

Ignoring Clustering and Nesting in Cluster Randomized Trials Renders Conclusions Unverifiable [Response to Letter]

Yibeltal Siraneh ¹, Mirkuzie Woldie ², Zewdie Birhanu³

¹Department of Health Policy and Management, Faculty of Public Health, Jimma University, Jimma, Ethiopia; ²Department of Global Health and Population, Harvard T.H. Chan School of Public Health, Boston, MA, USA; ³Department of Health, Behavior and Society, Faculty of Public Health, Jimma University, Jimma, Ethiopia

Correspondence: Yibeltal Siraneh, Email yibeltal_siraneh@yahoo.com

Dear editor

We appreciate the readers' interest and we welcome the insightful comment regarding our study, "Effectiveness of Positive Deviance Approach to Promote Exclusive Breastfeeding Practice: A Cluster Randomized Controlled Trial."¹ Although reaching a consensus may not be possible in this regard, we really appreciate this contribution and we value their perspectives. Worth reflecting on though is their title "Ignoring clustering and nesting in cluster randomized trials renders conclusions unverifiable"² which is judgmental and discouraging to other researchers who handle statistics differently based on current knowledge. Rather, it would be good to reframe their letter heading, "letter to the editor regarding the paper."

Although we understand their perspective, we still believe our sample size estimation and methods of analysis are justifiable and acceptable. That is why the reviewers involved as part of the basic requirement of peer review process who did not raise similar issues. Hence, we would like to provide our response to the letter as outlined below.

As noted in the letter by Siddique et al,² considering number of clusters per condition, average cluster size, and intra-cluster correlation coefficient (ICC)³ are recommended in sample size estimation for a cluster Randomized Controlled Trial (cRCT). However, there are also other ways of sample size calculation considering effect size rather than cluster size if ICC value for the outcome of interest in the study is not available for the study setting and/or unreliable. Therefore, we estimated the sample size considering effect size of 15% and determined the number of individuals to be enrolled in the respective clusters based on where they live. In this method, the threshold probability for rejecting the null hypothesis (Type I error rate) and the probability of failing to reject the null hypothesis under the alternative hypothesis (Type II error rate) were considered to estimate the sample size⁴⁻⁶ that would adjust the overestimation/inflation or underestimation of the confidence interval, statistical significance or type 1 error rate. Otherwise, if clustering and nesting are the main interest of the study and able to proceed with any number of cluster size, considering ICC and related parameters to determine the number of clusters as a sample and to enroll every eligible individuals without restriction who are found within the clusters would need considering your suggestion. As you know, there is no straightforward way of calculating the effective sample size for reaching an accurate conclusion, rather using available alternative ways that would serve to test our hypothesis is acceptable.⁷ There might be a statistical argument related with a cluster study, but no specific rule exists to prohibit use of a different way of sampling.⁸

Regarding clustering and nesting effect consideration during analysis, unweighted and variance-weighted cluster-level analysis, mixed models with degree of freedom correction, and general linear regression model (GLM) with a small-sample correction can maintain the type I error rate at less or equal to 5% in most cases, whereas uncorrected approaches lead to inflated type I error rates. However, these analysis yield low power (<50% in some cases) when fewer than 20 clusters are randomized with failure to reach the expected 80% power.⁹ We found that the generalized linear mixed

modelling approach is the most consistent when few clusters are available. We also found that none of the common analysis methods for such trials were both unbiased and maintain 5% type I error rate in the presence of few clusters such as three to six. This is because if the number of clusters are less than six (or by recommendation less than ten), ICC value would be less than 10% which might not need clustering and nesting analysis (such as multi-level modeling) since the effect would be minimal or nil.¹⁰

In our study, randomization was geographic/community based cluster randomization, the analysis was done at individual level (to see response difference) and arm/group level (to see the effect on proportion difference and relative risk). In determining the magnitude of the intervention effect, the “two arms were compared” in terms of the cumulative proportion of EBF practice, not at each cluster level. Note that the intervention was not provided in group like educational sessions, rather one-to-one counseling and support was provided by the positive deviants. In fact, determining the ICC may be relevant to see how much variation existed to suggest further multi-level modeling/analysis if it is >10%. If one is concerned about cluster level analysis, the experts suggest that the number of clusters should be greater than six to reasonably determine the ICC and to conduct the analysis at cluster level.¹⁰ In fact, the variability within a cluster and between the clusters should be considered during analysis, and the allocation unit is preferred to be the unit of analysis. However, we determined the sample size at individual level (did not account for cluster sampling initially) who were recruited within the randomized clusters. During the effect analysis, adjustment was done for significantly different baseline characteristics since those variables would affect the ICC. We did not consider ICC during sample size determination since we are not able to determine the number of cluster as a sample rather the number of subjects to be enrolled in each cluster. Hence, we did not report ICC others do. It is difficult to define ICC for the time to event data which related with repeated measure of outcome at different time. Adjusting for the intra-cluster correlation is necessary, but it reduces statistical power, and we did not expect much variation among such small number of clusters, rather between the arms. In general, the power is increased more easily by increasing the number of clusters rather than the cluster size unlike ours which could be one of the limitations.

