

EPI 203: Epidemiologic Methods
Problem Set 9: Confounding and Interaction I

NAME: _____

Due: November 14, 2017 at 1:30 pm section
Possible points: 32

For the questions marked “FOR DISCUSSION IN SECTION ONLY:”, please familiarize yourself with these items before section such that we can have a rewarding discussion. Also, please be sure to complete all of the required problem set questions.

Note: *For any DAGs you draw which show unconventional causal relationships (i.e., ones that most persons would question), please make sure to explain your thinking.*

1. PRACTICE QUESTIONS (please attempt to answer practice questions 1a-c, but it is not required that you type your answers into this document for the practice questions). Note: Question 1d is NOT a practice question and MUST be answered.

Recently, there have been several notable causal inferences espoused regarding the following exposures and outcomes/conditions. For each:

- i. describe one factor that you would consider as a “common cause” of a confounding path (i.e., the source of confounding, either causing positive or negative confounding) when planning an observational study regarding the putative causal association between the exposure and the outcome;
- ii. for the common cause you list, state whether this would cause positive or negative confounding and explain why;
- iii. for each scenario, draw a directed acyclic graph (DAG). Feel free to hand draw the DAG, use the drawing tools in MS Word (i.e., text boxes and arrows), or use dagitty.net.

(a) PRACTICE QUESTION

Exposure: cell phone use while driving;
Outcome: motor vehicle accidents

[Practice question; click to see answer](#)

(b) PRACTICE QUESTION

Exposure: enrollment in the TICR Program at UCSF;
Outcome: success in clinical/epidemiologic research

[Practice question; click to see answer](#)

(c) PRACTICE QUESTION

Exposure: scoring more runs in the first inning of a baseball game than one’s opponent;
Outcome: victory in the baseball game

(Subject matter note: Baseball games have 9 innings. In each inning, each team has the opportunity to score runs (points), and after 9 innings, the team with the most runs is the winner of the game.)

[Practice question; click to see answer](#)

(d) Exposure: use of cloth diapers (via a putative mechanism of less absorption that makes child uncomfortable and more apt to become “potty” (toilet) trained);

Outcome: earlier achievement of “potty” (toilet) training among children* (1 pt)

**Advertisements promoting this claim are plentiful (e.g., <http://www.easydoesitdiaperservice.com/why-cloth>)*

2. Advertisements from hearing aid companies are heavily promoting hearing tests and early use of hearing aids on the basis of recent data from observational (cohort) studies that found that older adults with hearing loss were more likely to develop dementia than older adults without hearing loss. This association was found even after controlling for age, sex, race, education, diabetes mellitus, smoking, and hypertension; each of the variables was measured accurately and, where relevant, controlled for incorporating continuous variable depictions of current and prior historical/cumulative levels. Assuming no measurement error in exposure or outcome, selection bias, or sampling error, can you think of another explanation for why this association might be spurious (i.e., biased)? Explain your response with a DAG. (1 pt) (Note: no special subject matter knowledge of hearing loss or dementia is needed to answer this question.)

3. A case-control study was conducted to assess whether drinking coffee (specifically, the ingestion of the components in coffee) is causally related to the occurrence of coronary artery disease (CAD). Coffee consumption was measured as either “High Coffee” use or “Low Coffee” use. Prevalent instances of CAD were used. The results, stratified on smoking status, are summarized below:

Non-smokers
(never smoked)

	CAD	No CAD
High coffee	40	90
Low coffee	65	150

Smokers
(past/current)

	CAD	No CAD
High coffee	150	65
Low coffee	45	20

(adapted from Kleinbaum 2003)

- (a) Calculate the crude odds ratio (for the association between coffee drinking and CAD) and the two stratum-specific odds ratios. Show your calculations. (2 pts)
- (b) Assess the following aims:
- whether smoking is related to coffee consumption among those without CAD
 - whether smoking is related to CAD among subjects with low coffee consumption
 - whether smoking is related to CAD among subjects with high coffee consumption

Calculate the measures of associations for each aim above. Show your calculations. (3 pts)

- (c) If you knew nothing else about this clinical/biological system (i.e., had no other data or information available to you), are the data given in the question and your calculations in item (b) sufficient evidence for you to conclude that smoking is a confounder in the relationship between coffee use and CAD? Explain your answer. (1 pt)
- (d) Drawing upon all of your own prior causal knowledge about this system (i.e., including external knowledge), show the DAG that you believe best depicts the system under study. Justify your selection. (1 pt)

