Sample Size Calculation for Clustered Binary Data with Nonparametric Methods Using Different Weighting Schemes

Chul Ahn

Department of Clinical Sciences

UT Southwestern Medical Center

Dallas, TX

Fan Hu

Department of Statistical Science

Southern Methodist University

Dallas, TX

William R. Schucany

Department of Statistical Science

Southern Methodist University

Dallas, TX

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

email: Chul.Ahn@UTSouthwestern.edu

Sample Size Calculation for Clustered Binary Data with Nonparametric Methods Using Different Weighting Schemes

Summary

We propose a sample size calculation approach for testing a proportion using the weighted sign test when binary observations are dependent within a cluster. Sample size formulas are derived with nonparametric methods using three weighting schemes: equal weights to observations, equal weights to clusters, and optimal weights that minimize the variance of the estimator. Sample size formulas are derived incorporating intracluster correlation and the variability in cluster sizes. Simulation studies are conducted to evaluate a finite sample performance of the proposed sample size formulas. Empirical powers are generally close to nominal levels. The number of clusters required increases as the imbalance in cluster size increases and the intracluster correlation increases. The estimator using optimal weights yields the smallest sample size estimate among three estimators. For small values of intracluster correlation the sample size estimates derived from the optimal weight estimator are close to that derived from the estimator assigning equal weights to observations. For large values of intracluster correlation, the optimal weight sample size estimate is close to the sample size estimate assigning equal weights to clusters.

Keywords: Proportion test; Intracluster correlation; Optimal weights; Simulation; Cluster size imbalance.

1 INTRODUCTION

Clustered binary data often arise in medical studies such as dental and radiologic studies in which observations are taken from multiple observations of each subject (called 'clusters'). For example, as many as 60 lesions may be observed through positron emission tomography (PET) in one patient since PET offers the possibility of imaging the whole body [1]. A major issue in the analysis of clustered data is the dependence among observations within each cluster. The degree of dependence is usually measured by the intracluster correlation coefficient. Application of the traditional statistical methods such as t-tests or chi-square tests developed for independent observations is invalid since observations within the same subject tend to be dependent. In other words, clustered binary data should be analyzed using statistical methods that take account for the dependence of within-subject observations.

While parametric methods have been extensively studied for the analysis of clustered binary data for the last three decades [2-6], nonparametric statistical methods have recently received increasing attention for the analysis of clustered data [7-12]. Larocque [7] accounted for clustering in a signed-rank test with a variance estimate that is based on the sums of squares over independent clusters. To account for clustering, Rosner *et al.* [8] adjusted the variance in the signed-rank test by estimating it from the ranks of absolute observations with a common intracluster correlation. Datta and Satten [12] derived a signed-rank test for clustered data in which the distribution of pairwise differences within a cluster depends on cluster size.

Jung *et al.* [13] proposed a sample size calculation method for testing a proportion in clustered binary data using parametric statistical methods. Hu *et al.* [14] derived the sample size formula for testing a proportion in clustered binary data with a sign test assigning equal weights to observations. In this paper we focus on sample size estimation for testing a proportion using a sign test with different weighting schemes; equal weights to observations, equal weights to clusters, and optimal weights that minimize the variance of the estimator. Here, the sample size refers to the number of clusters. Noether [15] discussed sample size determinations for some common traditional nonparametric tests under the assumption of independence among observations, in which Noether's sample size formula for a sign test is a special case of the sample size formulas presented in this paper.

We propose a nonparametric sample size calculation method for testing a proportion in clustered binary data incorporating the intracluster correlation and the distribution of cluster sizes. In this paper our discussion is limited to testing a one-sample proportion in clustered binary data. In Section 2,we review weighted sign test statistics for testing a binomial proportion. In Section 3, we derive sample size formulas using three different weighting schemes. In Section 4, we apply sample size formulas to the design of a dental study. In Section 5, we conduct extensive simulation studies to evaluate the performance of the sample size formulas and to investigate the effects of cluster size imbalance and intracluster correlation.