Instead of mean calculation as you suggested, we determined proportion for the dichotomous outcome at the arm level considering cluster level variance is negligible for such small number of clusters. Although such small number of clusters per arm is not advisable for such community based intervention trials with cluster randomization, the analysis could be at the individual or cluster level.¹¹ In addition to this, we followed the recommended method of analysis to determine differences in proportion at arm level; for two independent samples (2 arms) with categorical/nominal outcome chi-square test or fisher-exact test, and for continuous outcomes independent sample *t*-test were used.

Regarding the model used to estimate the relative risk (magnitude of the effect), there are a variety of methods, generally termed multilevel, hierarchical, or random-effects methods that can be used for such analysis. They are based on a wide variety of statistical models, such as the generalized linear mixed model, generalized estimating equations, and hierarchical Bayesian models. These models could be viewed as extensions to the GLM. GLM consists of Univariate, Bivariate, Multi-variate and repeated measures options. The GLM considering repeated measures provides analysis of variance when the same measurement is made several times on each subject or case. Using this procedure, we can test null hypotheses about the effects of both the between-subjects factors and the within-subjects factors that can adjust the variability and over time factors. It is possible to investigate interactions between factors as well as the effects of individual factors. However, GLM has no custom method, with SPSS data analysis tool, to use Log-binomial (for estimates of proportions) and/or poisson regression (for count response), which are the best and direct estimators of Relative Risk (RR). Hence, we calculated RR from the Adjusted Odds Ratio since the outcome is common. In fact, this could be another limitation. The good thing was that confounders were controlled through Multiple Analysis of Covariance (MANCOVA), which are GLM methods blending ANOVA and regression considering the repeated measures at three point in time. Given the impact of the extent of the intra-cluster correlation on the power of the study, the ICC for each primary outcome should also be analyzed or reported to assess the magnitude of the clustering. However, if a cluster level (in our case at arm level) analysis is being undertaken, the concept of the intra-cluster correlation is less relevant as each cluster provides a single data point. In some situations, especially if it is believed that the intervention will significantly affect the ICC (we assumed it might not affect), it will be useful to report for both arms.¹²

In general, CRTs with fewer clusters per arm are not advisable. A useful rule of thumb is increasing the number of clusters offers more statistical power than increasing the number of individuals per cluster, although logistical and economic implications should be considered alongside statistical considerations as it was the case for us. We understand that the appropriate analysis will depend on the study design, the number of clusters, and the number of individuals per cluster. Adjustment for cluster level covariates is straightforward, while adjustment for individual level covariates was also done in our study. Due to the need of buffering zone, and resource constraint to include all the clusters in the town (clusters' eligibility issue), we preferred to calculate the sample size considering effect size at individual and arm level (with participant eligibility criteria) without considering ICC and average cluster size. Besides, there was no evidence of ICC value previously reported for positive deviance approach on the same outcome variable that lead us to use such method of sample size estimation. If we took recommended values of these to calculate the sample size as you suggested, we were unable to follow the cluster sample to be determined since the total clusters in the town are 17, 5 clusters were considered as a buffer zone, and we were able to study only 50% of the remaining due to time and cost. Hence, we preferred to estimate the sample size of subjects to be enrolled within pre-determined and randomized clusters. However, the sample size per cluster was fortunately greater than 10 subjects that helped us to see the difference. After calculating the total sample size per arm/group, we allocated proportionate to the source population size of each cluster. Therefore, we believe that the sample size (n=260) is large enough to test the primary outcome, and it is possible to conduct individual and group based analysis for cluster randomization with the possibility of individual sample size determination. Of course, our study could be also limited due to the small number of clusters.

With regard to the data sharing request, we have a plan to do other publications as the letter rightly states. However, we also question the reasonableness of the request as the Editor-in-Chief said, "The request was part of the basic principle of peer-review process as it was passed through, and he left open the agenda for discussion with the readers." We also question whether such request to re-analyze the data using different methods other than stated in the method of data analysis is reasonable. At the beginning, the readers were commenting as the design of the study and sample size determination is not appropriate and the method of analysis is incorrect or invalid. This implies that they wanted to use different method of analysis without accepting the sample size estimated. Lately, they updated their primary request and implied that they only want to reproduce following the same procedure to test the research hypothesis without stating the reason for reproduction. Most importantly, we learnt lately that the journal's data sharing policy stated "data sharing should always be consistent with the terms of consent signed by participants." This means separate informed consent for "data sharing" is needed. However, we took informed consent to the purpose of our study alone and informed participants about the anonymity without separate mention of sharing the raw dataset to others.

