Clinical Biostatistics

Suggested Answers: Folate supplementation

a) *This trial is described as 'randomised' (Comment). Was it and to what problems might the method of allocation used lead?* We are told that 'The tablets were kept in numbered drawers and distributed in sequence.' This suggests that we have sequential allocation, not random. It means that the person recruiting subjects to the trial knew in advance what treatment the next subject would receive. This may lead to recruitment bias, where subjects recruited to active treatment differ in some way from subjects recruited to placebo.

b) *The trial is described as 'double blind'? What features of the design suggest that this is not true?* Each treatment was in tablets of a different colour. Any researcher seeing the tablets would know what treatment the subject was on. Also, because of the sequential allocation, the recruiter would know what each subject was on.

c) *What other feature of the reported trial design suggests that we are not being given a full account of this trial?* There were 4 treatment groups receiving placebo and two receiving different doses of folate. Hence there were four times as many on placebo, which was in four different colours, as on either active treatment. It is difficult to believe that these four 'placebo' groups all received the same treatment.

d) *For 5 mg folate and all mortality, the hazard ratio was 1.20 (Table). What is a hazard ratio, and what does 1.20 tell us? Why was this method used?* Hazard ratios are used in survival analysis. The hazard means the instantaneous risk of experiencing an event and changes over time. Here an event means a death. The hazard ratio is the ratio of the hazard in the folate group to the hazard in the control group. We have to assume that this is a constant independent of time, called the assumption of proportional hazards. 1.20 tells us that women given 5mg folate had a risk of death at any time 1.20 times that for women in the control group. Survival analysis is used because we are dealing with the time to an event, not everyone experiences the event during the follow-up of the study (i.e. some subjects are still alive) and because subjects were not all recruited at the same time follow-up time is not the same for everyone.

e) *For 5 mg folate and all mortality, the 95% confidence interval for the hazard ratio was 0.84 to 1.71 (Table). What is a confidence interval and what does '0.84 to 1.71' tell us?* We are using the women in this trial as a sample from which we wish to estimate the hazard ratio for the whole population which these women represent. The hazard ratio in the sample will not be exactly the same as the hazard ratio in the population, due to sampling error. We produce an interval estimate, a range of values which we estimate will include the population hazard ratio. The interval is chosen so that intervals calculated from 95% of possible samples would include the population value. We estimate that in the population from which these women come, the risk of death for women taking folate would be between 0.84 and 1.71 times the risk for women not taking folate.

f) *For 5 mg folate and all mortality, the P value for the hazard ratio was 0.33 (Table). What is a P value and what does '0.33' tell us?* The is the result of a significance test. It is the probability of getting data as far from what we would expect, if there were no difference in risk of death between women taking folate and women not taking folate, as the data which we have observed. The probability of getting a hazard ratio as far from 1.00 or further as is 1.20 is 0.33. This is a fairly large probability, much larger than 0.05, so the data are consistent with the null hypothesis of no difference. There is little or no evidence for a difference in mortality in the population.

g) *For 5 mg folate and all mortality, why is the hazard ratio, 1.20, not in the middle of its confidence interval, 0.84 to 1.71? As the confidence interval does not include zero, is it possible for the difference between folate and placebo to be not significant?* The hazard ratio is a ratio, so the value it would have if there were no difference in mortality is 1.00. This is the value the hazard ratio would have if the null hypothesis were true. 1.00 is inside the 95% confidence interval, so the difference is not significant.

h) *What is meant by 'adjusted hazard ratio' (Table)? What method was used to adjust it?* Survival and risk of death may be influenced by several variables. We wish to allow for these in the analysis by a regression method. This reduces the variability between subjects and corrects for chance imbalances between the randomised groups. Here they have adjusted for several variables which may be related to the risk of early death, such as cigarette smoking. We adjust the hazard ratio for the effects of these other variables by Cox regression, also called proportional hazards regression.

i) *What assumptions about the data must we make for the adjustment analysis? Which one is clearly violated here?* We must assume that the risk of an event is the same for censored subjects as for non-censored subjects, i.e. that anyone not followed up to death had the same risks as those who were followed up to death. We must assume that the proportional hazards model applies, i.e. that the ratio of the risk of death in the folate group to the risk of death in the placebo group is the same at all times. There must be sufficient data for the method of fitting the coefficients and the large sample z tests and confidence intervals to be valid. A rule of thumb is that there should be at least 10 events per variable, preferably 20. We have eight adjusting variables: maternal age, smoking, height, weight, social class, systolic blood pressure, parity, and gestational age. The treatment has three categories so requires two dummy variables, 10 variables in all. We need at least 100 events and this is not true for breast cancer (29) or for cardiovascular mortality (40).

k) *What feature of the significance tests in this trial should make us wary? What could we do about this?* There are a lot of significance tests, 32 to be precise. This means that we might expect to get significant differences even if the null hypotheses are all true and folate has no effect on the risk of death. If we expect one in 20 tests to be significant, we would expect one or two to be significant out of 32 tests. These tests are not independent, so these calculations do not apply exactly, but they are a guide. We have one significant difference, which could be spurious. We could use the Bonferroni correction and multiply all the P values by the number of tests. The smallest P value is 0.02 and $0.02 \times 32 = 0.64$ which is not significant. We have no way of knowing how many other causes of death were analysed and not reported.

l) *How good is the evidence from this paper that taking folates increases the risk of breast cancer?* Very weak. Although the hazard ratio for breast cancer mortality is the largest at 2.02, for 5mg folate versus placebo after adjustment, the effect is not significant (P=0.10). When we allow for the many significance tests there is no good evidence for any effect of folate on mortality from any cause. In addition, there are features of the trial design which may mean that the groups were not comparable in first place.