Causal Inference Methods and Case Studies

STAT24630

Jingshu Wang

Lecture 16

Topic: Assessing unconfoundedness, sensitivity analysis

- Assessing unconfoundedness
  - Negative control outcome
  - Negative control treatment
  - Robustness to subset unconfoundedness
- Sensitivity analysis
  - Bound under no assumptions
  - Bound for the smoking example
  - Model-based analysis
- Textbook Chapters 21 & 22

Unconfoundedness and balance

- Unconfoundedness property: $W_i \perp \left(Y_i(0), Y_i(1)\right) \mid \boldsymbol{X}_i$
- This is an untestable assumption: we can never test for the unconfoundedness property as it is an assumption on the partially unmeasured potential outcomes

- We assess balancing of covariates and test for $W_i \perp \boldsymbol{X}_i \mid e(\boldsymbol{X}_i)$
- What we really care about is the balance of potential outcomes:

$$W_i \perp \left(Y_i(0), Y_i(1)\right) \mid e(\boldsymbol{X}_i)$$

  within strata of observed covariates, potential outcomes corresponding to both treatment conditions need to be balanced between groups

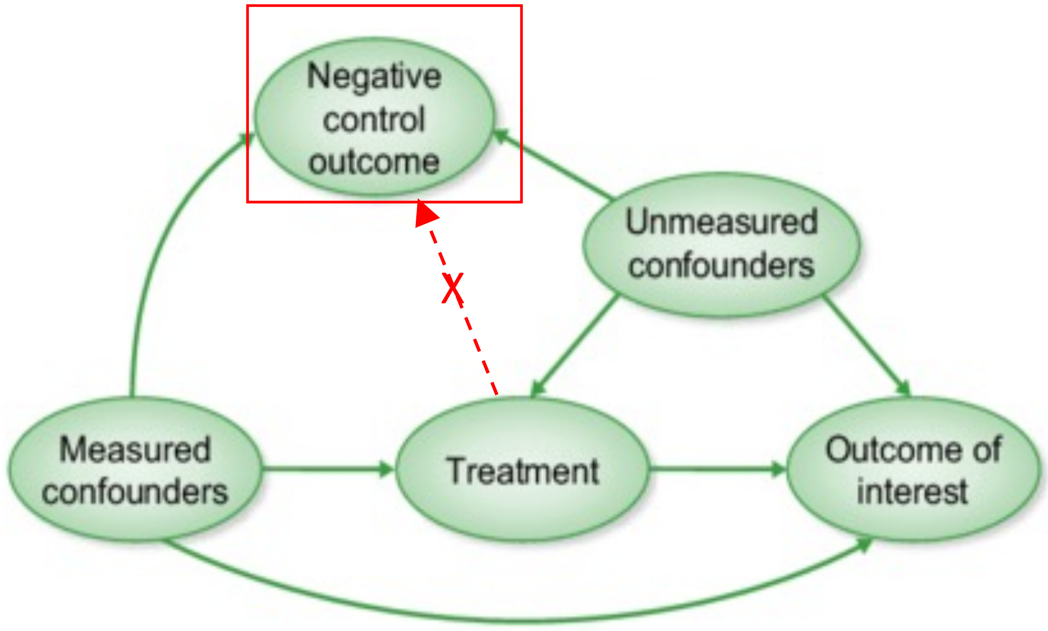
- Covariate balancing is a necessary, but not sufficient condition, especially when there are unmeasured confounding pre-treatment covariates