2 STATISTICAL METHOD

Let X_{ij} be the binary random variable (X_{ij}=1 for success, -1 for failure) for the jth observation in the ith cluster, $j = 1, ..., n_i$ and , $i = 1, ..., m$. We can express the total as the difference between the total number of successes and the total number of failures for each cluster by coding X_{ij}=1 and -1 for success and failure, respectively [16, 17]. Cluster size (n_i) may vary at random from a certain distribution with mass function $f(.)$. We assume the common correlation model [18], that is, observations in a cluster are assumed to be exchangeable in the sense that, given n_i, $X_{i1}, ..., X_{in_i}$ have a common marginal response probability $P(X_{ij} = 1) = p(0 < p < 1)$ and a common intracluster correlation coefficient $\rho = corr(X_{ij}, X_{ij'})$ for $j \neq j'$. We test the null hypothesis $H_0 : p = p_0$, versus $H_1 : p = p_1$ for $p_0 \neq p_1$.

Let S_i be the difference between the total number of successes and the total number of failures in the ith cluster ($S_i = \sum_{j=1}^{n_i} X_{ij}$), and $(w_1, w_2, ..., w_m)$ be a sequence of cluster weights such that $w_i \geq 0$ and $(1/m)\sum_{i=1}^{m} w_i n_i = 1$. Then we have a class of statistics used in this sign test

$$T = \sum_{i=1}^{m} w_i S_i = \sum_{i=1}^{m} w_i(n_i^+ - n_i^-), \tag{1}$$

where n_i^+ and n_i^- are the total numbers of successes and failures in the ith cluster, respectively. The expected value of T under the null hypothesis is given by

$$\mu_0(T) = \sum_{i=1}^{m} w_i E(S_i|H_0) = m(2p_0 - 1). \tag{2}$$

Under the null hypothesis, the variance of T is

$$\sigma_0(T)^2 = \sum_{i=1}^{m} w_i^2 Var(S_i|H_0) = 4p_0(1-p_0)\sum_{i=1}^{m} w_i^2 n_i\{1 + (n_i - 1)\rho\}, \tag{3}$$

which can be consistently estimated by

$$\widehat{\sigma_0(T)}^2 = 4p_0(1-p_0)\sum_{i=1}^{m} w_i^2 n_i\{1+(n_i-1)\hat{\rho}\}, \tag{4}$$

where $\hat{\rho}$ can be estimated by the ANOVA method [19]. Through simulation, Ridout *et al.* [20] evaluated the performance of various estimators of ρ for clustered binary data under the common-correlation model, $\rho = \mathrm{corr}(X_{ij}, X_{ij'})$ for $j \neq j'$. They showed that the ANOVA estimator performed well through simulation.

The weighted sign test is defined by

$$Z = \frac{\sum_{i=1}^{m} w_i(n_i^+ - n_i^-) - m(2p_0-1)}{\sqrt{4p_0(1-p_0)\sum_{i=1}^{m} w_i^2 n_i\{1+(n_i-1)\hat{\rho}\}}}, \tag{5}$$

which is asymptotically normal with mean 0 and variance 1.

The choice of $w_i = m/\sum_{i=1}^{m} n_i$ provides equal weights to observations, and the test statistic (1) can be expressed as

$$T_u = \frac{m}{\sum_{i=1}^{m} n_i}\sum_{i=1}^{m}(n_i^+ - n_i^-), \tag{6}$$

and its corresponding test statistic from (5) is

$$Z_u = \frac{\sum_{i=1}^{m}(n_i^+ - n_i^-) - n(2p_0-1)}{\sqrt{4p_0(1-p_0)\sum_{i=1}^{m} n_i\{1+(n_i-1)\hat{\rho}\}}}, \tag{7}$$

where $n = \sum_{i=1}^{m} n_i$.

