Clinical Biostatistics

Suggested Answers: Specimen Examination, 2006

1. *The difference in the proportions of women experiencing persistent occiput position was not statistically significant. Was does this mean and what could we conclude from it?* We are testing the null hypothesis that in the population the proportions of women who would experience persistent occiput position is the same whether given the exercise or not. The probability that we would get a difference as large as that observed if the null hypothesis were true is greater than 0.05. We conclude that we have failed to show any evidence for a differences between the treatments. We should not conclude that the treatments are the same.

2. *The difference in risk was 0.3%, 95% confidence interval –1.8% to 2.4%. What is meant by 'difference in risk was 0.3%' and '95% confidence interval –1.8% to 2.4%'?* The risk is the proportion of women who experience persistent occiput position. The 'difference in risk was 0.3%' means that the difference between the proportions in the treatment groups, intervention minus control, was 0.3 percentage points. The 95% confidence interval means that we estimate from the sample that in the population from which the trial subjects were drawn the difference in percentages experiencing persistent occiput position is between 1.8 percentage points lower for women given the intervention and 2.4 percentage points higher, than for women not given the intervention. This is calculated in such a way that 95% of samples would give confidence intervals including the difference for the population

3. *The authors say that '. . . we found that nulliparity was associated with an increased risk of occiput posterior position at birth (odds ratio 2.5, 95% confidence interval 1.9 to 3.3)'. What is an odds ratio and how can 2.5 be interpreted?* Odds is the ratio of the number of women with occiput posterior position at birth divided by the number of women without occiput posterior position at birth. The odds ratio is the odds for nulliparous women divided by the odds for multiparous women. Odds ratio 2.5 means that the number of nulliparous women with occiput posterior position at birth for every nulliparous woman who does not have occiput posterior position at birth is 2.5 times the number of multiparous women with occiput posterior position at birth for every multiparous woman who does not have occiput posterior position at birth.

4. *Explain why the odds ratio 2.5 is not in the centre of its confidence interval 1.9 to 3.3.* The confidence interval for the odds ratio was calculated on the logarithmic scale. Log odds follows an approximately Normal distribution and has a standard error which is easily calculated. The 95% confidence interval for the log odds is found using this and is symmetrical. We then antilog to get to the natural scale. This makes the confidence interval asymmetrical.

5. *Adjustment for parity was done using logistic regression (Methods). What is 'logistic regression' and why was it used here?* Logistic regression is a method of prediction used when the outcome variable is dichotomous. We want to predict the probability that a woman would experience occiput posterior position from treatment and parity. We do this using the log odds as the outcome variable, because this can take any value. It is used here because the outcome, presence of occiput posterior position, is dichotomous.

6. *In the Table, mode of delivery is shown using five categories. What method of analysis should be used to test the null hypothesis that mode of delivery is unaffected by the exercise?* We have two categorical variables and there is no particular order to the categories. We should therefore use a chi-squared test for a contingency table. The numbers are large so the condition for this test to be valid will be met.

7. *What can we deduce about the distribution of the duration of labour (Table) and why?* The standard deviation is greater than half the mean. This tells us that the distribution must be skew to the right, otherwise we would expect to get observations less than the mean minus two standard deviations, and therefore less than zero. This is impossible for duration of labour, which cannot be negative.

8. *What method should we use to estimate the confidence interval for the difference in duration of labour (Table), and why?* This is a comparison of two means of a continuous variable. The sample is very large, so we can use the large sample Normal or z method.

9. *The authors say 'We designed the study to have an 80% power to detect a clinically significant 50% reduction in fetal occiput posterior position at delivery from 5% to 2.5% by using a two sided method with α set at 0.05'. What is meant by 'two sided' here and why did they do this?* 'Two sided' means that when doing significance tests we consider the probabilities of extreme observations both above and below the null hypothesis value. This is because were our intervention to cause harm we would want our statistical method to detect it.

10. *The authors conclude that the trial did not support the intervention of hands and knees posturing with pelvic rocking exercise for achieving spontaneous rotation from occiput posterior to occiput anterior position. What evidence is there for this conclusion and do you agree?* This is a randomised trial so the groups should be comparable. The trial is large and the confidence interval for the difference is narrow. At most the intervention could only reduce the risk of fetal occiput posterior position by 1.8 percentage points from 8%. The groups appear very similar on all outcome measurements. Inevitably the trial could not be blind but the outcome is objective and unlikely to be influenced by psychological factors. The conclusion seems well founded.