4. Consider the following abstract:

RISK FACTORS FOR CERVICAL INTRAEPITHELIAL NEOPLASIA
IN SOUTHWESTERN AMERICAN INDIAN WOMEN

The authors assessed risk factors for cervical intraepithelial neoplasia (CIN) among southwestern American Indian women using case-control methods. Cases were New Mexico American Indian women with biopsy-proven CIN grade I (n = 190), CIN grade II (n = 70), or CIN grade III (n = 42) cervical lesions diagnosed between November 2004 and October 2007. Controls were American Indian women from the same Indian Health Service clinics with normal cervical epithelium (n = 326). All subjects underwent interviews and laboratory evaluations. Interviews focused on history of sexually transmitted diseases and sexual behavior (i.e., number of sexual partners). Laboratory assays included polymerase chain reaction-based tests for cervical human papillomavirus infection, tests for gonorrhea and chlamydia, wet mounts, and serologic assays for antibodies to *Treponema pallidum*, herpes simplex virus, and hepatitis B and C viruses. In multiple logistic regression analysis, the strongest correlate for CIN among American Indian women was any human papillomavirus infection (OR = 5.8; 95% CI: 3.3, 10.0). Unlike previous research, this study found no strong associations between CIN and sexual behavior (number of partners) after adjustment for all the above other factors associated with CIN.

(Substantive note: CIN is a precursor to cervical cancer. There are 3 grades of CIN (grade I, II, and III). Human papillomavirus is a causal determinant of CIN.)

The authors felt it was notable that unlike previous research, this study found no strong association between CIN and sexual behavior (as measured by number of sexual partners). Assuming this was not because of chance, selection bias, or measurement bias, can you explain why no association was found? Draw a DAG to support your explanation. (2 pts)

5. Observational studies are often performed to evaluate already-licensed drugs for new indications (i.e., new clinical uses). The term “confounding by indication” refers to the specific type of confounding that may occur in such observational studies because those patients who are prescribed drugs are generally different from those who are not prescribed the drug. Specifically, the reasons why certain patients with a given disease/condition are prescribed drugs, and others are not, may be causally related to the outcome of interest. Consider an observational study to evaluate a particular antibiotic for its ability to prevent post-operative infections following coronary artery bypass surgery. (Substantive note: coronary artery bypass surgery is performed for persons with coronary artery disease due to atherosclerosis, i.e., some degree of blockage of the arteries of the heart.)
- (a) Name one factor that you would want to measure as a confounding factor that might result in an underestimation of the effect of a truly effective antibiotic in preventing infection if you did not adjust for it. Explain your answer and draw a DAG. (1 pt)
- (b) Name another factor that you would want to measure as a confounding factor that might result in an overestimation of the effect of a truly ineffective antibiotic in preventing infection if you did not adjust for it. Explain your answer and draw a DAG. (1 pt)

6. You have been awarded a lucrative grant from a large food manufacturing corporation to study whether the consumption of irradiated food is causally related to the development of a variety of malignancies (cancers) via a putative mechanism of the irradiated food causing mutation of human genes. This is to address public concern about the safety of food irradiation. The design will be observational (case-control), and you decide to conduct the study within the Kaiser Health Care network (to take advantage of the various measurements that are made during the course of clinical care and the fact that Kaiser rigorously performs cancer screening to detect cancers early). You plan on including a measure of body mass index (BMI is available in the Kaiser medical records system at the start of the observation period, prior (i.e., assume that BMI is available at least many years prior to a cancer diagnosis) to the onset of any cancer for any patient) in your analysis, but the company's chief executive officer has insisted you *not* include BMI because he is concerned that you might find that eating his product is associated with obesity. How would you respond to the executive officer's request not to include BMI in your analysis? Draw a DAG to support your response. (2 pts)

