

NAME: \_\_\_\_\_ Student ID #: \_\_\_\_\_

**University of Guelph  
Department of Population Medicine**

**Midterm Examination I – Epidemiology (POPM\*3240)**

Course Coordinator: Dr. Andria Jones-Bitton  
October 6, 2014

This exam is out of 31 marks and is worth 20% of the total course grade

# EXAM VERSION: A

## INSTRUCTIONS:

1. Write your name and ID number on the top of EVERY page of the exam AND on the Computer Test Scoring Sheet.
2. Communication with anyone other than the instructors or invigilators during the exam is NOT permitted.
3. Scientific or desk calculators may be used, but *cannot* be shared.
4. The use of other electronic devices like computers, cell phones and PDAs is NOT permitted. Cell phones must be turned OFF during the exam.
5. ALL pages of the exam AND the Computer Test Scoring Sheet must be handed in at the end of the exam period. This includes the formulae sheet!
6. **PART A: MULTIPLE CHOICE QUESTIONS (MCQ)**
  - Fill in the personal information on the Computer Test Scoring sheet
  - Each question is worth one (1) mark
  - There is only one correct answer for each MCQ - choose the most correct answer and circle on the exam page
  - Enter the correct answer for each MCQ on the Computer Test Scoring sheet
  - Use only an HB #2 pencil to fill in the Computer Test Scoring sheet. Make heavy black marks that fill the circle completely. Erase cleanly any answer you wish to change. Make no stray marks on the answer sheet.
7. **PART B: SHORT ANSWER QUESTIONS**
  - Answer ALL questions in the space provided on (the front of) the exam pages  
\* DO NOT WRITE ON THE BACK OF EXAM PAGES \*
  - Please write legibly (we cannot mark things we cannot read)
  - For calculations, show your work. Retain values to 3 decimal places throughout your calculation work, but report end result to once decimal place
  - Answers written in pencil will not be eligible for any re-grade request

**Formulae and Information you may find useful:**

$$n = \frac{Z_{\alpha}^2 * \sigma^2}{L^2}$$

$$n = \frac{Z_{\alpha}^2 * p * q}{L^2}$$

$$n = \frac{(Z_{\alpha}\sqrt{2pq} - Z_{\beta}\sqrt{p_1q_1 + p_2q_2})^2}{(p_1 - p_2)^2}$$

$$n = 2 * \left[ \frac{(Z_{\alpha} - Z_{\beta})^2 * \sigma^2}{(\mu_1 - \mu_2)^2} \right]$$

$$\begin{aligned} RD &= P(D+ | E+) - P(D+ | E-) \\ &= [a/(a+b)] - [c/(c+d)] \end{aligned}$$

$$APe = RD / P(D+ | E+)$$

$$\begin{aligned} PAR &= p(D+) - p(D+ | E-) \\ &= [(a+c)/n] - [c/(c+d)] \end{aligned}$$

$$PAF = PAR / p(D+)$$

**PART A: MULTIPLE CHOICE QUESTIONS – 12 marks**

1. Edward Jenner is best known for:

- A. His work on cholera outbreaks in London
- B. Conducting one of the first randomized controlled trials, exploring treatments for scurvy
- C. Conducting one of the first vaccinations, for smallpox**
- D. Making one of the first known records of occupational exposures, namely, cancer of the scrotum in chimney sweepers
- E. Edward Jenner is not known for any of the above

2. The time period between when an animal is exposed to a disease-causing agent and when pathological changes can be detected is known as the \_\_\_\_\_ period, and interventions instituted at this point are known as \_\_\_\_\_ interventions.

- A. Latent, Primary
- B. Latent, Secondary**
- C. Incubation, Primary
- D. Incubation, Secondary
- E. Incubation, Tertiary

3. Assume the numbers of Lyme disease cases reported in Canada between 2010-2013 were published by the Public Health Agency of Canada to be:

- 2010: 128 cases
- 2011: 132 cases
- 2012: 140 cases
- 2013: 139 cases

Based on this information, Lyme disease is an example of which of the following disease patterns in Canada?

- A. Sporadic
- B. Endemic**
- C. Epidemic
- D. Pandemic
- E. Seasonal

4. Which of the following statements is CORRECT? The p-value gives the probability of:

- A. Obtaining a result as, or more, extreme than the one observed
- B. The null hypothesis arising by chance alone
- C. The null hypothesis being true
- D. The observed results arising if the null hypothesis is true**
- E. More than one of the above is correct

5. You are designing a study to determine whether a new feed ration will improve the average daily weight gain (kg) of pigs, compared to the current feed ration. To calculate an appropriate sample size for this study, which of the following formulae should you use?