11. *In the section on effect of intervention, the authors wrote 'Overall, 320 (95%) women in the intervention group breast fed initially compared with 324 (96%) in the control group (relative risk 0.99, 95% confidence interval 0.84 to 1.16, P = 0.44; table 3).' What is meant by 'risk' and 'relative risk'? What conclusions could we draw from this analysis?* The risk of an event is equal to the proportion of individuals in a particular category who will experience it, or the probability that an individual chosen at random will experience it. The relative risk is the ratio of the risk of breast feeding among women allocated to the intervention by counselling to the risk among women allocated to control. We estimate that the risk of breast feeding may be increased by up to 16% or reduced by as much as 16%, but there is no evidence that the intervention has any effect on breast feeding at all.

12. *In the section on association between counselling uptake and feeding behaviour, the authors wrote 'At six weeks, 76% (51/67) of those visited were still breast feeding compared with 64% (92/143) of those who telephoned and 60% (75/126) of those not in contact (χ^2 for trend = 4.89, P = 0.027).' What is a chi-squared test for trend and what is being tested here? Why is the number of degrees of freedom not given for this chi-squared test? What could we conclude?* The chi-squared test for trend is used when the categories in a contingency table are ordered. It is a test of the null hypothesis that there is no linear relationship between the proportion of women breast feeding and the degree of contact. This will be coded in equal intervals, e.g. 1 for visit, 2 for telephone, and 3 for no contact. The degrees of freedom is always one, so there is no need to give it. We can conclude that there is good evidence that the proportion breast feeding increased as the amount of contact increased. Whether this is because those who wanted to breast feed were more likely to maintain contact or because contact helped to maintain breast feeding we cannot say.

13. *In the section on effect of intervention, the authors wrote 'Kaplan-Meier survival analysis confirmed that the duration of breast feeding was not significantly different between the women in the intervention and control groups (median 110 days v 96 days; log rank statistic 0.58; P = 0.445).' What is 'Kaplan-Meier survival analysis' and why was it used here? What might it add to the comparisons at six weeks and four months?* Kaplan-Meier survival analysis is used for data which take the form of time to an event, here ceasing breast feeding. Not all women have to be followed to the event. Here, women stop breast feeding at different times after birth and some will not be followed up until the moment when breast feeding ceased, either because they broke of contact with the study or they were still breast feeding at the analysis point. The method uses the full data to estimate the proportion of women still breast feeding at each time after birth. The analysis takes into account all the follow-up, not just the follow-up up to a particular time such as six weeks, and also enables us to use data from women who drop out. It thus uses the data more efficiently than an analysis at a given time after birth.

14. *The authors used Cox regression to compare the association between group allocation and feeding duration taking intention into account. The estimated hazard ratio (chance of stopping breast feeding in intervention group to chance of stopping in control group) was 0.893 (0.717 to 1.112) when intention was not taken into account and 0.886 (0.712 to 1.104) when it was. What is a hazard ratio and what can we conclude from it here? Why was Cox regression used?* The hazard is the instantaneous rate at which events happen, stopping breast feeding in this study. It usually depends on time since the start of the follow-up, which is birth here. A hazard ratio is the ratio of the hazards in two groups. Here it is the ratio of the rate of stopping breast feeding in the intervention group to the rate in the control group. We assume the hazard ratio is constant over time, even though the hazards themselves may vary. Cox regression or proportional hazards regression predicts the hazard ratio from one or more predictor variables; in this study these are the randomised group and whether the mother intended to breast feed. It was used to see whether adjusting for intent altered the hazard rate, because intent to breast feed was different in the two randomised groups.

15. *Do the data support the conclusions given in the abstract? How could the conclusions be improved?* The conclusions are that **w**omen valued the support of a counselor in breast feeding, but the intervention did not significantly increase breast feeding rates, perhaps because some women did not ask for help. There is good evidence for the conclusion that **w**omen valued the support of a counselor in breast feeding, as 73% found her very helpful. The conclusion that the intervention did not significantly increase breast feeding is supported, as the rates are not significantly different at the beginning, after 6 weeks, or after 4 months, and over the entire follow-up using survival analysis. The conclusion that this is perhaps because some women did not ask for help is not supported. Those who did not ask for help were less likely to breast feed, but they are a minority of the intervention group and breast feeding rates are very similar in the two groups. The conclusions would be improved by drawing a conclusion about the population, rather than the sample, which is was 'not significant' is. They could conclude, for example, that after six weeks the proportion of intervention women breast feeding is estimated to be between 0.84 and 1.24 times that in the absence of the intervention.