Published in final edited form as:

Drug Inf J. 2012 July 2; 46(4): 428–433. doi:10.1177/0092861512449818.

Relative Efficiency of Unequal Versus Equal Cluster Sizes for the Nonparametric Weighted Sign Test Estimators in Clustered Binary Data

Chul Ahn, PhD¹, Fan Hu, PhD², and Seung-Chun Lee, PhD³

¹ Department of Clinical Sciences, UT Southwestern Medical Center, Dallas TX USA

² Department of Statistical Science, Southern Methodist University, Dallas, TX USA

³ Department of Statistics, Hanshin University, Osan, Korea

Abstract

We consider analysis of clustered binary data from multiple observations for each subject in which any two observations from a subject are assumed to have a common correlation coefficient. In the weighted sign test on proportion in clustered binary data, three weighting schemes are considered: equal weights to observations, equal weights to clusters and the optimal weights that minimize the variance of the estimator. Since the distribution of cluster sizes may not be exactly specified before the trial starts, the sample size is usually determined using an average cluster size without taking into account any potential imbalance in cluster size even though cluster size usually varies among clusters. In this paper we investigate the relative efficiency (RE) of unequal versus equal cluster sizes for clustered binary data using the weighted sign test estimators. The REs are computed as a function of correlation among observations within each subject and the various cluster size distributions. The required sample size for unequal cluster sizes will not exceed the sample size for an equal cluster size multiplied by the maximum RE. It is concluded that the maximum RE for various cluster size distributions considered here does not exceed 1.50, 1.61 and 1.12 for equal weights to observations, equal weights to clusters and optimal weights, respectively. It suggests sampling 50%, 61% and 12% more clusters depending on the weighting schemes than the number of clusters computed using an average cluster size.

Keywords

intraclass correlation coefficient; variable cluster sizes; sample size

1. Introduction

In this paper we focus on clustered binary data which are made from multiple observations for each subject (called a cluster). In this case, observations from each subject are correlated although those from different subjects are independent. For example, in a radiologic study, each subject may contribute multiple lesions to the study and an observation is made from each lesion. Any two distinct observations from each subject are often assumed to have an equal correlation coefficient, which will be defined more rigorously in next section.

Since the distribution of cluster sizes may not be exactly specified before the trial starts, the sample size is usually determined using an average cluster size without taking into account any potential imbalance in cluster size even though cluster size usually varies among clusters. The method was called the average size method by Manatunga *et al.* (2001). Throughout this paper, the sample size refers to the number of clusters. Since the number of observations frequently varies among clusters, the relative efficiency (RE) can be used to compute the sample size needed for unequal cluster sizes. The required sample size for varying cluster size can be obtained by multiplying the RE to the required sample size for an average cluster size. The required sample size for unequal cluster sizes does not exceed the required sample size for an equal cluster size multiplied by the maximum RE. In this paper, we will investigate the RE and maximum RE for various cluster size distributions, which can be used to determine the required sample size for varying cluster sizes.

Relative efficiency (RE) of unequal versus equal cluster sizes was investigated for binary outcomes with mixed effects logistic regression models using first-order marginal quasi-likelihood (MQL) estimation and second-order penalized quasi-likelihood (PQL) estimation in cluster randomized trials (Candel and van Breukelen, 2010). Candel and van Breukelen (2010) illustrated that the losses of efficiency due to variable cluster sizes was at most 14% and 25% using MQL and PQL, respectively. That is the loss of efficiency can be compensated by sampling 14% or 25% more clusters.

Sample size formulas were derived for a single-arm design using nonparametric weighted sign tests for binary outcomes by Hu *et al.* (2010) and Ahn *et al.* (2011). In this paper, we will investigate the RE and maximum RE of unequal versus equal cluster sizes in the estimation of a response rate in a single-arm design with clustered binary outcomes, which can be used to compute the required sample size for varying cluster sizes. The RE will be computed for weighted sign test estimators that assign equal weights to observations, equal weights to clusters, and the optimal weights that minimize the variance of the estimator.

The rest of the paper is organized as follows. Section 2 presents statistical methods for the weighted sign test statistics, and Section 3 presents the formulas for the RE of equal cluster size with respect to unequal cluster sizes. Section 4 provides the computation of the RE to a real example. Finally, we conclude with a discussion.

