Published in final edited form as:

Comput Stat Data Anal. 2014 January ; 69: . doi:10.1016/j.csda.2013.07.037.

A GEE Approach to Determine Sample Size for Pre- and Post-Intervention Experiments with Dropout

Song Zhang,

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

Jing Cao, and

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

Chul Ahn

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

Abstract

Pre- and post-intervention experiments are widely used in medical and social behavioral studies, where each subject is supposed to contribute a pair of observations. In this paper we investigate sample size requirement for a scenario frequently encountered by practitioners: All enrolled subjects participate in the pre-intervention phase of study, but some of them will drop out due to various reasons, thus resulting in missing values in the post-intervention measurements. Traditional sample size calculation based on the McNemar's test could not accommodate missing data. Through the GEE approach, we derive a closed-form sample size formula that properly accounts for the impact of partial observations. We demonstrate that when there is no missing data, the proposed sample size estimate under the GEE approach is very close to that under the McNemar's test. When there is missing data, the proposed method can lead to substantial saving in sample size. Simulation studies and an example are presented.

1 Introduction

Pre- and post-intervention experiments have been widely used in medical and social behavioral studies (Spleen et al. 2012, Rossi et al. 2010, Wajnberg et al. 2010, Knudtson et al. 2010, Zieschang et al. 2010). One distinct feature of a pre-post study is that each patient contributes a pair of observations (observations at pre-intervention and post-intervention). Thus statistical inference needs to account for within-subject correlation. The McNemar's test (McNemar 1947) has been the most widely used approach to detecting the intervention effect on a binary outcome in pre-post studies. Sample size calculation for studies involving the McNemar's test has been explored by many researchers. Miettinen (1968) and Connor (1987) derived sample size formulas through a conditional procedure based on the approximately normal distribution of the McNemar's test statistic given the number of discordant pairs. Shork & Williams (1980) presented an exact formula for the unconditional case. Lachin (1992) compared different unconditional sample size expressions relative to the exact power function. Lu & Bean (1995) investigated sample size requirement for one-sided

© 2013 Elsevier B.V. All rights reserved.

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, Chul.Ahn@UTSouthwestern.edu.

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.

equivalence of sensitivities based on the McNemar's test. Other literatures related to this topic include but are not limited to Cochran (1950), Royston (1993), Selicato & Muller (1998), and Julious et al. (1999). The existing literatures, however, have not addressed the issue of incomplete observations frequently encountered by practitioners. Specifically, some subjects might participate in the pre-intervention phase of the study, but then drop out of the study, resulting in missing values for post-intervention measurements. Thus the pre-intervention measurements are observed in all subjects, but the post-intervention measurements are likely to be missing in some subjects. The McNemar's test could not use incomplete pair of observations, so they have to be excluded from analysis. Accordingly, in practice, to account for dropout from study, researchers have first estimated the sample size under complete observations (denoted by n_0), and then calculated the final sample size by $n = n_0/q$, where q is the expected proportion of subjects who complete both the pre- and post-intervention assessments (i.e., $1 - q$ is the dropout rate). Such an adjustment for missing data might be unsatisfactory. We will show that the specific impact of dropout on sample size depends on factors such as the pre-intervention response rate, the post-intervention response rate (or equivalently, the intervention effect), and the within-subject correlation. These factors, however, are ignored by the traditional adjustment for missing data, which might lead to an unnecessarily inflated sample size and waste in clinical resources.

To utilize information from incomplete pairs, we use a mixed logistic regression model approach instead of the McNemar's test. The intervention effect is represented by a regression coefficient and estimated by the generalized estimating equation (GEE) method (Liang & Zeger 1986). The GEE method has been widely used to model correlated data and accommodate missing values in longitudinal and clustered studies (Zeger et al. 1988, Norton et al. 1996). Sample size calculation based on the GEE approach has been explored by many researchers. For example, Liu & Liang (1997) developed a sample size formula based on a generalized score test. Rochon (1998) proposed a sample size formula using a non-central version of the Wald χ^2 test statistics. Jung & Ahn (2005) investigated sample size calculation to detect rate of changes between two treatment groups. In this paper we present a closed-form sample size formula based on the GEE method that appropriately accounts for incomplete observations in pre-post studies. We also explore the connection between the sample sizes under the GEE approach and the McNemar's test. We demonstrate that with complete data, the sample size estimated under GEE is very close to that under the McNemar's test. When subjects are likely to drop out of study, however, the proposed approach can lead to substantial saving in sample size compared to that based on the McNemar's test with the traditional adjustment for missing values.

The paper is arranged as follows. In Section 2 we briefly review the sample size formula based on the McNemar's test. In Section 3 we present the GEE sample size approach based on mixed logistic regression models. Simulation studies and an example are presented in Section 4 and 5, respectively. We discuss potential extensions in Section 6.

2 Sample Size Based on the McNemar's Test

In a pre- and post-intervention study, let y_{it} (1: Yes and 0: NO) denote the binary response from the i th ($i = 1, \dots, n$) subject at time t ($t = 0$: pre-intervention and $t = 1$: post-intervention). The data can be summarized by Table 1, with cell counts

$n_{uv} = \sum_{i=1}^n I(y_{i0}=u, y_{i1}=v)$ for $u = 0, 1$ and $v = 0, 1$. The row totals and column totals are denoted by $n_{i\cdot}$ and $n_{\cdot j}$, respectively. We define joint probabilities by $h_u = P(y_0 = u, y_1 = v)$, and marginal probabilities for pre- and post-intervention responses by p_i : $p_0 = P(y_0 = 1) = h_{10} + h_{11}$ and $p_1 = P(y_1 = 1) = h_{01} + h_{11}$.

