**Methods in Epidemiologic Research**

**Sample Problems**

**Chapter 12 – Validity**

## Non-response Bias

1. Lack of response can lead to bias. Use the data in Example 12.1 to set up the population in the spreadsheet response_bias.

   (a) Is non-response associated with the exposure categories in the source population?

   (b) Is the outcome frequency (ie D+) associated with the exposure categories in the source population?

   (c) Is there any bias if the outcome frequency is the same in those who do and don't respond conditional on exposure (ie in each of the exposure categories)?

   (d) Is there bias if the outcome frequency is higher (or lower) in the responders than in the non-responders conditional on exposure? In which direction is the bias if it exists?

   (e) Is there bias if the non-response is higher in the non-exposed group? In which direction is the bias if it exists? (Since you have already found out that there is no bias if the disease risks are equal in responders and non-responders, do this problem with unequal disease risks.)

   (f) Does the magnitude of a selection bias depend on the size of the non-response rate (it is actually a risk, not a rate)?

   (g) If the non-response rate is equal in the two exposure catgeories, is there any bias in the estimate of the RR or OR? First, do this exercise with data that produce the same RR in the responders and non-responders. Then repeat the exercise with different RR in the two groups.

## Information Bias

2. Information bias (called misclassification with categorical outcomes) can also lead to biased measures of association. Begin this exercise by setting up the population shown in Example 12.3 in the spread sheet misclassification_bias.

   (a) Pretend this is your study population and obtain both the odds ratio and risk ratio on these data as if you had done a risk-based cohort study.

   (b) Vary both the Se and Sp of exposure individually, but always keep the errors non-differential. Which ( Se or Sp of exposure) has a bigger impact on the outcome?

   (c) What happens to the measures of effect when both the Se and Sp of exposure are less than 100% but non-differential?

3. Repeat 2(b), and 2(c) but now use differential misclassification of exposure.

   (a) What are the impacts on the OR?

   (b) What pattern of differential Se misclassification increases the measures?

   (c) What pattern of differential Sp misclassification increases the measures?

   (d) Can you find a combination of Se and Sp that will increase the measure of association? What is it?

   (e) What combination of Se and Sp likely mimics recall bias in a case-control study?

4. Now we turn our attention to misclassifying disease status in a cohort or cross-sectional study. Using the same starting population as in question 2 above:

   (a) Vary both the Se and Sp of disease individually, but always keep the errors non-differential. Which (Se or Sp of disease) has a bigger impact on the outcome?

   (b) What happens to the measures when both are less than 100% but non-differential?

   (c) What combination of Se and Sp likely mimics detection bias? What direction of bias does this produce in the measures of association?

Use the attached sheet to record some of your answers:

**Information Bias Calculations**

| *Quest.* | *Se E* | *Sp E* | *Se D* | *Sp D* | *Obs. OR* | *Bias* |
|---|---|---|---|---|---|---|
| 2 (b) | | | | | | |
| 2 (b) | | | | | | |
| 2 (b) | | | | | | |
| 2 (b) | | | | | | |
| 2 (c) | | | | | | |
| 2 (c) | | | | | | |
| 3 (a) | | | | | | |
| 3 (a) | | | | | | |
| 3 (a) | | | | | | |
| 3 (a) | | | | | | |
| 3 (d) | | | | | | |
| 3 (d) | | | | | | |
| 3 (e) | | | | | | |
| 4 (a) | | | | | | |
| 4 (a) | | | | | | |
| 4 (a) | | | | | | |
| 4 (a) | | | | | | |
| 4 (b) | | | | | | |
| 4 (b) | | | | | | |
| 4 (c) | | | | | | |