2. Statistical Method

Let n be the total number of clusters in an experiment, and m_i be the number of observations in the i th cluster ($i=1, \dots, n$), which is random with the probability mass function $f(\cdot)$. Here we assume that the cluster size m_i 's are small compared to n , so that asymptotic theories are applied with respect to n . For the j th observation ($j=1, \dots, m_i$) of cluster i , let X_{ij} be a binary random variable indicating a response ($X_{ij}=1$) or no response ($X_{ij}=0$). We coded response as 1 and no response as -1 due to merits of this coding scheme (Molenberghs and Ryan, 1999; Cox and Wermuth, 1994). We assume that observations in a cluster are exchangeable in the sense that, given m_i , $X_{i1}, \dots, X_{im_i}$ have a common intraclass correlation coefficient, $\rho = \text{corr}(X_{ij}, X_{ij'})$ for $j \neq j'$, and a marginal response probability $P(X_{ij}=1) = p$, $0 < p < 1$, for all i and j . This model is often called the common correlation model (Mak 1988). Note that the model includes the beta-binomial model, the correlated binomial model (Kupper and Haseman 1978), those of Donner, Birkett and Buck (1981) and George and Kodell (1996) as special cases. We test the null hypotheses $H_0: p = p_0$ versus $H_1: p = p_1$ for $p_1 > p_0$. Let

$y_i = \sum_{j=1}^{m_i} X_{ij}$ be the difference between the total number of responses and the total number of non-responses in the i th cluster. Under the exchangeability assumption, given m_i , we have

$$E(y_i) = m_i(2p - 1) \quad \text{and} \quad \text{var}(y_i) = 4m_i p(1 - p) \{1 + (m_i - 1)\rho\}.$$

Let $(w_1, \dots, w_n)$ be a set of weights assigned to clusters such that $w_i \geq 0$ and $(1/n) \sum_i w_i m_i = 1$. Then we have a class of statistics given by

$$T = \sum_{i=1}^n w_i y_i = \sum_{i=1}^n w_i (m_i^+ - m_i^-), \quad (1)$$

where m_i^+ and m_i^- are the total number of responses and non-responses in the i th cluster, respectively. The variance of T is given by

$$\text{Var}(T) = 4p_0(1 - p_0) \sum_{i=1}^n w_i^2 m_i \{1 + (m_i - 1)\widehat{\rho}\}. \quad (2)$$

The weighted sign test is defined by

$$z = \frac{\sum_{i=1}^n w_i (m_i^+ - m_i^-) - n(2p_0 - 1)}{\sqrt{4p_0(1 - p_0) \sum_{i=1}^n w_i^2 m_i \{1 + (m_i - 1)\widehat{\rho}\}}},$$

which follows the standard normal distribution as $n \rightarrow \infty$, where $\widehat{\rho}$ is the estimate of ρ that can be obtained by the ANOVA method (Donald and Donner, 1990). The simulation studies of Ridout *et al.* (1999) showed that the ANOVA estimator performed well under the common-correlation model, $\rho = \text{corr}(X_{ij}, X_{ij'})$.

Assigning equal weights to each individual observation ($w_i = n / \sum_{i=1}^n m_i$), the statistic (1) becomes

$$T_u = \frac{n}{\sum_{i=1}^n m_i} \sum_{i=1}^n (m_i^+ - m_i^-), \quad (3)$$

and the variance of T_u is

$$\text{Var}(T_u) = 4p_0(1 - p_0) \frac{n^2}{\sum_{i=1}^n m_i^2} \sum_{i=1}^n m_i \{1 + (m_i - 1)\widehat{\rho}\}. \quad (4)$$

With equal weights to each cluster ($w_i = 1/m_i$), the statistic (1) becomes the nonparametric statistic of Datta and Satten (2008)

$$T_c = \sum_{i=1}^n \left(\frac{m_i^+ - m_i^-}{m_i} \right), \quad (5)$$

and its corresponding variance of the test statistic, $\text{Var}(T_c)$, is

$$Var(T_c) = 4p_0(1-p_0) \sum_{i=1}^n \{1 + (m_i - 1)\widehat{\rho}\} / m_i. \quad (6)$$

