Solutions

Exercise 1

(1) (i) T. The main purpose of cluster trials is to minimise contamination, (ii) F. Usually an individually randomised trial with the parameters will have more power, (iii) Often one cannot blind patients to the intervention in a cluster trial, (iv) Often it is more effective (or the only method) to deliver the intervention through clusters.

(1) (i) F. Usually field trials have fewer, but larger clusters than cohort trials, (ii) Field trials may be known as cluster–cluster trials since both the randomisation and analysis are at the cluster level, (iii) F. Usually field trials obtain their subjects by randomly sampling from the population, (iv) T. Because they use random sampling they may be more representative of the population.

Exercise 2

Nourhashemi *et al*.

(1) Yes, (2) Longitudinal unmatched, (3) Memory clinic, (4) 50 clinics, (5) 1131 patients, (6) A comprehensive specific care plan, (7) Change in the Alzheimer's Disease Cooperative Study activities of daily living at 12 and 24 months, (8) Yes, 12 and 24 months, (9) 24 months.

Skarbinski *et al*.

(1) Yes, (2) Cross-sectional unmatched, (3) Health facility, (4) 60 health facilities, (5) 1540, (6) Malaria rapid diagnostic tests + TGS versus TGS alone, (7) Recommended treatment and overtreatment, (8) Yes.

How to Design, Analyse and Report Cluster Randomised Trials in Medicine and Health Related Research, First Edition. Michael J. Campbell and Stephen J. Walters.

Companion website: http://www.wiley.com/go/cluster_randomised

Steill *et al*.

(1) Yes, (2) Cross-sectional matched, (3) University and community emergency departments in Canada, (4) 12, (5) 11 824, (6) Active strategies to implement the Canadian C-Spine Rule, (7) Diagnostic imaging rate of the cervical spine, (8) No, (9) 2 years.

Mermin *et al*.

(1) No – individuals were randomised, (2) three-arm parallel group, (3) People with HIV who were members of the Aids Support Organisation, (4) 1094, (5) 1094, (6) Three different monitoring arms: a viral load arm (clinical monitoring, quarterly CD4 counts and viral load measurements), CD4 arm (clinical monitoring and CD4 counts) or clinical arm (clinical monitoring alone), (7) Serious morbidity (newly diagnosed AIDS defining illness) and mortality, (8) Yes.

Exercise 3

(1) (i) T. $(1 + (10 - 1) \times 0.1)$, (ii) T. $(0.1 = 10\%)$, (iii) F. It is only when the design effect is small that intracluster correlation can be ignored), (iv) T. It is often assumed that the design effect is the same in the intervention and control arms, but it is possible that the intervention may affect the ICC, and this could be checked by computing the ICC separately for the two arms.

(2) A – We do not need the median values – we only need the mean difference.

(3) E – Power is the probability of not making a Type II error.

(4) E – The outcome is binary so we do not need the standard deviation.

(5) B – If you increase the power, then you increase the sample size.

Exercise 4

(i) T.

(ii) F – if the clusters are of different size, we need to find the weights and these depend on the ICC.

(iii) F – we could use the baseline summary measures in an analysis of covariance.

(iv) T.

Exercise 5

(i) T. The standard error can be made 'robust' to clustering, (ii) F. The population-averaged model treats the clustering as a 'nuisance' whereas the cluster-specific model tries to model it, (iii) F. A random-effects model is also a cluster-specific model, (iv) T. With continuous outcomes the random-effects and population-averaged models give similar estimates.

Exercise 6

(i) F. This is explained in Tables 6.1 and 6.2, (ii) T. This is also explained in Tables 6.1 and 6.2, (iii) T, (iv) T. Relative risks can be misleading when absolute risks are small.

Exercise 7

(1) D – The trial protocol which is written before the data are collected will not contain any study results.

(2) D – We do not need to know the correlation coefficient but do need to know the ICC.

(3) E – We need to know the variance or standard deviation of the outcome, 18-month HBA1c level, which is not reported.