In conclusion, we would like to reiterate that there is no single resource that provides a summary of such methods, rather there are different approaches that are well established for sample size estimation as well as methods of data analysis for cCRTs.^{13,14}

Funding

The authors have not received any financial support for this communication.

Disclosure

The authors report no conflict of interest in this communication.

References

1. Siraneh Y, Woldie M, Birhanu Z. Effectiveness of positive deviance approach to promote exclusive breastfeeding practice: a cluster randomized controlled trial. *Risk Manag Healthc Policy*. 2021;14:3483–3503. doi:10.2147/RMHP.S324762
2. Siddique AB, Jamshidi-Naeini Y, Golzarri-Arroyo L, Allison DB. Ignoring clustering and nesting in cluster randomized trials renders conclusions unverifiable [Letter]. *Risk Manag Healthc Policy*. 2022;15:1895–1896. doi:10.2147/RMHP.S391521
3. Campbell MK, Piaggio G, Elbourne DR, Altman DG. Consort 2010 statement: extension to 82 cluster randomised trials. *BMJ*. 2012;345:e5661. doi:10.1136/bmj.e5661
4. Hulley SB, Cummings SR, Browner WS, Grady D, Newman TB. *Designing Clinical Research: An Epidemiologic Approach*. 4th ed. Philadelphia, PA: Lippincott Williams & Wilkins; 2013:Appendix 6A, 73.

5. Fleiss JL, Tytun A, Ury HK. A simple approximation for calculating sample sizes for comparing independent proportions. *Biometrics*. 1980;36:343–346. doi:10.2307/2529990
6. Chow S-C, Shao J, Wang H. *Sample Size Calculations in Clinical Research*. 2nd ed. Boca Raton: Chapman & Hall/CRC; 2008:Section 3.2.1, 58.
7. Serdar CC, Cihan M, Yücel D, Serdar MA. Sample size, power and effect size revisited: simplified and practical approaches in pre-clinical, clinical and laboratory studies. *Biochem Med*. 2021;31(1):010502. doi:10.11613/BM.2021.010502
8. Wears RL. Advanced statistics: statistical methods for analyzing cluster and cluster-randomized data. *Acad Emerg Med*. 2002;9(4):330–341. doi:10.1197/aemj.9.4.33079
9. Leyrat C, Morgan KE, Leurent B, Kahan BC. Cluster randomized trials with a small number of clusters: which analyses should be used? *Int J Epidemiol*. 2018;47(1):321–331. doi:10.1093/ije/dyx169
10. Barker D, D'Este C, Campbell MJ, et al. Minimum number of clusters and comparison of analysis methods for cross sectional stepped wedge cluster randomised trials with binary outcomes: a simulation study. *Trials*. 2017;18:119. doi:10.1186/s13063-017-1862-2
11. Yudkin PL, Moher M. Putting theory into practice: a cluster randomised trial with a small number of clusters. *Stat Med*. 2001;20:341–349. doi:10.1002/1097-0258(20010215)20:3<341::AID-SIM796>3.0.CO;2-G
12. Campbell MK, Elbourne DR, Altman DG, Altman DG. CONSORT statement: extension to cluster randomised trials. *BMJ*. 2012;345:e5661. doi:10.1136/bmj.e5661
13. Clare Marie Robinson, School of Medicine and Dentistry, Queen Mary University of London, PhD Dissertation; Sample size calculations for cluster randomised trials, with a focus on ordinal outcomes Available from: https://qmro.qmul.ac.uk/xmlui/bitstream/handle/123456789/12913/Robinson_Clare_PhD_Final_160516.pdf?sequence=1&isAllowed=y. Accessed October 25, 2022.
14. Van Breukelen GJ, Candel MJ. Calculating sample sizes for cluster randomized trials: we can keep it simple and efficient! *J Clin Epidemiol*. 2012;65(11):1212–1218. doi:10.1016/j.jclinepi.2012.06.002

Dove Medical Press encourages responsible, free and frank academic debate. The content of the Risk Management and Healthcare Policy 'letters to the editor' section does not necessarily represent the views of Dove Medical Press, its officers, agents, employees, related entities or the Risk Management and Healthcare Policy editors. While all reasonable steps have been taken to confirm the content of each letter, Dove Medical Press accepts no liability in respect of the content of any letter, nor is it responsible for the content and accuracy of any letter to the editor.

Risk Management and Healthcare Policy

Dovepress

Publish your work in this journal

Risk Management and Healthcare Policy is an international, peer-reviewed, open access journal focusing on all aspects of public health, policy, and preventative measures to promote good health and improve morbidity and mortality in the population. The journal welcomes submitted papers covering original research, basic science, clinical & epidemiological studies, reviews and evaluations, guidelines, expert opinion and commentary, case reports and extended reports. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/risk-management-and-healthcare-policy-journal>