The optimal weight that minimizes the variance of T in (1) can be obtained by minimizing $Var(T)$ subject to $(1/n) \sum_{i=1}^n w_i m_i = 1$. The weight that minimizes variance is

$$w_i^* = \frac{n\{1 + (m_i - 1)\rho\}^{-1}}{\sum_{i=1}^n m_i\{1 + (m_i - 1)\rho\}^{-1}}, \quad (7)$$

and the corresponding variance is

$$Var(T_0) = 4p_0(1-p_0) \sum_{i=1}^n w_i^{*2} m_i \{1 + (m_i - 1)\widehat{\rho}\}, \quad (8)$$

where w_i^* is given in (7), in which the weight depends on ρ .

When the cluster sizes are not equal among clusters, one may replace m_i by an advanced estimate of the average cluster size, which was called the average size method by Manatunga *et al.* (2001). The average size method generally underestimates the variance of the estimator and consequently underestimates required sample size (Donner and Klar, 2000). When the variance is computed assuming equal cluster sizes with $m_i = \bar{m}$, the variance of T is given by

$$Var(T_e) = 4p_0(1-p_0) \frac{n\{1 + (\bar{m} - 1)\widehat{\rho}\}}{\bar{m}}, \quad (9)$$

When $m_i = m$, all the weighted estimators have the same variance, $Var(T_u) = Var(T_c) = Var(T_o) = Var(T_e)$. When $\rho = 0$, $Var(T_e) = Var(T_u) = Var(T_o)$ while $Var(T_c) = Var(T_c) = Var(T_o)$ for $\rho = 1$. In this paper, the weighted test statistics corresponding to T_u, T_c, T_o and T_e are given by Z_u, Z_c, Z_o and Z_e , respectively.

Let m be the random variable with mean Θ and variance τ^2 corresponding to the cluster size, and $E[\cdot]$ is the expectation with respect to the distribution of the cluster size.

For large n , $\sqrt{n}(\widehat{p}_i - p)$, $i = u, c, o, e$, is approximately normal with mean 0 and variance

$$\begin{aligned} \sigma_u^2 &= 4p(1-p) \frac{E[m\{1+(m-1)\rho\}]}{E(m)^2} \\ \sigma_c^2 &= 4p(1-p) E\left[\frac{1+(m-1)\rho}{m}\right] \\ \sigma_o^2 &= 4p(1-p) \frac{1}{E[m\{1+(m-1)\rho\}^{-1}]} \\ \sigma_e^2 &= \frac{4p(1-p)E[1+(m-1)\rho]}{E[m]} \end{aligned}$$

Error! Bookmark not defined. When cluster size is constant, i.e. $m_i = m$, all variances are equal. Let $K = \{1, 2, \dots, K\}$ denote the support of the mass function $f(k)$ of the cluster size. According to the definition of the RE of two tests given by Noether (1955), the RE of T_e with respect to T_u is attained by

$$RE_u = \frac{\sigma_u^2}{\sigma_e^2} = \frac{E[m\{1+(m-1)\rho\}]}{E[m]^2} \frac{E[m]}{E[1+(m-1)\rho]} \\ = \left(\frac{\sum_{k=1}^K k\{1+(k-1)\rho\}f(k)}{\left[\sum_{k=1}^K kf(k)\right]^2} \right) \left(\frac{\sum_{k=1}^K kf(k)}{\sum_{k=1}^K \{1+(k-1)\rho\}f(k)} \right)$$

The RE of T_e with respect to T_c is given by

$$RE_c = \frac{\sigma_c^2}{\sigma_e^2} = E\left[\frac{1+(m-1)\rho}{m}\right] \frac{E[m]}{E[1+(m-1)\rho]} \\ = \left(\sum_{k=1}^K \frac{1+(k-1)\rho}{k} f(k) \right) \left(\frac{\sum_{k=1}^K kf(k)}{\sum_{k=1}^K \{1+(k-1)\rho\}f(k)} \right)$$

The RE of T_e with respect to T_o is given by

$$RE_o = \frac{\sigma_o^2}{\sigma_e^2} = \frac{1}{E[m\{1+(m-1)\rho\}^{-1}]} \frac{E[m]}{E[1+(m-1)\rho]} \\ = \left(\sum_{k=1}^K \frac{k}{1+(k-1)\rho} f(k) \right)^{-1} \left(\frac{\sum_{k=1}^K kf(k)}{\sum_{k=1}^K \{1+(k-1)\rho\}f(k)} \right)$$