The choice of $w_i = 1/n_i$ assigns equal weights to clusters, and the test statistic (1) becomes the nonparametric statistic of Datta and Satten [12]

$$T_c = \sum_{i=1}^{m}\left(\frac{n_i^+ - n_i^-}{n_i}\right), \tag{8}$$

and its corresponding test statistic from (5) is

$$Z_c = \frac{\sum_{i=1}^{m}\left(\frac{n_i^{(+)} - n_i^{(-)}}{n_i}\right) - m(2p_0-1)}{\sqrt{4p_0(1-p_0)\sum_{i=1}^{m}\{1+(n_i-1)\hat{\rho}\}/n_i}}. \tag{9}$$

The variance of T, $\sigma_0(T)^2$, is minimized under the constraint of $(1/m)\sum_{i=1}^{m} w_i n_i = 1$ when w_i is given by

$$w_i = \frac{m\{1+(n_i-1)\rho\}^{-1}}{\sum_{i=1}^{m} n_i\{1+(n_i-1)\rho\}^{-1}}. \tag{10}$$

That is, the choice of the above w_i provides the minimum variance of T. We refer the resulted statistic and the test statistic using optimal weights as T_o and Z_o, respectively. Note that the optimal weights, w_i, depends on ρ. The test statistic Z_o has the same form as (5) except that w_i needs to be estimated. We reject H_0 if the absolute value of the test statistic, $(Z_i, i = u, c, o)$, is larger than $z_{1-\alpha/2}$, which is the $100(1-\alpha/2)$ percentile of the standard normal distribution.

3 SAMPLE SIZE DETERMINATION

Noether [15] proposed a sample size formula for some common nonparametric tests such as a sign test. He showed that the sample size or the power (1-β) of the test can be estimated from the following equation.

$$\left\{\frac{\mu_1(T) - \mu_0(T)}{\sigma_0(T)}\right\}^2 = (z_{1-\alpha/2} + r z_{1-\beta})^2, \tag{11}$$

where $r = \sigma_1(T)/\sigma_0(T)$.

We will extend Noether's nonparametric sample size formula to obtain sample size formula for clustered binary data. Noether [15] stated that it will be often appropriate to assume that $\sigma_1(t)$ is closse to $\sigma_0(T)$ for alternatives that do not differ too much from the alternative hypothesis. Here we assume that cluster sizes n_i's are small relative to m so that asymptotic results can be attained with respect to m. The expected value of T under the alternative distribution is

$$\mu_1(T) = \sum_{i=1}^{m} w_i E(S_i|H_1) = m(2p_1 - 1).$$

Since n_i's are independent and identically distributed random variables, from (3) we obtain three limits,

$$\frac{1}{m}\sigma_0(T_u)^2 \to 4p_0(1-p_0)\{(1-\rho)E[N] + E[N^2]\rho\}/E[N]^2$$
$$\frac{1}{m}\sigma_0(T_c)^2 \to 4p_0(1-p_0)\{(1-\rho)E[1/N] + \rho\}$$
$$\frac{1}{m}\sigma_0(T_o)^2 \to 4p_0(1-p_0)\frac{1}{E[N\{1+(N-1)\rho\}^{-1}]},$$

as $m \to \infty$, where N is the random variable with mean θ and variance τ^2 corresponding to the cluster size, and $E[.]$ is the expectation with respect to the distribution of the cluster size. Then we have

$$\sigma_0(T_u)^2 \to 4mp_0(1-p_0)\left\{\frac{1-\rho}{\theta} + \rho + \frac{\tau^2}{\theta^2}\rho\right\}$$
$$\sigma_0(T_c)^2 \to 4mp_0(1-p_0)\{(1-\rho)E[1/N] + \rho\}$$
$$\sigma_0(T_o)^2 \to 4mp_0(1-p_0)\frac{1}{E[N\{1+(N-1)\rho\}^{-1}]}.$$

