



Published in final edited form as:

Contemp Clin Trials. 2009 July ; 30(4): 317–320. doi:10.1016/j.cct.2009.03.007.

Effect of Imbalance and Intracluster Correlation Coefficient in Cluster Randomization Trials with Binary Outcomes When the Available Number of Clusters is Fixed in Advance

Chul Ahn^a, Fan Hu^b, Celette Sugg Skinner^a, and Daniel Ahn^c

^a Department of Clinical Sciences, UT Southwestern Medical Center, Dallas, TX

^b Department of Statistical Science, Southern Methodist University, Dallas, TX

^c Department of Internal Medicine, University of Texas Medical Branch, Galveston, TX

Summary

In some cluster randomization trials, the number of clusters cannot exceed a specified maximum value due to cost constraints or other practical reasons. Donner and Klar [1] provided the sample size formula for the number of subjects required per cluster when the number of clusters cannot exceed a specified maximum value. The sample size formula of Donner and Klar assumes that the number of subjects is the same in each cluster. In practical situations, the number of subjects may be different among clusters. We conducted simulation studies to investigate the effect of the cluster size variability (κ) and the intracluster correlation coefficient (ρ) on the power of the study in which the number of available clusters is fixed in advance. For the balanced case ($\kappa = 1.0$), *i.e.*, equal cluster size among clusters, the sample size formula yielded empirical powers close to the nominal level even when the number of available clusters per group (k^*) is as small as 10. The sample size formula yielded empirical powers close to the nominal level when the number of available clusters per group (k^*) is at least 20 and the imbalance parameter (κ) is at least 0.8. Empirical powers were close to the nominal level when ($\rho \leq 0.02$, $\kappa \geq 0.8$, and $k^* = 10$) or ($\rho \leq 0.02$, $\kappa = 0.6$, and $k^* = 20$).

Keywords

Intracluster Correlation; Sample size; Binary Outcomes; Varying cluster size

1. Introduction

Cluster randomization trials have received increasing attention in the evaluation of non-therapeutic interventions among healthcare researchers. Groups (called ‘clusters’) of subjects, rather than subjects themselves, are randomized to different interventions, and all subjects within a cluster receive the same intervention. In such studies, inferences are often applied at the subject level while randomization is done at the cluster level.

Correspondence should be sent to: Chul Ahn, Ph.D., Department of Clinical Sciences, UT Southwestern Medical Center, 5323 Harry Hines Blvd, E5.506, Dallas, TX 75390-9066.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

The sample size determination in cluster randomized cluster trials has been mainly concentrated on the sample size estimate for the number of clusters required per intervention group. In some cluster randomization trials such as the ongoing GOODNEWS (Genes, Nutrition, Exercise, Wellness, and Spiritual Growth) and the CRIS (Cancer Risk Intake System) trials, the number of clusters cannot exceed a specified maximum value due to practical considerations such as the fixed expense of maintaining staff to collect data and conduct intervention in each cluster. In those trials, the sample size requirement should be provided for the number of subjects per cluster instead of the number of clusters per group. Donner and Klar [1] and Taljaard *et al.* [2] provided the sample size formula for the number of subjects required per cluster assuming the same number of subjects in each cluster. However, cluster sizes vary among clusters in many cluster randomization studies. It is expected that the variation in cluster size will affect the power of the study. We conduct the simulation studies to investigate the effect of the cluster size variability and the intracluster correlation coefficient on the power of the study in which the number of available clusters is fixed in advance.

Section 2 presents the review of the sample size estimates for the cluster size when the number of clusters is fixed in advance, and Section 3 illustrates how the cluster size can be estimated using the example of the ongoing CRIS intervention trial. Simulation studies are presented in Section 4, and the final section is devoted to a discussion.

2. Statistical Methods

Suppose that we are interested in comparing the proportions of responses between two groups using a simple randomization trial. Let p_1 and p_2 be the proportions of responses in groups 1 and 2. With the two-sided significance level of α and power of $1 - \beta$, the required sample size to test $H_0 : p_1 = p_2$ versus $H_1 : p_1 \neq p_2$ is given by the following equation [3, Section 9.1].

$$n = \frac{(z_{1-\alpha/2} + z_{1-\beta}) [p_1(1-p_1) + p_2(1-p_2)]}{(p_1 - p_2)^2}. \quad (1)$$

In cluster randomization trials with an equal cluster size (m), the total number of subjects per group can be obtained by multiplying the standard sample size estimate (n) from a simple randomization trial by the variance inflation factor $[1 + (m-1)\rho]$, where m is the cluster size, and ρ is the intracluster correlation coefficient [1]. Thus, the number of clusters required per group is computed by $k = n[1 + (m-1)\rho]/m$ and the total number of subjects required per group is $km = n[1 + (m-1)\rho]$.