It is of interest to test hypotheses $H_0: p_0 = p_1$ versus $H_1: p_0 \neq p_1$, which is equivalent to $H_0: h_{10} = h_{01}$ versus $H_1: h_{10} \neq h_{01}$. The McNemar's test is based on statistic

$$T_{MN} = \frac{|n_{01} - n_{10}|}{\sqrt{n_{01} + n_{10}}}.$$

We reject H_0 if $T_{MN} > z_{1-\alpha/2}$. Here α is the significance level and $z_{1-\alpha/2}$ is the $100(1-\alpha/2)$ th percentile of the standard normal distribution. To detect an intervention effect of $\delta = p_1 - p_0 = h_{01} - h_{10}$ with power $1 - \beta$ and type I error α , one of the existing sample size formulas is

$$n_{MN} = \frac{\left[z_{1-\alpha/2} \sqrt{h_{01} + h_{10}} + z_{1-\gamma} \sqrt{h_{01} + h_{10} - (h_{01} - h_{10})^2} \right]^2}{(h_{01} - h_{10})^2}. \quad (1)$$

Details of derivation can be found in Connor (1987). The design parameters required by (1) are h_{10} and h_{01} . Equivalently, the power analysis software PASS 11 (NCSS LLC., Kaysville, Utah) requires the difference $(h_{01} - h_{10})$ and proportion of discordant pairs $(h_{01} + h_{10})$.

3 Sample Size Based on the GEE Approach

The GEE approach models y_{it} by a logistic regression model. It is assumed that $P(y_{it} = 1) = p_{it}$ and

$$\log \left(\frac{p_{it}}{1 - p_{it}} \right) = \beta_1 + \beta_2 t = X'_{it} \beta. \quad (2)$$

Thus $\beta_1 = \log \left(\frac{p_{i0}}{1 - p_{i0}} \right)$ models the log-transformed baseline odds and

$\beta_2 = \log \left(\frac{p_{i1}}{1 - p_{i1}} \right) - \log \left(\frac{p_{i0}}{1 - p_{i0}} \right)$ is the log-transformed odds ratio between the post- and pre-intervention responses, representing the intervention effect. To facilitate discussion, we present the model in a matrix form, with $X_{it} = (1, t)$ being the vector of covariates and $\beta = (\beta_1, \beta_2)$ being the vector of regression parameters. Because y_{i0} and y_{i1} are observed from the same subject, we use $\rho = \text{Corr}(y_{i0}, y_{i1})$ to measure within-subject correlation. **It is also called the Phi-coefficient, calculated as the Pearson product-moment correlation coefficient between two binary variables** (McNemar 1962):

$$\rho = \frac{E(y_{i0}y_{i1}) - E(y_{i0})E(y_{i1})}{\sqrt{\text{Var}(y_{i0})\text{Var}(y_{i1})}} = \frac{h_{11} - p_0p_1}{\sqrt{p_0(1-p_0)p_1(1-p_1)}}.$$

We assume the responses to be independent across different subjects, $\text{Corr}(y_{it}, y_{it'}) = 0$ for $i \neq i'$.

The primary interest is to test the null hypothesis $H_0: \beta_2 = 0$ versus the alternative hypothesis $H_1: \beta_2 \neq 0$. We make statistical inference based on a consistent estimator using the GEE method. Under an independent working correlation structure, the estimator $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2)$ is obtained by solving $S_{\hat{\beta}}(\hat{\beta}) = 0$,

$$\mathbf{S}_n(\beta) = \sum_{i=1}^n \sum_{t=0}^1 [y_{it} - p_{it}(\beta)] X_{it}.$$

Here $p_{it}(\beta) = \exp(X'_{it}\beta) / [1 + \exp(X'_{it}\beta)]$ is implied by (2). The Newton-Raphson algorithm can be employed to obtain a numerical solution. At the $(m+1)$ th iteration,

$$\hat{\beta}^{(m+1)} = \hat{\beta}^{(m)} + \mathbf{A}_n^{-1}(\hat{\beta}^{(m)}) \mathbf{S}_n(\hat{\beta}^{(m)}) / n,$$

where

$$\mathbf{A}_n(\beta) = -\frac{1}{n} \cdot \frac{\partial \mathbf{S}_n(\beta)}{\partial \beta} = \frac{1}{n} \sum_{i=1}^n \sum_{t=0}^1 p_{it}(\beta) [1 - p_{it}(\beta)] X_{it} X'_{it}.$$

Liang & Zeger (1986) showed that $\sqrt{n}(\hat{\beta} - \beta)$ is approximately normal with mean zero and the variance is consistently estimated by $\Sigma_n = \mathbf{A}_n^{-1}(\hat{\beta}) \mathbf{V}_n(\hat{\beta}) \mathbf{A}_n^{-1}(\hat{\beta})$, where

$$\mathbf{V}_n(\hat{\beta}) = \frac{1}{n} \sum_{i=1}^n \left[\left(\sum_{t=0}^1 \hat{\varepsilon}_{it} X_{it} \right) \left(\sum_{t=0}^1 \hat{\varepsilon}_{it} X_{it} \right)' \right]$$

with $\hat{\varepsilon}_{it} = y_{it} - p_{it}(\hat{\beta})$. We reject H_0 if $|\sqrt{n}\hat{\beta}_2 / \hat{\sigma}_2| > z_{1-\alpha/2}$, where $\hat{\sigma}_2^2$ is the (2, 2)th element of Σ_n .