Note that the limiting quantities of $\sigma_0(T_u)^2$ and $\sigma_0(T_c)^2$ are linear in ρ. The convergent quantity of $\sigma_0(T_c)^2$ is dependent on the distribution of cluster sizes only through its harmonic mean, $E[1/N]$ while that of $\sigma_0(T_u)^2$ depends on the distribution of cluster sizes through its first and second moments. With a power of $1-\beta$, the sample size estimates from (11) $(m_i, i = u, c, o)$ to test $H_0 : p = p_0$ versus $H_1 : p = p_1$ are

$$m_u = \frac{(z_{1-\alpha/2} + z_{1-\beta})^2}{(p_1 - p_0)^2}\left\{\frac{1-\rho}{\theta} + \rho + \frac{\tau^2}{\theta^2}\rho\right\} p_0(1-p_0) \tag{12}$$

$$m_c = \frac{(z_{1-\alpha/2} + z_{1-\beta})^2}{(p_1 - p_0)^2}\{(1-\rho)E[1/N] + \rho\} p_0(1-p_0) \tag{13}$$

$$m_o = \frac{(z_{1-\alpha/2} + z_{1-\beta})^2}{(p_1 - p_0)^2}\frac{1}{E[N\{1+(N-1)\rho\}^{-1}]} p_0(1-p_0). \tag{14}$$

When cluster size is constant, the sample size formulas (12) - (14) are identical. If all responses are independent ($\rho = 0$) and all cluster sizes are equal to 1, then all the sample size formulas (12)-(14) then reduce to Noether's formula [15].

4 EXAMPLE

We use the data of Hujoel *et al.* [21] to illustrate the sample size estimation for clustered binary data with the weighted sign test using different weighting schemes. An enzymatic diagnostic test was performed to investigate whether a site was infected by either *treponema denticola* or *bacteroides gingivalis*. An antibody assay was used as the gold standard to determine infected sites against the two organisms. There were different numbers of infected sites per subject. Table 1 shows the number of true positive test results (S_i) and the number of infected sites (n_i) in 29 subjects. Table 2 presents the observed distribution and the projected distribution of cluster size (N).

Suppose we want to use the above data as pilot data to design a similar experiment to test the hypothesis $H_0 : p = 0.6$ versus $H_1 : p = 0.7$. From these data, we obtain the intracluster correlation coefficient estimate $\hat{\rho} = 0.2$ using the ANOVA method. Thus, we assume the intracluster correlation coefficient of $\hat{\rho} = 0.2$ for a new experiment. We get $E[N] = 4.9$ by both the observed distribution and the projected distribution. Hence, the estimated sample sizes required in the experiment for 80% and 90% are $m_o = 70$ and $m_o = 95$ for T_o, respectively. Similarly, we obtain $m_u = 71$ (and 95) for T_u and $m_c = 71$ (and 95) for T_c for the power of 80% (and 90%) of power. Although the difference is small, m_o is the smallest.

5 SIMULATION STUDY

In this section, we report the results of a simulation study to investigate the performance of sample size formulas in terms of empirical powers. Since cluster sizes frequently exhibit considerable variations in medical studies, cluster sizes were generated using a negative binomial distribution truncated at zero as in Donner and Koval [22] and Ahn *et al.* [23]. We refer the reader to Ahn *et al.* [23] or Johnson *et al.* [24] for details on the truncated negative binomial distribution.

The imbalance in cluster size is measured by the quantity $\kappa = 1/(1+\sigma^2/\mu^2)$, where μ and σ are the mean and standard deviation of cluster size. The variation in cluster size decreases as κ increases. All cluster sizes are equal when $\kappa = 1$. Cluster sizes are generated from the truncated negative binomial distribution with mean cluster size of μ =5, 10, and 20, and the imbalance parameter of κ =0.6, 0.8, and 1.0, which corresponds to severe, moderate, and no variability. We used ρ values of 0.05, 0.1, 0.3, and 0.5. Here, we test the null hypothesis $H_0 : p = p_0$ against the alternative hypothesis $H_1 : p = p_1$ with $\alpha = 0.05$, $1 - \beta = 0.9$ and (p_0, p_1)=(0.5, 0.7), (0.6, 0.7) and (0.7, 0.9). The required sample size (m) under three weighting schemes is estimated using sample size formulas (12)-(14) for given values of p_0, p_1, ρ, κ, μ, α, and β. The correlated binary data are generated by the method of Lunn and Davies [25] conditional on the estimated number of clusters and mean cluster size.

