Clinical Biostatistics

Suggested Answers: Categorical Data

1. *What is Pearson's χ^2 test and why was it used to compare epidural analgesia between the treatment groups?* This is the chi-squared test for a contingency table. We calculate the frequencies which we would expect to get if the null hypothesis of no relationship between treatment and epidural were true. We compare the observed and expected frequencies to give a chi-squared statistic, which would follow a Chi-squared distribution if the null hypothesis were true. This is a large sample test for the association between two categorical variables. Treatment and type of anaesthesia are both categorical and the numbers are large enough for all expected frequencies to exceed 5.

2. *In Table 1, for the labour in water group, the mean, SD, and range for cervical dilation between start of labour and randomisation (cm/hour) were 0.27 (0.24, –0.22 - 0.72). What can we conclude about this distribution and why?* The mean is very close to the middle of the range (which is actually at 0.25, because (0.72 – 0.22)/2 = 0.50/2 = 0.25). This suggests that the distribution is symmetrical. The standard deviation is similar to the mean, so a symmetrical distribution would have many observations below zero (two SD below the mean). This is possible here, because the variable, which is a difference, can be negative.

3. *Women randomised to immersion in water had a lower rate of epidural analgesia than women allocated to augmentation (relative risk 0.71) (Abstract). What is meant by risk and relative risk here and what does '0.71' tell us?* 'Risk' means the proportion of women who have epidural analgesia out of those who are undergoing delivery, i.e. who are at risk. Relative risk means the proportion of the labour in water group who have epidural analgesia divided by the proportion of the augmentation group who have epidural analgesia. 0.71 tells us that in this sample the risk of epidural was lower in women allocated to labour in water than in women allocated to augmentation.

4. *For epidural analgesia, the number needed to treat for benefit (NNT) was 5 (Abstract). What does this mean?* This means that for every 5 women we treat with labour in water rather than augmentation, we reduce the number of women requiring an epidural by one. It is calculated by one over the difference between treatment groups in the proportion of women requiring an epidural.

5. *For amniotomy or oxytocin, or both, the number needed to treat to benefit one woman was 4 (95% CI 3 to 9) (Table 2). Why is the NNT not in the middle of its confidence interval?* The confidence interval for NNT is calculated from the confidence interval for the difference between treatment groups in the proportion of women requiring amniotomy or oxytocin. This confidence interval is symmetrical, being calculated by the observed difference minus or plus 1.96 standard errors. We then take the reciprocals to get the NNT and its confidence interval. This destroys the symmetry.

6. *For any operative delivery, the number needed to treat to benefit one woman was 98 (95% CI 98 (3 to ∞ to NNTH −5)) (Table 2). What does this mean?* The difference between the proportions was not significant and the confidence interval for the difference included zero. The reciprocal of zero is infinity. The NNT could be as low as 3, or take any value above 3, as high as we like. NNTH means the number of needed to treat to harm. It may be that the treatment would cause an decrease in the number of women avoiding an operative delivery, a negative difference in proportions. The number of women required to produce one extra operative delivery would be 5 or more, any value up to infinity being possible.

7. *For admission to neonatal unit in first 10 days, 6 (12%) of the labour in water group were admitted compared to none of the augmentation group (P = 0.013, exact test) (Table 4). Why was an exact method used for this test?* Chi-squared tests are large sample tests and require at least 80% of the expected frequencies to exceed 5. In a table with only 4 frequencies, like this one, at least 80% means all of them. If only 6 babies were admitted, the two expected frequencies for this row of the table cannot both exceed 5, as their total is 6. The condition is not met. Instead we use a test with exact probabilities which involves no large sample approximations.

8. *In the discussion, the authors say 'Our findings suggest that delaying augmentation in association with a supportive environment (water immersion) . . . may reduce the need for epidural analgesia without increasing labour length or operative deliveries.' Why are they so cautious?* The P value (Table 2) is 0.056, so the evidence is only weak for an effect of labour in water on need for epidural anaesthesia. Of course, they could have made this statement without doing the study.