias: Let $\boldsymbol{y}$ denote the outcome for which there is a known effect of zero. For a particular unobserved covariate $\boldsymbol{u}$, what unaffected outcome $\boldsymbol{y}$ would be useful in detecting hidden bias from $\boldsymbol{u}$?
- Precise statement of results in:
  - "The Role of Known Effects in Observational Studies"
- Basic result:

null effect outcomes

- Power of the test of hidden bias: Let $\boldsymbol{y}$ denote the outcome for which there is a known effect of zero. For a particular unobserved covariate $\boldsymbol{u}$, what unaffected outcome $\boldsymbol{y}$ would be useful in detecting hidden bias from $\boldsymbol{u}$?
- Precise statement of results in:
  - "The Role of Known Effects in Observational Studies"
- Basic result: The power of the test of whether $\boldsymbol{y}$ is affected by the treatment increases with the strength of the relationship between $\boldsymbol{y}$ and $\boldsymbol{u}$. If one is concerned about a particular unobserved covariate $\boldsymbol{u}$, one should search for an unaffected outcome $\boldsymbol{y}$ that is strongly related to $\boldsymbol{u}$.

takeaway

takeaway

- Having a detailed understanding of how your intervention functions, what the causal pathway includes and excludes, will give you more data sources that may validate or refute your hypothesis.

takeaway

- Having a detailed understanding of how your intervention functions, what the causal pathway includes and excludes, will give you more data sources that may validate or refute your hypothesis.
- Coherence is trying to flesh out your hypothesis.

takeaway

- Having a detailed understanding of how your intervention functions, what the causal pathway includes and excludes, will give you more data sources that may validate or refute your hypothesis.
- Coherence is trying to flesh out your hypothesis.
- Known null effects may help to address unobserved confounding

a second control group

TWO PROBLEMS
TWO CONTROLS

Design of Observational Studies: chapter 5.2.2
Rosenbaum, “The Role of a Second Control Group in an Observational Study”

structure of argument

structure of argument

- In an RCT the control and treatment groups are created from a pool of study participants. The assignment to C or T is due to a researcher-directed mechanism (e.g., flipping a coin, or matched pairs). Importantly: all participants can receive C or T.

structure of argument

- In an RCT the control and treatment groups are created from a pool of study participants. The assignment to C or T is due to a researcher-directed mechanism (e.g., flipping a coin, or matched pairs). Importantly: all participants can receive C or T.
- In an observational study there are possibly many different reasons for people to have not received the treatment.

structure of argument

- In an RCT the control and treatment groups are created from a pool of study participants. The assignment to C or T is due to a researcher-directed mechanism (e.g., flipping a coin, or matched pairs). Importantly: all participants can receive C or T.
- In an observational study there are possibly many different reasons for people to have not received the treatment.
- In some situations there are discernable subgroups within the non-treatment group, each subgroup being identifiable by the reason for the subgroup not receiving the treatment.

structure of argument

- In an RCT the control and treatment groups are created from a pool of study participants. The assignment to C or T is due to a researcher-directed mechanism (e.g., flipping a coin, or matched pairs). Importantly: all participants can receive C or T.
- In an observational study there are possibly many different reasons for people to have not received the treatment.
- In some situations there are discernable subgroups within the non-treatment group, each subgroup being identifiable by the reason for the subgroup not receiving the treatment.
- In some subset of these situations these subgroups will be open to critiques of bias when compared to the treatment group, but at least two of the subgroups will differ in the nature of their bias.

structure of argument

- In an RCT the control and treatment groups are created from a pool of study participants. The assignment to C or T is due to a researcher-directed mechanism (e.g., flipping a coin, or matched pairs). Importantly: all participants can receive C or T.
- In an observational study there are possibly many different reasons for people to have not received the treatment.
- In some situations there are discernable subgroups within the non-treatment group, each subgroup being identifiable by the reason for the subgroup not receiving the treatment.
- In some subset of these situations these subgroups will be open to critiques of bias when compared to the treatment group, but at least two of the subgroups will differ in the nature of their bias.
- The contrast of these two control groups with the treatment group may strengthen your analysis.

second control group: army toxicity

- Example:

second control group: army toxicity

- Example: The army is interested in the long term effects of exposure to a list of specific chemical agents that were suspected of being toxic. Relatively few soldiers were exposed to these chemicals.

second control group: army toxicity

- Example: The army is interested in the long term effects of exposure to a list of specific chemical agents that were suspected of being toxic. Relatively few soldiers were exposed to these chemicals.
- At first pass, one might think to compare these exposed (“treated”) service members to service members who were not exposed at all.

second control group: army toxicity

- Example: The army is interested in the long term effects of exposure to a list of specific chemical agents that were suspected of being toxic. Relatively few soldiers were exposed to these chemicals.
- At first pass, one might think to compare these exposed ("treated") service members to service members who were not exposed at all.
- Complicating that comparison, though, is that the army sorted people into jobs which exposed them or to jobs which did not.

