

## Clinical Biostatistics

### Suggested Answers to revision exercise: Trial of physiotherapy for low back pain

1. *'Responders were older (mean age 43 (SD 15) v 37 (13);  $P < 0.001$ ), less likely to smoke (39 (22%) v 50 (47%);  $P < 0.001$ ), . . .'* (Results). What methods mentioned in the Methods section could they have used to calculate these two  $P$  values and why? (N.B. the two  $P$  values were calculated using different methods.) Age is a continuous variable so we could use an independent t test (or two sample or two group or unpaired t test). The standard deviations are similar so the pooled variance form would be suitable. If we did not think that the distributions were Normal, we could use the Mann-Whitney U test. Either answer would be acceptable. Whether the subject smokes is a categorical variable, so we could use a chi-squared test for a contingency table. (We could also use the large sample Normal method for the age and Fisher's exact test for the smoking, but these are not mentioned in the Methods section of the paper.)
2. *'Responders were older (mean age 43 (SD 15) v 37 (13);  $P < 0.001$ ), less likely to smoke (39 (22%) v 50 (47%);  $P < 0.001$ ), . . .'* (Results). What is a  $P$  value and what do these  $P$  values tell us? A  $P$  value is the result of a significance test. It enables us to assess the strength of the evidence against the null hypothesis of no difference. It is the probability of getting data as different as are those we have observed from what we would expect if the null hypothesis were true. If the  $P$  value is small, which we usually take to be less than 0.05, there is evidence against the null hypothesis. Here the  $P$  values are very small and we have very strong evidence against the null hypothesis for each. We would conclude that there is very strong evidence that responders had a greater mean age and were less likely to smoke than non-responders.
3. *'We found no differences between the groups in change in scores on the Oswestry disability index at 12 months (mean difference  $-1.04$ , 95% confidence interval  $-3.7$  to  $1.59$ ) . . .'* (Results and Table 2). What is a confidence interval and what does this one tell us? Using a sample, we are trying to estimate the difference in change in disability index in this population as it would be between patients treated by physiotherapy and patients treated by advice. A confidence interval is a range of values within which we estimate a population value to lie. It is calculated so that 95% of confidence intervals will include the population value. We estimate that in the population the mean change in disability index is between 3.70 points less and 1.59 points greater in patients treated by physiotherapy. (A negative change means that the disability has reduced, so the trial sample patients did better on physiotherapy.)
4. *The authors say that they found no differences between the groups in change in scores on the Oswestry disability index. What do they mean by this and is this a reasonable statement?* They clearly did find a difference in their sample. They mean that they found no evidence for a difference in the population. This is correct, because the difference is not significant and the confidence interval includes zero. They cannot conclude that there is no difference, however, only that any difference which exists is small, between  $-3.7$  to  $1.59$  points.

5. *In Table 2, the authors quote effects sizes. What do they mean by this and why is it a useful thing to do for a disability index?* The effect size means the difference divided by the standard deviation of the variable, e.g. for disability index after 2 months the effect size, 0.24, = the mean, -2.65, divided by the standard deviation of the baseline measurement, 11 (Table 1). This is useful because the disability index is not measured in any recognised units and we can judge better by how much the patient's disability fell. Although the index is a 0 to 100 scale (Outcome measures) these patients will not use all of it.
6. *In Table 2, the authors give P values for a disability index score adjusted for score at baseline and several other variables. What does 'adjusted' mean here and why did they do it?* 'Adjusted' means that regression analysis has been used to estimate what the difference would be if all subjects had the same values for baseline score and for age, sex, smoking status and time since first episode of pain. This reduces the variability in the outcome variable, disability index, and so increases the power of the study and the precision of the estimated treatment difference. It also adjusts for any chance imbalance between the treatment groups in these variables.
7. *In Tables 2 and 3, the authors give P values for analysis of covariance. What is 'analysis of covariance' and what assumptions must we make about the data?* 'Analysis of covariance' is a name given to multiple regression analysis when there are both categorical and continuous predictor variables. We must assume that the deviations from the regression estimates follow a Normal distribution and have uniform variance.
8. *In Table 4, the authors give relative risks for perceived benefit. What is a 'relative risk' and what does a relative risk of benefit of 1.28 tell us?* The risk of benefit in a group is the proportion of patients in that group who report benefit. The relative risk is the proportion reporting benefit in the physiotherapy group divided by the proportion reporting benefit in the advice group. We estimate that the proportion reporting benefit in the physiotherapy group is 1.28 times the proportion reporting benefit in the advice group.
9. *In Table 4, the authors give P values for perceived benefit versus no perceived benefit, adjusted using logistic regression. What is 'logistic regression' and why was it used here?* 'Logistic regression' is a method for predicting a proportion, or for predicting the probability of reporting benefit, from a set of predicting variables. Because proportions must lie between zero and one, we use the log odds as the outcome variable. This can take any value. It was used because the outcome variable, reporting benefit, is dichotomous.
10. *The authors conclude that 'Routine physiotherapy seemed to be no more effective than one session of assessment and advice from a physiotherapist.' (Abstract). Do the data support this conclusion?* The improvement is at most 3.7 points on the disability index. As this has a range 0 to 100 (Outcome measures) the difference is small. The gain is therefore small. It is also only a third of a standard deviation of the baseline disability index. Patients were more likely to think they benefited from the physiotherapy than the advice and had better mental health scores on SF-36 after 2 months, so they felt better even if they were not much better physically.