We investigate RE_u , RE_c and RE_o for various cluster size distributions which are displayed in Table 1. Four cluster size distributions are considered for the cluster size of five and ten, respectively. f_1 and f_5 are symmetric. f_2, f_3, f_6 and f_7 are asymmetric. f_4 and f_8 are uniformly distributed. Here, we label f_1 to f_8 as “narrow, centered”, “narrow, pushed to 1”, “narrow, pushed to 5”, “narrow, uniform”, “wide, centered”, “wide, pushed to 1”, “wide, pushed to 5”, and “wide, uniform”. The RE_u , RE_c , and RE_o are plotted in Figures 1 to 3. Here, RE_o , RE_c and RE_o , RE_u . The sample size is usually computed using an average cluster size without taking into account any potential imbalance in cluster size since the cluster size distribution may not be exactly specified before the trial starts. If the sample size is computed under a naive assumption using the average cluster size, the results should be inflated by a factor of RE to account for the loss of efficiency that is produced by disparate cluster sizes. Table 2 shows the maximum REs for cluster size distributions f_1 to f_8 , respectively. The required sample size for unequal cluster sizes does not exceed the required sample size for an equal cluster size multiplied by the maximum RE. The maximum RE for various cluster size distributions considered here are less than or equal to 1.50, 1.61 and 1.12 for equal weights to observations, equal weights to clusters and optimal weights, respectively. Under the optimal weights, the maximum RE for various cluster size distributions considered here does not exceed 1.12, which suggests sampling 12% more clusters than the number of clusters computed using an average cluster size.

While RE_c decreases exponentially as ρ increases, RE_u increases exponentially as ρ increases. While RE_o is closer to RE_u than RE_c for small values of ρ , RE_o is closer to RE_c than RE_u for large values of ρ . The values of RE_o are always less than or equal to those of RE_c and RE_u . For $\rho=0$, $RE_u=RE_o=1$ regardless of the cluster size distribution, and

$$RE_c = E(m) E\left(\frac{1}{m}\right), \text{ For } \rho=1, RE_c=RE_o=1 \text{ regardless of the cluster size distribution, and} \\ RE_u = E(m^2)/E(m)^2.$$

4 Example

Here we provide, as an example, the estimation of the sensitivity and the specificity of an enzymatic diagnostic test (Hujuel *et al.*, 1990). An enzymatic diagnostic test was employed to decide whether a site was infected by two organisms, *treponema denticola* and *bacteroides ginggivalis*. Each subject contributed a different number of infected sites which were determined by the gold standard (an antibody assay against the two organisms). In a sample of 29 subjects for positive test results the relative frequency of cluster size is given by

$$(f(2), f(3), f(4), f(5), f(6)) = (2/29, 1/29, 7/29, 12/29).$$

Suppose we want to design an experiment to test the hypothesis $H_0: p=0.6$ versus $H_1: p=0.7$ based on the above data. From the observed data, we get the intracluster correlation coefficient estimate $\hat{\rho}=0.2$ using the ANOVA method, and estimate the mean of m to be 4.9. We will use these estimated values in computations for the sample size of the next experiment with the sample size formula for the nonparametric sign test (Ahn *et al.*, 2011)

$$n_u = \frac{(z_{1-\alpha/2} + z_{1-\beta})^2}{(p_1 - p_0)^2} \left\{ \frac{1 + (E(m) - 1)\rho}{E(m)} \right\} p_0(1 - p_0).$$

We obtain a sample size of 69 and 92 at 80% and 90% power, respectively at a 5% significance level using an average size method. The frequency of cluster size from the above data is used to compute the RE as a cluster size distribution. The RE is displayed in Figure 4. For the positive test results, the maximum REs are 1.06, 1.09, and 1.02 using equal weights to observations, equal weights to clusters and optimal weights, respectively. It suggests sampling 6%, 9% and 2% more clusters depending on the weighting schemes than the number of clusters computed using an average cluster size.