We conducted 10,000 experiments for each parameter combination. The intracluster correlation coefficient was estimated using the ANOVA estimator. The empirical powers were computed as the proportion of rejecting H_0 by the test statistics ($Z_i, i = u, c, o,$) in 10,000 samples .

Table 3 shows empirical powers and estimated sample sizes for testing $H_0 : p = 0.6$ vs. $H_1 : p = 0.7$ at significance level of 5% and 90% power. In general, these empirical powers are close to the nominal power of 90% for all three weighted estimators. All the empirical powers are within 2% of the nominal level. All the estimators provide the same results when all cluster sizes are equal ($\kappa = 1$).

Tables 4 and 5 present the sample size estimates and the corresponding empirical powers for testing $H_0 : p = 0.5$ versus $H_1 : p = 0.7$, and $H_0 : p = 0.7$ versus $H_1 : p = 0.9$, respectively. Table 4 shows that all the empirical powers lie between 86% and 92% for testing $H_0 : p = 0.5$ versus $H_1 : p = 0.7$. Table 5 shows that all test statistics are generally underpowered for testing $H_0 : p = 0.7$ versus $H_1 : p = 0.9$. That is, when p_1 is close to 1.0, the test statistics generally yields empirical powers lower than the nominal level. When a large sample size is required, for example testing $H_0 : p = 0.6$ versus $H_1 : p = 0.7$, empirical powers are close to the nominal level of 90%. Tables 3-5 shows that all three estimators generally yield similar empirical powers even though m_o is always less than or equal to m_c or m_u. If ρ is small and cluster size varies among clusters, m_o is closer to m_u. However, as ρ increases, m_o becomes closer to m_c. When $\kappa = 1$, estimated sample sizes for all three methods of assigning weights are equal.

Overall, The sample size estimates increase due to clustering effect for all three estimators as the intracluster correlation ρ increases. As κ decreases, cluster sizes vary more and more severely, the required number of clusters increases. When the mean cluster size μ increases, the required sample size decreases.

6 DISCUSSION

In this article, we introduced sample size formulas for testing proportions using the weighted sign test when binary observations are dependent within a cluster. We illustrated the sample size calculation using the data of an enzymatic diagnostic test. We evaluated the finite sample performance of the test statistics ($Z_i, i = u, c, o$), and the sample size formulas (12) - (14) for the proportion test. All the empirical powers for testing $H_0 : p = 0.6$ vs. $H_1 : p = 0.7$ are within 2% of the nominal level of 90% power for all three weighted estimators. All the empirical powers for testing $H_0 : p = 0.5$ vs. $H_1 : p = 0.7$ lie within 4% of the nominal level of 90% power for all three weighted estimators. However, all three weighted estimators become underpowered for testing $H_0 : p = 0.7$ versus $H_1 : p = 0.9$. That is, when p_1 is close to 1.0, the test statistics generally yields empirical powers less than the nominal level.

An optimal weighted estimator yields the smallest sample size estimate among three weighted estimators. The optimal sample size estimate (m_o) is closer to m_u for smaller values of ρ while m_o is closer to m_c for larger values of ρ. When $\kappa = 1$, all three estimators yield the same sample size. The required sample size increases as the intracluster correlation ρ increases. As the variability in cluster size increases, the required number of clusters increases. When the mean cluster size μ increases, the required sample size decreases.

As a future study, we will investigate the asymptotic relative efficiency (ARE) of the weighted sign test using optimal weights with respect to equal weights to observations or clusters. The ARE will be computed for both the extent of correlation among observations within clusters and the various moments of the distribution of the number of observations. We will investigate at which distribution of the number of observations the maximum ARE occurs and the what the maximum value is.