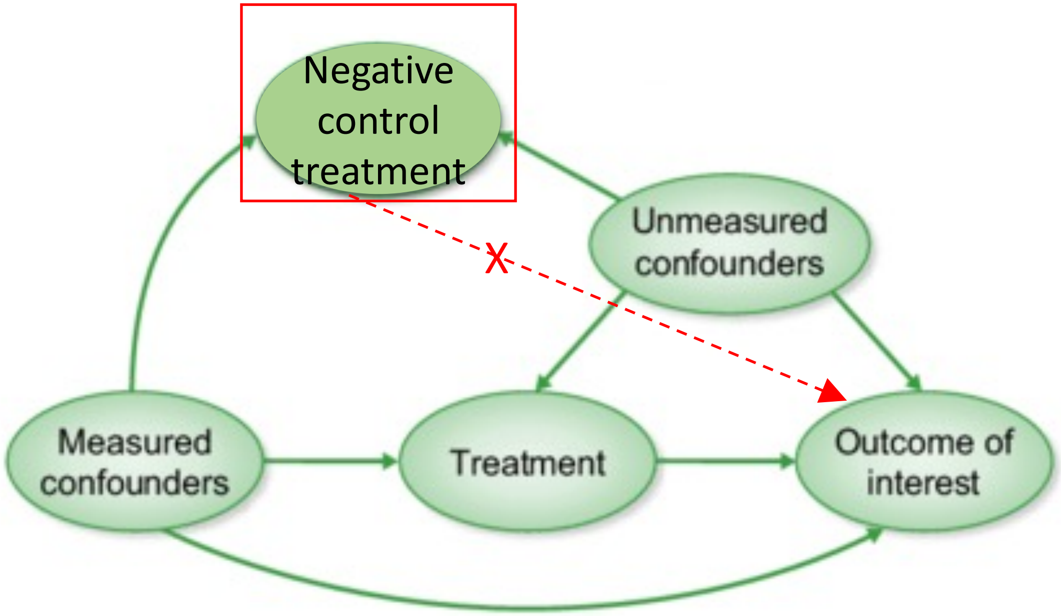
Assessing unconfoundedness

- We can not test for unconfoundedness but we can assess the credibility of the unconfoundedness assumption indirectly

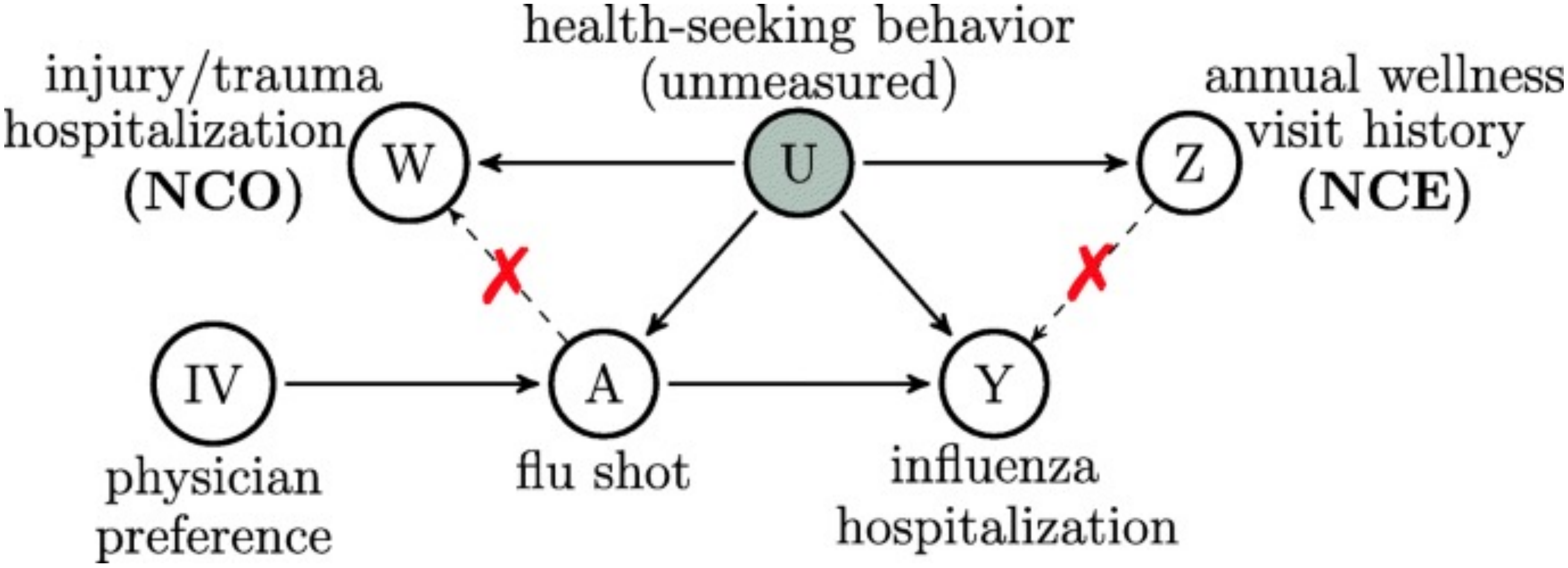
- Three approaches
  - **Negative control outcome**: choose proxy of the real outcome that
    1. Share a similar set of possible unmeasured confounding variables with the real outcome
    2. We know a priori that the treatment have zero causal effect on the proxy
  - **Negative control treatment**: choose new “treatment” that
    1. Share a similar set of possible unmeasured confounding variables with the real treatment
    2. We know a priori that the new “treatment” has zero causal effect on the outcome
  - **Assess robustness of the ATE estimation** given different sets of pre-treatment covariates

