Clinical Biostatistics

Specimen answers to Examination, Tuesday 20th June 2006

Questions 1 to 6 are about the paper 'Randomised placebo controlled multicentre trial to assess short term clarithromycin for patients with stable coronary heart disease: CLARICOR trial'.

1. *The authors report (results, paragraph 1) that all tablets were reported taken by 90.0% (1954 patients) in the clarithromycin arm and 93.7% (2061) in the placebo arm, P < 0.0001. What do they mean by 'P<0.0001' and what can we conclude from this?* This is the result of a significance test. It is asking whether the data are consistent with the null hypothesis that, in the population from which these patients come, the proportion of subjects taking all tablets would be the same for clarithromycin and placebo. P is the probability of observing a difference as far from zero as this if the null hypothesis were true. We can conclude that there is very good evidence that the difference in the population is not zero.

2. *In the second paragraph of the results, the authors say 'The primary outcome (all cause mortality or non-fatal cardiac outcomes) did not differ significantly between the clarithromycin and placebo arms (15.8% v 13.8%; hazard ratio 1.15, 95% confidence interval 0.99 to 1.34; P = 0.08).' What is meant by 'primary outcome' and why is it important to identify one? When should this be done?* When we run a clinical trial, there may be many possible outcome variables which we might wish to compare between treatment groups. If we were to carry out significance tests for all of these, the probability that at least one will be significant even if all the null hypotheses are true is much greater than the nominal 0.05. To avoid this problem, we choose a primary outcome variable. If there is no significant difference for this variable, there is not good evidence that the treatment is effective, whatever the other outcome variables are like. Having a primary outcome variable avoids the problem of multiple testing. The primary variable should be chosen preferably at the trial design stage and always before any data are analysed.

3. *What is meant by 'did not differ significantly' and what should we conclude about the risk of the primary outcome following clarithromycin?* This means that the difference is not significant, because the probability is greater than the convention cut-off value of 0.05. We conclude that there is insufficient evidence of a difference between the treatments in the population. As the P value is between 0.05 and 0.1, we could conclude that there is weak evidence, but not enough on its own to say that there is a difference. We should ***never*** conclude that there is not difference.

4. *What is a hazard ratio and how can we interpret a hazard ratio = 1.15?* A hazard is the rate at which events happen, so that the probability of an event happening in a short time interval is the length of time multiplied by the hazard. Although the hazard may vary with time, we assume that the hazard in one group is a constant multiple of the hazard in the other group. This multiple is the hazard ratio. Here the hazard in the clarithromycin group is 1.15 times the hazard in the placebo group. Hence we conclude that, at any given time, patients given clarithromycin had 1.15 times the risk of an event as did patients given placebo.

5. *Figure 2 shows a Kaplan Meier estimate. What is this? What would you conclude from Figure 2?* A Kaplan Meier estimate is an estimate of event-free survival rate, so that at any given follow-up time we estimate the proportion of subjects who would not yet have experienced an event. The Kaplan Meier estimate allows for some subjects not having been followed until the event and for some of these censored patients having shorter follow-up than others who experience an event. At each time where an event takes place, we calculate the proportion of the subjects followed to that time who do not have an event. If we multiply these together cumulatively, we get the proportion who would have remained event-free if all subjects could have been followed to this time. Figure 2 shows the proportions who have had an event, one minus the Kaplan Meier proportion, called the failure function. We could conclude that the event rates appeared similar for the first year then appeared to diverge, the clarithromycin group experiencing more events than the placebo group.

6. *The authors report that the tertiary outcome was significantly more frequent in the clarithromycin arm than in the placebo arm (P = 0.03), the number of non-fatal tertiary outcomes was insignificantly increased by 16% (P = 0.09), all cause mortality was significantly higher (P = 0.03), as was cardiovascular mortality (P = 0.01), and non-cardiovascular mortality and non-classified mortality did not differ significantly between groups. Why should we not conclude from these P values that there is good evidence of increased risk with clarithromycin and what method can we use to examine these P values?* These were not the primary outcome variable. We are likely to find some variables where there is a significant difference even when the treatments are identical. We could adjust the P values using the Bonferroni correction to test the composite null hypothesis that the treatment groups do not differ on any outcome variable. We would count the number of relevant tests and multiply the individual P values by this number. If any were less than 0.05 after this, we would have a significant difference between the groups.

7. *In Table 1, what is meant by 'Difference in percentage points between groups'?* The difference between two percentages should not be reported as a percentage, but as percentage points. For example, the difference between 20% and 15% is not 5%, but 5 percentage points. This is because a difference of 5% could be interpreted as 5% of 20%, i.e. 1 percentage point, the other group being 19%.

8. *For any antidepressant medication, 24 month follow-up, what is '(8.69 to 19.14)' and what can we conclude from it?* This is a 95% confidence interval for the difference in percentage on medication in the population. From the sample, we estimate that in the population the difference is between 8.69 and 19.14 percentage points. A 95% confidence interval is calculated so that for 95% of possible samples the confidence interval would include the difference for the population.

9. *What method would be used for the calculation of the P values in Table 1 and why?* This a comparison of two dichotomous variables between two independent groups. A chi-squared test for a contingency table could be used, because the sample size is large and all the expected frequencies will exceed five.

10. *In Table 1, what is P = 0.7592 testing for the baseline? Is this a meaningful thing to do?* This is testing the null hypothesis that in the population the proportion of patients on antidepressant medication before treatment would be

the same in patients given usual care and patients given the intervention. As the patients were allocated randomly to intervention or usual care, this null hypothesis is true. If the test yielded a significant difference, it would be the one in 20 which arise by chance. The test tells us nothing and is therefore not meaningful.

11. *In Table 2, the 24 month follow-up of the overall quality of life score has mean 6.08 and standard deviation 2.22 for the usual care group. What do these numbers mean and how can we interpret them?* The mean is the average value, found by adding all the scores and dividing by the number of subjects. The standard deviation is a measure of variation. We subtract the mean from each score, square and add. We divide by the degrees of freedom, the number of observations minus one, to get the estimate of the variance. The square root of the variance is the standard deviation. The majority of observations will be within one standard deviation of the mean and nearly all (about 95%) within two standard deviations of the mean.

12. *In Table 2, the 24 month follow-up of the overall quality of life score has a 95% confidence interval for the difference (0.03 to 0.49), P = 0.0296. What method would be used to calculate these and why?* We could use the large sample Normal distribution or *z* method for the difference between two means. This is because the samples are large, much greater than 50 in each group, and the subjects are independent. The authors in their method section refer to the two sample t method, but this would be equivalent in such a large sample.