(4) D – The outcome, 18-month HBA1c level, is continuous so a marginal or population-averaged linear regression model is appropriate to analyse the data.

(5) A – the outcome, 18-month HBA1c level, is now binary so a random-effects logistic regression model is appropriate to analyse the data.

Exercise 8

(1) The design is a cohort unmatched cluster randomised controlled trial.

(2) It is single blind (p. 2 bottom of column 1). The people who were blind were 'specially trained researchers measured outcome and were blind to group allocation'.

(3) The unit of randomisation is class.

(4) 40 classes.

(5) 652 children.

(6) Opaque envelopes.

(7) 30 schools. School should be allowed for and were (see p. 4 bottom of first column).

(8) Aerobic fitness after one school year.

(9) $\text{DE} = 1 + (13 - 1) \times 0.06 = 1.72$.

(10) From formula 3.13, amended for 90% power $n = 2 \times 21/0.5^2 = 168$ in total.

(11) $168 \times 1.72 = 289$ children. With 13 children per cluster (class) $289/13 = 22.2$ and rounding up to the nearest even number means 24 clusters are required (12 per group). This does not correspond with the 40 clusters that the authors suggest.

(12) $\text{ICC} = \left(\frac{\sigma_B^2}{\sigma_B^2+\sigma_W^2}\right) = \frac{0.25^2\sigma_W^2}{0.25^2\sigma_W^2+\sigma_W^2} = \frac{1}{1+16} = \frac{1}{17} = 0.0588.$

(13) The methods used are mixed models (linear and logistic regressions).

(14) Yes, groups are comparable (see second para second column p. 4).

(15) (a) Aerobic fitness was assessed by the 20-m shuttle run test, where children run back and forth for 20 m with an initial running speed of $8.0\,\mathrm{km\,h^{-1}}$ and a progressive $0.5\,\mathrm{km\,h^{-1}}$ increase in the running speed every minute. Results were expressed as stages; one stage corresponding to the running time of 1 min. So 0.32 is the difference, in stages run on the shuttle test, between intervention and control at follow-up, after allowing for age, sex, language region and class and is approximately the same as the unadjusted observed mean difference of $4.6 - 4.3 = 0.3$ after the intervention.

(b) We are 95% confident that true estimate lies between 0.07 and 0.57.

(c) The probability of getting this result (or one more extreme) under the null hypothesis, of no difference in outcome between the intervention and control groups, is 0.01. The ratio of the between-groups variance to the total variance is 0.07.

(d) Outcome SD = 1.7 in both the intervention and control groups. Thus, the standardised effect size is $0.32/1.7 = 0.2$. This is much smaller than the standardised effect size of 0.5, which was the prior estimate. The study is in fact much bigger than planned, but perhaps fortunately so.

(e) $\sigma_B^2/1.7^2 = 0.07$. Thus, $\sigma_B^2 = 0.20$ and $\sigma_W^2 = 1.7^2 - 0.20 = 2.69$. Thus, the ratio of SDs is $\sqrt{(0.2/2.69)}$=0.27, which is pretty close to the assumption of 0.25.

(f) The odds ratio of 0.65 implies that the odds of children being overweight in the intervention group is 0.65 times the odds of being overweight in the control group, that is, children are less likely to be overweight in the intervention compared to the control group at follow-up.

(g) The crude-unadjusted odds ratio estimate is $11.0/(100 - 11.0)/\{14.9/(100 - 14.9)\} = 0.70$. After allowing for baseline covariates, the adjusted estimate is 0.65. Since the BMI is based on ≥90th centile one might have expected 10% in both groups (i.e. intervention not overweight and control slightly overweight).

(16) There are 342 children in intervention compared to 310 children in control. This could have arisen by chance ($X^2 = 1.57$, $P > 0.10$), but there is some slight evidence of recruitment bias. The percentages with aerobic fitness measurements at both measurements 302/342 = 88% and 269/310 = 87% are very similar. Thus, there is no evidence of selective dropout.