7. Consider the following data:

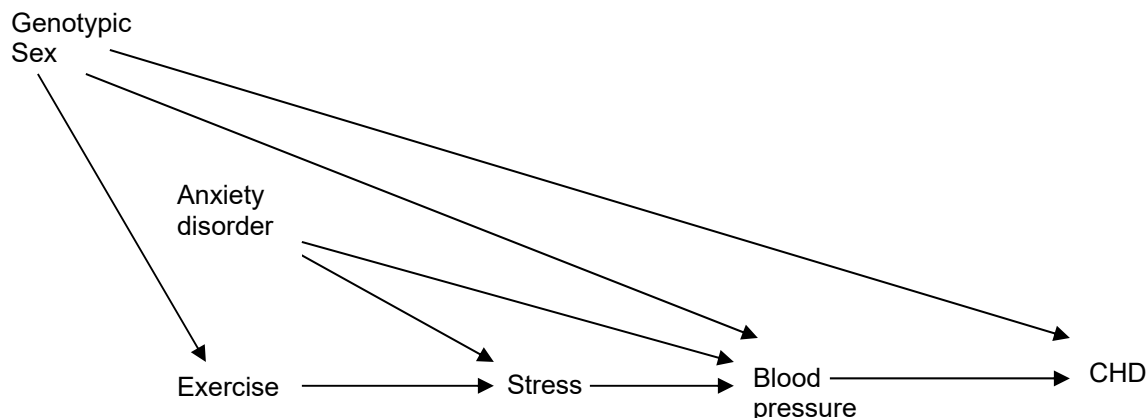
The Pima-Papago group of Native Americans has long been known to have a higher prevalence (compared to Caucasians) of type 2 (non-insulin-dependent) diabetes mellitus. In a sample of 4,920 individuals self-reporting as Pima-Papago but who also have varying extents of Caucasian racial background, there is a very strong negative (inverse) association between a particular genetic haplotype for immunoglobulin (Gm3;5,13,14) and the occurrence of type 2 diabetes (crude prevalence ratio = 0.27, 95% confidence interval 0.18-0.40).

One might conclude from this observation that the presence of this haplotype protects against the occurrence of type 2 diabetes mellitus (DM). Investigations using other populations which were more homogeneous in terms of racial background could not replicate this finding regarding the protective effect of the haplotype. Can you think of a way that confounding could explain the association between the Gm haplotype and prevalence of type 2 DM in the study described? Draw a DAG to support your response. (2 pts)

8. A cohort study has been conducted to assess whether regular jogging (running) prevents sudden cardiac death in middle-aged men. The goal is to see if jogging should be recommended to middle-aged men as a means to prevent sudden cardiac death. The data analyst notes that 30% of joggers had pulse rates below 60 beats/minute (which was defined as a slow pulse rate) at their study enrollment exams, versus only 10% of sedentary (non-jogging) subjects. A slow pulse rate at enrollment is also found to be a protective factor for sudden cardiac death (rate ratio = 0.6) among both joggers and non-joggers. Should the investigator stratify upon pulse rate at enrollment when evaluating the causal total effect of jogging on the risk of sudden cardiac death? Why or why not? Draw one or more DAGs to explain your answer. (2 pts)

9. There is growing interest in whether inflammation is important in the pathogenesis of coronary artery disease (CAD) and whether the presence of certain inflammatory diseases outside of the heart can influence the risk of development of CAD. Specifically, there is interest in the role of periodontal disease (inflammation of the gums and other soft tissue in the mouth). To address this, a cohort study of persons free of CAD was performed to evaluate the causal relationship between periodontal disease and the incident occurrence of CAD. The unadjusted (crude) analysis found that those persons with periodontal disease at baseline have a higher rate of CAD than persons without periodontal disease. Assessment of plasma level of C-reactive protein (CRP) was also performed at baseline. [NOTE: CRP level is known to be elevated in many inflammatory conditions and has also been associated with the development of coronary artery disease. CRP is best thought of as a biological marker for the unmeasured construct of inflammation.] In this cohort study, after adjustment for CRP level, the association between periodontal disease and coronary artery disease was no longer present. Using a DAG, sketch out two scenarios that would explain why adjustment for CRP resulted in no longer seeing an effect for periodontal disease. In which scenario would you predict that treating periodontal disease would reduce the incidence of CAD? (2 pts)

10. Consider the following causal DAG (i.e., assume that all edges are accurate and all common causes between any two nodes are accounted for) and use it to answer the following questions (external knowledge is not necessary). In each question, do not consider adjustment for purposes of decreasing variance.