Negative control treatments and negative control outcomes





Negative control treatments and negative control outcomes



Shi, X., Miao, W., & Tchetgen, E. T. (2020). A selective review of negative control methods in epidemiology. *Current epidemiology reports*, *7*(4), 190-202.

Negative control outcome (pseudo-outcome)

- One common way to find a good proxy of the outcome is the lagged outcome
  - E.x., outcome is the earning 1 year after treatment, lagged outcome is the earning 1 year before treatment

- The idea: the lagged outcome Y_i^{lag}, can be considered a proxy for $Y_i(0)$ and, given it is observed before the treatment, it is unaffected by the treatment

- By definition, the lagged outcome is also a pre-treatment covariate
  - Define $\boldsymbol{X}_i^r = \boldsymbol{X}_i \backslash Y_i^{lag}$, we test for the independence

$$H_0\text{: } W_i \perp Y_i^{lag} \mid \boldsymbol{X}_i^r$$

- In general, negative control outcome satisfies that $Y_i^{lag}(0) \equiv Y_i^{lag}(1)$, so we always observe its potential outcomes

- If we do not reject H_0, it suggests that the unconfoundedness assumption is plausible.

The Imbens-Rubin-Sacerdote lottery data

Table 21.1. ***Summary Statistics for Selected Lottery Sample for the IRS Lottery Data***

Variable	Label	All ($N = 496$) Mean	All ($N = 496$) (S.D.)	Non-Winners ($N_t = 259$) Mean	Winners ($N_c = 237$) Mean	[t-Stat]	Nor Dif
Year Won	(X_1)	6.23	(1.18)	6.38	6.06	−3.0	−0.27
Tickets Bought	(X_2)	3.33	(2.86)	2.19	4.57	9.9	0.90
Age	(X_3)	50.22	(13.68)	53.21	46.95	−5.2	−0.47
Male	(X_4)	0.63	(0.48)	0.67	0.58	−2.1	−0.19
Years of Schooling	(X_5)	13.73	(2.20)	14.43	12.97	−7.8	−0.70
Working Then	(X_6)	0.78	(0.41)	0.77	0.80	0.9	0.08
Earnings Year -6	(Y_{-6})	13.84	(13.36)	15.56	11.97	−3.0	−0.27
Earnings Year -5	(Y_{-5})	14.12	(13.76)	15.96	12.12	−3.2	−0.28
Earnings Year -4	(Y_{-4})	14.21	(14.06)	16.20	12.04	−3.4	−0.30
Earnings Year -3	(Y_{-3})	14.80	(14.77)	16.62	12.82	−2.9	−0.26
Earnings Year -2	(Y_{-2})	15.62	(15.27)	17.58	13.48	−3.0	−0.27
Earnings Year -1	(Y_{-1})	16.31	(15.70)	18.00	14.47	−2.5	−0.23
Pos Earnings Year -6	($Y_{-6} > 0$)	0.69	(0.46)	0.69	0.70	0.3	0.03
Pos Earnings Year -5	($Y_{-5} > 0$)	0.71	(0.45)	0.68	0.74	1.6	0.14
Pos Earnings Year -4	($Y_{-4} > 0$)	0.71	(0.45)	0.69	0.73	1.1	0.10
Pos Earnings Year -3	($Y_{-3} > 0$)	0.70	(0.46)	0.68	0.73	1.4	0.13
Pos Earnings Year -2	($Y_{-2} > 0$)	0.71	(0.46)	0.68	0.74	1.6	0.15
Pos Earnings Year -1	($Y_{-1} > 0$)	0.71	(0.45)	0.69	0.74	1.2	0.10

