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?

Yes, 0.1 in E- versus 0.3 in E+

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

Yes, 0.25 in E+ versus 0.12 in E-

(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)?

No, the measures are the same in the source and study population (RR=2.08)

(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?

Scenario A – Risk higher in responders

Population size	10000
Probability of E+	0.1
Probability of non-responder in E+	0.3
Probability of non-responder in E-	0.1
Risk of D+ in E+ (responding)	0.3
Risk of D+ in E+ (non-responding)	0.2
Risk of D+ in E- (responding)	0.15
Risk of D+ in E- (non-responding)	0.1

If the risk of D+ is higher in the E+ responders than in the E+ non-responders, (and also in the comparable E- groups) the bias is positive

$RR_{target} = 1.86$

$RR_{cohort} = 2.00$

Scenario B – Risk higher in non-responders

Population size	10000
Probability of E+	0.1
Probability of non-responder in E+	0.3
Probability of non-responder in E-	0.1
Risk of D+ in E+ (responding)	0.2
Risk of D+ in E+ (non-responding)	0.3
Risk of D+ in E- (responding)	0.1
Risk of D+ in E- (non-responding)	0.15

If the risk of D+ is lower in the E+ responders than in the E+ non-responders, (and also in the comparable E- groups) the bias is negative.

$RR_{target} = 2.19$

$RR_{cohort} = 2.00$

(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.)

Scenario A was modified so the non-response was higher in the non-exposed group as follows:

Population size	10000
Probability of E+	0.1
Probability of non-responder in E+	0.1
Probability of non-responder in E-	0.3
Risk of D+ in E+ (responding)	0.3
Risk of D+ in E+ (non-responding)	0.2
Risk of D+ in E- (responding)	0.15
Risk of D+ in E- (non-responding)	0.1

Scenario A – modified (non-response higher in non-exposed)

Yes, there is still bias but now the bias is negative

$RR_{target} = 2.15$

$RR_{cohort} = 2.00$

(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)?

Population size	10000
Probability of E+	0.1
Probability of non-responder in E+	0.7
Probability of non-responder in E-	0.4
Risk of D+ in E+ (responding)	0.3
Risk of D+ in E+ (non-responding)	0.2
Risk of D+ in E- (responding)	0.15
Risk of D+ in E- (non-responding)	0.1

Scenario A – modified to have larger non-response rates (risks)

Yes, the bias is larger if the non-response rate is higher. However, it is interesting to note that even with quite large non-response rates (70% and 40%), the bias is not particularly large.

$RR_{target} = 1.77$

$RR_{cohort} = 2.00$

(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.

Scenario A was modified so that there was 30% non-response in both groups, but both responders and non-responders had a RR of 2.00.

Population size	10000
Probability of E+	0.1
Probability of non-responder in E+	0.3
Probability of non-responder in E-	0.3
Risk of D+ in E+ (responding)	0.3
Risk of D+ in E+ (non-responding)	0.2
Risk of D+ in E- (responding)	0.15
Risk of D+ in E- (non-responding)	0.1

There is no bias in the RR when the non-response is equal in the two groups. (There is a very small bias in the OR, but it would be too small to be of much concern)

$RR_{target} = 2.00$

$RR_{cohort} = 2.00$

This finding is of considerable importance. It means that any study in which there is equal follow-up in the two study groups and this results in equal levels of non-response in the two groups, we do not expect there to be any non-response bias provided that the effect of the factor (ie the RR) is the same in the non-responders and the responders.

Scenario A was modified so that there was 30% non-response in both groups, but the exposure had a RR of 2.00 in the responders and 8.00 in the non-responders.

Population size	10000
Probability of E+	0.1
Probability of non-responder in E+	0.3
Probability of non-responder in E-	0.3
Risk of D+ in E+ (responding)	0.3
Risk of D+ in E+ (non-responding)	0.8
Risk of D+ in E- (responding)	0.15
Risk of D+ in E- (non-responding)	0.1

There is substantial bias in the RR (bias toward the null in this case, but it could be in either direction).

$RR_{target} = 3.33$

$RR_{cohort} = 2.00$

This last finding highlights the importance of getting some data about non-responders in order to determine if they were likely to be fundamentally different than the responders. If you are reasonably convinced that the non-responders were similar to the responders and that the exposure of interest would probably have produced an equal effect in the non-responders, then there will not be any selection bias. However, if the two groups differ substantially, a bias will be present.

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.

   RR=3 and OR=3.86

   (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?

   Sp for an equal change. (This is because there are more non-exposed than exposed)

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

   They just reduce the OR even further (greater bias toward the null)

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

   (a) What is the impact on the OR?

   Depends on the actual values

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

   As the Se for exposure in D+ is increased relative to the Se for exposure in the D-, the bias toward the null decreases and eventually starts moving away from the null.

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

   As the Sp for exposure in D- is increased relative to the Sp for exposure in the D+, the bias toward the null decreases and eventually starts moving away from the null.

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

   Se(E|D+)=.7; Se(E|D-)=.5; Sp(E|D+)=.7; Sp(E|D-)=.9 is one such combination.

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

   Perhaps a higher sensitivity of exposure in the D+ (cases) and a lower specificity of exposure in the D+ (cases) as above.

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?

   Sp has a much bigger impact (again, this is because there are more D- individuals)

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

   They combine to reduce the measures even more drastically

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

Detection bias can arise from problems in either Se or Sp, but perhaps problems with Se are more common (or easier to think about). Detection bias is a problem if the proportion of D+ detected in the E- and E+ groups are different .... so a setting of non-differential misclassification of disease might mimic detection bias.

Information Bias Calculations

Quest.	*Se E*	*Sp E*	*Se D*	*Sp D*	*Obs. OR*	*Bias*
2 (b)	.8	.9	1	1	2.57	- med
2 (b)	.5	.9	1	1	1.93	- lg
2 (b)	.9	.8	1	1	2.44	- med
2 (b)	.9	.5	1	1	1.76	- lg
2 (c)	.9	.9	1	1	2.85	- med
2 (c)	.7	.7	1	1	1.66	- lg
3 (a)	.7 .7	.7 .7	1	1	1.66	- lg
3 (a)	.9 .7	.7 .7	1	1	2.64	- med
3 (a)	.7 .9	.7 .7	1	1	1.35	- v lg
3 (a)	.7 .7	.7 .9	1	1	3.32	- sm
3 (d)	.7 .7	.5 .9	1	1	4.74	+ med
3 (d)	.9 .9	.8 .99	1	1	4.82	+ med
3 (e)	.9 .7	.8 .8	1	1	3.04	- sm
4 (a)	1	1	.8	.9	2.19	- lg
4 (a)	1	1	.5	.9	1.73	- v lg
4 (a)	1	1	.9	.8	1.88	- lg
4 (a)	1	1	.9	.5	1.39	- v lg
4 (b)	1	1	.9	.7	1.64	- lg
4 (b)	1	1	.7	.9	2.04	- lg
4 (c)	1	1	.9 .7	.9	4.21	+ sm