We occasionally encounter situations in which the number of clusters cannot exceed a specified maximum value due to cost constraints or other practical reasons. In those cases, we need to specify the sample size requirement in terms of the number of subjects per cluster instead of the number of clusters. Donner and Klar [1] and Taljaard *et al.* [2] provided the sample size formula for the number of subjects per cluster when there is an upper limit in the number of clusters to be studied in a cluster randomization trial. Since the number of clusters is usually small in community intervention trials, the use of critical values $z_{1-\alpha/2}$ and $z_{1-\beta}$ in Equation (1) instead of critical values $t_{1-\alpha/2}$ and $t_{1-\beta}$ corresponding to the t-distribution underestimates the required sample size. To adjust for underestimation, one cluster per group may be added when the sample size is determined with a 5% level of significance, and two clusters per group with a 1% level of significance [4, p. 104]. In cluster randomization trials with an equal cluster size (m) at the 5% level of significance, the number of clusters per group is given by

$$k=1+n[1+(m-1)\rho]/m, \quad (2)$$

where n is the number of subjects required for each group under simple randomization trials.

When the number of clusters per group is fixed by k^* , the number of subjects per cluster is given by the following equation assuming that the number of subjects is the same in all the clusters

$$m^* = \frac{(1-\rho)n}{(k^* - 1 - \rho n)}. \quad (3)$$

The total number of subjects required for each group is obtained by multiplying k^* by m^* . The number of subjects (m^*) is obtained only when the denominator of Equation (3) is positive. That is, m^* can be computed only when k^* is greater than $(1 + \rho n)$.

3. Example

The innovative cancer risk intake system (CRIS) elicits responses to questions about personal and familial risk, intent, and perceived barriers to risk-appropriate testing. A cluster randomization trial conducted in primary-care clinics will determine efficacy of CRIS for facilitating participation in risk-appropriate colorectal cancer testing. Physicians are randomly allocated to either a risk-based innovative cancer risk intake system (CRIS) group or a comparison group. Based on assignment of his or her physician, each patient will be assigned either to the CRIS intervention or a comparison group, in which patients and physicians will receive non-tailored print outs that are simple reminders about testing, but are not risk-based; nor will they list or address patient barriers to testing. Patients will be seen by 28 physicians at the primary care clinics of the University of Texas Southwestern Medical Center. Twenty eight physicians will be randomly allocated to the CRIS group or a comparison group. The primary outcome of the trial is participation in risk-appropriate colorectal cancer testing (yes/no, 1=participation in any risk-appropriate testing, 0=non participation) at month 12 after the start of the intervention. Pilot and published data suggests that about 20% of the comparison group will participate in appropriate colorectal cancer testing [5]. We expect that the CRIS group will have at least 32% participation rate. With a type I error rate of 5% and a power of 80%, we can detect at least 12% difference between two study groups using 442 patients (221 in each group) when we ignore the clustering structure of study subjects. To account for clustering of patients within attending physician, we have assumed $\rho=0.02$. Although an estimate of ρ is difficult to know in advance of a study, we believe this is an appropriate and conservative estimate based on our own and others' research findings. A recent survey article reviewing outcomes from 31 studies found a median ρ of 0.01 with an inter-quartile range of 0 to 0.03 [6]. With $\rho=0.02$, $k^*=14(=28/2)$, and $n = 206$, the required number of subjects per cluster is

$$m^* = \frac{(1-\rho)n}{k^* - 1 - \rho n} = \frac{(1-0.02) * 206}{14 - 1 - 0.02 * 206} = 22.7 \quad (4)$$

The required number of subjects per cluster is 23. The total number of subjects needed for the study is 644 ($=23*28$). We anticipate being unable to audit electronic medical records for about 8% of patients due to relocation or death by 12-month follow-up. Therefore, we will need a

final sample size of 700 subjects, that is, 350 subjects in each study group to provide sufficient power for the planned analyses accounting for 8% loss to follow-up.

4. Simulation Study

We conducted a simulation study to investigate the performance of the sample size formula for the number of subjects given in Equation (3) when the number of available clusters is fixed in advance. Equation (3) provides the sample size estimate for the number of subjects per cluster assuming that the number of subjects is constant among all the clusters. Since cluster sizes are not constant in many cluster randomization studies, we examine the effect of intracluster correlation coefficient and cluster size variability on the power of the study based on the ongoing actual clinical trial in which the number of available clusters is fixed in advance.

Let k^* be the number of clusters in each group. For each cluster, m_{ij} , the cluster size of the i^{th} cluster in group j , is generated using a negative binomial distribution truncated below 1 with the probability density function

$$p(m) = \frac{(s+m-1)! Q^{-s} (P/Q)^m}{(s-1)! m! (1-Q^{-s})}, \quad (5)$$

where Q^{-1} is the probability of response, $Q = 1 + P$, and s is the number of responses.