Now we derive the sample size required to reject H_0 given $\mathbf{b} = (b_1, b_2)$, with type I error and power $1 - \gamma$. Let σ_2^2 be the true variance of the GEE estimator of β_2 . The sample size n is obtained by solving $P(|\sqrt{nb_2}/\sigma_2| > z_{1-\alpha/2}) = 1 - \gamma$. The required sample size is

$$n = \frac{\sigma_2^2 (z_{1-\alpha/2} + z_{1-\gamma})^2}{b_2^2}. \quad (3)$$

Note that the marginal probabilities defined in Section 2 can be equivalently presented as $p_t = p_{it}(\mathbf{b})$ for $t = 0, 1$. The following theorem presents a closed-form for σ_2^2 , which implies a closed-form GEE sample size formula under complete observations.

Theorem 1. Define $\tau_t^2 = p_t(1 - p_t)$. As $n \rightarrow \infty$, σ_2^2 has a closed-form,

$$\sigma_2^2 = \frac{\tau_0^2 + \tau_1^2 - 2\rho\tau_0\tau_1}{\tau_0^2\tau_1^2} \quad (4)$$

Proof. See Appendix A.

3.1 Sample Size in the Presence of Missing Data

The GEE sample size approach can be readily extended to accommodate missing data. First we define $y_{ij} = 1/0$ to indicate whether y_{ij} (the post-intervention response of subject i) is

observed/missing. The proportion of patients with complete observations is $P(i=1) = q$, and the dropout rate is $1 - q$. The following theorem shows that σ_2^2 also has a closed-form under incomplete observations.

Theorem 2. Under dropout rate $1 - q$, as $n \rightarrow \infty$, σ_2^2 has a closed-form,

$$\sigma_2^2 = \frac{\tau_0^2 + q\tau_1^2 - 2q\rho\tau_0\tau_1}{q\tau_0^2\tau_1^2}.$$

Proof. See Appendix B.

Thus the general sample size formula that accounts for potential missing values in the post-intervention measurements is

$$n_{GEE} = \frac{(\tau_0^2 + q\tau_1^2 - 2q\rho\tau_0\tau_1)(z_{1-\alpha/2} + z_{1-\gamma})^2}{q\tau_0^2\tau_1^2b_2^2}. \quad (5)$$

It is obvious from (5) that besides type I error and power $(1 - \gamma)$, the design parameters required by GEE sample sizes are q and (p_0, p_1) . Importantly, the last three factors (p_0, p_1, γ) all affect how missing data, characterized by q , affect the sample size. Let $n_{GEE}(1)$ be the GEE sample sizes under complete data ($q = 1$). Here the notation $n_{GEE}(q)$ emphasizes that the sample size is calculated under a completion rate of q . Recall that the traditional adjustment for missing data would require a sample size of $n_{GEE}(1)/q$. It is straightforward to show that the GEE approach leads to a saving in sample size compared with that based on traditional adjustment for missing data when $q > 1/(2\rho)$. That is, $n_{GEE}(q) < n_{GEE}(1)/q$ when $q > 1/(2\rho)$.

Note that depending on the values of p_0 and p_1 , the constraint that $h_u \geq 0$ (for $u = 0, 1$; $\rho = 0, 1$) implies a valid range of ρ , $\rho \in (\rho_L, \rho_U)$, with

$$\rho_L = \max \left\{ -\sqrt{\frac{p_0p_1}{(1-p_0)(1-p_1)}}, -\sqrt{\frac{(1-p_0)(1-p_1)}{p_0p_1}} \right\}, \quad \rho_U = \min \left\{ \sqrt{\frac{p_0(1-p_1)}{p_1(1-p_0)}}, \sqrt{\frac{p_1(1-p_0)}{p_0(1-p_1)}} \right\}.$$

Thus the condition $q > 1/(2\rho)$ should be considered within range (ρ_L, ρ_U) . For each combination of (p_0, p_1) , we calculate

$$R = \min \left\{ \frac{\tau_1/(2\tau_0) - \rho_L}{\rho_U - \rho_L}, 1 \right\}.$$

Thus a larger value of R (close to 1) means that, under this particular (p_0, p_1) , most of the valid values satisfy the condition for $n_{GEE}(q) < n_{GEE}(1)/q$, or it is very likely that the proposed GEE approach is superior to traditional adjustment for missing values. On the other hand, a smaller value of R (close to 0) means that few of the valid values satisfy the condition for $n_{GEE}(q) < n_{GEE}(1)/q$. In Figure 1 we plot the R values under various combination of (p_0, p_1) . We first notice that the R value under (p_0, p_1) equals that under $(1 - p_0, 1 - p_1)$, which is obvious from the properties of ρ_L, ρ_U and $1/(2\rho)$. The R values tend to be larger for two scenarios: 1) $p_0 > p_1$ and p_0 close to 1 (the lower right corner); 2) $p_0 < p_1$ and p_0 close to 0 (the upper left corner). For example, when $(p_0 = 0.8, p_1 = 0.2)$ or $(p_0 = 0.2,$

