

# PHS 2000B Problem Set 4

## Mediation

Due Monday, March 3, 2025

Name:

*We encourage students to help each other out on problem sets. While the final product that you turn in must be your own individual work that you have written up in isolation from your partners, you can still seek the help of your peers if you get stuck on a particular question. We request that you write the names of your collaborators below.*

Collaborators:

## 1 Counterfactual notation and assumptions (34 points)

### 1.1 Mediator as natural value

In a classic paper, Robins and Greenland (1992) define counterfactual direct and indirect effects of mediation using the example of the effect of smoking ( $A$ ) and hyperlipidemia ( $M$ ) on incidence of cardiovascular disease ( $Y$ ). The question is to what extent are well-documented effects of smoking on cardiovascular disease risk mediated through its effects on serum lipid levels? Consider the following population of individuals, and imagine we knew all the potential outcomes. Let  $M^a$  be the counterfactual value of hyperlipidemia (1: yes, 0: no) under smoking (1: yes, 0: no) level  $a$  and let  $Y^{a,m}$  be the counterfactual value of the cardiovascular event (1: yes, 0: no) under smoking level  $a$  and hyperlipidemia level  $m$ .

Type	$M^{a=0}$	$M^{a=1}$	$Y^{a=0,m=0}$	$Y^{a=1,m=0}$	$Y^{a=0,m=1}$	$Y^{a=1,m=1}$
1	1	1	1	0	0	0
2	1	0	0	0	0	0
3	0	0	0	1	0	1
4	1	0	1	1	0	1
5	0	1	1	1	1	0
6	0	0	0	0	0	1

1. Using the counterfactual notation given above write a mathematical expression for each of the following effects and provide a one sentence interpretation of its meaning in the context of the smoking-hyperlipidemia mediation study. For the purpose of this assignment, please provide **all** estimates in this assignment on the **difference** scale. (6 points)

(i) Pure direct effect

(ii) Total indirect effect

2. Write out the assumptions in notation needed to identify the pure direct effect and total indirect effect. Describe these assumptions in words. (4 points)
3. Suppose that in a population there were 100 individuals of each of types 1, 2, and 3 and that there were 200 individuals of each of types 4, 5, and 6. For this population, calculate the average (show all work): (6 points)

- (i) Pure direct effect
- (ii) Total indirect effect

## 1.2 Mediator as a well-defined intervention

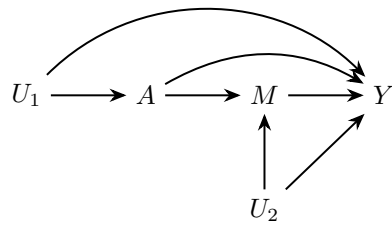
Patients indicated for bariatric surgery often undergo an extensive preoperative program in the months leading up to surgery which includes psychological evaluation, nutritional evaluation, a weight loss plan, and other medical clearances, screenings, and counseling (e.g., smoking cessation). Not all patients who undergo this preoperative program will receive bariatric surgery. There may also be some patients who do not undergo the preoperative program and advance directly to bariatric surgery, and some patients who are indicated for bariatric surgery may never undergo the preoperative program or surgery. Bariatric surgery is an expensive, invasive, and permanent procedure; thus, alternatives to bariatric surgery for this population are of interest. We are interested in the effect of enrollment in the preoperative program on the 10-year risk of a major adverse cardiovascular event (MACE; composite of acute myocardial infarction, stroke, or death) that is not mediated through bariatric surgery.