The Imbens-Rubin-Sacerdote lottery data

Pseudo-Outcome	Remaining Covariates	Selected Covariates	Est	$(\widehat{\text{s.e.}})$
Y_{-1}	$X_1,\ldots,X_6,Y_{-6},\ldots,Y_{-2},Y_{-6}>0,\ldots,Y_{-2}>0$	X_2,X_5,X_6,Y_{-2}	-0.53	(0.58)
$\frac{Y_{-1}+Y_{-2}}{2}$	$X_1,\ldots,X_6,Y_{-6},\ldots,Y_{-3},Y_{-6}>0,\ldots,Y_{-3}>0$	X_2,X_5,X_6,Y_{-3}	-1.16	(0.71)
$\frac{Y_{-1}+Y_{-2}+Y_{-3}}{3}$	$X_1,\ldots,X_6,Y_{-6},Y_{-5},Y_{-4},Y_{-6}>0,Y_{-5}>0,Y_{-4}>0$	X_2,X_5,X_6,Y_{-4}	-0.39	(0.77)
$\frac{Y_{-1}+\ldots+Y_{-4}}{4}$	$X_1,\ldots,X_6,Y_{-6},Y_{-5},Y_{-6}>0,Y_{-5}>0$	X_2,X_5,X_6,Y_{-5}	-0.56	(0.89)
$\frac{Y_{-1}+\ldots+Y_{-5}}{5}$	$X_1,\ldots,X_6,Y_{-6},Y_{-6}>0$	X_2,X_5,X_6,Y_{-6}	-0.49	(0.87)
$\frac{Y_{-1}+\ldots+Y_{-6}}{6}$	$X_1,\ldots,X_6$	X_2,X_5,X_6	-2.56	(1.55)
Actual outcome Y	$X_1,\ldots,X_6,Y_{-6},\ldots,Y_{-1},Y_{-6}>0,\ldots,Y_{-1}>0$	X_2,X_5,X_6,Y_{-1}	-5.74	(1.14)

← Worse balance as no previous earnings are controlled

Negative control treatment (pseudo-treatment)

- One common case of negative control treatment is when there are multiple control groups
- Suppose we have two control groups and one treatment group $G_i \in \{c_1, c_2, t\}$ [e.g., ineligibles, eligible nonparticipants and participants]

$$W_i = \begin{cases} 0 & \text{if } G_i = c_1, c_2, \\ 1 & \text{if } G_i = t. \end{cases}$$

- We test for

$$G_i \perp\!\!\!\perp Y_i(0) \,\Big|\, X_i, G_i \in \{c_1, c_2\}$$

which is equivalent to

$$G_i \perp\!\!\!\perp Y_i^{\text{obs}} \,\Big|\, X_i, G_i \in \{c_1, c_2\}.$$

Define pseudo-treatment for the lottery data

- One option is to have a comparison control group, of individuals who did not play the lottery at all
- Then we can compare between the “losers” and non-lottery players
- This comparison group is good because “losers” and non-lottery players can be substantially different due to various reasons (so they may share the same unmeasured confounders with that between “losers” and “winners”)

- However, we do not have such data

- Here, we split the winners into two subgroups
    - Median yearly prize for the winners is $31,800
    - We treat the winners with yearly prize less than $30,000 as the other group of control
    - Treat the winners with yearly prize larger than $30,000 as the treated group

Pseudo-treatment analysis for the lottery data

Table 21.3. ***Estimates of Average Treatment Effect on Transformations of Pseudo-Outcome for Subpopulations for the IRS Lottery Data***

Pseudo-Outcome	Subpopulation	Est	$(\widehat{\text{s. e.}})$
$\mathbf{1}_{Y_{-1}=0}$	$Y_{-2}=0$	-0.05	(0.04)
$\mathbf{1}_{Y_{-1}=0}$	$Y_{-2}>0$	-0.04	(0.03)
Y_{-1}	$Y_{-2}=0$	-1.46	(0.92)
Y_{-1}	$Y_{-2}>0$	-0.59	(0.58)
		statistic	p-value
Combined statistic (chi-squared, df 4)		5.51	(0.24)