A. 
$$n = \frac{Z_{\alpha}^2 * \sigma^2}{L^2}$$

---

B. 
$$n = \frac{Z_{\alpha}^2 * p * q}{L^2}$$

---

C. 
$$n = \frac{(Z_{\alpha}\sqrt{2pq} - Z_{\beta}\sqrt{p_1q_1 + p_2q_2})^2}{(p_1 - p_2)^2}$$

---

D. 
$$n = 2 * \left[ \frac{(Z_{\alpha} - Z_{\beta})^2 * \sigma^2}{(\mu_1 - \mu_2)^2} \right]$$

---

E. None of these formulae should be used

---

6. An investigator randomly selects 100 old-age homes across Canada and then randomly samples 25% of the residents from each home and tests them for Methicillin-resistant *Staphylococcus aureus* (MRSA). This is an example of:

- A. **Multistage sampling**
- B. Cluster sampling
- C. Stratified random sampling
- D. Systematic random sampling
- E. Simple random sampling

7. A medical company has developed a new test for antibodies to the *Ostertagia ostertagi* parasite, and wants to know how well it compares with their competitor's test. They calculate the Kappa statistic. How can Kappa help the company make a decision?

- A. It cannot help, because statistical tests cannot be used to compare medical tests
- B. It cannot help, because you need a gold standard to compare tests
- C. It can help by showing which tests performs better
- D. It can help by providing an indication of how well the tests agree beyond chance**
- E. It can help by providing an indication of the sensitivity of the tests when used together

8. You own a sheep farm (lucky you!) and want to increase your flock size to meet the growing demand for lamb in Ontario. You decide to purchase sheep from a nearby producer who is retiring. Being smart, you want to test his sheep first, so that you do not introduce Johne's disease to your existing flock. Two tests are available. Which approach would be *best in this scenario*?

- A. Use the one test with the higher sensitivity
- B. Use the one test with the higher specificity
- C. Use two tests, and interpret them in series to achieve higher specificity
- D. Use two tests, and interpret them in series to achieve higher sensitivity
- E. Use two tests, and interpret them in parallel to achieve higher sensitivity**

9. Commonly used techniques to reduce confounding effects include all of the following EXCEPT:

- A. Matching
- B. Stratification
- C. Increasing sample size**
- D. Multivariable regression analysis
- E. Randomization

10. You are conducting a study to determine the prevalence of Q-fever infection amongst farm workers who work with cattle, sheep and goats in the province. Based on available literature, you suspect approximately 45% of farm workers to have evidence of infection. You want your prevalence estimate to have a precision of +/- 5% and 95% confidence. The minimum required sample size for your study is:

- A. 20
- B. 195
- C. 353
- D. 381**
- E. 1902

NAME: \_\_\_\_\_ Student ID #: \_\_\_\_\_

11. When assessing causality, all of the following should be considered, EXCEPT:

- A. **Disease prevalence**
- B. Consistency with other knowledge
- C. Strength of association
- D. Alternative explanations
- E. Biological plausibility

12. Which of the following is FALSE?

- A. Sub-clinical infection occurs when a pathogen establishes itself in a host, but clinical signs are not visible
- B. **A chi-square test should be used to test whether there's a significant difference in mean values across two populations**
- C. Hand-washing to prevent fecal-oral contamination is a form of primary prevention
- D. If enough people in a population have been vaccinated, herd immunity can confer protection to non-vaccinated individuals
- E. A sampling frame is a list of individuals in a source population

**PART B: SHORT ANSWER – 19 marks**

**Question 13 (6 marks).** You are evaluating a screening test for a newly discovered disease. In order to determine its effectiveness, you administered it to 1000 workers, 100 of whom truly had the disease. A negative test result occurred in 30 of the people who had the disease. A total of 150 people who did not have the disease tested positive for it. (Hint: construct a 2x2 table)

	Disease +	Disease -	
Test +	70	150	220
Test -	30	750	880
	100	900	1000

a. Your boss wants to know how good the test is in classifying diseased individuals. Name (1 mark), calculate (1 mark) and interpret in words (1 mark) an appropriate measure that will tell your boss what she wants to know.

**Sensitivity =  $70/100 = 70.0\%$ , which means the test will correctly identify 70% of disease individuals. Said alternately, 70% of all truly diseased individuals will test positive (or any other correct interpretation).**

*1 mark for correct measure (SN), 1 mark for correct calculation, and 1 mark for correct interpretation of their calculation (even if their mathematical calculation was wrong) -- if just gave definition of SN and did not interpret the actual value: 0.5 mark (question asked student to “interpret in words”, not just to “define” SN).*

b. Calculate (1 mark) and interpret in words (1 mark) the positive predictive value of this test.

**Positive Predictive Value =  $70/220 = 31.8\%$ , which means that only approximately 32% of test-positive individuals will actually have the disease (or any other correct interpretation).**

*1 mark for correct calculation, 1 mark for correct interpretation of their PPV (even if their mathematical calculation was wrong) -- if just gave definition of PPV and did not interpret the actual value: 0.5 mark (question asked student to “interpret in words the PPV of this test” i.e. in relation to the scenario/question; not just to define PPV).*

c. You are travelling to a country that has a higher prevalence of this disease. What effect will this have on the positive predictive value of the test, if any? (1 mark).

