

Sample size determination for quadratic inference functions in longitudinal design with dichotomous outcomes

Youna Hu and Peter X.-K. Song^{*†}

Quadratic inference functions (QIF) methodology is an important alternative to the generalized estimating equations (GEE) method in the longitudinal marginal model, as it offers higher estimation efficiency than the GEE when correlation structure is misspecified. The focus of this paper is on sample size determination and power calculation for QIF based on the Wald test in a marginal logistic model with covariates of treatment, time, and treatment–time interaction. We have made three contributions in this paper: (i) we derived formulas of sample size and power for QIF and compared their performance with those given by the GEE; (ii) we proposed an optimal scheme of sample size determination to overcome the difficulty of unknown true correlation matrix in the sense of minimal average risk; and (iii) we studied properties of both QIF and GEE sample size formulas in relation to the number of follow-up visits and found that the QIF gave more robust sample sizes than the GEE. Using numerical examples, we illustrated that without sacrificing statistical power, the QIF design leads to sample size saving and hence lower study cost in comparison with the GEE analysis. We conclude that the QIF analysis is appealing for longitudinal studies. Copyright © 2012 John Wiley & Sons, Ltd.

Keywords: average risk; clinical trial; correlation structure; GEE; marginal model; power

1. Introduction

Longitudinal clinical studies are undertaken extensively in biomedical sciences. In a longitudinal clinical study, repeated measurements are recorded at prescheduled time points during a study period. One primary objective of a clinical study is to compare effects of test treatments with those of controlled treatments.

When outcome variables of interest are binary, the marginal logistic model is popular for assessing the population-average effect of a test treatment. For the details regarding marginal generalized linear models for longitudinal data, refer to [1, 2]. Researchers have widely applied the method of generalized estimating equations (GEE), proposed by Liang and Zeger [3], to estimate and infer treatment effects in the marginal logistic model. Accordingly, researchers have studied sample size determination in the GEE-based design in the literature; for example, [1, Chapter 2; 4–6], among others. It is worth pointing out that Diggle *et al.* [1], Pan [4], and Rochon [6] considered a very simple longitudinal model that contains only a single treatment covariate; neither time covariate nor treatment–time interaction covariate is included in the design. Jung and Ahn [5] considered all these covariates but treated the working correlation structure as independent. Although these existing tools have addressed basic needs in the longitudinal study design, sample size and power calculation under more general scenarios, say, longitudinal models with both covariates of time and treatment–time interaction with various working correlations considered, are clearly of great importance in clinical research.

One objective of this paper is to establish a new scheme of sample size and power calculation on the basis of the quadratic inference functions (QIF) approach [7]. QIF is a powerful alternative method of estimation and inference to the popular GEE. It has been shown in the literature that QIF is superior to the GEE with respect to efficiency gain and robustness against outliers [8]. We will base our sample

Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, U.S.A.

^{*}Correspondence to: Peter X.-K. Song, Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109-2029, U.S.A.

[†]E-mail: pxsong@umich.edu

size and power calculation on the Wald test in marginal logistic models. In the literature, Demidenko [9] and Hsieh [10], among others, have investigated the Wald-test-based design in the logistic model for cross-sectional data. Unfortunately, these existing formulas cannot be easily modified to suit longitudinal studies because of the presence of within-subject correlation. In addition, we compare QIF and GEE in terms of their sample size and power. We also examine how the sample size is affected by varying the cluster size (i.e., the number of repeated measurements or follow-up visits per subject). As shown in our numerical examples, the GEE sample size may escalate substantially with an increased number of follow-up visits, whereas the QIF sample size remains nearly unchanged.

In the longitudinal design, a fundamental obstacle is the lack of knowledge about the true correlation matrix. Although, in practice, one correlation structure (e.g., compound symmetry (CS)) can be used to determine the sample size, the result is then subject to risk of either overestimation or underestimation of the needed sample size. In particular, once the data are collected, the true correlation structure can be obtained, which is very likely to be different from the one used initially in the design. To overcome this difficulty, we propose an optimal strategy of sample size determination from a perspective of minimal average risk. We postulate that the true correlation matrix is an element of a class of correlation matrices that are governed by a certain prior Wishart distribution. The hyperparameter in the prior distribution may be a correlation matrix specified according to previously acquired knowledge from a pilot study, from similar studies in the literature, or simply from an assumption of noninformative independent correlation. Then, we obtain the optimal sample size that is derived by minimizing the average risk under such prior distribution.

