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

Due: 11/17/15 at 1:30 pm section
Possible points: 34

*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. Recently, there have been several notable causal hypotheses generated regarding the following exposures and outcomes/conditions. For each:
   i. describe one factor that you would consider as a confounding factor (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 confounding factor 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, or use dagitty.net.

(a) Exposure: cell phone use while driving;
Outcome: motor vehicle accidents (1 pt)

(b) Exposure: enrollment in the TICR Program at UCSF;
Outcome: success in clinical/epidemiologic research (1 pt)

(c) Exposure: scoring more runs in the first inning of a baseball game than one's opponent;
Outcome: victory in the baseball game (1 pt)

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

(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)

2. 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 whether:
i. smoking is related to coffee consumption among those without CAD.
ii. smoking is related to CAD among subjects with low coffee consumption
iii. smoking is related to CAD among subjects with high coffee consumption

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)

3. 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 grade I (n = 190), grade II (n = 70), or grade III (n = 42) cervical lesions diagnosed between November 1994 and October 1997. 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 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 risk factor 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 the above other factors associated with CIN.

(Substantive note: CIN is a precursor to cervical cancer. Human papillomavirus is a causal agent 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)

4. Observational studies are often performed to evaluate already-licensed drugs for new indications (i.e., new clinical uses). 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 an 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)

5. 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 due to the rarity of most cancers). The company's chief executive officer has insisted you *not* measure body mass index (BMI) as part of your study as he is concerned that you might find that eating his product is associated with obesity during the course of your analysis. How would you respond to the executive officer's request not to measure BMI? Draw a DAG to support your response. (2 pts)

6. Consider the following data:

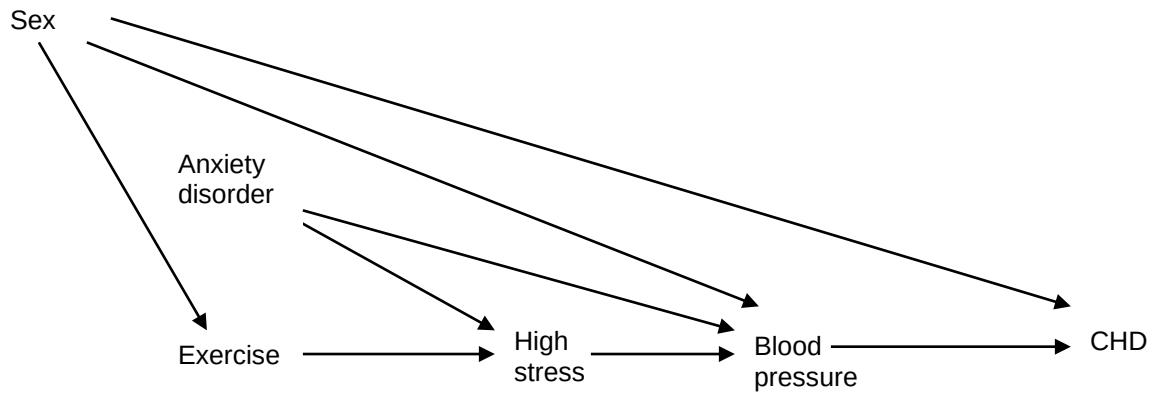
The Pima-Papago group of Native Americans has long been known to have a high prevalence of type 2 (non-insulin-dependent) diabetes mellitus. In a sample of 4,920 individuals self-reporting as Pima-Papago but who also have varying extents of Caucasian racial background, there is a very strong negative (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. 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)

7. 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 in evaluating the causal 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)

8. 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. To address this, a cohort study was performed to evaluate the causal association between periodontal disease (inflammation of the gums and other soft tissue in the mouth) 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)

9. Consider the following DAG to answer the following questions (external knowledge is not necessary):



(a) Which factor(s) would you control for if you want to know the causal effect of exercise on blood pressure? (1 pt)

(b) If you want to know the causal association between high stress and blood pressure would you consider sex to be a confounder? Explain. (1 pt)

(c) If you wanted to know the causal association between high stress and blood pressure, are there any other options to prevent confounding by sex other than conditioning on 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 sex on CHD (coronary heart disease), irrespective of mechanism? (1 pt)

(e) After submitting a manuscript in which you examine the causal association between 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 sex and blood pressure. The reviewer suggests controlling for this factor in the analysis. Do you agree? Explain. (1 pt)

10. 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 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 overall differences in age or sex distributions between studies). Other than unmeasured confounding (which is always a possibility in observational work), describe how the discrepancy in the inference between the observational study and the clinical trial may have occurred? (2 pts)

11. 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.

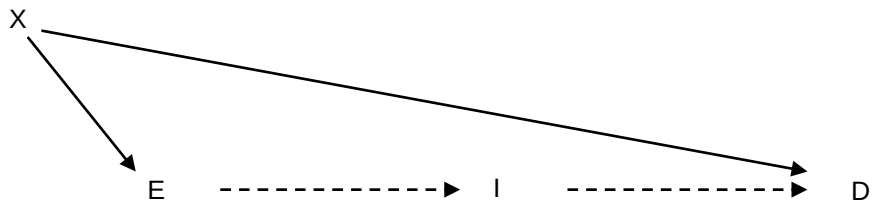
(a) Using dagitty.net, draw the DAG that depicts this system. Export your DAG as a JPG file and insert it into your homework or print it as a PDF (and attach it to your homework). (2 pts)

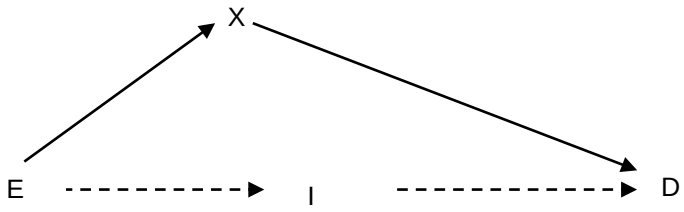
(b) Save your DAG on dagitty.net and show the URL where it is saved (the URL that dagitty.net that gave you)? (1 pt)

(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)

12. **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.