- (a) Which factor(s) would you control for if you want to know the total causal effect of exercise on blood pressure? (1 pt)
- (b) If you want to know the causal association between stress and blood pressure would you consider genotypic sex to be a confounder? Explain. (1 pt)
- (c) If you wanted to know the causal association between stress and blood pressure, are there any other options to prevent confounding by genotypic sex other than conditioning on genotypic sex (e.g., by stratification)? Explain briefly. (1 pt)
- (d) What would you control for if you want to estimate the total causal effect of genotypic sex on CHD (coronary heart disease)? (1 pt)
- (e) After submitting a manuscript in which you examine the causal association between genotypic sex and blood pressure, a reviewer suggests that CHD may be a confounder of the association because prior studies have shown that it is associated with both genotypic sex and blood pressure. The reviewer suggests controlling for this factor in the analysis. Do you agree? Explain. (1 pt)

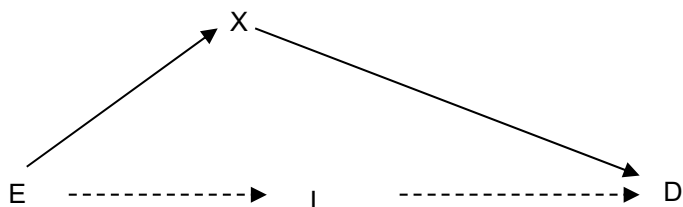
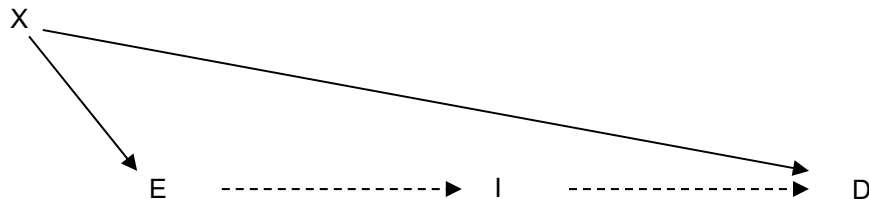
11. Data from a very large population-based cohort study of dietary habits were analyzed for the association between potassium dietary intake and the occurrence of stomach cancer. Outcomes in this study were diagnosed during the course of routine clinical care by personal private practice physicians (and not through serial research-level screening). After adjustment for age (measured as a continuous variable), sex, and weight change during the study (also measured as a continuous variable), the authors found an association between potassium intake and stomach cancer. This prompted a large randomized clinical trial of a low potassium diet versus a normal potassium diet, which found no difference in stomach cancer incidence between diets.

Assume that in both studies measurement characteristics of exposed and non-exposed (i.e., potassium intake) was the same, all of the outcome measurements were reproducible and fully specific and that the large sample sizes precluded sampling error (and ensured perfect randomization in the clinical trial). National cancer registries were examined in both studies to exclude the possibility of losses to follow-up. Assume also that in the trial those assigned to the low potassium diet did adhere to it (and there was perfect adherence to the normal potassium diet). Finally, both studies were done among urban adults in the East Coast of the United States (with no differences in the distributions of age, sex, race, socioeconomic status or several metrics of self-reported health between studies). Other than unmeasured confounding (which is always a possibility in observational work) or measurement error/bias, describe how the discrepancy in the inference between the observational study and the clinical trial may have occurred? (2 pts)

12. Consider a study investigating whether neighborhood violence is causally related to the occurrence of incident heart attacks. You believe that:
- a person's race/ethnicity and his/her income causally contribute to his/her neighborhood's violence;
 - income and physical activity are causally related to the occurrence of heart attack;
 - income casually influences physical activity;
 - neighborhood violence causally influences physical activity; and
 - race/ethnicity causally influences income.
- (a) Using dagitty.net, draw the DAG that depicts this system. Export your DAG as a JPG file and insert it into your homework or print it as a PDF (and attach it to your homework). (2 pts)
- (b) Save your DAG on dagitty.net and show the URL where it is saved (i.e., provide the URL that dagitty.net gave you)? (1 pt)
- (c) What is the minimal factor or set of factors you need to control for in order to estimate the total causal effect of neighborhood violence on incident heart attacks? (1 pt)

13. FOR DISCUSSION IN SECTION ONLY:

Two investigators want to study the effect of E on D as mediated thru I. They had a disagreement about which of the two below DAGs was operative.