$p_1 = 0.8$), we have $L = -0.763$, $U = 0.327$, and $1/(2 - \rho) = 0.656$. Since $U < 1/(2 - \rho)$, the condition for the GEE method to be superior to the traditional adjustment for missing values holds for all valid values of ρ and we have $R = 1$. On the other hand, the R values tend to be smaller for two scenarios: 1) $p_0 > p_1$ and p_0 is close to 0 (the lower left corner beneath the diagonal line); 2) $p_0 < p_1$ and p_0 is close to 1 (the upper right corner above the diagonal line). For example, when $(p_0 = 0.02, p_1 = 0.01)$ or $(p_0 = 0.98, p_1 = 0.99)$, we have $L = -0.014$, $U = 0.704$, $1/(2 - \rho) = 0.253$ and $R = 0.372$. In general, among all the combinations of (p_0, p_1) , as p_0 and p_1 range from 0 to 1, 45% of them have $R = 1$ and 75% of them have $R > 0.75$.

3.2 Relationship between n_{MN} and n_{GEE}

Under complete observations ($q = 1$), the design parameters requested by n_{GEE} and n_{MN} are different: (p_0, p_1, ρ) for n_{GEE} while (h_{10}, h_{01}) for n_{MN} . Given (p_0, p_1, ρ) , we can easily derive $(h_{00}, h_{01}, h_{10}, h_{11})$. First we have $E(y_{it}) = p_t$, $\text{Var}(y_{it}) = p_t(1 - p_t)$ for $t = 0, 1$, and

$\text{Cov}(y_{i0}, y_{i1}) = \rho \sqrt{p_0(1 - p_0)p_1(1 - p_1)}$. Then,

$$\begin{aligned} h_{11} &= E(y_{i0} \cdot y_{i1}) = E(y_{i0})E(y_{i1}) + \text{Cov}(y_{i0}, y_{i1}) = p_0p_1 + \rho \sqrt{p_0(1 - p_0)p_1(1 - p_1)}, \\ h_{00} &= (1 - p_0)(1 - p_1) + \rho \sqrt{p_0(1 - p_0)p_1(1 - p_1)}, \\ h_{10} &= p_0 - h_{11} = p_0(1 - p_1) - \rho \sqrt{p_0(1 - p_0)p_1(1 - p_1)}, \\ h_{01} &= p_1 - h_{11} = p_1(1 - p_0) - \rho \sqrt{p_0(1 - p_0)p_1(1 - p_1)}. \end{aligned}$$

Given (h_{10}, h_{01}) , however, we can not derive (p_0, p_1, ρ) . That is, n_{GEE} requires more information than n_{MN} in sample size calculation. Theoretically, for every specification of (h_{10}, h_{01}) for n_{MN} , there exists an infinite number of corresponding combinations of (p_0, p_1, ρ) , each leading to a different calculation of n_{GEE} . In practice, due to the integer constraint, the real difference between n_{MN} and n_{GEE} might be small. This issue will be further explored by simulation studies in Section 4.

4 Simulation Study

In Table 2 we compare the performance of n_{MN} and n_{GEE} when there is no missing data. We specify power $1 - \beta = 0.8$ and type I error $\alpha = 0.05$. In the first column, we list (h_{10}, h_{01}) , the trial configurations for n_{MN} . In the second column, we list five possible combinations of (p_0, p_1, ρ) for n_{GEE} , all of which correspond to a particular (h_{10}, h_{01}) . In the third column, the sample sizes based on the McNemar's test (n_{MN}) are identical within each (h_{10}, h_{01}) . In the fourth column, the sample sizes based on the GEE approach (n_{GEE}) are calculated based on (p_0, p_1, ρ) , so they are slightly different although they correspond to the same (h_{10}, h_{01}) . Despite differences in calculation approach and trial parameters, we observe that n_{MN} and n_{GEE} are very close to each other. To assess performance, for each (p_0, p_1, ρ) , we generate 5000 data sets each containing $n_{MN}(n_{GEE})$ pairs of binary observations, and compute empirical power and type I error as the proportion of times that the null hypothesis is rejected based on the McNemar's test (GEE approach). The empirical type I errors are close to the nominal value 0.05. The empirical powers are slightly larger than the nominal level when the sample size is relatively small, and approaches to the nominal level as sample size increases.

