

Statistical Reasoning in Public Health II, 2009

1. In a 2005 New England Journal of Medicine Article¹ researchers reported the results of a randomized clinical trial to investigate the efficacy of low dose aspirin therapy in potentially reducing cardiovascular disease in women age 45 and older. As per the researchers’:

“We randomly assigned 39,876 initially healthy women 45 years of age or older to receive 100 mg of aspirin on alternate days (ie: 100 mg of aspirin every other day) or placebo and then monitored them for 10 years for a first major cardiovascular event (i.e., nonfatal myocardial infarction, nonfatal stroke, or death from cardiovascular causes).”

19,934 women were randomized the low dose aspirin group as compared to 19,942 to the placebo group.

The main findings were given as follows:

“During follow-up, 477 major cardiovascular events were confirmed in the aspirin group, as compared with 522 in the placebo group.”

(Please note: the researchers were only concerned with whether a woman developed cardiovascular disease within the 10-year follow-up period, and not the time until developing cardiovascular disease in this period.)

Answer the following based on the given data:

- (a) Calculate both the odds ratio and relative risk for low dose aspirin therapy (relative to placebo) and its association with cardiovascular disease in the 10-year follow-up period. Are they both allowable statistics to estimate given the study design? (3 points; 1 for each estimate, 1 for discussing “allowability”)
- (b) Are the relative risk and odds ratios from part (a) similar in magnitude? Why, or why not? (2 points, 1 for each question)
- (c) Use the *csi* command in Stata to compute 95% CIs for both the relative risk and odds ratio (for the odds ratio, you can use *csi* command with the option “, or”). How do the confidence intervals compare in terms of the values of the endpoints, and the inclusion/exclusion of the null values for no association? (3 points; 1 for each CI, 1 for comparison)
- (d) What statistical test can be used to compare the low dose aspirin and placebo groups to determine if there are differences in 10-year cardiovascular disease risks. What does the null hypothesis of no association mean in terms of the true relative risk? What does it mean in terms of the true odds ratio? Why do you only need one statistical test to test both of these hypotheses? (3 points: 1 for each portion)
- (e) Report a p-value from this test discussed in part (d). Is the p-value consistent with your confidence intervals from part (c)? (1 point)

¹ Ridker P, et al. A Randomized Trial of Low-Dose Aspirin in the Primary Prevention of Cardiovascular Disease in Women. *New England Journal of Medicine*. 2005; 352,13; 1293-1304

- (f) Calculate/Report the corresponding risk difference (difference in proportions) and its corresponding 95% confidence interval. (2 points)
- (g) Interpret both the estimated (sample) relative risk and the estimated (sample) risk difference each from this study. (2 points)
- (h) Is it likely that the association you estimated between the 10 year risk of CVD and aspirin in previous sections is confounded by other characteristics of the women enrolled in the study? Justify your answer. (1 point)
- (i) Is it possible that there is a differential association between the 10 year risk of CVD and aspirin for women ≥ 65 years old and women < 65 years old? Why/why not? (1 point)
- (j) What do these overall study results show regarding the relationship between low dose aspirin therapy and the 10 year risk of developing cardiovascular disease in women 45 and older? (1 point)

Hypothetical computation time:

Assume the same number of women were randomized to the aspirin and placebo groups as in the study referenced in parts (a)-(j). However, suppose the baseline risk of developing cardiovascular disease in the 10-year follow-up period in the placebo group is 50%, and the relative risk estimate of *cardiovascular disease* for women who received low dose aspirin compared to women who did not is the same value you report in part (a). In this scenario:

- (k) What is the 10-year risk of cardiovascular disease in the sample of women who received *low dose aspirin therapy*? (1 point)
- (l) What is the risk difference for the 10-year cardiovascular disease risk for the women who received low dose aspirin therapy compared to those who did not? (1 point)
- (m) What is the odds ratio comparing the odds of cardiovascular disease in the 10 year follow-up period for the women who received *low dose aspirin therapy* compared to those who did not? (1 point)
- (n) How do the attributable risk and odds ratio estimates compare to what you computed in sections (e) and (a)? If they differ, give a reason for the differences. (2 points)
- (o) Give a p-value testing for the association between low dose aspirin and 10-year risk of cardiovascular disease in this hypothetical scenario.

2. The following table summarizes the smoking status and lung cancer status of 1,418 subjects studied in a landmark case-control study done in England in the late 1940's by British researchers Richard Doll and A. Bradford Hill². (This is a classic example of applied biostatistics and epidemiology that has made a huge impact on the medical and public health worlds – I encourage you to read the full article which is posted on the course website)

Disease Group	Smoker	Non-smoker
Lung Carcinoma Patients	688	21
Control Patients	650	59

- Using the data in the table, estimate the odds ratio of lung cancer (lung carcinoma) for smokers relative to non-smokers. (1 point)
- Use the `cci` command in Stata to estimate a 95% confidence interval for the odds ratio. Interpret the results in a sentence. (2 points: 1 for CI, 1 for interpretation)
- What statistical test(s) is appropriate for testing for a statistical association between lung cancer and smoking? What is the p-value for testing this? Is this consistent with the results in part (b)? (2 points: for test and p-value, 1 for “consistency” question)
- Why did it make sense to do this study as a case-control study as opposed to a prospective, observational study? (hint: lung cancer is a relatively rare disease) Is the estimated odds ratio a reasonable estimate of the relative risk of lung cancer for smokers compared to non-smokers in the study population? (2 points: 1 for each question)

The following table is a verbatim copy of Table IV from the article – notice that the column referring to smoking is the 2nd column.