REFERENCES

1. Gonen M, Panageas KS, Larson SM. Statistical Issues in Analysis of Diagnostic Imaging Experiments with Multiple Observations per Patient. *Radiology* 2001;**221**:763-767.

2 Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika* 1986;**73**:13-22.

3. Donner A, Birkett N, Buck C. Randomization by cluster: sample size requirements and analysis. *American Journal of Epidemiology* 1981;**114**:906-914.

4. Rao JNK, Scott AJ. A simple method for the analysis of clustered binary data. *Biometrics* 1992;**48**:577585.

5. Lee E, Dubin N. Estimation and sample size considerations for clustered binary responses. *Statistics in Medicine* 1994;**13**:1241-1252.

6. Jung SH, Ahn C. Estimation of response probability in correlated binary data: a new approach. *Drug Information Journal* 2000;**34**:599 604.

7. Larocque D. The Wilcoxon signed-rank test for cluster correlated data. *Statistical Modeling and Analysis for Complex Data Problems*, P. Duchesne and B. Rémillard (eds), 2005; 309-323. New York: Springer.

8. Rosner B, Glynn RJ, Lee MLT. (2006a). The Wilcoxon signed rank test for paired comparisons of clustered data. *Biometrics* 2006;**62**:185-192.

9. Rosner B, Glynn RJ, Lee MLT. (2006b). Extension of the rank sum test for clustered data: two-group comparison with group membership defined at the subunit level. *Biometrics* 2006;**62**:1251-1359.

10. Gerard PD, Schucany WR. An enhanced sign test for dependent binary data with small numbers of clusters. *Computational Statistics and Data Analysis* 2007;**51**:4622-4632.

11. Larocque D, Nevalainen J, Oja H. A weighted multivariate sign test for cluster-correlated data. *Biometrika* 2007;**94**:267-283.

12. Datta S, Satten GA. A signed-rank test for clustered data. *Biometrics* 2008;**64**:501-507.

13. Jung SH, Kang SH, Ahn, C. Sample size calculation for clustered binary data. *Statistics in Medicine* 2001;**20**:1971-1982.

14. Hu F, Schucany WR, Ahn C. Nonparametric sample size estimation for sensitivity and specificity with multiple observations per subject. *Drug Information Journal* in press.

15. Noether GE. Sample size determination for some common nonparametric tests. *Journal of the American Statistical Association* 1987;**82**:645-647.

16. Cox DR, Wermuth N. A note on the quadratic exponential binary distribution. *Biometrika* 1994;**81**:403-408.

17. Molenberghs G, Ryan LM. An Exponential family model for clustered multivariate binary data. *Environmetrics* 1999;**10**:279-300.

18. Mak TK. Analysing intraclass correlation for dichotomous variables. *Applied Statistics* 1988;**37**:344-352.

19. Donald A, Donner A. A simulation study of the analysis of sets of 2 x 2 contingency tables under cluster sampling: Estimation of a common odds ratio. *J. Amer. Statist. Assoc.* 1990;**85**:537-543.

20. Ridout MS, Demetrio CG, Firth D. Estimating intraclass correlation for binary data. *Biometrics* 1999;**55**:137-148.

21. Hujoel P, Moulton L, Loesche W. Estimation of sensitivity and specify of site-specific diagnostic tests. *Journal of Periodontal Research* 1990;**25**:193-196.

22. 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.

23. Ahn C, Hu F, Skinner CS. Effect of imbalance and intracluster correlation coefficient in cluster randomized trials with binary outcomes. *Computational Statistics and Data Analysis* 2009;**53**:596-602.

24. Johnson NL, Kotz S, Kemp AW. *Univariate Discrete Distributions*, 2nd edition. New York: Wiley, 1992.

25. Lunn AD, Davies SJ. A note on generating correlated binary variables. *Biometrika* 1998;**85**:487-490.

Table 1: Pilot data of S_i/n_i from $m = 29$ subjects (clusters).