In Table 3 we evaluate the performance of n_{GEE} in accommodating missing data under various design configurations: pre-intervention rate p_0 from 0.1 to 0.3; treatment effect $p_1 - p_0$ from 0.05 to 0.2; within-subject correlation ρ from 0 to 0.3; and proportion of

complete observations q from 1 to 0.6. In each cell we present calculated sample size n_{GEE} , empirical power $1 - \beta$, and empirical type I error α . We have several observations from Table 3. First, with all other factors fixed, a larger within-subject correlation leads to a smaller sample size requirement. This property is also straightforward from (5). Second, a larger treatment effect leads to a smaller sample size requirement. The impact of treatment effect, however, also depends on the baseline response rate p_0 . For example, with $\rho = 0.05$ and no missing data ($q = 1$), the required sample sizes range from 967 to 1380 for $p_0 = 0.3$ but from 490 to 696 for $p_0 = 0.1$. Third, in Table 3 we have explored scenarios where the sample sizes can be as small as 50 and as large as 1821, and the empirical powers and type I errors are generally close to their nominal values. It is assuring for practitioners to know that the desired power and type I error are maintained over a wide range of sample size. Fourth, the proposed GEE sample size approach can lead to substantial saving in sample size compared with traditional adjustment approach for missing data. For example, under ($p_0 = 0.1$, $p_1 = 0.15$, $\rho = 0.3$), the required sample size for $q = 1$ and $q = 0.6$ are 490 and 682, respectively. If we follow the traditional adjustment approach for missing data, the sample size under $q = 0.6$ would be 816.

5 Example

Ahmet et al. (2011) reported that the prevalence rate of nocturnal hypoglycemia (NH) is 68% in pediatric patients with type I diabetes. NH refers to a symptom of low blood sugar levels at night, a common problem for children and adolescents with type I diabetes. An investigator plans to conduct a clinical trial to investigate if a new compound improves the status of NH using the pre-post design. We assume that 68% of pediatric type I diabetes patients have NH at baseline ($p_0 = 0.68$). Based on the pilot study, it is expected that 40% of NH patients will shift from NH at pre-treatment to normal at post-treatment ($h_{10} = 0.4 \times 0.68 = 0.27$), and 35% of normal patients will shift from normal at pre-treatment to NH at post-treatment ($h_{01} = 0.32 \times 0.35 = 0.11$). From $p_0 = 0.68$, $h_{10} = 0.27$, and $h_{01} = 0.11$, we can easily obtain $h_{00} = 0.21$, $h_{11} = 0.41$, $p_1 = 0.52$, and $\rho = 0.23$. To test the hypotheses $H_0: p_0 = p_1$ versus $H_1: p_0 \neq p_1$ with 80% power at 5% significance level, the number of subjects needs to be determined.

Assuming no missing data, the required sample size based on the McNemar's test can be obtained using Equation (1), $n_{MN} = 116$. To employ the GEE sample size approach, first we have $b_1 = 0.75$ and $b_2 = -0.67$. Under no dropout ($q = 1$), the sample size calculated based on Equation (5) is $n_{GEE}(1) = 115$. When 20% of patients drop out after pre-treatment, the sample size increases to $n_{GEE}(0.8) = 132$. If a traditional naive adjustment is made for missing data, the sample size is $n_{GEE}(1)/0.8 = 144$. The proposed GEE approach leads to an 8.4% saving in sample size.

6 Discussion

In this study we present a closed-form sample size formula for pre- and post-intervention studies where a portion of subjects might fail to provide post-intervention measurements. A mixed logistic regression model is employed to account for the pre-post correlation, and statistical inference is obtained through the GEE approach. We show that the proposed GEE method leads to sample size estimates very similar to those based on the McNemar's test under no missing data. When there is missing data, the GEE method is advantageous in utilizing incomplete observations. Our simulation studies suggest that in scenarios where the calculated sample sizes are large, the empirical power and type I error are close to their nominal levels. When the calculated sample sizes are small, however, the empirical powers tend to be larger than the nominal level. There are two possible explanations. First, the normal approximation might be unsatisfactory when sample size is small. Second, the

integer constraint require us to round the sample size solution upward to the nearest integer, and the effect of rounding is more noticeable for smaller sample sizes.

Mathematically, the sample size formula we have derived is the application of Jung & Ahn (2005)'s approach to a special scenario of longitudinal study where each subject has two measurements and missing values can only occur at the second measurement. Our contribution includes the novel implementation of the GEE method in pre- and post-intervention experimental design where the traditional McNemar-based sample size approaches have not been able to appropriately account for missing data. Furthermore, under the special scenario, the closed-form sample formula is drastically simplified, which allows deeper insight on the impact of various design factors (response rates, within-subject correlation, and missing proportion) on sample size. It also allows us to theoretically derive the condition under which the proposed method is superior to the traditional adjustment for missing data.

The pre- and post-intervention study we have investigated can also be perceived as a special case of paired experiments, which compare the responses observed from the same subjects under two diagnostic devices. For example, in Romeo et al. (2010), each patient was examined for brain involvement in myotonic dystrophy type 1 by single photon emission tomography (SPECT) and positron emission tomography (PET), and the diagnostic performance of SPECT and PET was compared. One important feature of paired experiments is that missing values can occur for either of the diagnostic devices. In this paper, we consider a special scenario where the measurements are complete for one device (corresponding to pre-intervention) and missing values only occur in the other device (corresponding to post-intervention). In future research, we will extend the sample size formula to paired experiments to account for the possibility of missingness in both devices.

Gönen (2004) investigated sample size calculations for the McNemar's test where there is a clustered structure among subjects. Extending the proposed approach to such scenarios with incomplete observations will be the topic of our future research.

