Published in final edited form as:
Comput Stat Data Anal. 2009 January 15; 53(3): 596–602. doi:10.1016/j.csda.2008.09.007.

Effect of Imbalance and Intracluster Correlation Coefficient in Cluster Randomized Trials with Binary Outcomes

Chul Ahn,
Department of Clinical Sciences, UT Southwestern Medical Center, Dallas, TX

Fan Hu, and
Department of Statistical Science, Southern Methodist University, Dallas, TX

Celette Sugg Skinner
Department of Clinical Sciences, UT Southwestern Medical Center, Dallas, TX

Summary

Cluster randomization trials are increasingly popular among healthcare researchers. Intact groups (called 'clusters') of subjects are randomized to receive different interventions and all subjects within a cluster receive the same intervention. In cluster randomized trials, a cluster is the unit of randomization and a subject is the unit of analysis. Variation in cluster sizes can affect the sample size estimate or the power of the study. Guittet *et al.* (2006) investigated the impact of an imbalance in cluster size on the power of trials with continuous outcomes through simulations. In this paper, we examine the impact of cluster size variation and intracluster correlation on the power of the study for binary outcomes through simulations. Because the sample size formula for cluster randomization trials is based on a large sample approximation, we evaluate the performance of the sample size formula with small sample sizes through simulation. Simulation study findings show that the sample size formula (m_p) accounting for unequal cluster sizes yields empirical powers closer to the nominal power than the sample size formula (m_a) for the average cluster size method. The differences in sample size estimates and empirical powers between m_a and m_p get smaller as the imbalance in cluster sizes gets smaller.

Keywords

Cluster randomization; Sample size; Cluster size imbalance; Power

1. Introduction

Over the past three decades cluster randomized trials have received increasing attention among healthcare researchers. Intact groups (called 'clusters') of subjects are randomized to receive 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. In cluster randomization trials, subjects within each cluster may be dependent, although subjects from different clusters are assumed independent.

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, email: Chul.Ahn@UTSouthwestern.edu.

The sample size formula for the case of equal cluster sizes can be obtained by multiplying the standard sample size formula from an individual randomization by the variance inflation factor (also known as the design effect) $[1 + (n - 1)\rho]$, where n denotes the cluster size and ρ is a measure of intracluster correlation. The above sample size formula for cluster randomized trials assumes an equal cluster size, which is optimal but rarely encountered in practice. In practice, cluster randomization trials exhibit high parities in cluster sizes due to variation in recruitment rates and loss to follow-up rates among clusters (Taljaard *et al.*, 2007), and due to natural variation in the actual size of the clusters, such as families, schools, or health care practices.

At the planning stage, the commonly used sample size formula is to replace the cluster sizes n with an advance estimate of the average cluster size $\bar{n}$ without taking into account any potential imbalance in cluster size. An imbalance in cluster size reduces the power of the trial (Donner and Klar, 2000) and has to be taken into account for the sample size estimation. Taljaard *et al.* (2007) have presented sample size formulas to account for potential attrition in cluster randomization trials in which the plan is to enroll the same number of subjects in each cluster prior to randomization. The sample size formulas of Taljaard *et al.* (2007) are useful for trials in which advance estimates of the distribution of cluster sizes are unknown or in which the plan is to enroll a constant number of subjects per clusters, but enrollment rates vary among clusters. Guittet *et al.* (2006) who investigated the impact of cluster size variation on a proper power of the trials with continuous outcomes through simulation showed that an imbalance in cluster size can highly influence the power in the case of severe imbalance, particularly in the case of a small number of clusters and/or high intracluster correlation. Kang *et al.* (2003) presented sample size formula for dichotomous outcomes in cluster randomization trials with varying cluster size. In this paper we will investigate the impact of small numbers of clusters, various intracluster correlation coefficients, and varying cluster sizes on the power of trials with binary clustered outcomes through simulation.

2. Statistical Methods

Suppose that we are interested in comparing the proportions of responses between two intervention groups from an individual 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 per group (m_s) to test H_0: $p_1 = p_2$ versus H_1: $p_1 \neq p_2$ is given by

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

where $z_{1-\alpha/2}$ is the $100(1 - \alpha/2)$ percentile of the standard normal distribution.

