

## Unit 2 – Introduction to Probability

### Week #4 - Practice Problems

**Due Monday October 13, 2008**

*These exercises, as you will see, emphasize introductory probability as it applies to epidemiology. Take your time with these. They are a little more involved conceptually. But hopefully well worth it!*

1. **Before you begin:** Parts “a” – “c” invite you to pause in your thinking. Parts “d” – “f” are straightforward application of the ideas of relative risk and odds ratio. See lecture notes pp 39-47 and/or the text pp 630-64.

In introductory epidemiology, one of the study designs that are introduced is the **prospective cohort study**. In this type of study involving two groups, the investigator enrolls pre-set and known numbers of participants into each of the two groups that are generically described as “exposed” and “not exposed” and follows them forward to a designated end of the observation period, at which point some outcome is measured.

Consider the following prospective cohort study. A total of 1500 **never smoker** consenting heart attack survivors aged 60-65 are enrolled as “non-exposed”. An equal number, 1500 **current smoker** heart attack survivors aged 60-65 are enrolled as “exposed”. All are followed for a full 10 years and the occurrence of death recorded. Following are the data.

		Vital Status at 10 Years		
		Dead	Alive	
Exposure Status	Current Smoker	<b>40</b>	<b>1460</b>	<b>1500</b>
	Never Smoker	<b>10</b>	<b>1490</b>	<b>1500</b>
		50	2950	3000

- (a) Is it possible to estimate the probability of 10 year survival on the basis of these data?
- (b) Is it possible to estimate the relative risk of 10 year mortality that is associated with current cigarette use?
- (c) Is it possible to estimate the probability that a randomly selected person with a vital status of “Alive” at 10 years is a current smoker?
- (d) Using these data, estimate the relative risk of 10 year mortality that is associated with current cigarette use.
- (e) Using these data, estimate the relative odds of 10 year mortality that is associated with current cigarette use. (Note – This question is asking you to compute an **odds ratio**).

- (f) Using these data, estimate the relative odds of a **current smoker** notation for non-survivors relative to survivors.
- (g) What do you notice about your answers to “e” and “f”?
- (h) How do your answers to “e” and “f” compare to your answer to “d”?

2. **Before you begin:** *This question is really an elaboration of the thinking that was developed in question 1.*

Another study design that is introduced in introductory epidemiology is the **retrospective case-control study**. This is, by definition, a study that compares two groups. Here, the investigator enrolls pre-set and known numbers of participants into each of the two groups defined by disease status; “cases” are the enrollees with disease, “controls” do not have the disease under investigation. Retrospective review of the histories of all study participants is performed to identify the subsets in each of the case and control groups who have a history of the exposure of interest.

Consider the following retrospective case-control study of the association between coffee consumption and tumors of the lower urinary tract. The investigator enrolls 30 consenting cases that are patients with one or more tumors of the lower urinary tract. For comparison purposes, he/she also enrolls 100 consenting controls who have no such tumors. Following are the data.

		Tumors of Lower Urinary Tract		
		Yes	No	
History of Coffee consumption	5+ cups/day	20	44	64
	<1 cup/day	10	56	66
		30	100	130

- (a) Is it possible to estimate the probability of one or more tumors of the lower urinary tract on the basis of these data?
- (b) Is it possible to estimate the relative risk of one or more tumors of the lower urinary tract that is associated with consumption of 5 or more cups of coffee/day?
- (c) Using these data, estimate the relative odds of **high coffee consumption (5+ cups/day)** among cases, relative to controls.
- (d) Using these data, estimate the relative odds of **tumors of the lower urinary tract** among high coffee consumers (5+ cups/day), relative to non-coffee drinkers.
- (e) What do you notice about your answers to “c” and “d”?

3. **Before you begin:** *This last question is a bit of a weaving of a story. If you follow along, step by step, you will end up seeing for yourself a truly marvelous result. It is the result of page 46 of your lecture notes!*

Now consider a fully **cross-sectional study design**, this time with generic counts “a”, “b”, “c”, and “d”. In this design, the investigator does not do any formal enrollment. Counts are accumulated by observation. CDC surveillance programs are examples

		Disease	
		Yes	No
History of Exposure	Yes	<b>a</b>	<b>b</b>
	No	<b>c</b>	<b>d</b>

- Using the letters “a”, “b”, “c”, and “d”, what is the formula for estimating relative odds of the **event of exposure** for persons with disease, compared to that for persons without disease?
- Using the letters “a”, “b”, “c”, and “d”, what is the formula for estimating relative odds of the **event of disease** for exposed persons, compared to that for non-exposed persons?
- Using the letters “a”, “b”, “c”, and “d”, what is the formula for estimating relative risk of the **event of disease** for exposed persons, compared to that for non-exposed persons?
- What happens to your formula in your answer to #3b when the counts of disease (a and b) are very very small? Comment.