Appendix A

Proof of Theorem 1

As $n \rightarrow \infty$, it is straightforward that

$$\begin{aligned} \mathbf{A}_n(\hat{\beta}) &\rightarrow \mathbf{A}(\mathbf{b}) = \sum_{t=0}^1 \tau_t^2 X_{it} X'_{it} = \begin{pmatrix} \tau_0^2 + \tau_1^2 & \tau_1^2 \\ \tau_1^2 & \tau_1^2 \end{pmatrix}, \\ \mathbf{V}_n(\hat{\beta}) &\rightarrow \mathbf{V}(\mathbf{b}) = \sum_{t=1}^1 \sum_{t'=0}^1 \rho_{tt'} \tau_t \tau_{t'} X_{it} X'_{it'} = \begin{pmatrix} \tau_0^2 + 2\rho\tau_0\tau_1 + \tau_1^2 & 2\rho\tau_0\tau_1 + \tau_1^2 \\ \tau_1^2 & \tau_1^2 \end{pmatrix}. \end{aligned}$$

Here we have defined $\rho_{tt} = 1$ if $t = t'$, and $\rho_{tt} = \rho$ otherwise. The derivation uses the fact that $\text{Var}(y_{it}) = p_t(1 - p_t)$ and $\text{Corr}(y_{it}, y_{it'}) = \rho_{tt'}$. Thus the variance matrix of the GEE estimator, $\mathbf{V}_n(\hat{\beta})$, approaches to

$$\Sigma = \mathbf{A}^{-1}(\mathbf{b}) \mathbf{V}(\mathbf{b}) \mathbf{A}^{-1}(\mathbf{b}).$$

Because

$$\mathbf{A}^{-1}(\mathbf{b}) = \frac{1}{\tau_0^2 \tau_1^2} \begin{pmatrix} \tau_1^2 & -\tau_1^2 \\ -\tau_1^2 & \tau_0^2 \tau_1^2 \end{pmatrix},$$

a few steps of algebra show that the (2, 2)th element of $\mathbf{A}^{-1}(\mathbf{b})$ is

$$\sigma_2^2 = \frac{\tau_0^2 + \tau_1^2 - 2\rho\tau_0\tau_1}{\tau_0^2 \tau_1^2}.$$

Thus we complete the proof.

Appendix B

Proof of Theorem 2

With the inclusion of i_b , the updated equations in the Section are

$$\begin{aligned} \mathbf{S}_n(\beta) &= \sum_{i=1}^n \{ [y_{i0} - p_{i0}(\beta)]X_{i0} + \delta_i [y_{i1} - p_{i1}(\beta)]X_{i1} \}, \\ \mathbf{A}_n(\beta) &= \frac{1}{n} \sum_{i=1}^n \{ p_{i0}(\beta)[1 - p_{i0}(\beta)]X_{i0}X_{i0}' + \delta_i p_{i1}(\beta)[1 - p_{i1}(\beta)]X_{i1}X_{i1}' \}, \\ \mathbf{V}_n(\hat{\beta}) &= \frac{1}{n} \sum_{i=1}^n [(\hat{\varepsilon}_{i0}X_{i0} + \delta_i \hat{\varepsilon}_i X_{i1})(\hat{\varepsilon}_{i0}X_{i0} + \delta_i \hat{\varepsilon}_i X_{i1})'] \end{aligned}$$

As $n \rightarrow \infty$, we have

$$\mathbf{A}_n(\hat{\beta}) \rightarrow \mathbf{A}(\mathbf{b}) = \begin{pmatrix} \tau_0^2 + q\tau_1^2 & q\tau_1^2 \\ q\tau_1^2 & q\tau_1^2 \end{pmatrix}, \quad \mathbf{V}_n(\hat{\beta}) \rightarrow \mathbf{V}(\mathbf{b}) = \begin{pmatrix} \tau_0^2 + 2q\rho\tau_0\tau_1 + q\tau_1^2 & q\rho\tau_0\tau_1 + q\tau_1^2 \\ q\rho\tau_0\tau_1 + q\tau_1^2 & q\tau_1^2 \end{pmatrix}.$$

Following similar derivation in Appendix A, we have

$$\sigma_2^2 = \frac{\tau_0^2 + q\tau_1^2 - 2q\rho\tau_0\tau_1}{q\tau_0^2 \tau_1^2}.$$

References

- Ahmet A, Dagenais S, Barrowman N, Collins C, Lawson M. Prevalence of nocturnal hypoglycemia in pediatric type 1 diabetes: A pilot study using continuous glucose monitoring. *Journal of Pediatrics*. 2011; 159:297–302. [PubMed: 21414634]
- Cochran WG. The comparison of percentages in matched samples. *Biometrika*. 1950; 37(4):256–266. [PubMed: 14801052]
- Connor RJ. Sample size for testing differences in proportions for the paired-sample design. *Biometrics*. 1987; 43(1):207–211. [PubMed: 3567305]
- Gönen M. Sample size and power for mcnemar's test with clustered data. *Statistics in Medicine*. 2004; 23(14):2283–2294. [PubMed: 15236431]