second control group: army toxicity

- Example: The army is interested in the long term effects of exposure to a list of specific chemical agents that were suspected of being toxic. Relatively few soldiers were exposed to these chemicals.
- At first pass, one might think to compare these exposed ("treated") service members to service members who were not exposed at all.
- Complicating that comparison, though, is that the army sorted people into jobs which exposed them or to jobs which did not.
- The army used medical examinations – which were not well documented – to sort some individuals out of high-exposure jobs.

second control group: army toxicity

- Example: The army is interested in the long term effects of exposure to a list of specific chemical agents that were suspected of being toxic. Relatively few soldiers were exposed to these chemicals.
- At first pass, one might think to compare these exposed ("treated") service members to service members who were not exposed at all.
- Complicating that comparison, though, is that the army sorted people into jobs which exposed them or to jobs which did not.
- The army used medical examinations – which were not well documented – to sort some individuals out of high-exposure jobs. This leaves the comparison between exposed and strictly unexposed potentially biased due to baseline conditions.

second control group: army toxicity

second control group: army toxicity

- A second control group was constructed using service members who were in jobs which exposed them to chemical agents…

second control group: army toxicity

- A second control group was constructed using service members who were in jobs which exposed them to chemical agents, but not the specific list of chemical agents under consideration.

second control group: army toxicity

- A second control group was constructed using service members who were in jobs which exposed them to chemical agents, but not the specific list of chemical agents under consideration. These other chemical agents were thought to have little or no longer term effects.

second control group: army toxicity

- A second control group was constructed using service members who were in jobs which exposed them to chemical agents, but not the specific list of chemical agents under consideration. These other chemical agents were thought to have little or no longer term effects. Thus this group is thought to have received an “ineffective dose” of the exposure.

second control group: army toxicity

- A second control group was constructed using service members who were in jobs which exposed them to chemical agents, but not the specific list of chemical agents under consideration. These other chemical agents were thought to have little or no longer term effects. Thus this group is thought to have received an “ineffective dose” of the exposure.
- Each of these control groups is problematic…

second control group: army toxicity

- A second control group was constructed using service members who were in jobs which exposed them to chemical agents, but not the specific list of chemical agents under consideration. These other chemical agents were thought to have little or no longer term effects. Thus this group is thought to have received an “ineffective dose” of the exposure.
- Each of these control groups is problematic: the first group is open to critiques of baseline differences in medical conditions…

second control group: army toxicity

- A second control group was constructed using service members who were in jobs which exposed them to chemical agents, but not the specific list of chemical agents under consideration. These other chemical agents were thought to have little or no longer term effects. Thus this group is thought to have received an "ineffective dose" of the exposure.
- Each of these control groups is problematic: the first group is open to critiques of baseline differences in medical conditions; the second group has individuals who were potentially exposed to actively toxic chemical agents.

second control group: army toxicity

- A second control group was constructed using service members who were in jobs which exposed them to chemical agents, but not the specific list of chemical agents under consideration. These other chemical agents were thought to have little or no longer term effects. Thus this group is thought to have received an “ineffective dose” of the exposure.
- Each of these control groups is problematic: the first group is open to critiques of baseline differences in medical conditions; the second group has individuals who were potentially exposed to actively toxic chemical agents.
- But the first control group is unlikely to suffer from the bias encounter in the second control group, and vice versa.

second control group: army toxicity

- The hope is that the two control groups will not differ from each other in a meaningful way.

second control group: army toxicity

- The hope is that the two control groups will not differ from each other in a meaningful way.
- A rejection of a test of equivalency between the control groups is a strong warning sign of potential bias.

second control group: army toxicity

- The hope is that the two control groups will not differ from each other in a meaningful way.
- A rejection of a test of equivalency between the control groups is a strong warning sign of potential bias.
- A non-rejection may arise for several reasons. A false-negative would be problematic.

second control group: army toxicity

- The hope is that the two control groups will not differ from each other in a meaningful way.
- A rejection of a test of equivalency between the control groups is a strong warning sign of potential bias.
- A non-rejection may arise for several reasons. A false-negative would be problematic.
- The hope is that the control reservoir (i.e., the ratio of controls to treated observations) is large enough that we can reach adequate levels of statistical power for our tests.

second control group: army toxicity

- The hope is that the two control groups will not differ from each other in a meaningful way.
- A rejection of a test of equivalency between the control groups is a strong warning sign of potential bias.
- A non-rejection may arise for several reasons. A false-negative would be problematic.
- The hope is that the control reservoir (i.e., the ratio of controls to treated observations) is large enough that we can reach adequate levels of statistical power for our tests.
- Precise statements of how this argument works statistically, as well as a couple more examples from the literature, can be found in "The Role of a Second Control Group in an Observational Study"

structure of argument

- In an RCT the control and treatment groups are created from a pool of study participants. The assignment to C or T is due to a researcher-directed mechanism (e.g., flipping a coin, or matched pairs). All participants can receive either T or C.
- In an observational study there are possibly many different reasons for people to have not received the treatment.
- In some situations there are discernable subgroups within the non-treatment group, each subgroup being identifiable by the reason for the subgroup not receiving the treatment.
- In some subset of these situations these subgroups will be open to critiques of bias when compared to the treatment group, but at least two of the subgroups will differ in the nature of their bias.
- The contrast of these two control groups with the treatment group may strengthen your analysis.

fin.