Assess robustness to subset unconfoundedness

- Partition the pre-treatment covariates into two sets $\boldsymbol{X}_i = (X_i^p, \boldsymbol{X}_i^r)$
- Estimate ATE under the subset unconfoundedness: $W_i \perp \left(Y_i(0), Y_i(1)\right) \mid \boldsymbol{X}_i^r$
- If the estimated ATE differs substantially with the estimated ATE after adjusting for the full $\boldsymbol{X}_i$, then either X_i^p is an important confounder, or the unconfoundedness assumption does not hold
- If, we are somewhat sure that the X_i^p would not be an important confounder given $\boldsymbol{X}_i^r$, then if we see substantially differences, we may doubt the plausibility of unconfoundedness
- Example: $\boldsymbol{X}_i$ contains multiple lagged outcomes

Results on the lottery data

- The data contains lagged outcomes of previous six years: $Y_{i,-1}, \cdots, Y_{i,-6}$ and other time-invariant pre-treatment covariates (denoted as $\boldsymbol{V}_i$)
- Let $X_i^p = Y_{i,-1}$ and $\boldsymbol{X}_i^r = (Y_{i,-2}, \cdots, Y_{i,-6}, \boldsymbol{V}_i)$

- Estimate propensity score using $\boldsymbol{X}_i^r$ to get trimmed sample
- Then we use subclassification with propensity scores estimated using the full $\boldsymbol{X}_i$ and the subset $\boldsymbol{X}_i^r$

- We get

$$\hat{\tau}_{\mathrm{sp}}^{X} = -6.94 \ \ (\widehat{\mathrm{s.e.}} = 1.20), \qquad \hat{\tau}_{\mathrm{sp}}^{X^{\mathrm{r}}} = -5.92 \ \ (\widehat{\mathrm{s.e.}} = 1.16)$$

which are quite similar

Sensitivity analysis

- Most often, validity of unconfoundedness can not be easily checked. Alternatively, one should check sensitivity of a causal analysis to unconfoundedness
- **Sensitivity analysis** aims at assessing the bias of causal effect estimates when the unconfoundedness assumption is assumed to fail in some specific and meaningful ways
- Sensitivity is different from testing – unconfoundedness is intrinsically non-testable, more of a “insurance” check
- Sensitivity analysis in causal inference dates back to the Hill-Fisher debate on causation between smoking and lung cancer, and first formalized in Cornfield (1959, JNCI)

Bounds under no assumptions

- Consider a simple case where: 1. no covariates; 2. binary outcome
- We are interested in the ATE

$$\tau_{\mathrm{sp}} = \mu_{\mathrm{t}} - \mu_{\mathrm{c}},$$

where

$$\mu_{\mathrm{t}} = \mathbb{E}\left[Y_i(1)\right] = p \cdot \mu_{\mathrm{t},1} + (1-p) \cdot \mu_{\mathrm{t},0},$$

and

$$\mu_{\mathrm{c}} = \mathbb{E}\left[Y_i(0)\right] = p \cdot \mu_{\mathrm{c},1} + (1-p) \cdot \mu_{\mathrm{c},0}.$$

$$\mu_{t,1} = \mathbb{E}[Y_i(1)|W_i = 1]$$

$$\mu_{t,0} = \mathbb{E}[Y_i(1)|W_i = 0]$$

$$\mu_{c,1} = \mathbb{E}[Y_i(0)|W_i = 1]$$

$$\mu_{c,0} = \mathbb{E}[Y_i(0)|W_i = 0]$$

$$p = P(W_i = 1)$$

Identifiable from observed data

Bound the unknown $\mu_{t,0}$ and $\mu_{c,1}$ by $[0, 1]$ as the outcome is binary

Bounds under no assumptions

- So we get the bounds

$$\mu_t \in \left[p \cdot \mu_{t,1}, p \cdot \mu_{t,1} + (1-p)\right]$$
$$\mu_c \in \left[(1-p) \cdot \mu_{c,0}, (1-p) \cdot \mu_{c,0} + p\right]$$

- The the bound of ATE $\tau = \tau_{sp} = \mu_t - \mu_c$ is

$$\tau \in \left[p \cdot \mu_{t,1} - (1-p) \cdot \mu_{c,0} - p, p \cdot \mu_{t,1} + (1-p) - (1-p) \cdot \mu_{c,0}\right]$$

- Unfortunately, because we don't have any assumptions at all, this bound is not very informative:
  we can easily show that $\tau^{upper} - \tau^{lower} \equiv 1$, so the bound always covers 0.