Disease Group	Number of Non-smokers	Number of Smokers	Probability Test
Males:			
Lung Carcinoma Patients (n = 649)	2 (0.3%)	647	p = 0.0000064 (exact method)
Control Patients (n = 649)	27 (4.2%)	622	
Females:			
Lung Carcinoma Patients (n = 60)	19 (31.7%)	41	.01 < p < .02 (chi-squared)
Control Patients (n=60)	32 (53.3%)	28	

- Using the data in Table IV estimate the odds ratio of lung cancer for smokers relative to non-smokers separately for males and females. (2 points)
- Use Stata to compute 95% confidence intervals for each of the odds ratios in part (e). (2 points)
- How do the odds ratio and corresponding confidence intervals for males and females compare? How do they both compare to the odds ratio and confidence interval from

² Doll R, Hill A. Smoking and Carcinoma of the Lung, A Preliminary Report. *British Medical Journal*. Sep 30 1950, 739-747.

parts (a) and (b)? (2 points)

- h) Do these results suggest that sex confounded the overall relationship between smoking and lung cancer and/or do these results suggest an interaction between sex and smoking? (2 points)
3. Suppose you are charged with both implementing and evaluating a new drug treatment program in Baltimore City. One of the measurable goals of the new program is to improve the one-year relapse rate (defined as percentage of graduates who relapse within one year after completing treatment) compared to the one year relapse in the city's largest current treatment program. You are only concerned with the percentages, not the time to relapse. Your research suggests that 31% of the patients in the current program relapse within in one year.

Your evaluation plan will be to randomize subjects in equal numbers to both the new treatment program and the largest current program – after “graduation” from the respective programs you will track graduates for up to a year to ascertain whether each relapses.

- a) Suppose you only interested in adopting the new program in favor of the current program if it reduces the absolute percentage of graduates relapsing within one-year by at least 10% (ie: the new program's one year relapse rate is 21% maximum). How many subjects will need to be enrolled in each program to detect (at least) this difference with 80% power, and alpha level 5% (.05)? How many subjects will need to be enrolled in each program to detect (at least) this difference with 90% power, and alpha level 5% (.05)? (4 points)
- b) Suppose you only interested in adopting the new program in favor of the current program if it reduces the absolute percentage of graduates relapsing within one-year by 5% (ie: the new program's one year relapse rate is 26% maximum). How many subjects will need to be enrolled in each program to detect (at least) this difference with 80% power, and alpha level 5% (.05)? How many subjects will need to be enrolled in each program to detect (at least) this difference with 90% power, and alpha level 5% (.05)? (4 points)
- c) How does the necessary sample size vary with changes in the detectable difference of interest and changes in the desired power? (2 points)

Sample Quiz Questions: For each question, please choose one answer and provide a short justification (1-2 sentences) for each. (1 point for each correct answer choice, 1 point for correct justification)

4. A case-control study is performed to study the relation between cardiovascular disease and smoking. The cases were **200 persons** with cardiovascular disease, and the controls were **260 persons** without disease. Suppose 40 of the cases were smokers and 20 of the controls were smokers. What statements can be made about the estimated odds ratio comparing the odds of cardiovascular disease for smokers to non-smokers?
- a) The estimate of the odds ratio is 2.0.
 - b) The estimate of the odds ratio is 3.0.
 - c) The estimate of the odds ratio is 5.2.
 - d) The estimate of the odds ratio is 6.0.
 - e) The odds ratio cannot be estimated from this study or with the information provided.
5. Referring to the information given in question 5, what is an estimate of the percent of smokers in the population of persons from which the cases and controls were sampled that have cardiovascular disease? In other words, among all smokers in this population, about what percent have cardiovascular disease?
- a) The percent of smokers is about 10%
 - b) The percent of smokers with disease is about 20%
 - c) The percent of smokers with disease is about 33%
 - d) The percent of smokers with disease is about 67%
 - e) The percent can be estimated from the information given but none of the above answers is correct
 - f) The percent cannot be correctly estimated from this study or with the information provided.
6. A study with 2000 subjects was performed to assess the efficacy of a flu shot in preventing the flu in a 3-month period following administration of the shot. Patients were **randomized** to receive either a flu shot or placebo. The sample relative risk of getting the flu in three month period following treatment was .60 comparing those on treatment to those on placebo. It is a fact that smokers have higher flu risk than non-smokers. If the relative risk of the flu for those in the treatment group versus the placebo was adjusted for smoking status, the adjusted relative risk will most likely be:
- a) substantially greater than 0.60
 - b) about equal to 0.6
 - c) substantially less than 0.60