**Using the same test in an area with higher prevalence of the disease will increase (1 mark – no part marks) the PPV.**

NAME: \_\_\_\_\_ Student ID #: \_\_\_\_\_

**Question 14 (6 marks).** An investigator wants to determine the relationship between pregnant women smoking during pregnancy and low birth weights in their babies. He follows 1000 women through their pregnancies. His findings are detailed below.

	Low Birth Weight	Normal Birth Weight	
Smoker	65	100	165
Non-smoker	35	800	835
	100	900	1000

a. Name (1 mark), calculate (1 mark) and interpret in words (1 mark) a measure of the strength of association between smoking during pregnancy and low birth weight.

**Relative Risk (RR) =  $(65/165) / (35/835) = 0.393 / 0.042 = 9.3$**

**Therefore, the risk of low birth weight babies amongst women who smoke during pregnancy is approx. 9 times higher than for women who don't smoke during pregnancy.**

Or,

**Odds Ratio (OR) =  $65 \cdot 800 / 100 \cdot 35 = 14.8$**

**Therefore, the odds of low birth weight babies amongst women who smoke during pregnancy is approx. 15 times higher than amongst women who don't smoke during pregnancy.**

*1 mark for naming the correct measure (RR: risk ratio, relative risk or OR: odds ratio) – no part marks, 1 mark for correct calculation, and 1 mark for correct interpretation of the RR or OR (even if their mathematical calculation was wrong).*

b. What proportion of low birth weight babies amongst women who smoke during pregnancy is due to smoking during pregnancy? (2 marks).

**Looking for Attributable Proportion (Exposed)**

**$APe = [P(D+ | E+) - P(D+ | E-)] / P(D+ | E+)$   
 $= [(65/165) - (35/835)] / (65/165)$   
 $= (0.393 - 0.042) / 0.393$   
 $= 0.893 \rightarrow 89.3\%$**

*1 mark for knowing question is asking for APe (no part marks), 1 mark for correct calculation (1/2 mark if some elements of calculation correct but answer ultimately wrong)*

c. The premier of Ontario wants to know how much of the occurrence of low birth weight babies could be avoided in the province if measures could be taken to completely prevent smoking during pregnancy. What measure would you calculate for her? (1 mark – calculation NOT required)

**Population Attributable Risk or Population Attributable Fraction (1 mark). If just gave acronym (i.e. PAR or PAF) – ½ mark. No marks for incorrect name.**

NAME: \_\_\_\_\_ Student ID #: \_\_\_\_\_

**Question 15 (4 marks).** In 2010, there was a foodborne outbreak of Listeriosis, a disease caused by the bacterium, *Listeria monocytogenes*. To help identify the cause of the outbreak, a case-control study was done whereby a group of cases (i.e. people with Listeriosis) and a group of controls (i.e. people without Listeriosis) were identified and asked about their recent food consumption. One of the suspect foods was deli meat produced by Maple Leaf Foods. It seems reasonable that cases would be more likely to more thoroughly recall deli meat consumption than controls (i.e. controls may be more likely to forget that they had consumed deli meat).

a. What type of bias may be present in this study? Why? Be very specific in your answer (2 mark).

**Differential misclassification bias – “differential recall bias” also accepted (1 mark), but it should be noted that most recall biases are considered “differential”.**

**½ mark for ‘information bias’, ‘misclassification bias’**

**Errors in exposure misclassification (i.e. deli meat consumption status) differed by disease status (1 mark).**

b. What impact do you think this may have on the measure of association? Why? (2 marks).

	Cases	Controls
Ate deli meat	a	b ↓
Didn't eat deli meat	c	d ↑

**This misclassification would lead to a bias away from the null / an overestimated RR or OR (1 mark) because it would lead to higher numbers in the “d” cell (D- | E-) and/or lower numbers in the “b” cell (D- | E+) of the above 2x2 table due to misclassification of controls (1 mark).**

Note: it is not sufficient to say “differential misclassification bias can bias towards or away from the null”. As discussed in class and in the notes, we must think about how the bias is likely to have affected the results.

Note: only controls were misclassified; there is no erroneous inflation in the “a” cell (D+ | E+) because, as stated in the question, it was the controls who were misclassified – specifically, they were more likely to be incorrectly classified due to poor recall.

NAME: \_\_\_\_\_ Student ID #: \_\_\_\_\_

**Question 16 (3 marks).** An investigator found a positive association between use of mouthwash and oral cancer, and was eager to report his findings. A journal reviewer recommended that the paper be rejected because of failure to look at the potential confounding effect of smoking. The investigator is confused. Explain why is smoking likely to confound the observed relationship between mouthwash and oral cancer? (3 marks)

**Smoking is very likely a confounder in the observed relationship between mouthwash and oral cancer, because it is: (1) associated with oral cancer, (2) associated with use of mouthwash, and (3) not a consequence of using mouthwash (1 mark each criteria for max. 3 marks)**

*Extra explanation for student review (not required for answer):* This means that the investigator's study findings are likely false – there is likely not an association between use of mouthwash and oral cancer – it just looked that way because of not accounting for the effect of smoking.