3/6, 2/6, 2/4, 5/6, 4/5, 5/5, 4/6, 3/4, 2/4, 3/4, 5/5, 4/4, 6/6, 3/3, 5/6, 1/2, 4/6, 0/4, 5/6, 4/5, 4/6, 0/6, 4/5, 3/5, 0/2, 2/6, 2/4, 5/5, 4/6

Table 2: Distribution of n_i's.

	m				
	2	3	4	5	6
Relative frequency	2/29	1/29	7/29	7/29	12/29
Projected mass function	0.05	0.05	0.25	0.25	0.4

Table 3: Empirical powers (%) and sample size estimates in parentheses for testing $H_0 : p = 0.6$ vs. $H_1 : p = 0.7$ with $\alpha = 0.05$ and β=0.1 from 10,000 simulations.

κ	ρ^a	μ^b	Z_u	Z_c	Z_o
0.6	0.05	5	91 (66)	92 (99)	91 (64)
		10	89 (43)	90 (62)	90(40)
		20	89(32)	90(40)	89(28)
	0.1	5	90(84)	92(106)	91(77)
		10	90(62)	90(72)	90(53)
		20	88(51)	90(51)	89(41)
	0.3	5	90(154)	91(136)	90(121)
		10	89(137)	89(109)	90(100)
		20	89(129)	90(93)	90(88)
	0.5	5	89(224)	90(166)	90(159)
		10	89(212)	90(146)	90(142)
		20	90(206)	90(135)	90(133)
0.8	0.05	5	89(61)	90(76)	90(60)
		10	89(38)	89(44)	89(37)
		20	88(27)	89(28)	88(25)
	0.1	5	90(74)	90(84)	89(71)
		10	89(52)	90(55)	90(49)
		20	88(41)	89(39)	89(37)
	0.3	5	90(124)	90(119)	90(112)
		10	89(107)	90(96)	89(93)
		20	88(99)	90(84)	89(83)
	0.5	5	90(174)	90(153)	89(151)
		10	89(162)	89(137)	90(136)
		20	88(156)	89(128)	90(128)
1	0.05	5	90(58)	90(58)	90(58)
		10	89(35)	90(35)	89(35)
		20	89(24)	89(24)	89(24)
	0.1	5	90(68)	90(68)	90(68)
		10	89(46)	89(46)	89(46)

a: ρ is an intracluster correlation coefficient

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

Table 3 (Continued): Empirical powers (%) and sample size estimates in parentheses for testing $H_0 : p = 0.6$ vs. $H_1 : p = 0.7$ with $\alpha = 0.05$ and β=0.1 from 10,000 simulations.

κ	ρ^a	μ^b	Z_u^c	Z_c^c	Z_o^c
		20	89(35)	89(35)	89(35)
	0.3	5	89(106)	89(106)	89(106)
		10	89(89)	89(89)	89(89)
		20	89(81)	89(81)	89(81)
	0.5	5	90(144)	90(144)	90(144)
		10	90(132)	90(132)	90(132)
		20	90(126)	90(126)	90(126)

a: ρ is an intracluster correlation coefficient

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

Table 4: Empirical powers (%) and sample size estimates in parentheses for testing $H_0 : p = 0.5$ vs. $H_1 : p = 0.7$ with $\alpha = 0.05$ and β=0.1 from 10,000 simulations.

κ	ρ^a	μ^b	Z_u	Z_c	Z_o
0.6	0.05	5	89 (17)	91 (26)	90 (17)
		10	87 (11)	89 (16)	90(11)
		20	89(9)	90(11)	86(7)
	0.1	5	91(22)	92(28)	90(20)
		10	88(16)	90(19)	88(14)
		20	87(13)	88(13)	89(11)
	0.3	5	90(40)	91(35)	90(31)
		10	90(36)	89(28)	90(26)
		20	89(33)	89(24)	89(23)
	0.5	5	90(58)	91(43)	90(41)
		10	89(55)	90(38)	90(37)
		20	89(53)	89(35)	89(34)
0.8	0.05	5	88(16)	90(20)	89(16)
		10	88(10)	90(12)	88(10)
		20	86(7)	91(8)	88(7)
	0.1	5	88(19)	90(22)	89(19)
		10	89(14)	89(14)	88(13)
		20	88(11)	87(10)	88(10)
	0.3	5	89(32)	90(31)	89(29)
		10	89(28)	90(25)	89(24)
		20	89(26)	89(22)	89(22)
	0.5	5	90(45)	91(40)	90(39)
		10	90(42)	89(35)	89(35)
		20	89(40)	89(33)	89(33)
1	0.05	5	89(15)	89(15)	89(15)
		10	87(9)	87(9)	87(9)
		20	86(6)	86(6)	86(6)
	0.1	5	90(18)	90(18)	90(18)
		10	88(12)	88(12)	88(12)