Ultimately, they decided it did not matter which was operative because they were going to stratify for X in either case. Comment on their decision.

ANSWERS TO PRACTICE QUESTIONS 1a-c:

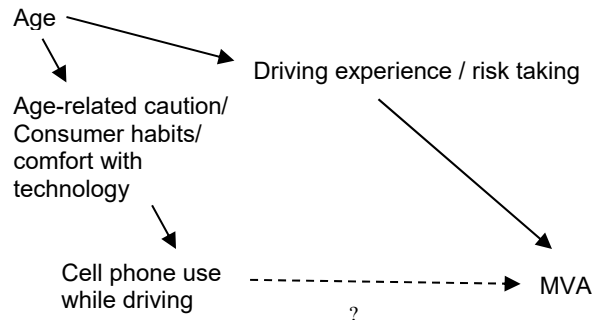
1.

(a) Answer (examples):

Note: This is not a complete list of correct answers, other answers may be acceptable

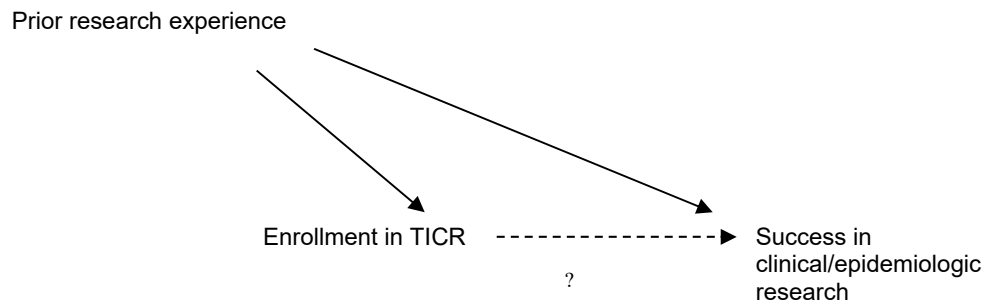
- Age: younger drivers are at higher risk for MVAs (perhaps because of more reckless driving) and are more apt to be using cell phones while driving. (positive confounding)
- Socioeconomic status: socioeconomic status may be causally related to cell phone use while driving (either in a positive or negative direction depending upon the population) and also ownership of cars with more safety features (e.g., lane change safety monitoring); the presence of this confounding may serve to accentuate or attenuate a possible association (again, depending upon the population). (positive or negative confounding)
- Children in the car: drivers with children in the car are more apt to be cell phone users because of the need to schedule and coordinate the children's activities; children in the car may themselves be distracting and be causally related to occurrence of an MVA. (positive confounding)

Note: Being distracted while on the cell phone would not be a correct response because this is an intermediary factor (mediator) on the causal pathway under investigation between cell phone use and motor vehicle accidents.



(b) Answer (examples):

- Prior experience in research: Prior experience in research causes individuals to see what they are lacking and causes them to seek additional training. Prior experience in research is also causally related to making correct decisions about research projects and collaborators and ultimately is causally related to ultimate success in research. (positive confounding)
- Hard-working personality type: Persons with inherent hard-working personality traits are driven to obtain all of the training they can get, including training in clinical research. Hard-working personality traits are also causally related to success in research. (positive confounding)
- Intelligence: Native intelligence leads people to know that formal training in research methods is important and hence intelligence drives people to seek more training. Native intelligence is also causally related to high quality work in all career endeavors, including clinical/epidemiologic research. (positive confounding)



(c) Answer (example):

- Overall effectiveness (e.g., higher batting averages and more runs batted in) of the hitters on the baseball team
- Teams with overall more effective hitters will, on average, score more runs in the first inning of a baseball game than their opponents and also each subsequent inning. Scoring more runs overall is causally related to winning the entire game (positive confounding)

Note: This question makes us recall the adage in many sports that scoring first in a game has special causal significance for winning. The implication of this adage for many sportsmen is that the team should do all they can to score first because this will cause victory to occur. However, it is unlikely that this adage is causal. Hence, it is likely not useful to be pulling out all of the stops to score first because this will more often not bring victory. It is more likely that the overall effectiveness of the hitters will dictate the number of runs scored.

