# **Biostatistics 513**

### Homework 2 - due 4/15/13

Please submit your assignment in Word or PDF format by 9:30 am (PDT) to the Biost 513 Assignment Dropbox at <u>http://canvas.uw.edu/</u>. You may go to "Assignments" at the left side menus on your Canvas page and then follow the instructions to upload your assignment.

*NOTE:* Unless explicitly stated, direct computer output is not desired. Typically only part of the computer output is asked for (such as a confidence interval) and then proper interpretation of the statistics is requested.

DATA: The data for these exercises can be found on the class web page: http://courses.washington.edu/b513/ in the Homework directory.

**Stata.help2**: Key Stata commands that are useful for these exercises are described in the text file stata.help2, also available in the homework directory.

#### Counterfactuals

 A study was recently undertaken to prevent mother to child transmission (MTCT) of HIV and is to be conducted in Lusaka, Zambia. The goal is to find ways to increase the number of HIV-infected pregnant women to use nevirapine (NVP), a drug that substantially reduces the risk of MTCT. NVP is given to the mother in a single dose at the onset of labor; a dose is also given to the infant shortly after birth. The drug is safe - side effects (from a single dose) are rare and seldom severe – and has been proven effective in a randomized trial. However, there have been some difficulties in providing the drug to HIV-positive women in Africa. The current standard of care is to provide the drug to all women who test HIV positive (this is called "targetted" therapy). However, many women refuse to be tested (there are often negative social consequences to testing HIV positive) and, therefore, are not eligible to receive NVP. As a result, an alternative strategy of providing NVP has been suggested for high prevalence settings. Under this alternative standard, all women who test positive for HIV and all women who refuse testing would be offered NVP (this is called "combined" therapy). We are planning a trial to determine which approach provides better "NVP coverage" (as measured by NVP in cord blood samples collected from HIV-infected women at delivery).

A key limiting factor in the design of this study is that women cannot be individually randomized to one approach or the other. Rather, the only feasible approach is to randomize entire health clinics to one approach or the other. Under this scenario, women would receive whichever approach is offered by the clinic they attend. Now here's where things get tricky! The NVP pill is offered to a woman at two points in time: antenatally (to take when labor begins) and again at labor and delivery (LD) (in case she forgot, lost or wasn't able to take the pill at the onset of labor). The problem is that a woman may go to one clinic for antenatal care but a different clinic for labor and delivery. Thus, clinics are to be randomized to one treatment arm or the other, but women are not uniquely associated with a single clinic. A given woman may receive a combination of the treatments. So, how should we define the treatment effect? (Note: for the <u>primary</u> analysis, interest is in the <u>overall</u> treatment effect; not separating out the antenatal component from the LD component).

Assume that half the antenatal clinics are randomize to targetted therapy and half to combined and assume also that half the LD clinics are separately randomize to targetted therapy and half to combined. The outcome, Y, is the presence of NVP in the cord blood of

# **Biostatistics 513**

an HIV-infected woman. Assume that for each HIV-infected woman in Lusaka we know the outcome and which clinics she attended for antenatal care and LD.

- a) List the (4) potential (counterfactual) outcomes for any given HIV-infected woman and express the causal effect of the intervention in terms of these potential outcomes. Choose any measure of the effect (i.e. RD, RR, OR) that you think is appropriate.
- b) Describe (briefly) how one could estimate the causal effect that you listed in (a) using the data that will be available from the study.
- c) What key assumption do you have to make to estimate the causal effect in (a) from the available data?
- d) What common assumption of most statistical inference procedures (i.e. confidence intervals and hypothesis tests) may be violated in this study (hint: are women attending the same clinic likely to be more alike than women attending different clinics?)?

#### **Stratified Tables**

- 2) Beitler and Landis (1985) present data from a randomized, controlled clinical trial conducted at eight clinics. The purpose of the study was to evaluate the effect of a topical cream in curing nonspecific infections. The binary response variable was classified as favorable or unfavorable response to treatment. The data are in the file landis.dat on the class web page. Documentation is in landis.txt. The do-file landis.do will read in the data.
  - (a) Summarize the response to treatment in a 2 x 2 table that ignores clinic. What is the success rate for the drug group and the control group? Calculate and interpret a summary odds ratio. Interpret the  $\chi^2$  statistic.
  - (b) Summarize the response to treatment after stratifying on clinic. Present a table that shows: estimates of p<sub>0i</sub>, the success rate for the control group at clinic i; and p<sub>1i</sub>, the success rate for the drug group at clinic i; the estimated relative risk for each clinic; and the estimated odds ratio for each clinic. Summarize what this table suggests.
  - (c) Compute the Mantel-Haenszel test. State the null and alternative hypotheses for this example. Interpret the  $\chi^2$  statistic and corresponding p-value.
  - (d) Compute the test of homogeneity of odds ratios. State the null and alternative hypotheses for this example. Interpret the  $\chi^2$  statistic and corresponding p-value.
  - (e) Compute the Mantel-Haenszel odds ratio estimate and corresponding confidence interval. Interpret this CI and compare these conclusions from those obtained in part (a).
  - (f) Is clinic a confounder for the association between the response to treatment and the intervention arm (ie. exposure groups are drug and control)? Justify your answer.
  - (g) Summarize the results of your analysis. Specifically, do you conclude that the drug is effective? Can you state a measure of the effectiveness of the drug? Can you state an estimate of the probability that a favorable response is obtained for a patient that you are advising?

3) The following data were taken from a recent study: "Dietary Iron and Coronary Heart Disease: A Study from Greece" by Tzonou et al. (1998) AJE. Because biochemical data suggest that iron is involved in lipid peroxidation and animal experiments have indicated that iron can promote ischemic myocardial injury, these findings have prompted the study of iron in relation to coronary heart disease (CHD). The following variables are reported in Tzonou et al.:

FEMALE:	0 = male
	1 = female
AGE:	1 = <= 49
	2 = 50-59
	3 = 60-69
	4 =>= 70
IRON: estimated monthly iron intake (mg)	
	1 = <= 250
	2 = 251 - 300
	3 = 301-350
	4 = 351 - 400
	5 = > 400
CASE:	0 = control
	1 = case
COUNT: number of subjects	

The goal of the study was to assess the relationship between elevated iron consumption and the risk of CHD. A total of 338 cases were obtained and 570 community controls were also recruited. A dichotomous exposure variable was created:

NEWIRON=1 if IRON > 350 mg/month NEWIRON=0 if IRON  $\leq$  350 mg/month

The files athens.dat, athens.doc and athens.do on the course website contain the data, documentation and stata commands to read the data, respectively.

**Question**: Is there evidence that iron consumption (defined using NEWIRON) is associated with CHD? Support your conclusions with appropriate analyses and justify the methods that you use in support of your conclusion.