

Problem Set 11: Confounding and Interaction III**Due: December 8, 2015 at 1:30 pm section****Possible points: 33 (plus 2 extra credit)**

*Note: 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.*

For questions 1-12, use the data from a case-control study of incident esophageal cancer conducted between 1975 and 1980 among adults in 4 urban areas in the U.S. The study was population-based and used incidence density sampling to select controls. In the dataset “esophageal_cancer.dta”, found on the course website, there are 1,956 observations (participants) and 4 variables, which are defined as follows:

- cancer, where 0 is no esophageal cancer and 1 is cancer.
- age 44, where 0 is $\leq$ age 44 (i.e., 21-44 years) and 1 is $>$ age 44. This refers to age at cancer diagnosis (cases) or at time of control selection (controls).
- alcohol, where 0 is $\leq$ 39 grams per day and 1 is $>$ 39 grams per day. This refers to average daily use in the 2 years prior to diagnosis (or prior to control selection). This was derived from a complex survey estimating number, volume, and alcohol content of alcoholic beverages consumed in representative time periods. [Note: a standard drink contains 10 grams of alcohol. Standard drink sizes of beer, wine, and spirits can be found at: http://www.nhmrc.gov.au/files_nhmrc/file/your_health/healthy/alcohol/std-drinks-large.jpg]
- tobacco, where 0 is $\leq$ 19 grams per day and 1 is $>$ 19 grams per day. This refers to average daily use in the 2 years prior to diagnosis (or prior to control selection). This was derived from a complex survey estimating number and type of cigars, cigarettes, and pipes smoked in representative time periods. One cigarette contains up to 1 gram of tobacco.

The objective of your analysis will be to determine whether alcohol use causes esophageal cancer. You will need to contend with two other factors, age and tobacco use, to attempt to understand the causal direct relationship between alcohol use and esophageal cancer (i.e., independent of age and tobacco use). [Clinical note: Tobacco is known to be a causal determinant of a variety of cancers in the head, neck, and thorax, including the esophagus. For review, see: IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. Volume 83: Tobacco Smoke and Involuntary Smoking.]

1. Draw a DAG depicting how you conceive the relationships between alcohol use, tobacco use, age, and esophageal cancer. Include, if you see fit, other nodes that may be relevant to the system as common causes of alcohol use, tobacco use, and/or age. List the minimal sufficient adjustment set or sets for this DAG. Note: You don't have to evaluate any external data or literature. Just use your personal knowledge and what has been provided to you in this document. (2 pts)

2. Briefly describe your analysis plan. Specifically, at a minimum, describe:
- what analysis steps you will perform to fulfill the stated objective
 - what type of interaction you will look for (additive and/or multiplicative)
 - what parameters/rules/logic you will use to decide whether to report or ignore interaction
 - how you will decide which variables to include in adjusted summary estimates. (4 pts)

3. Show a 2 x 2 table describing the crude association between alcohol use and esophageal cancer. State the measure of association and the 95% confidence interval. In a clear, informative, and easily comprehensible sentence, describe the crude association (in terms of the point estimate only) between alcohol use and esophageal cancer. (1 pt)

4. Show a 2 x 2 table describing the association between age and esophageal cancer among subjects who consume less than or equal to 39 grams per day of alcohol. State the measure of association and describe the association (in terms of the point estimate only) in a clear, informative, and easily comprehensible sentence. Repeat this by showing the association in those who use more than 39 grams per day of alcohol. (2 pts)

5.

- (a) Show a 2 x 2 table describing the association between age and alcohol use among subjects without esophageal cancer. State the measure of association and describe it in a clear, informative, and easily comprehensible sentence. (1 pt)
- (b) What do you predict about whether the point estimate for the association between alcohol use and esophageal cancer, adjusted for age, will be substantially different than the crude estimate for the association between alcohol use and esophageal cancer? (1 pt)

6.

- (a) For the moment, disregard the presence of tobacco use. Determine whether age is an effect modifier on the multiplicative scale when evaluating the odds ratio for the association between alcohol use and esophageal cancer. Will you report that multiplicative statistical interaction is present? (1 pt)
- (b) Assuming for a moment that no multiplicative statistical interaction is present, what is the association between alcohol use and esophageal cancer adjusted for age? (1 pt)
- (c) Finally, forgetting about the prior assumptions in part (b), if age were the only additional factor you had to contend with, how would you report the relationship between alcohol use and esophageal cancer -- do you prefer to report the crude odds ratio, the two stratum-specific odds ratios, or the adjusted odds ratio? Justify your preference. (1 pt)

7. If tobacco use were the only additional factor you had to contend with (i.e., disregarding age), how would you report the relationship between alcohol use and esophageal cancer? Do you prefer to report the crude odds ratio, the two stratum-specific odds ratios, or the adjusted odds ratio? Consider only the multiplicative scale for interaction. Justify your preference. (2 pts)