Similarly, in a sample of 21 subjects for negative test results the relative frequency of cluster size is given by

$$(f(2), f(3), f(4), f(5), f(6)) = (8/21, 2/21, 9/21, 1/21, 1/21)$$

and is also used to compute the RE as a cluster size distribution. The RE is displayed in Figure 4. For the negative test results, the maximum REs are 1.12, 1.14, and 1.03 using equal weights to observations, equal weights to clusters and optimal weights, respectively. It suggests sampling 12%, 14% and 3% more clusters depending on the weighting schemes than the number of clusters computed using an average cluster size.

5 Discussion

In this paper, we investigate the RE and maximum RE of unequal versus equal cluster sizes for clustered binary data in a single-arm clinical trial design, which has been widely used for pilot studies and early-stage clinical trials. Candel and van Breukelen (2010) investigated the losses of efficiency using mixed effects logistic regression model under cluster randomized trials such as community intervention trials in which the cluster sizes are usually large. In contrast to the study design for community intervention trials, the cluster sizes we encounter in pilot clinical trials and early-stage clinical trials are usually small.

The standard method for calculating sample size for clustered binary outcomes assumes an equal cluster size since the number of observations may not be exactly specified before the trial starts. However, the number of observations frequently varies among clusters.

Since equal cluster sizes yield the minimum required sample size (*number of clusters*), a good question is how many more clusters are needed by variation of cluster size. This paper addresses the RE of unequal versus equal cluster sizes for clustered binary outcomes. The RE is investigated numerically for a range of cluster size distributions with weighted sign test estimators. Formulas are presented for computing the RE as a function of the cluster size and the intraclass correlation, which can be used to adjust the sample size for varying cluster sizes. Focusing on variance estimation under different weighting schemes, this paper shows that the maximum REs of unequal versus equal cluster sizes do not exceed 1.50, 1.61 and 1.12 for equal weights to observations, equal weights to clusters and optimal weights, respectively. It suggests sampling 50%, 61% and 12% more clusters depending on the weighting schemes than the number of clusters computed using an average cluster size.

Under the weights minimizing the variance, the maximum RE for various cluster size distributions considered here does not exceed 1.12, which suggests sampling 12% more clusters than the number of clusters computed using an average cluster size. The study will be underpowered if one assumes equal cluster size for sample size determination, but ends up having unequal cluster sizes. The sample size should be increased by the amounts shown previously in order to compensate for the decreased efficiency that is produced by disparate cluster sizes. The increase in the sample size can range from very little (3% to 12%) for the optimal weighting scheme to 50% or more for the other weighting schemes. Failure to adjust is to ignore relevant components of the data.

We regard RE_u , RE_c and RE_o as functions of cluster size distribution, $f(k)$. Research is in progress to find what the maximum value is, and where the maximum of RE_u , RE_c and RE_o occurs from $K=\{1,2,\dots,K\}$. Here, $K=\{1,2,\dots,K\}$ denotes the support of the mass function $f(k)$ of the cluster size.

Acknowledgments

This work was supported in part by NIH grants UL1 RR024982, P30CA142543, P50CA70907, and DK081872.

REFERENCES

1. Ahn C, Hu F, Schucany WR. Sample size considerations for clustered binary data with nonparametric methods using different weighting schemes. *Statistics in Biopharmaceutical Research*. 2011; 3(1):65–72. [PubMed: 21339864]
2. Cox DR, Wermuth N. A note on the quadratic exponential binary distribution. *Biometrika*. 1994; 81:403–408.
3. Datta S, Satten GA. A signed-rank test for clustered data. *Biometrics*. 2008; 64:501–507. [PubMed: 17970820]
4. Donald A, Donner A. A simulation study of the analysis of sets of 2×2 contingency tables under cluster sampling: Estimation of a common odds ratio. *J. Amer. Statist. Assoc.* 1990; 85:537–543.
5. Donner A, Birkett N, Buck C. Randomization by cluster: sample size requirements and analysis. *American Journal of Epidemiology*. 1981; 114:906–914. [PubMed: 7315838]
6. Donner, A.; Klar, N. Design and analysis of cluster randomization trials in health research. Oxford University Press; 2000.
7. George EO, Kodell RL. Tests of independence, treatment heterogeneity, and doserelated trend with exchangeable binary data. *Journal of the American Statistical Association*. 1996; 91:1602–1611.