The truncated negative binomial distribution is known to provide a reasonable fit to group sizes generated in several application areas [7]. The mean and variance are $\mu = sP/(1 - P_0)$ and $\sigma^2 = \mu[1 + P - sPP_0/(1 - P_0)]$, where $P_0 = (1 + P)^{-s}$ [8]. The measure of imbalance is given by $\kappa = 1/(1 + v^2)$, where $v = \sigma/\mu$. The relationship between variance and the imbalance parameter can be written as $\sigma^2 = \mu^2(1 - \kappa)/\kappa$. As κ decreases, the variance of the cluster size increases.

The parameter values are chosen based on the CRIS study design described in the previous section. The numbers of clusters per group are fixed at 10, 20 or 30. The values of ρ used are 0.01, 0.02, 0.05 and 0.1. We allow the cluster size to vary randomly with the imbalance parameter of $\kappa=0.6, 0.8$ and 1.0, which corresponds to severe, moderate and no variability in cluster size. We use the response probabilities of $(p_1, p_2)=(0.2, 0.3)$ and $(0.2, 0.4)$. We ran 10,000 experiments for each combination of parameters.

The simulation study is conducted in the following ways. For prespecified values of k^* , (p_1, p_2) , and ρ , we estimate the cluster size (m^*) using Equation (3), and then use the estimated m^* as the mean cluster size (μ) to allow the variability in cluster size among clusters. Cluster sizes are generated from the negative truncated binomial distribution with the mean cluster size of μ and the imbalance parameter of $\kappa=0.6, 0.8$, and 1.0. Conditional on cluster sizes, the binary outcomes are generated with the method of Lunn and Davies [9].

Donner and Klar [1] showed that the adjusted one degree of freedom chi-square statistic to test $H_0 : p_1 = p_2$ is given by

$$\chi_A^2 = \sum_{j=1}^2 \frac{M_j(\widehat{P}_j - \widehat{P})^2}{C_j \widehat{P}(1 - \widehat{P})}, \quad (6)$$

where $M_j = \sum_{i=1}^{k^*} m_{ij}$ is the total number of subjects in group j , \hat{P}_j is the proportion of responses over all clusters in group j , and \hat{P} is the proportion of responses over all clusters across groups, and the clustering correction factor C_j is given by

$$C_j = \frac{\sum_{i=1}^{k^*} m_{ij} [1 + (m_{ij} - 1)\hat{\rho}]}{\sum_{i=1}^{k^*} m_{ij}}, \quad (7)$$

Here, k^* is the number of clusters per group, and $\hat{\rho}$ is the estimate of the intracluster correlation coefficient, which can be estimated by the ANOVA method [10]. We reject the null hypothesis $H_0 : p_1 = p_2$ when χ_A^2 is larger than $\chi^2(1, 1 - \alpha)$, where $\chi^2(1, 1 - \alpha)$ is the 100(1 - α) percentile of χ^2 distribution with 1 degree of freedom. We compute empirical powers as the proportion of samples rejecting $H_0 : p_1 = p_2$ among 10,000 samples.

Table 1 presents the empirical powers when the numbers of clusters (k^*) are 10, 20 and 30 in each group. The sample size estimate obtained from Equation (3) yields empirical powers close to the nominal level of 90% when $k^* \geq 20$ and the imbalance parameter ($\kappa \geq 0.8$). The sample size estimate from Equation (3) also yields empirical powers close to the nominal level when ($\rho \leq 0.02$, $\kappa \geq 0.8$, and $k^* = 10$) or ($\rho \leq 0.02$, $\kappa = 0.6$, and $k^* = 20$). The sample size estimate from Equation (3) yields empirical powers close to the nominal level for balanced case ($\kappa = 1.0$) even when k^* is as small as 10. Note that the sample size estimate for the number of subjects cannot be computed from Equation (3) when $k^* > (1 + \rho n)$. Thus, Table 1 does not show the sample size estimates and empirical powers when $k^* > (1 + \rho n)$.

5. Discussion

We investigate the effect of the cluster size variability and the intracluster correlation coefficient on the power of the study given a fixed number of clusters per group since cluster sizes are usually different among clusters in practical situations. The number of subjects per cluster cannot be computed from Equation (3) when $k^* > (1 + \rho n)$.

When cluster size is the same among all clusters ($\kappa = 1.0$), the sample size estimate yields empirical powers close to the nominal level even for k^* as small as 10. The sample size estimate yields empirical powers close to the nominal level when $k^* \geq 20$ and $\kappa \geq 0.8$. Empirical powers are close to the nominal level when ($\rho \leq 0.02$, $\kappa \geq 0.8$, and $k^* = 10$) or ($\rho \leq 0.02$, $\kappa = 0.6$, and $k^* = 20$).