a: ρ is an intracluster correlation coefficient

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

Table 4 (Continued): Empirical powers (%) and sample size estimates in parentheses for testing $H_0 : p = 0.5$ vs. $H_1 : p = 0.7$ with $\alpha = 0.05$ and β=0.1 from 10,000 simulations.

κ	ρ^a	μ^b	Z_u	Z_c	Z_o
		20	87(9)	87(9)	87(9)
	0.3	5	89(28)	89(28)	89(28)
		10	89(23)	89(23)	89(23)
		20	89(21)	89(21)	89(21)
	0.5	5	89(37)	89(37)	89(37)
		10	89(34)	89(34)	89(34)
		20	90(33)	90(33)	90(33)

a: ρ is an intracluster correlation coefficient

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

Table 5: Empirical powers (%) and sample size estimates in parentheses for testing $H_0 : p = 0.7$ vs. $H_1 : p = 0.9$ with $\alpha = 0.05$ and β=0.1 from 10,000 simulations.

κ	ρ^a	μ^b	Z_u	Z_c	Z_o
0.6	0.05	5	86 (12)	89 (17)	83 (11)
		10	84 (8)	88 (11)	81(7)
		20	83(6)	86(7)	81(5)
	0.1	5	86(15)	89(19)	86(14)
		10	84(11)	88(13)	85(10)
		20	82(9)	86(9)	80(7)
	0.3	5	87(27)	89(24)	87(21)
		10	86(24)	87(19)	86(18)
		20	86(23)	85(16)	87(16)
	0.5	5	88(39)	89(29)	89(28)
		10	87(37)	88(26)	88(25)
		20	87(36)	88(24)	87(23)
0.8	0.05	5	86(13)	90(20)	86(11)
		10	85(7)	87(8)	85(7)
		20	83(5)	84(5)	84(5)
	0.1	5	86(13)	87(15)	86(13)
		10	82(9)	86(10)	85(9)
		20	80(7)	83(7)	82(7)
	0.3	5	87(22)	87(21)	87(20)
		10	85(19)	86(17)	84(16)
		20	84(17)	85(15)	85(15)
	0.5	5	87(30)	88(27)	87(26)
		10	87(28)	87(24)	88(24)
		20	86(27)	86(22)	86(22)
1	0.05	5	85(10)	85(10)	85(10)
		10	85(6)	85(6)	86(6)
		20	86(5)	86(5)	86(5)
	0.1	5	85(12)	86(12)	85(12)
		10	83(8)	83(8)	84(8)

a: ρ is an intracluster correlation coefficient

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

Table 5 (Continued): Empirical powers (%) and sample size estimates in parentheses for testing $H_0 : p = 0.7$ vs. $H_1 : p = 0.9$ with $\alpha = 0.05$ and β=0.1 from 10,000 simulations.

κ	ρ^a	μ^b	Z_u	Z_c	Z_o
		20	78(6)	78(6)	78(6)
	0.3	5	87(19)	87(19)	88(19)
		10	86(16)	87(16)	86(16)
		20	83(14)	83(14)	84(14)
	0.5	5	87(25)	87(25)	87(25)
		10	88(23)	87(23)	88(23)
		20	87(22)	86(22)	87(22)

a: ρ is an intracluster correlation coefficient

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