In this paper we assume that an equal number of clusters is assigned in each intervention group. Let n_{ik} denote the cluster size of the ith cluster of the kth treatment, $i = 1, \ldots, m$ and $k = 1, 2$, where m is the number of clusters in each intervention group. For the ith cluster of the kth treatment, let Y_{ijk} denote the binary outcome variable of the jth subject, $j = 1, \ldots, n_{ik}$ with $E(Y_{ijk}) = p_{ijk}$ that is expressed as

$$g(p_{ijk}) = \alpha + \beta X_{ijk} \quad (2)$$

where $g(p) = log(p/(1-p))$ is the logit-transformation, and X_{ijk} is the indicator variable denoting the intervention group indicator. Here, X_{ijk} is fixed within cluster, that is, $X_{ijk} = X_{ik}$ for all j. But, X_{ijk} can vary between clusters. We assume that subjects in a cluster are exchangeable in

the sense that, given n_{ik}, Y_{i1k}, …, Y have a common marginal response probability $P(Y_{ijk} = 1) = p_k (0 < p_k < 1)$ and a common intraclass correlation coefficient, $\rho = corr(Y_{ijk}, Y_{ij'k})$ for $j \neq j'$.

When all clusters have an equal cluster size ($n_{ik} = n$), the variance of $\hat{p_1} - \hat{p_2}$ can be estimated by

$$s_c^2 = [\widehat{p}_1(1 - \widehat{p}_1) + \widehat{p}_2(1 - \widehat{p}_2)] \frac{[1 + (n - 1)\rho]}{mn} \tag{3}$$

where m is the number of clusters in each group, $\widehat{p}_k = \sum_{i=1}^{m} \sum_{j=1}^{n} Y_{ijk}/(nm)$. Therefore, the sample size formula for the number of clusters (m_c) can be obtained by multiplying the standard sample size formula (m_s) from an individual randomization by $[1 + (n - 1)\rho]/n$, where n is the cluster size, and ρ is an intracluster correlation coefficient. That is, $m_c = m_s[1 + (n - 1)\rho]/n$. Note that m_s is the sample size estimate under individual cluster randomization. That is, m_s is the sample size estimate when the cluster size (n) is equal to one.

The commonly used sample size formula for the number of clusters to test the null hypothesis H_0: $p_1 = p_2$ vs. H_1: $p_1 \neq = p_2$ for unequal cluster sizes is to replace the cluster sizes n with an advance estimate of the average cluster size $\bar{n}$. Manatunga *et al.* (2001) refer to use of average cluster size as the 'average cluster size method', which is likely to underestimate the required number of clusters (Donner and Klar, 2000). The sample size estimate for the 'average cluster size method' can be written as

$$m_a = (z_{1-\alpha/2} + z_{1-\beta})^2 \frac{p_1(1 - p_1) + p_2(1 - p_2)}{(p_1 - p_2)^2} \frac{[1 + (\bar{n} - 1)\rho]}{\bar{n}} \tag{4}$$

Let θ, τ^2 and $\gamma = \tau/\theta$ be the mean, variance and the coefficient of variation of the cluster size, respectively. Kang *et al.* (2003) provided the formulas for the variance of $\hat{p_1} - \hat{p_2}$ and the sample size estimate for a cluster randomization trial with an unequal cluster size. The variance of $\hat{p_1} - \hat{p_2}$ taking account of unequal cluster sizes can be estimated by

$$s_p^2 = \sum_{k=1}^{2} \widehat{p}_k(1 - \widehat{p}_k) \frac{\sum_{i=1}^{m} n_{ik}[1 + (n_{ik} - 1)\widehat{\rho}]}{(\sum_{i=1}^{m} n_{ik})^2} \tag{5}$$

The sample size formula for an unequal cluster size is given by

$$m_p = (z_{1-\alpha/2} + z_{1-\beta})^2 \frac{[p_1(1 - p_1) + p_2(1 - p_2)]}{(p_1 - p_2)^2} [(1 - \rho)\frac{1}{\theta} + \rho + \rho\gamma^2]. \tag{6}$$

When cluster sizes are equal for all clusters, τ becomes zero, and $m_p = m_a = m_c$. When the sample size for the number of clusters is sufficiently large, the null hypothesis H_0: $p_1 = p_2$ is rejected with significance level α if

$$Z=|\frac{\widehat{p}_1-\widehat{p}_2}{s_p}|>z_{1-\alpha/2} \tag{7}$$

where $z_{1-\alpha/2}$ is the $100(1-\alpha/2)$ percentile of the standard normal distribution.