Simulation studies show that empirical powers are not close to the nominal power when there is severe imbalance ($\kappa = 0.6$) in cluster size. Further research is needed to estimate the cluster size in case of severe imbalance when the available number of clusters is fixed in advance. Cluster randomization trials have received increasing attention among healthcare researchers over the past decades. Cluster randomized trials have been conducted for binary, continuous and survival outcomes. It is needed to investigate the effect of the cluster size variability and the intracluster correlation coefficient on the cluster size for continuous and survival outcomes when the number of available clusters is fixed in advance.

Acknowledgments

This work was supported in part by NIH grants UL1 RR024982, R01 CA122330, and R01 HL087768.

References

1. Donner, A.; Klar, N. Design and analysis of cluster randomization trials in health research. Oxford University Press; 2000.
2. Taljaard M, Donner A, Klar N. Accounting for expected attrition in the planning of community intervention trials. *Statistics in Medicine* 2007;26:2615–2628. [PubMed: 17068842]
3. Pocock, SJ. *Clinical Trials: A practical approach*. New York: John Wiley; 1983.
4. Snedecor, G.; Cochran, W. *Statistical methods*. Vol. 8th. Ames, Iowa: Iowa State University Press; 1989.
5. Skinner C, Rawl S, Moser B, et al. Impact of the Cancer Risk Intake System on patient-clinician discussions of tamoxifen, genetic counseling, and colonoscopy. *Journal of General Internal Medicine* 2005;20:360–365. [PubMed: 15857495]
6. Adams G, Gulliford MC, Ukoumunne OC, Eldridge S, Chinn S, Campbell MJ. Pattern of intracluster correlation from primary care research to inform study design and analysis. *Journal of Clinical Epidemiology* 2004;57:785–794. [PubMed: 15485730]
7. Donner A, Koval A. A procedure for generating group sizes from a one-way classification with a specified degree of imbalance. *Biometrical Journal* 1987;29:181–187.
8. Johnson, NL.; Kotz, S. *Discrete Distributions*. Wiley; New York: 1969. *Distributions in Statistics*.
9. Lunn AD, Davies SJ. A Note on Generating Correlated Binary Variables. *Biometrika* 1998;85:487–490.
10. Donner A, Eliasziw M. Methodology for inferences concerning familial correlations: a review. *Journal of Clinical Epidemiology* 1991;44:449–55. [PubMed: 2010789]

Table 1

Empirical powers(%) and sample size estimates for cluster sizes from 10,000 simulations when the number of clusters (k^*) is fixed in advance.

k^*	κ	ρ	(p_1, p_2)	
			(0.2,0.3)	(0.2,0.4)
10	0.6	0.01	76.3(76)	82.6(14)
		0.02	72.4(312)	82.1(15)
		0.05	— [*] (—)	74.7(27)
		0.10	—(—)	—(—)
	0.8	0.01	85.2(76)	87.1(14)
		0.02	82.3(312)	87.1(15)
		0.05	—(—)	83.7(27)
		0.10	—(—)	—(—)
	1.0	0.01	88.1(76)	91.8(14)
		0.02	88.2(312)	88.7(15)
		0.05	—(—)	88.9(27)
		0.10	—(—)	—(—)
20	0.6	0.01	86.0(26)	89.2(6)
		0.02	81.2(34)	89.9(7)
		0.05	—(—)	84.6(8)
		0.10	—(—)	77.8(12)
	0.8	0.01	86.8(26)	90.2(6)
		0.02	87.4(34)	91.8(7)
		0.05	—(—)	90.8(8)
		0.10	—(—)	85.8(12)
	1.0	0.01	90.9(26)	92.2(6)
		0.02	88.7(34)	94.3(7)
		0.05	—(—)	92.0(8)
		0.10	—(—)	89.8(12)
30	0.6	0.01	88.2(16)	93.6(4)
		0.02	85.7(18)	91.7(4)
		0.05	77.2(39)	89.7(5)
		0.10	—(—)	87.7(6)
	0.8	0.01	90.7(16)	91.3(4)
		0.02	89.1(18)	90.5(4)
		0.05	86.1(39)	90.2(5)
		0.10	—(—)	90.6(6)
	1.0	0.01	90.2(16)	91.2(4)
		0.02	90.2(18)	91.1(4)
		0.05	90.6(39)	91.8(5)
		0.10	—(—)	92.5(6)

* denotes that cluster sizes cannot be computed since $k^* \leq (1 + \rho n)$

Number in parenthesis corresponds to average cluster size (m^*), and the corresponding variance of the cluster size can be computed using the formula $\sigma^2 = (m^*)^2(1 - \kappa)/\kappa$.