Reviewer's report

Title: Sample size calculations for cluster randomised trials with a fixed number of clusters

Version: 1 Date: 8 March 2011

Reviewer: Monica Taljaard

Reviewer's report:

The authors provide practical advice for planning cluster randomized trials when the number of clusters available for randomization is limited. This situation frequently arises in health services research where the number of units (e.g., medical practices, health professionals or geographic regions) that can feasibly be randomized is fixed, but there is some flexibility in increasing the number of subjects that can be enrolled within each cluster. It is well-known that there is a greater relative effect on power of increasing the number of clusters versus increasing the cluster size, and thus, existing sample size formulae for cluster trials are most easily used to calculate the required number of clusters given a specified average cluster size. This paper provides useful formulae that will allow investigators to quickly determine whether a study is feasible given the fixed number of clusters, the assumed intracluster correlation coefficient and the desired effect size. The authors then provide formulas that allow the investigator to determine what the feasible minimum effect size (or maximum achievable power) is. Although some of the results in this paper have been previously published (e.g., Donner and Klar, 2000, Design and analysis of cluster randomization trials in health research), the authors present different versions of these formulas to determine the required number of clusters, detectable difference and achievable power, and provide practical guidance that is likely to be of great interest to statisticians and cluster randomization trialists. However, there are a few errors and omissions in the paper. Once corrected, I believe this paper will be a great addition to the cluster trials literature.

Major Compulsory Revisions

1. The study design assumptions underlying the sample size calculation formulas are not presented anywhere in the paper. For example, the authors should state that their formulas apply to a two-arm, parallel design.

2. Related to (1), it will be helpful to the reader to explain if/how the formulas can be applied / modified to be used in other designs, such as stratified designs or parallel designs with more than two arms.

3. The underlying statistical assumptions, including the null and alternative hypotheses are not presented in the paper, e.g., “The response Yi, i=1,2 is assumed to be normally distributed with unknown mean μi and common variance σ^2, and the aim is to test the hypothesis that H0: μ1=μ2, at the
two-sided 100# percent level of significance with power 1-beta.” Similar assumptions should be stated for the dichotomous outcome.

4. The authors use \( z_{2*beta} \) to refer to the upper 100\( \times \)2\( \times \)beta standard normal percentile corresponding to power = 1-beta. I believe this is an error in notation and should be \( z_{beta} \). This error occurs throughout the manuscript, i.e., formula #1, unnumbered formula above #2, formula #3, #7, #9, unnumbered formula below 9, as well as formula #11.

5. The authors interchangeably use either a greek letter or English notation to refer to the effect size. Why the difference in notation?

6. Formula #6 is not correct, as the variance of the mean difference, for example, is \( [(2*sigma^2)/km] \times VIF \), rather than \( sigma^2 \times VIF \) as presented in the paper.

7. In formula 8, one cluster is added to each arm, but this is not explained in the paper. I believe the authors may be referring to the suggestion by Snedecor and Cochran to add one cluster per intervention group when a 5% significance level is used, or 2 per group assuming \( \# = 1\% \), in order to account for using critical values from the normal rather than the t-distribution. If the authors want to make this addition to their formulas, this should be explained and then used consistently throughout the paper.

8. Many cluster randomized trials include only a small number of clusters which limits the degrees of freedom for conducting hypothesis tests. Yet, the sample size calculation formulas presented here use critical values from the normal distribution. The authors should add that the use of the sample size calculation formulas presented in this paper will underestimate the required sample size unless the degrees of freedom are large, i.e., unless the total number of clusters is <30.

9. In several places of the manuscript, the authors use the words “making allowance for equal cluster sizes”. This is an odd choice of words, as the formulas are in fact assuming equal cluster sizes, and hence, NOT making allowance for Unequal cluster sizes. The authors should mention the implication of violation of this assumption of equal cluster sizes, which is that the formulas are likely to underestimate the required sample size, especially if the variation among the cluster sizes is large. Please see: “Planning a cluster randomized trial with unequal cluster sizes: practical issues involving continuous outcomes”, Lydia Guittet, Philippe Ravaud and Bruno Giraudieu, BMC Medical Research Methodology 2006, 6:17 doi:10.1186/1471-2288-6-17.

10. The authors’ formula #18 is in fact identical to formula 5.4 provided in the Donner and Klar textbook (page 61) (except for the addition of 1, which needs to be explained or deleted). Likewise, formula #19 is identical to the guideline provided in the Donner and Klar textbook on page 61. The authors should state this in the manuscript.

11. Point number (b) in Paragraph: “CRCT with limited clusters: Practical Advice”. I am concerned that readers may misread the statement that the
available number of clusters per arm “will be sufficient” when the given inequality is satisfied. Likewise, in the example, the statement is made that “20 clusters per arm will provide sufficient power for this design”. The statement is potentially misleading, as whether or not power is sufficient depends on whether or not a sufficient number of subjects can be enrolled per cluster. Please choose the wording more carefully, e.g., explain that this “power MAY be sufficient, given a large enough cluster size”.