3. Simulation Study

We conducted a simulation study to investigate performance of the sample size formula of m_a and m_p. In an ongoing community intervention trial called CRIS (Cancer Risk Intake System, CA R01 1223301), cluster sizes (the number of study patients per primary care physicians) are expected to be skewed. So, we generate the cluster sizes using a truncated negative binomial distribution, which has been previously used to generate unequal cluster sizes with specific imbalance parameter values (Donner and Hauck, 1986; Donner *et al.*, 1989; Ahn, 1997; Zhou and Donner, 2004). Cluster size is generated using the negative binomial distribution truncated below 1 (Donner and Koval, 1987), specifically with probability density function

$$P(n)=\frac{(s+n-1)!Q^{-s}(P/Q)^n}{(s-1)!n!(1-Q^{-s})}, \tag{8}$$

where $Q = 1 + P$, $i = 1, \ldots, m$.

The mean and variance of the above cluster size distribution are $\mu = sP/(1 - P_0)$ and $\sigma^2 = \mu[1 + P - sPP_0/(1-P_0)]$, where $P_0 = (1+P)^{-s}$ (Johnson nd Katz, 1969). The measure of imbalance in cluster size is given by $\kappa = 1/(1 + v^2)$, where $v = \sigma/\mu$. When κ is equal to 1, all the cluster sizes are equal. As κ decreases, the variance of the cluster size increases. The properties of the measure of imbalance is given by Ahrens and Pincus (1981).

Simpson *et al.* (1995) reviewed the characteristics of 21 cluster randomized primary prevention trials that were published from 1990 to 1993. The average cluster size ranged from 1.5 teenagers from family to about 350 subjects per community. In most community intervention trials, values for ρ are quite small, often ranging between 0.001 and 0.05 with relatively large cluster sizes. For example, in an adolescent tobacco use prevention trial (Murray *et al.*, 1992), the average cluster size is 190. In the recently completed PROSPECT (Prevention of Suicide in Primary Care Elderly: Collaborative Trial, Bruce *et al.*, 2004) and the ongoing CRIS (Cancer Risk Intake Systems) cluster randomization trials, the average cluster size ranged between 20 and 25. In simulation, we used ρ values of 0.05 and 0.10, and average cluster sizes (μ) of 10, 25, 50, 100 and 300 for community intervention settings. Larger values for ρ are often found in family, dental, or ophthalmologic studies with smaller cluster sizes. For example, in three dental studies in which the unit of sampling was subject and the unit of analysis was tooth (Banting *et al.*, 1985; Donner and Banting, 1987, 1988), the intracluster correlation values ranged between 0.354 and 0.432. We used μ values of 5, 10 and 20, and ρ values of 0.25, 0.50, and 0.75 for family or dental settings. We allow very high values of ρ since the study of Guittet *et al.* (2006) showed that an imbalance in cluster size can highly influence the power in the case of severe imbalance, particularly in the case of small number of clusters and/or high intracluster correlation coefficient for continuous outcomes. Here, we investigate the impact of cluster size variation, small number of clusters, and high intracluster correlation coefficient on the power of the study. The simulation covers a range of ρ values larger than those usually observed in cluster randomized trials. Cluster sizes are generated from the negative truncated

binomial distribution with mean cluster sizes specified above, and the imbalance parameter of κ=0.6, 0.8, and 1.0. Simulations are conducted with no variability in cluster size (κ=1), moderate variability (κ=0.8) and high variability (κ=0.6). The variance is extremely large for values of $\kappa \leq 0.4$. The required number of clusters in each intervention group is estimated using m_a and m_p in Equations (4) and (6) for given values of (p_1, p_2), ρ, κ, μ, α and β. Conditional on cluster size and the estimated number of clusters, the binary outcomes are generated with the method of Lunn and Davies (1998), which generates the correlated binary data in a simple and efficient way.

We conduct five thousand experiments for each parameter combination, and compute empirical powers as the proportion of samples rejecting H_0: $p_1 = p_2$ by Equation (7) among 5,000 samples. Tables 1–3 present the empirical powers of the sample size formula of m_a and m_p with the two-sided significance level of α=5% and the nominal power of 1−β=90% for κ=0.6, 0.8 and 1.0, respectively. Tables 1–3 report on the response probabilities of (p_1, p_2)=(0.2, 0.3), (0.2, 0.4), (0.5, 0.6) and (0.5, 0.7). In each cell, the first and second rows present the empirical power (and the sample size estimate) for m_a and m_p, respectively.

