## Section 1: Binary data (one sample).

For binary outcomes, such as being sick or healthy, we usually calculate a proportion. Answer the following questions:

1. What is the proportion of disease, if we had observed 250 sick and 1000 healthy persons?
2. For what values of a proportion (p) uncertainty is greatest? That is, those values of p that gives the widest confidence intervals. While keeping everything else constant.

Formula for 95% confidence interval around proportion (p) is given by

## Section 2: Binary data (2 sample)

Ninety-nine pregnant women, with dystocia (difficult childbirth or labour), were allocated at random to receive immersion in water in a birth pool (Intervention group: Labour in water 49 women) or standard augmentation for dystocia (Control group: Augmentation 50 women) in a randomised controlled trial to evaluate the impact of labouring in water during the first stage of labour. The main outcome was use of epidural analgesia at any stage of labour. The results are shown in table-1

Table 1: Results from the intervention

|  |  |  |
| --- | --- | --- |
|  | Epidural analgesia  at any stage of labour | |
|  | No | Yes |
| Intervention group  (Labour in water) | 26 | 23 |
| Control group  (Augmentation) | 17 | 33 |

1. Report the proportion of women that had an epidural analgesia in each of the two groups?
2. Report and interpret the relative risk along with 95% confidence interval, of using epidural analgesia for the labour in water group compared with the Augmentation?
3. Report and interpret the risk difference along with 95% confidence interval, for the use of epidural analgesia for Labour in water group compared to Augmentation group.
4. Can you conclude that there is a relationship between the treatment and the outcome? Based on results from
   * 1. Q2. relative risk (with 95% CI)
     2. Q3. risk difference (with 95% CI)
5. Which effect measure (RR or RD) is more appropriate to use? Please motivate

## Section 3: Hypothesis testing

A GP have 6000 patients and refer 27 patients to neurology in one year. In the health authority region (HAR), there are 1400 neurology referrals from a population of 500 000.

1. Is this GP’s proportion of referrals to neurology (27/6000) **different** to the HAR proportion of referrals to neurology (1400/500000)?

Perform hypothesis testing based on your judgement of choosing either *Relative Risk* or *Risk Difference*.

Let us make this exercise more interesting, what if we take a smaller sample ONLY for the health authority region but with same proportion.

GP = 27 referrals out of 6000, HAR = 14 referrals out of 5000

1. Redo the analysis (from question 1) but using this smaller sample size.
2. Did your conclusions from (1) change? If yes, then why?

## Question 4: Diagnostic tests

Three tests (A, B and C) for the diagnosis of cancer were assessed against a ‘gold standard’ taken to be 100% accurate. Their sensitivities were **A** 90%, **B** 85%, **C** 80%. Their specificities were **A** 100%, **B** 90%, **C** 95%. All three tests carried the same cost, and none was associated with any side effects. Mark *right* or *wrong*.

It follows that:

1. In these circumstances test A will always be preferable to test B
2. In these circumstances test C will always be preferable to test A
3. In cancer diagnostics, test B will always be preferable to test C.

(What is more important in cancer diagnosis? *Sensitivity* or *specificity.* )

1. Test C detects a higher proportion of cases than test B
2. There are no false positive results with test A
3. The predictive value of test B will depend on the prevalence of the disease in the population to which it is applied