- Julious SA, Campbell MJ, Altman DG. Estimating sample sizes for continuous, binary, and ordinal outcomes in paired comparisons: Practical hints. *Journal of Biopharmaceutical Statistics*. 1999; 9(2):241–251. [PubMed: 10379691]
- Jung S, Ahn CW. Sample size for a two-group comparison of repeated binary measurements using gee. *Statistics in Medicine*. 2005; 24(17):2583–2596. [PubMed: 16118812]
- Knudtson EJ, Lorenz LB, Skaggs VJ, Peck JD, Goodman JR, Elimian AA. The effect of digital cervical examination on group b streptococcal culture. *J Am Geriatr Soc*. 2010; 202(1):58.e1–58.e4.
- Lachin JM. Power and sample size evaluation for the mcnemar test with application to matched case-control studies. *Statistics in Medicine*. 1992; 11(9):1239–1251. [PubMed: 1509223]
- Liang K, Zeger SL. Longitudinal data analysis for discrete and continuous outcomes using generalized linear models. *Biometrika*. 1986; 84:3–32.
- Liu G, Liang K. Sample size calculations for studies with correlated observations. *Biometrics*. 1997; 53(3):937–947. [PubMed: 9290224]
- Lu Y, Bean JA. On the sample size for one-sided equivalence of sensitivities based upon mcnemar's test. *Statistics in Medicine*. 1995; 14(16):1831–1839. [PubMed: 7481214]
- McNemar Q. Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*. 1947; 12(2):153–157. [PubMed: 20254758]
- McNemar, Q. *Psychological Statistics*. New York: Wiley; 1962.
- Miettinen OS. The matched pairs design in the case of all-or-none responses. *Biometrics*. 1968; 24(2): 339–352. [PubMed: 5683874]
- Norton EC, Bieler GS, Ennett ST, Zarkin GA. Analysis of prevention program effectiveness with clustered data using generalized estimating equations. *Journal of Consulting and Clinical Psychology*. 1996; 64(5):919–926. [PubMed: 8916620]
- Rochon J. Application of GEE procedures for sample size calculations in repeated measures experiments. *Statistics in Medicine*. 1998; 17(14):1643–1658. [PubMed: 9699236]
- Romeo V, Pegoraro E, Squarzanti F, Soraru G, Ferrati C, Zucchetta P, Chierichetti F, Angelini C. Retrospective study on pet-spect imaging in a large cohort of myotonic dystrophy type 1 patients. *J Am Geriatr Soc*. 2010; 31(6):757–763.
- Rossi MC, Perozzi C, Consorti C, Almonti T, Foglini P, Giostra N, Nanni P, Talevi S, Bartolomei D, Vespasiani G. An interactive diary for diet management (dai): a new telemedicine system able to promote body weight reduction, nutritional education, and consumption of fresh local produce. *Diabetes Technol Ther*. 2010; 12(8):641–647. [PubMed: 20615106]
- Royston P. Exact conditional and unconditional sample size for pair-matched studies with binary outcome: A practical guide. *Statistics in Medicine*. 1993; 12(7):699–712. [PubMed: 8511446]
- Selicato GR, Muller KE. Approximating power of the unconditional test for correlated binary pairs. *Communications in Statistics Part B: Simulation and Computation*. 1998; 27(2):553–564.
- Shork M, Williams G. Number of observations required for the comparison of two correlated proportions. *Communications in Statistics - Simulation and Computation*. 1980; 9(4):349–357.
- Spleen AM, Kluhsman BC, Clark AD, Dignan MB, Lengerich EJTAHCTF. An increase in HPV-related knowledge and vaccination intent among parental and non-parental caregivers of adolescent girls, age 9–17 years, in appalachian pennsylvania. *J Cancer Educ*. 2012; 27(2):312–319. [PubMed: 22131065]
- Wajnberg A, Wang KH, Aniff M, Kunins HV. Hospitalizations and skilled nursing facility admissions before and after the implementation of a home-based primary care program. *J Am Geriatr Soc*. 2010; 58(6):1144–1147. [PubMed: 20487075]
- Zeger SL, Liang K, Albert PS. Models for longitudinal data: A generalized estimating equation approach. *Biometrics*. 1988; 44(4):1049–1060. [PubMed: 3233245]
- Zieschang T, Dutzi I, Müller E, Hestermann U, Grünendahl K, Braun AK, Hüger D, Kopf D, Specht-Leible N, Oster P. Improving care for patients with dementia hospitalized for acute somatic illness in a specialized care unit: a feasibility study. *Int Psychogeriatr*. 2010; 22(1):139–146. [PubMed: 19602306]

Figure 1.

The horizontal axis shows the values of p_0 and the vertical axis shows the values of p_1 . The plot represents the values of R at various combinations of (p_0, p_1) with a grayscale. A brighter area represents a larger value of R . We also include contour lines in the figure. R values on the diagonal line ($p_0 = p_1$) are not evaluated.

Table 1
Summarization of Pre- and Post-Intervention Observations

		Post-intervention	
pre-intervention		NO	YES
NO		n_{00}	n_{01}
YES		n_{10}	n_{11}
		$n_{0\cdot}$	$n_{1\cdot}$
		$n_{\cdot 0}$	$n_{\cdot 1}$
		n	