The sample size estimates and empirical powers of m_a and m_p are the same in Table 3 when all cluster sizes are equal (κ = 1). When there is any variability in cluster sizes (κ=0.6 and 0.8), the empirical powers of m_p are much closer to the nominal power of 90% than those of m_a. Note that the sample size estimates for the number of clusters using m_a do not depend on the values of κ even though the empirical powers of m_a increase as κ increases. The required number of clusters using m_p decreases as increases while that using m_a does not depend on the values of κ. The sample size estimates from m_a are smaller than those from m_p. The differences in sample size estimates between m_a and m_p get smaller as κ increases, that is, as the variability in cluster sizes decreases. As the intracluster correlation ρ increases, the required sample size for the number of clusters per intervention group increases. The required sample size for the number of clusters per intervention group decreases as the mean cluster size μ increases. The sample size estimates are smaller when the absolute difference between p_1 and p_2 gets larger. For all the parameter combinations of ρ=(0.25, 0.50, 0.75), μ=(5, 10, 20) and κ=(0.6, 0.8, 1.0), m_p yields the empirical powers within 2% of the nominal power of 90%. For all the parameter combinations of ρ=(0.05, 0.10), μ=(10, 25, 50, 100, 300), and κ=(0.6, 0.8, 1.0), m_p yields empirical powers between 85% and 91% except two cases of (ρ, μ, κ)=(0.05, 300, 0.8) and (0.05, 300, 1.0) when (p_1, p_2)=(0.2, 0.4).

4. Example

An ongoing innovative cancer risk intake system (CRIS) 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 comparison group or a risk-based innovative cancer risk intake system (CRIS) group that delivers patient-tailored print outs based on personal risk factors and perceived barriers to colon cancer testing. 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. 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). We anticipate about 20% of the comparison group will participate in appropriate testing based on pilot and published data (Skinner *et al.*,2005). We will assume that the CRIS group will yield at least 32% participation rate. Cluster sizes are expected to be unequal because a different number of patients will be recruited among physicians. With ρ = 0.02, 5% significance level and 80% power, the numbers of physicians needed for the trial will be 19 and 20 using m_a and m_p assuming that average number of patients recruited for each physician

is 23 and the corresponding variance is 60. If $\rho = 0.05$, we will have m_a=13 and m_p=14. That is, in order to detect p_1=0.32 and p_2=0.2 with 5% significance level, 80% power and ρ=0.05, the method (m_p) accounting for unequal cluster sizes assigns 14 physicians to each arm, and the average size method (m_a) assigns 13 physicians to each arm. That is, the total number of patients required for the trial will be 644(=14*23*2) using m_p and 598 (=13*23*2) using m_a.

5. Discussion

We investigated the effect of intracluster correlation and cluster size imbalance on the power of cluster randomization trials. We evaluated the performance of the sample size formula for the number of clusters with small sample sizes through simulation since the sample size formulas are derived using a large sample approximation. Simulation studies show that empirical powers of m_p are generally close to the nominal power of 90% even under the presence of imbalances (κ=0.6 or κ =0.8) and small sample size estimates. The sample size formula using the average cluster size method (m_a) yields empirical powers lower than the nominal power of 90%. The simulation study suggests that the design of cluster randomization trials should account for the cluster size variation.

Our simulation study shows that empirical powers are lower than the nominal power for some parameter combinations especially when the number of clusters, m_p are less than 10. Methods such as the adjusted chi-square approach, the ratio estimator approach and the method of GEE are no longer applicable for binary outcomes with a fairly small number of clusters per group (often 10 or less) since the large sample approximations underlying these procedures become questionable (Donner and Klar, 2000). Donner and Klar (2000) suggested the use of two-sample t-test or nonparametric procedures such as Wilcoxon rank sum test or Fisher's permutation procedures if the number of clusters per group is less than 10. Further studies are need to investigate the performance of these test procedures for the small number of clusters per group.

In this paper, we assume that subjects in a cluster are exchangeable in the sense that, given n_{ik}, Y_{i1k}, …, Y have a common marginal response probability $P(Y_{ijk} = 1) = p_k(0 < p_k < 1)$ and a common intraclass correlation coefficient. This assumption may not be reasonable when the intracluster correlation decreases as the cluster size decreases. It is necessary to evaluate the performance of sample size formula when the assumption of the exchangeable intracluster correlation is violated. Further study is needed to evaluate the effect of intracluster correlations when the intracluster correlation values are different between intervention groups.