12. Last paragraph in “CRCT with limited clusters: Practical Advice”. The statement: “where there are just 10 or 20 clusters available, power will be limited to about 50% or less”. This is a strong statement and is overly general as it assumes that the ICC will be greater than 0.05. However, many large cluster randomized trials have ICCs much smaller than that, e.g., ICCs of 0.001 or 0.01 in community intervention trials are not uncommon.

13. The text description of the results in the Figures is difficult to follow as it does not refer to a specific Figure #.

14. The authors refer to a Stata function –is this function available as an add-on and where can the reader obtain it?

15. Example: “so that the number of clusters per arm is fixed to m=20” should be k=20

16. Example: “Formula at equation 18 shows that under the assumption…” I believe this should be Formula 17 not 18.

17. Discussion: “Results show that with just 20 clusters available per arm...with an estimated intra-cluster correlation of around 0.08”. Why not assume an oft-used value for the ICC of 0.05?

18. The Figures are very difficult to read. The authors should try to increase the size of the Figures, and improve the scaling on the y- and x-axes by adding additional intervals and grid lines, if possible.

19. SES0.5 in Figure 1 is a typo – should be 0.05.

20. Figure 2 is depicting scenarios to detect increases of 0.02 (2%).... Why such small increases? I think few cluster randomised trials would be powered to detect such small increases. Why not choose a more realistic effect size, such as 10%?

21. The legends for the figures are potentially misleading. An improvement would be to add a statement that these are the best case scenarios, or maximum possible values assuming an infinite cluster size. In reality, cluster sizes will be limited by feasibility constraints and so power will often be much lower.

22. Figure 2 assumes baseline proportions ranging from 0 to 0.25. Why did the authors not extend the range to include a proportion of 0.5 as that provides the most conservative estimate for variance?

23. Table 1: last line of the table refers to k=189 and k=146. This should be
m=189 and m=146 according to your notation.

Minor Essential Revisions

24. The use of the term “deflated” when referring to the effect of clustering on the detectable difference is confusing. The term suggests a reduction in the effect size, but the authors must mean that clustering leads to an increase in the minimum detectable effect size. I would consider using a different adjective.

25. Introduction, third paragraph, last sentence: Please insert “an” prior to “increase in breast feeding rates…”

26. Introduction, fourth paragraph, first sentence: Since there could be “participants” at the individual level (e.g., patients) as well as at the cluster level (e.g., health professionals, i.e., the clusters), consider inserting “individual” prior to “participants per arm”.

27. I believe the plural form for “formula” is “formulas” or “formulae”. This incorrect usage of the plural form appears throughout the manuscript.

28. Text immediately above formula #6: Don’t start the sentence with “So that”. Could replace with “Thus,…”

29. Text immediately above formula #7: please delete “to” in the sentence “then to the required sample size”.

30. First sentence under “CRCT of fixed size: Achievable Power”: Please add “per arm” to “fixed sample size”.
   Also, “Equatations” should be equations.

31. CRCT with limited clusters: Minimum Detectable Difference”. I would not start the sentence with “Then”…

32. Text underneath formula #18: replace “making allowances for” with “assuming”.

33. “CRCT with limited clusters: Practical Advise” should be “Advice”.
   Also: “condition is not meet” should be “met”

34. Example: “The minimum detectable difference is 0.012” should be 0.12, i.e., 12%. Also 0.014 should be 0.14.

35. Discussion: “i.e., GP practices” should be e.g., GP practices.
   Also: straight forward should be straightforward

Discretionary Revisions

36. Abstract: Results:
   One of the neat results of this paper is the simple feasibility check to determine whether a study design is feasible. The authors might want to consider adding this simple result to the abstract, if possible.
37. Abstract: Conclusions: I don’t think the conclusion provided in the abstract really adequately captures this study’s contribution. The authors may want to consider revising the abstract to better represent their contributions in terms of simple guidance. This may include reiterating that when the number of clusters is fixed in advance, the study may not be feasible, unless the requirements for effect size or power are relaxed.

38. I think some of the formulas will be easier to read once they are rearranged, e.g., in formulas 1, 8 and 11, the authors could consider placing sigma^2 next to 2. This is more intuitive as it reflects the assumption of equal variances in the two arms, i.e., sigma1^2+sigma2^2= 2*sigma^2 assuming that sigma1^2=sigma2^2. (Again, this assumption has not been stated anywhere in the paper.)

39. The simple “rule” for determining feasibility is currently presented as n_I <k/rho. Since k is fixed, I think the rule is more intuitive as a requirement on the number of clusters, i.e., “the available number of clusters per arm (k) will be sufficient as long as k > rho*n_I”. That is the way I would try to remember this rule in future planned studies in any case!

40. Figures: The authors have chosen a range for the ICC between 0 and 0.1. Why was this choice made? Bigger ICCs are possible and the authors may consider expanding the range for the Figures.

41. Consider changing the order for Figures 3 and 4 to follow continuous first and dichotomous second. Adding additional tick marks and grid lines may be helpful to reader.

Level of interest: An article of importance in its field

Quality of written English: Needs some language corrections before being published

Statistical review: Yes, and I have assessed the statistics in my report.

Declaration of competing interests: I declare that I have no competing interests.