8. Hu F, Schucany WR, Ahn C. Nonparametric sample size estimation for sensitivity and specificity with multiple observations per subject. *Drug Information Journal*. 2010; 44:609–616. [PubMed: 22114363]
9. Hujoel P, Moulton L, Loesche W. Estimation of sensitivity and specificity of site-specific diagnostic tests. *Journal of Periodontal Research*. 1990; 25:193–196. [PubMed: 2197400]
10. Kupper LL, Haseman JK. The use of a correlated binomial model for the analysis of certain toxicological experiments. *Biometrics*. 1978; 34:69–76. [PubMed: 564719]
11. Mak TK. Analysing intraclass correlation for dichotomous variables. *Applied Statistics*. 1988; 37:344–352.
12. Manatunga AK, Hudgens MG, Chen S. Sample size estimation in cluster randomized studies with varying cluster size. *Biometrical Journal*. 2001; 43:75–86.
13. Molenberghs G, Ryan LM. An Exponential family model for clustered multivariate binary data. *Environmetrics*. 1999; 10:279–300.
14. Noether GE. On a theorem by Pitman. *The Annals of Mathematical Statistics*. 1955; 26:64–68.
15. Ridout MS, Demétrio CGB, Firth D. Estimating intraclass correlation for binary data. *Biometrics*. 1999; 55:137–148. [PubMed: 11318148]
16. Candel M, van Breukelen GJ. Sample size adjustments for varying cluster sizes in cluster randomized trials with binary outcomes analyzed with second-order PQL mixed logistic regression. *Stat Med*. 2010; 29:1488–1501. [PubMed: 20101669]

Figure 1.

The relationship between RE_u and ρ for the cluster size distribution “narrow, centered”, “narrow, pushed to 1”, “narrow, pushed to 5”, “narrow, uniform”, “wide, centered”, “wide, pushed to 1”, “wide, pushed to 10”, and “wide, uniform”.

Figure 2.

The relationship between RE_c and ρ for the cluster size distribution “narrow, centered”, “narrow, pushed to 1”, “narrow, pushed to 5”, “narrow, uniform”, “wide, centered”, “wide, pushed to 1”, “wide, pushed to 10”, and “wide, uniform”.

Figure 3.

The relationship between RE_o and ρ for the cluster size distribution “narrow, centered”, “narrow, pushed to 1”, “narrow, pushed to 5”, “narrow, uniform”, “wide, centered”, “wide, pushed to 1”, “wide, pushed to 10”, and “wide, uniform”.

Figure 4.
 RE_u , RE_c , and RE_o for positive and negative tests.

Table 1

Cluster size distribution.

K	dist.	$f(1)$	$f(2)$	$f(3)$	$f(4)$	$f(5)$	$f(6)$	$f(7)$	$f(8)$	$f(9)$	$f(10)$
5	f_1	0.1	0.2	0.4	0.2	0.1					
	f_2	0.4	0.3	0.15	0.1	0.05					
	f_3	0.05	0.1	0.15	0.3	0.4					
	f_4	0.2	0.2	0.2	0.2	0.2					
10	f_5	0.02	0.03	0.05	0.15	0.25	0.25	0.15	0.05	0.03	0.02
	f_6	0.3	0.2	0.15	0.11	0.08	0.06	0.04	0.03	0.02	0.01
	f_7	0.01	0.02	0.03	0.04	0.06	0.08	0.11	0.15	0.2	0.3
	f_8	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1

f_1 – f_8 : “narrow, centered”, “narrow, pushed to 1”, “narrow, pushed to 5”, “narrow, uniform”, “wide, centered”, “wide, pushed to 1”, “wide, pushed to 10”, “wide, uniform”.

Table 2

Maximum relative efficiency (RE) for cluster size distributions f_1 to f_8 .

Maximum RE	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8
RE_u	1.13	1.31	1.09	1.22	1.10	1.50	1.08	1.27
RE_c	1.20	1.33	1.18	1.36	1.17	1.61	1.18	1.58
RE_o	1.04	1.08	1.03	1.07	1.03	1.12	1.03	1.09

f_1 – f_8 : “narrow, centered”, “narrow, pushed to 1”, “narrow, pushed to 5”, “narrow, uniform”, “wide, centered”, “wide, pushed to 1”, “wide, pushed to 10”, “wide, uniform”.