4. A trial was conducted that randomized patients indicated for bariatric surgery to enroll in a preoperative program or not at time zero. After 6 months, patients were conditionally randomized to undergo bariatric surgery or not (conditional on how enrollment in the preoperative program was assigned). The 10-year risk of MACE was estimated. Assume no loss to follow-up, complete adherence to assigned treatment, and the consistency assumption holds.
  - (i) Draw the DAG that corresponds to the data generating process in this trial (define all symbols used in your graph). (3 points)
  - (ii) Using counterfactual notation, express the controlled direct effect (define all counterfactuals used). (1 point)
  - (iii) What are the steps to identify the controlled direct effect so as to estimate it using trial data? State any assumptions invoked for each step. (3 points)
  - (iv) The controlled direct effect (when the value of the mediator was set to 0), was -0.28. Interpret this estimate using counterfactual language. (2 points)
5. Another trial was conducted that randomized patients indicated for bariatric surgery to enroll in a preoperative program or not at time zero. Some patients underwent bariatric surgery while others elected not to undergo surgery. Patients who lost health insurance coverage during the study were less likely to undergo surgery, and lack of health insurance affects one's risk of MACE. Also, enrollment in the preoperative program increases the risk of losing insurance (e.g., preoperative program leads to less time working, loss of job, loss of insurance). Assume there is an unmeasured common cause of losing insurance and experiencing MACE. The 10-year risk of MACE was estimated. Assume no loss to follow-up, complete adherence to assigned treatment, and the consistency assumption holds.
  - (i) Draw a DAG that corresponds to the data generating process in this trial (define all symbols used in your graph). (3 points)
  - (ii) Based on the DAG, what are the steps to identify the controlled direct effect so as to estimate it using trial data? State any assumptions invoked for each step. (3 points)
  - (iii) What method(s) could you use to validly estimate the controlled direct effect? Can you use outcome regression alone? (1 point)
  - (iv) The controlled direct effect (when the value of the mediator was set to 0), was -0.12. Interpret this estimate using counterfactual language. (2 points)

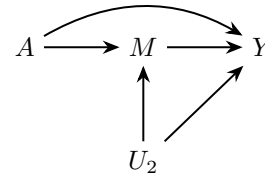
## 2 Using DAGs for causal estimand identification (6 points)

5. Below are possible DAGs of treatment (A), mediator (M), and outcome (Y) from hypothetical studies. Assume the mediator can be posited as a well-defined intervention in all studies. For each DAG, state which of the following effects can be identified (i) the controlled direct effect, (ii) the pure direct effect, and (iii) the total indirect effect. If an effect cannot be identified state the assumption(s) that is violated. Note that  $L$  is used to denote measured variables and  $U$  unmeasured variables (6 points)

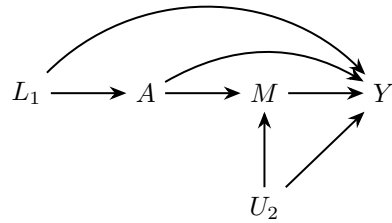
Note: When assessing assumptions, the counterfactual should be different for each effect. For example, for controlled direct effects, the counterfactual should consider  $M$  as directly intervenable (i.e., lower case  $m$ ). The counterfactuals for the pure direct effect and total indirect effect will have their own counterfactuals (i.e., involving  $M^a$ ).



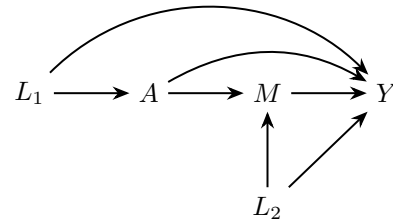
(a)



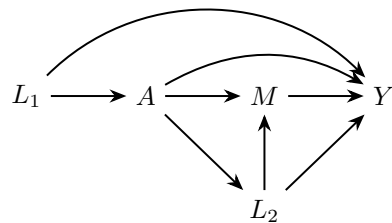
(b)



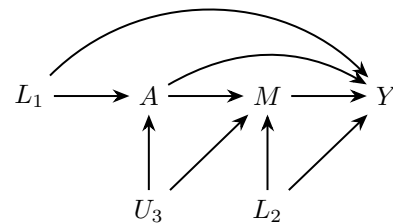
(c)



(d)



(e)



(f)

## 3 Mediation analysis in practice (6 points)

Review the article by Afshar et al. Consider only the main analysis (i.e., ignore the sub-analysis involving only flame injuries).

Afshar M, Netzer G, Mosier MJ, Cooper RS, Adams W, Burnham EL, Kovacs EJ, Durazo-Arvizu R, Klierthermes S. The Contributing Risk of Tobacco Use for ARDS Development in Burn-Injured Adults With Inhalation Injury. *Respir Care*. 2017 Nov;62(11):1456-1465. doi: 10.4187/respcare.05560. Epub 2017 Sep 12. PMID: 28900039; PMCID: PMC6373849.

1. Briefly summarize the methods for mediation analysis applied in this study (2-3 sentences). (2 points)
2. State the causal mediation estimand(s) of interest in the study, and use counterfactual variables to express these mathematically (define all counterfactuals used). (2 points)

3. What assumptions are required to estimate these estimands? How do the authors address each assumption? (2 points)