Homework 3, Due Friday September 23, 2016

1. A 20-year cohort study of British male physicians noted that the proportion per year who died from lung cancer was 0.00140 for cigarette smokers and 0.00010 for nonsmokers. The proportion who died from coronary heart disease was 0.00669 for smokers and 0.00413 for nonsmokers.

   (a) Describe the association of smoking with each of lung cancer and heart disease, using the ER, NNH, RR, and OR. Interpret each of these.

   (b) Which response (lung cancer or coronary heart disease deaths) is more strongly related to cigarette smoking in terms of the reduction in number of deaths that would occur with elimination of cigarettes? Explain. (Do not try to get AR for this one. Just answer it using the information in (a).)

   (c) Why are RR and OR so close to one another in both responses?

2. Problem 4.5 on page 42.

3. A study looked at the relationship between the presence of squamous cell carcinoma (yes, no) and smoking behavior. There were 3 categories of smoking behavior: nonsmoker, light smoker ($\leq$ 20 cigarettes per day) and heavy smoker ($>$ 20 cigarette per day). The estimated OR for females between the presence of carcinoma and smoking behavior (nonsmoker, light smoker) was 11.7. The corresponding OR for females between the presence of carcinoma and smoking behavior (nonsmoker, heavy smoker) was 26.1. Find the estimated OR between carcinoma and smoking behavior (light smoker, heavy smoker). You must justify any calculations.

4. A study exploring the relationship between exposure to a risk factor and a disease used a simple random sample of $100,000$ exposed individuals and a simple random sample of $200,000$ unexposed individuals. Disease status was subsequently determined for the individuals in the samples. The results are shown below.

	D	$\overline{D}$	
E	250	99,750	100,000
$\overline{E}$	125	199,875	200,000
	375	299,625	300,000

   They estimate $P(D) = 375/300,000 = 0.00125$ and $P(D|\overline{E}) = 125/200,000$ and then estimate attributable risk as

$$AR = \frac{0.00125 - 0.000625}{0.00125} = 0.5$$

   which they interpret by writing "The disease could be reduced by 50% if we could eliminate the risk factor."

   Give two major flaws with this interpretation for this particular problem. There is one you MUST give to get credit.

5. All of the problems at the end of Chapter 5 BUT you only need to identify the type of study: Population-based, Cohort, or Case-Control.

6. The data are given in Problem 6.5 on page 72.

   (a) Carry out a test of independence. Show the tables of expected counts and the standardized Pearson residuals. What do the results indicate about the association between heart disease and hair loss? Don't just say there is a lack of independence - I want some idea of where the independence is breaking down.

   (b) Estimate ER, RR, and OR. Construct approximate 95% confidence intervals in each case and interpret them.