DEPARTMENT OF EPIDEMIOLOGY, BIOSTATISTICS,
AND OCCUPATIONAL HEALTH

Ph.D. Comprehensive Examination

Epidemiology

Friday, 17 June 2011

8:45 AM – 12:00 PM

(Note: Time adjusted accordingly to allow students sufficient time to log in and to log out of the Securexam System)

1. This is a closed book exam. Bilingual dictionaries are allowed but cannot be shared. You may also bring with you a one-page (two-sided) set of notes.

2. This exam includes 14 questions. The number of points allocated to each question is indicated. You must answer all questions.

3. The exam lasts three hours. No extra time will be granted.

4. Desk calculators are allowed but cannot be shared. The use of any other electronic device such as computers, phones, or BlackBerry-like devices is strictly forbidden.

5. You may answer in English or in French.

Comprehensive Exam June 17, 2011 : Morning

Questions on: O'Donnell et al. Lancet 2010.

1. [8 marks] In view of how the cases were identified, list and describe 2 advantages and 2 disadvantages of the authors' strategy for selecting controls.

2. Selection bias.

a. [4 marks] Define selection bias in a case-control study.

b. [4 marks] Describe (or show numerically) how differential selection in a case-control study affects the direction of bias for the exposure effect.

3. The authors matched the cases and controls on age and sex.

a. [4 marks] List and explain 2 advantages of matching in a case-control study.

b. [4 marks] List and explain 2 disadvantages of matching in a case-control study.

c. [4 marks] In view of the objectives of the INTERSTROKE study, would you have matched on age and/or sex? Why or why not?

4. One of the risk factors under study was elevated blood pressure.

a. [4 marks] List 2 aspects related to the way blood pressure was measured in this study that could be potential sources of bias.

b. [4 marks] In which direction would the bias operate?

c. [4 marks] On p117 the authors note that using the combined self-report and measured hypertension variable "increased the strength of the associations" with stroke. Why do you think would this strengthen the association?

5. [4 marks] Why did the authors use data from the control group to establish the exposure tertiles for body mass index and waist-to-hip ratio?

6. [6 marks] Look at the evidence on smoking and risk of stroke provided in Figure 1. Taking this into account, describe 2 possible consequences of the authors' dichotomization of smoking on the effects estimated in Table 2.

Cusimano paper ("Ontario Study")

7. [6 marks] Describe in your own words the type of study design used in this paper and explain your choice of terms.

8. The authors stated [first column, page E59] "We calculated the odds of sustaining a bodychecking injury as the proportion of emergency department visits for hockey-related injuries that were due to bodychecking after the rule change divided by the proportion of visits for hockey-related injuries due to bodychecking before the rule change."

Refer to the Atom division row of Table 3.

a [1 mark] Ideally (i.e., if the information were available), in order to calculate an incidence rate [as was done in the Alberta vs. Quebec study] what would the 243 be divided by? the 158 divided by?

b [3 marks] The authors computed the OR as (243/253) / (158/360) = 2.2. In the context of the information the authors had access to, what roles do the 253 and 360 play?

c [1 mark] If the use of these numbers in this role is indeed valid, what parameter is the value of 2.2 a valid estimate of?

d [3 marks] Briefly describe one realistic scenario where the condition for validity is not satisfied and the 2.2 is not a valid estimate.

Emery paper ("Alberta vs. Quebec study")

9. "In each model, player-hours were included as an offset." [second paragraph of Statistical Analysis, page 2267]

a [3 marks] As carefully and clearly as you can, explain the term "offset".

b [1 marks] Why does Poisson regression software need an offset, i.e., why is the offset not included automatically?

10. "clustering by team effect was accounted for" [same paragraph; for a clearer wording, see footnote a under Table 3]

a [2 marks] What do the authors mean by clustering?

b [2 marks] What effect does clustering generally have on the point and interval estimates of the rate ratios?

c [2 marks] Describe one approach you would have used to investigate the extent of the clustering.

11. Refer to the results shown in Table 3 for the "all injury" endpoint, and think about preparing (coding) the dataset for the analysis, the statistical model, and running the models.

a [1 mark] Does each row i.e. 'observation' of the analysis dataset represent a team? or a subject? or something else?

b [2 mark] Write out the regression model that corresponds to the results in the Table.

c [1 mark] In the programming statement that you would use in your favorite statistical software package, what goes on the left hand side?

d [3 marks] Compute the values of the fitted coefficients corresponding to the incidence rate ratios for the "Province" and "Attitudes toward body checking" risk factors, along with their SE's.

e [3 marks] Imagine that the player size (low, high) was found to be an effect modifier so that instead of an overall incidence rate ratio of 3.26 for Province (Alberta vs. Quebec), the Poisson regression fitted rate ratio for Province was 1.5 in the low size group and 4.5 in the high size group. Provide the additional fitted coefficients from the Poisson regression model corresponding to this analysis.

12. Consider the design differences between the two studies (Cusimano, Emery).

a [3 marks] The Alberta vs. Quebec study used "any injury". Give two advantages and one disadvantage of using this as the endpoint (in contrast with the endpoint used in the Ontario study: "injury due to bodychecking)".

b [3 marks] Which of the two studies do you consider provides the more solid evidence with respect to the effect of the bodychecking policy? In at most 3 sentences, explain your reasoning.

13. The authors ("Limitations") conclude that selection bias due to teams not participating in the study was unlikely.

[2 marks] How can selection bias occur in a cohort study?

[2 marks] Do you think it could have happened here? Explain why / why not. If you think it could, explain its impact on the estimated rate ratio of 3.26.

<u>Follow-up to Ontario and Alberta-Quebec studies</u>

14. [6 marks] Suppose both sets of authors had pilot funding to get together and plan a new and more comprehensive study. (In at most four sentences) What advice on study design would you give them?