Jung *et al.* (2001) provided the sample size formula for the response probability in a one-sample clustered binary data using the equal weights to clusters, equal weights to subjects and optimal weights which yield the minimum variance. Using the above three weighting schemes, Ahn *et al.* (2003) compared the performance of three weighted chi-square statistics. The simulation study shows that the weighted chi-square statistic using an optimal weight yields higher empirical powers than the others. Guittet *et al.* (2006) showed that cluster size variation can highly influence the power of cluster-level adjusted two-sample t-test in the case of severe imbalance and an optimal weighting scheme yields the best performance in empirical powers in the case of severe imbalance for continuous outcomes. Further study is needed to compare the performance of different weighting schemes.

Acknowledgments

This work was supported in part by NIH grants UL1 RR024982, R01 CA122330 and R01 HL087768. We thank two anonymous reviewers for their constructive comments and helpful suggestions. We thank Sae Joon Ahn for computational assistance.

References

Ahn C. An evaluation of simple methods for the estimation of a common odds ratio in clusters with variable size. Computational Statistics and Data Analysis 1997;24:47–61.

Ahn C, Jung SH, Kang SH. An evaluation of weighted chi-square statistics for clustered binary data. Drug Information Journal 2003;37:91–99.

Ahrens H, Pincus R. On two measures of unbalancedness in a one-way model and their relation to efficiency. Biometrical Journal 1981;23:227–235.

Banting DW, Ellen RP, Fillery ED. A longitudinal study of root caries: baseline and incidence data. Journal of Dental Research 1985;64:1141–1144.

Bruce ML, Ten Have TR, Reynolds CF III, Katz II, Schulberg HC, Mulsant BH, Brown GK, McAvay GJ, Pearson JL, Alexopoulos GS. Reducing suicidal ideation and depressive symptoms in depressed older primary care patients: A randomized controlled trial. JAMA 2004;291:1081–1091. [PubMed: 14996777]

Donner A, Banting D. Analysis of site-specific data in dental studies. Journal of Dental Research 1987;67:1392–1395. [PubMed: 3183156]

Donner A, Banting D. Adjustment of frequently used chi-square procedures for the effects of site-to-site dependence in the analysis of dental data. Journal of Dental Research 1988;68:1350–1354. [PubMed: 2778178]

Donner A, Eliasziw M. Statistical implications of the choice between a dichotomous or continuous trait in studies of interobserver agreement. Biometrics 1994;50:550–555. [PubMed: 19405213]

Donner A, Hauck W. The large-sample relative efficiency of the Mantel-Haenszel estimator in the fixed-strata case. Biometrics 1986;42:537–545. [PubMed: 3567287]

Donner, A.; Klar, N. Design and Analysis of Cluster Randomization Trials in Health Research. Arnold; London: 2000.

Donner A, Koval J. A procedure for generating group sizes from a one-way classification with a specified degree of imbalance. Biometrical J 1987;29:181–187.

Donner A, Wells G, Eliasziw M. On two approximations to the F-Distribution: application to testing for intraclass correlation in family studies. The Canadian Journal of Statistics 1989;17:209–215.

Guittet L, Ravaud P, Giraudeau B. Planning a cluster randomized trial with unequal cluster sizes: practical issues involving continuous outcomes. BMC Med Res Methodol 2006;6(17):1–15. [PubMed: 16412232]

Johnson, NL.; Kotz, S. Discrete Distributions. Wiley; New York: 1969. Distributions in Statistics.

Lunn AD, Davies SJ. A Note on generating correlated binary variables. Biometrika 1998;85:487–490.

Jung SH, Kang SH, Ahn C. Sample size calculations for clustered binary data. Statistics in Medicine 2001;20:1971–1982. [PubMed: 11427953]

Kang SH, Ahn C, Jung SH. Sample size calculations for dichotomous outcomes in cluster randomization trials with varying cluster size. Drug Information Journal 2003;37:109–114.

Manatunga AK, Hudgens MG, Chen S. Sample size estimation in cluster randomized studies with varying cluster size. Biometrical Journal 2001;1:75–86.