8. Examine the association between alcohol use and esophageal cancer while jointly stratifying by age and tobacco use. Show how you did this. (Hint: use the “`mh odds`” command in Stata and show your output). Will you report or ignore multiplicative statistical interaction? (2 pts)

9. Contending with both age and tobacco use, what would you conclude regarding the effect of alcohol use on the development of esophageal cancer? In other words, what is your overall answer: which numerical measure(s) of association would you publish as your final estimate(s) of the effect of alcohol on esophageal cancer? Create a table with your results that you would include in your publication. Use your DAG, analysis plan and perhaps the findings from the earlier questions to guide you. (4 pts)

10. Assess for the presence of mechanistic interaction between alcohol use and tobacco use. If present, state if the interaction is positive or negative. (1 pt)

11. Consider the Nitz et al. article that was discussed in Journal Club (Association of Prostaglandin E Synthase ...J Clin Endo Met 2007).
- (a) Draw the DAG that you envision the authors were considering as they planned and implemented their study in order to address their primary research question. (1 pt)
 - (b) In the authors view, what is the chief concern regarding confounding in this study? In other words, what confounder did the authors spend the most text space in the article trying to contend with? (1 pt) How did the authors manage this confounder? (1 pt extra credit)

12. FOR DISCUSSION IN SECTION ONLY:

Rheumatoid arthritis (RA) is an autoimmune condition with several clinical manifestations outside of the joints, one of which is myocardial infarction (MI or “heart attack”). It is generally recognized that patients with RA who have the most severe inflammation, as evidenced by blood levels of inflammatory markers (e.g., the erythrocyte sedimentation rate (ESR) and C reactive protein (CRP)), are at greatest risk for myocardial infarction. There are a growing number of drugs available to treat RA including older drugs such as methotrexate and several newer biologic agents (e.g., anti-tumor necrosis factors). Whether or not a physician will prescribe one of the older drugs vs. a newer biologic agent depends upon a number of factors, including the severity of disease as evidenced by blood markers of inflammation, such as the ESR and CRP levels. (Patients with higher and more persistent levels of these inflammatory markers are more often treated with the biologic agents). In patients with RA, there is interest in whether use of the newer biologic agents is superior to the older agents in lowering the incidence of myocardial infarction. In other words, there is desire to know the total causal effect of the biologic agents (compared to non-biologic agents) in the occurrence of myocardial infarction. To address this, a group of investigators used the Kaiser clinic-based electronic medical record system to study a group of patients with RA. Kaiser is essentially a dynamic cohort with excellent measurement of prescribed drugs, medication adherence, inflammatory markers in the blood (ESR and CRP), myocardial infarction, and other known causal determinants of heart attack (e.g., blood pressure, hyperlipidemia, and smoking). The investigators controlled for ESR and CRP as it varied over time within and between patients with a proportional hazards mathematical regression model with time-varying covariates. Their findings showed no strong evidence to suggest the newer biologic agents were superior to the older agents in the prevention of myocardial infarction among patients with RA. Assuming no selection or measurement bias and no sampling error, can you see any problems in the analysis which was conducted?

13. Using DAGitty.net, create a DAG depicting a research question you are currently studying. Be sure to make it a causal DAG, meaning that all common causes of any two variables should be shown. If you are not working on a project that fits the spirit of what DAGitty.net does, then create a DAG based upon a research question in your field that you would like to investigate. Paste your DAG into your answer document or print out a separate copy and attach it. (Note: please also bring the DAG on a memory stick to your small group section, or record the URL where it is stored, such that a few DAGs can be projected and discussed during class). If you desire, please feel free to provide any narrative explanation that you feel would assist readers. Also, state the study design you will use to address your research question, and be sure to account for this design in the DAG, if warranted. (5 pts)

14. Refer to the article that will be discussed in the upcoming Journal Club to answer the following questions (Madsen et al. *NEJM* 347:1477-1482, 2002.).

Note that the authors clarify in their Methods Section that they are using “relative risk” to refer to “incidence rate ratios.” To avoid confusion in referring to the article, we will also use the term relative risk.

- (a) The authors report an adjusted relative risk of 0.92 (0.68-1.24) for vaccination on the occurrence of autism. They do not provide a crude relative risk, but they do give enough information to calculate one. Calculate the crude relative risk and show your calculation/output. (2 pts)
- (b) In the Results, the authors report that some children were censored because of emigration. Discuss whether you believe this will introduce an important degree of selection bias when estimating the association between vaccination and autism. (1 pt)
- (c) Consider possible misclassification of the outcome. Discuss the sensitivity and specificity of the measurement of the outcome and how, if at all, it affects the relative risk for vaccination on the occurrence of autism? (1 pt extra credit)