Table 2Comparison of n_{MN} with n_{GEE}

(h_{10}, h_{01})	(p_0, p_1, γ)	$n_{MN}(1 - \gamma, \gamma)$	$n_{GEE}(1 - \gamma, \gamma)$
(0.1, 0.30)	(0.15, 0.35, 0.24)	77(0.829, 0.053)	79(0.836, 0.044)
	(0.25, 0.45, -0.01)	77(0.823, 0.057)	76(0.830, 0.052)
	(0.35, 0.55, 0.17)	77(0.826, 0.051)	75(0.819, 0.053)
	(0.45, 0.65, 0.24)	77(0.827, 0.053)	75(0.823, 0.052)
	(0.55, 0.75, 0.24)	77(0.835, 0.048)	76(0.820, 0.055)
(0.1, 0.25)	(0.15, 0.30, 0.03)	120(0.810, 0.050)	122(0.821, 0.047)
	(0.25, 0.40, 0.24)	120(0.814, 0.051)	119(0.809, 0.047)
	(0.35, 0.50, 0.31)	120(0.816, 0.051)	118(0.810, 0.053)
	(0.45, 0.60, 0.33)	120(0.811, 0.049)	118(0.813, 0.049)
	(0.55, 0.70, 0.29)	120(0.804, 0.048)	119(0.821, 0.051)
(0.1, 0.20)	(0.15, 0.25, 0.08)	234(0.810, 0.048)	236(0.812, 0.049)
	(0.25, 0.35, 0.30)	234(0.805, 0.050)	233(0.807, 0.050)
	(0.35, 0.45, 0.39)	234(0.810, 0.050)	231(0.811, 0.051)
	(0.45, 0.55, 0.41)	234(0.810, 0.051)	231(0.808, 0.051)
	(0.55, 0.65, 0.39)	234(0.810, 0.047)	231(0.803, 0.050)
(0.1, 0.15)	(0.15, 0.20, 0.14)	783(0.801, 0.051)	785(0.797, 0.051)
	(0.25, 0.30, 0.38)	783(0.807, 0.048)	782(0.801, 0.052)
	(0.35, 0.40, 0.47)	783(0.807, 0.050)	780(0.801, 0.052)
	(0.45, 0.50, 0.50)	783(0.806, 0.050)	780(0.793, 0.050)
	(0.55, 0.60, 0.49)	783(0.798, 0.053)	780(0.804, 0.051)

Here (h_{10}, h_{01}) are design parameters for the McNemar sample size approach, and (p_0, p_1, γ) are the corresponding design parameters for the GEE sample size approach. We specify power $1 - \gamma = 0.8$ and type I error $\gamma = 0.05$. The empirical powers and type I errors are denoted by $1 - \gamma$ and γ , respectively.

Table 3Performance of GEE sample sizes: $n_{GEE}(1 - \alpha, \beta)$

(p_0, p_1)		$q = 1$	$q = 0.8$	$q = 0.6$
(0.10, 0.05)	0	696(0.807, 0.052)	768(0.806, 0.053)	887(0.783, 0.049)
	0.15	593(0.814, 0.052)	665(0.803, 0.047)	785(0.797, 0.045)
	0.30	490(0.810, 0.051)	562(0.805, 0.053)	682(0.796, 0.054)
(0.10, 0.20)	0	208(0.815, 0.046)	226(0.819, 0.049)	257(0.805, 0.051)
	0.15	178(0.818, 0.046)	197(0.816, 0.044)	228(0.801, 0.047)
	0.30	148(0.837, 0.046)	167(0.817, 0.051)	198(0.797, 0.054)
(0.10, 0.30)	0	69(0.838, 0.038)	74(0.828, 0.033)	83(0.818, 0.037)
	0.15	59(0.852, 0.041)	65(0.830, 0.042)	73(0.825, 0.043)
	0.30	50(0.867, 0.046)	55(0.847, 0.046)	64(0.823, 0.048)
(0.20, 0.25)	0	1099(0.804, 0.051)	1225(0.799, 0.052)	1436(0.798, 0.049)
	0.15	935(0.801, 0.051)	1061(0.791, 0.053)	1272(0.803, 0.046)
	0.30	771(0.795, 0.052)	897(0.800, 0.046)	1108(0.795, 0.052)
(0.20, 0.30)	0	298(0.804, 0.049)	330(0.804, 0.050)	384(0.802, 0.049)
	0.15	254(0.819, 0.054)	286(0.800, 0.051)	340(0.791, 0.053)
	0.30	210(0.816, 0.052)	242(0.807, 0.052)	295(0.789, 0.048)
(0.20, 0.40)	0	85(0.817, 0.049)	94(0.815, 0.049)	108(0.806, 0.050)
	0.15	73(0.831, 0.054)	81(0.823, 0.047)	96(0.809, 0.048)
	0.30	61(0.843, 0.056)	69(0.830, 0.046)	83(0.801, 0.045)
(0.30, 0.35)	0	1380(0.795, 0.050)	1546(0.797, 0.046)	1821(0.792, 0.051)
	0.15	1173(0.801, 0.054)	1339(0.795, 0.053)	1615(0.796, 0.046)
	0.30	967(0.795, 0.053)	1132(0.803, 0.049)	1408(0.800, 0.053)
(0.30, 0.40)	0	359(0.794, 0.048)	401(0.804, 0.055)	471(0.802, 0.049)
	0.15	306(0.796, 0.051)	348(0.808, 0.054)	417(0.798, 0.048)
	0.30	252(0.807, 0.048)	294(0.816, 0.046)	364(0.806, 0.055)
(0.30, 0.50)	0	96(0.806, 0.050)	107(0.809, 0.052)	125(0.807, 0.049)
	0.15	82(0.803, 0.046)	93(0.809, 0.050)	111(0.807, 0.048)
	0.30	68(0.826, 0.052)	79(0.814, 0.046)	97(0.811, 0.048)