Murray DM, Perry CL, Griffn G, Harty KC, Jacobs DR, Schmid L, Daly K, Pallonen U. Results from a statewide approach to adolescent tobacco use prevention. Preventive Medicine 1992;21:449–472. [PubMed: 1409487]

Simpson JM, Klar N, Donner A. Accounting for cluster randomization: A review of Primary prevention trials, 1990 through 1993. American Journal of Public Health 1995;85:1378–1382. [PubMed: 7573621]

Skinner CS, Rawl SM, Moser BK, Buchanan AH, Scott LL, Champion VL, Schildkraut JM, Parmigiani G, Clark S, Lobach DF, Bastian LA. Impact of the Cancer Risk Intake System on patient-clinician discussions of tamoxifen, genetic counseling, and colonoscopy. Journal of General Internal Medicine 2005;20(4):360–365. [PubMed: 15857495]

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]

Zhou G, Donner A. Confidence Interval Estimation of the Intraclass Correlation Coefficient for Binary Outcome Data. Biometrics 2004;60:807–811. [PubMed: 15339305]

Table 1

Empirical powers(%) and sample size estimates for the number of clusters per group in parentheses from 5,000 simulations for m_a and m_p when κ=0.6.

		(p_1, p_2)			
ρ^*	μ^{**}	**(0.2,0.3)**	**(0.2,0.4)**	**(0.5,0.6)**	**(0.5,0.7)**
0.05	10	84(57[a])	80(16)	83(75)	80(18)
		91(70[p])	87(19)	90 (92)	90(22)
	25	78(35)	76(10)	78(46)	74(11)
		90(48)	86(14)	89(63)	85(15)
	50	73(27)	69(8)	75(36)	68(9)
		89(40)	86(11)	90(53)	88(13)
	100	72(24)	68(7)	73(31)	66(8)
		90(37)	85(10)	89(48)	86(12)
	300	70(21)	61(6)	70(28)	60(7)
		90(34)	87(10)	88(45)	87(11)
0.10	10	80(74)	77(20)	80(98)	75(23)
		90(100)	88(27)	90(133)	89(32)
	25	76(53)	74(15)	75(71)	72(17)
		90(79)	89(22)	90(105)	88(25)
	50	72(46)	71(13)	72(61)	71(15)
		88(72)	89(20)	90(96)	86(23)
	100	71(43)	68(12)	72(57)	69(14)
		89(69)	89(19)	90(91)	89(22)
	300	70(41)	66(11)	71(54)	68(13)
		90(66)	88(18)	90(88)	89(21)
0.25	5	78(156)	82(43)	78(206)	78(49)
		91(221)	90(60)	91 (292)	89(69)
	10	76(127)	76(35)	75(168)	74(40)
		91(192)	90(52)	90(254)	89(60)
	20	71(112)	75(31)	73(149)	72(35)
		89(177)	88(48)	90(234)	89(55)
0.50	5	74(234)	75(64)	73(309)	73(73)

		(p_1, p_2)			
ρ^*	μ^{**}	**(0.2,0.3)**	**(0.2,0.4)**	**(0.5,0.6)**	**(0.5,0.7)**
		90(363)	90(99)	90 (481)	89(113)
	10	71(214)	72(58)	72(284)	72(67)
		90(334)	90(93)	89(455)	90(107)
	20	73(205)	72(56)	72(271)	71(64)
		89(334)	88(91)	91(442)	89(104)
0.75	5	73(312)	73(85)	73(412)	73(97)
		90(506)	89(137)	89(670)	90(158)
	10	72(302)	72(82)	72(400)	73(94)
		90(496)	89(134)	91(657)	90(155)
	20	73(297)	73(81)	72(393)	70(93)
		91(491)	90(133)	90(651)	90(153)

* ρ is an intracluster correlation coefficient

** μ is the mean cluster size of a truncated negative binomial distribution below 1

[a] The estimated number of clusters per intervention group for 90% power with m_a

[p] The estimated number of clusters per intervention group for 90% power with m_p

Table 2

Empirical powers(%) and sample size estimates for the number of clusters per group in parentheses from 5,000 simulations for m_a and m_p when κ=0.8.