Result on the lottery data

- Binary outcome: whether the earning after treatment is positive or not

- Estimated quantities: $\hat{p} = \frac{N_t}{N} = 0.4675$, $\hat{\mu}_{t,1} = \bar{Y}_t^{\text{obs}} = 0.4106$ and $\hat{\mu}_{c,0} = \bar{Y}_c^{\text{obs}} = 0.5349$

- Plug in these quantities into our bound:

$$\tau \in [-0.56, 0.44]$$

- The two-sample difference estimate: $\bar{Y}_t^{\text{obs}} - \bar{Y}_c^{\text{obs}} = -0.134$

Sensitivity analysis bound: a more useful example

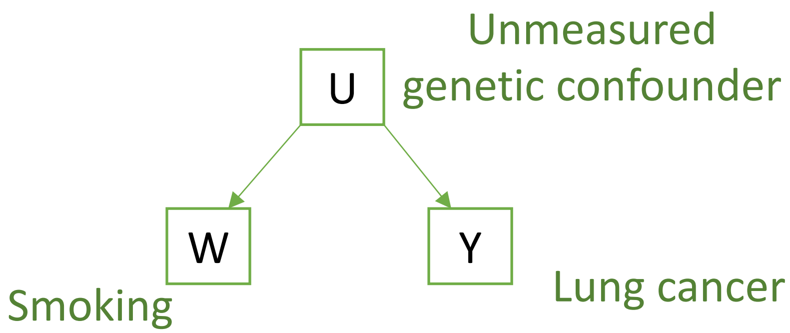
The smoking on lung cancer effect example (Cornfield et al. 1959 JNCI)

- Fisher argued the association between smoking and lung cancer may be due to a common gene that causes both

- Cornfield showed that if Fisher is right, we have $RR_{AU} \geq RR_{AY} \approx 9$
- Such a genetic confounder is too strong to be realistic

- Thus, smoking should have a causal effect on lung cancer

$$RR_{AY} = \frac{P[Y_i = 1 | W_i = 1]}{P[Y_i = 1 | W_i = 0]}$$



$$RR_{AU} = \frac{P[U_i = 1 | W_i = 1]}{P[U_i = 1 | W_i = 0]}$$

Sensitivity analysis bound: a more useful example

- Here, Y_i, U_i and W_i are all binary variables
- Define

$$p_0 = P[U_i = 1|W_i = 0], \qquad p_1 = P[U_i = 1|W_i = 1]$$

- If there is no causal effect of smoking on lung cancer, then

$$P[Y_i = 1|W_i = 0, U_i = 0] = P[Y_i = 1|W_i = 1, U_i = 0] = r_0,$$
$$P[Y_i = 1|W_i = 0, U_i = 1] = P[Y_i = 1|W_i = 1, U_i = 1] = r_1$$

- Then we have

$$RR_{AY} = \frac{P[Y_i = 1|W_i = 1]}{P[Y_i = 1|W_i = 0]} = \frac{r_0(1 - p_1) + r_1 p_1}{r_0(1 - p_0) + r_1 p_0}$$

- Let $p_1 \geq p_0$, then because we observe $RR_{AY} > 1$, then (from some math)

$$RR_{AY} = \frac{r_0(1 - p_1) + r_1 p_1}{r_0(1 - p_0) + r_1 p_0} \leq \frac{p_1}{p_0} = RR_{AU}$$

Another sensitivity analysis idea: base on a model

Idea:

observed ↓ ↙ unobserved

$$W_i \perp \big(Y_i(0), Y_i(1)\big) \mid \boldsymbol{X}_i, U_i$$

- How sensitive is our estimate of causal effect to the presence of U_i?

- A model-based approach (Rosenbaum and Rubin, 1983 JRSS-B)

  Assume that

$$U_i \sim \text{Bernoulli}(\pi)$$
$$\text{logit}(P[W_i = 1 \mid \boldsymbol{X}_i, U_i]) = r + \boldsymbol{X}_i^T \boldsymbol{\kappa} + \alpha U_i$$
$$\text{logit}(P[Y_i(w) = 1 \mid \boldsymbol{X}_i, U_i]) = \beta_w + \boldsymbol{X}_i^T \boldsymbol{b_w} + \delta_w U_i$$

Sensitivity parameters: $(\pi, \alpha, \delta_0, \delta_1)$

Set the sensitivity parameters to different values and see how estimates of causal effects change