		(p_1, p_2)			
ρ^*	μ^{**}	**(0.2,0.3)**	**(0.2,0.4)**	**(0.5,0.6)**	**(0.5,0.7)**
0.05	10	88(57[a])	86(16)	88(75)	86(18)
		89(62[p])	88(17)	90 (82)	89(20)
	25	86(35)	84(10)	86(46)	82(11)
		89(40)	88(11)	89(52)	89(13)
	50	84(27)	81(8)	83(36)	80(9)
		89(32)	86(9)	89(42)	85(10)
	100	82(24)	79(7)	83(31)	79(8)
		89(28)	85(8)	90(38)	86(9)
	300	81(21)	74(6)	81(28)	75(7)
		89(26)	83(7)	89(34)	85(8)
0.10	10	85(74)	84(20)	85(98)	84(23)
		89(84)	88(23)	91(111)	89(26)
	25	84(53)	84(15)	84(71)	82(17)
		90(63)	88(17)	90(83)	89(20)
	50	84(46)	81(13)	82(61)	82(15)
		89(56)	88(16)	88(74)	89(18)
	100	83(43)	80(12)	82(57)	83(14)
		90(53)	88(15)	90(69)	88(22)
	300	84(41)	77(11)	82(54)	80(13)
		90(50)	88(14)	89(66)	88(16)
0.25	5	86(156)	88(43)	85(206)	84(49)
		89(180)	91(49)	90 (239)	90(56)
	10	84(127)	83(35)	85(168)	84(40)
		90(151)	88(41)	90(200)	90(47)
	20	83(112)	83(31)	85(149)	82(35)
		91(137)	89(37)	91(181)	89(43)
0.50	5	84(234)	83(64)	82(309)	84(73)

		(p_1, p_2)			
ρ*	μ**	**(0.2,0.3)**	**(0.2,0.4)**	**(0.5,0.6)**	**(0.5,0.7)**
		89(282)	91(77)	90 (374)	89(88)
	10	85(214)	83(58)	83(284)	82(67)
		90(263)	90(71)	91(348)	90(82)
	20	83(205)	85(56)	85(271)	82(64)
		92(253)	91(69)	89(335)	89(79)
0.75	5	82(312)	84(85)	83(412)	82(97)
		90(384)	89(104)	90(509)	92(120)
	10	84(302)	82(82)	81(400)	83(94)
		89(375)	88(102)	90(496)	90(117)
	20	85(297)	83(81)	83(393)	83(93)
		89(370)	90(100)	90(400)	89(115)

* ρ is an intracluster correlation coefficient

** μ is the mean cluster size of a truncated negative binomial distribution below 1

[a] The estimated number of clusters per intervention group for 90% power with m_a

[p] The estimated number of clusters per intervention group for 90% power with m_p

Table 3

Empirical powers(%) and sample size estimates for the number of clusters per group in parentheses from 5,000 simulations for m_a and m_p when κ=1.0.

		(p_1, p_2)			
ρ*	μ**	**(0.2,0.3)**	**(0.2,0.4)**	**(0.5,0.6)**	**(0.5,0.7)**
0.05	10	90($57^{a,p}$)	88(16)	90(75)	89(18)
	25	90(35)	89(10)	90(46)	87(11)
	50	89(27)	88(8)	88(36)	87(9)
	100	90(24)	85(7)	90(31)	88(8)
	300	89(21)	83(6)	90(28)	87(7)
0.10	10	90(74)	89(20)	90(98)	88(23)
	25	89(53)	89(15)	91(71)	88(17)
	50	89(46)	88(13)	90(61)	88(15)
	100	89(43)	88(12)	90(57)	87(14)
	100	90(24)	85(7)	90(31)	88(8)
	300	90(41)	87(11)	90(54)	87(13)
0.25	5	91(156)	91(43)	90(206)	89(49)
	10	89(127)	90(35)	89(168)	89(40)
	20	90(112)	91(31)	91(149)	89(35)
0.50	5	90(234)	89(64)	89(309)	90(73)
	10	90(214)	89(58)	89(284)	89(67)
	20	89(205)	90(56)	90(271)	90(64)
0.75	5	90(312)	91(85)	89(412)	90(97)
	10	89(302)	89(82)	90(400)	90(94)
	20	90(297)	89(81)	88(393)	90(93)

* ρ is an intracluster correlation coefficient

** μ is the mean cluster size of a truncated negative binomial distribution below 1

^{a,p}The estimated number of clusters per intervention group for 90% power for both m_a and m_p

Note that m_a and m_p have the same sample size estimates and empirical powers when κ=1.