We organize this article as follows. Section 2 presents a brief introduction to the QIF method. We devote Section 3 to the sample size determination for the Wald test in the marginal logistic model in both QIF-based and GEE-based designs, in which we present an optimal strategy of sample size determination. Section 4 investigates, analytically and numerically, the relationship between sample size and cluster size (i.e., the number of follow-up visits). Section 5 contains some concluding remarks. We include technical details and information regarding the R package QIFSAMS in the appendices.

2. Quadratic inference functions

In a balanced longitudinal clinical trial, y_{ij} denotes the outcome of subject i at time point t_j . There are n subjects in the study and m repeated measurements planned to be collected from each of the n subjects. Thus, the total number of observations is $N = n \times m$. We further assume that the observations from different subjects are independent and those of the same subject are correlated. Both GEE and QIF methods postulate that the marginal mean, μ_{ij} , of the outcome y_{ij} , is a function of some covariates through a link function g , namely $g(\mu_{ij}) = \mathbf{x}'_{ij}\boldsymbol{\beta}$, where $\boldsymbol{\beta}$ is the regression coefficient. Here, \mathbf{x}' denotes the transpose of matrix \mathbf{x} . The variance of y_{ij} is a function of the mean $\text{var}(y_{ij}) = \phi V(\mu_{ij})$, where ϕ is the dispersion parameter. Vectors \mathbf{y}_i and $\boldsymbol{\mu}_i$, with elements y_{ij} and μ_{ij} , respectively, denote the longitudinal measurements and their mean for subject i . To obtain an estimate of $\boldsymbol{\beta}$, the GEE method solves

$$\sum_{i=1}^n \boldsymbol{\mu}_i' \mathbf{V}_i^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i) = 0, \tag{1}$$

where $\boldsymbol{\mu}_i = \partial \boldsymbol{\mu}_i / \partial \boldsymbol{\beta}$ is an $m \times p$ matrix and $\mathbf{V}_i = \mathbf{A}_i^{1/2} \mathbf{R}_i(\boldsymbol{\rho}) \mathbf{A}_i^{1/2}$ with \mathbf{A}_i being the diagonal matrix of the marginal variances, $\text{var}(y_{ij})$, and $\mathbf{R}_i(\boldsymbol{\rho})$ being the working correlation matrix.

The QIF method, proposed by Qu *et al.* [7], is derived from the fact that the inverse of the working correlation matrix $\mathbf{R}(\boldsymbol{\rho})$ can be approximated by a linear combination of several basis matrices:

$$\mathbf{R}(\boldsymbol{\rho})^{-1} \approx \sum_{l=0}^k a_l(\boldsymbol{\rho}) \mathbf{M}_l, \tag{2}$$

where $\mathbf{M}_0 = \mathbf{I}$ is the identity matrix, $\mathbf{M}_1, \dots, \mathbf{M}_k$ are known basis matrices with entries 0 or 1, and $a_0(\boldsymbol{\rho}), \dots, a_k(\boldsymbol{\rho})$ are unknown coefficients depending on the parameter $\boldsymbol{\rho}$. Expression (2) holds exactly for some commonly used working correlation structures. For example, the CS correlation structure corresponds to $\mathbf{R}(\boldsymbol{\rho})^{-1} = a_0(\boldsymbol{\rho})\mathbf{I} + a_1(\boldsymbol{\rho})\mathbf{M}_1^{\text{CS}}$, where the entries of \mathbf{M}_1 are 0 along the diagonal and 1 elsewhere. And the AR-1 correlation structure can be written as $\mathbf{R}(\boldsymbol{\rho})^{-1} = a_0(\boldsymbol{\rho})\mathbf{I} + a_1(\boldsymbol{\rho})\mathbf{M}_1^{\text{AR1}} +$

$a_2(\boldsymbol{\rho})\mathbf{M}_2^{\text{AR1}}$, where \mathbf{M}_1 has 1 on the two main off diagonals and 0 elsewhere and \mathbf{M}_2 has 1 on the two corner components of the diagonal.

Plugging expression (2) into (1) leads to a linear combination of the elements of an extended score vector

$$\bar{\mathbf{g}}_n(\boldsymbol{\beta}) = \frac{1}{n} \sum_{i=1}^n \mathbf{g}_i(\boldsymbol{\beta}) \approx \frac{1}{n} \begin{pmatrix} \sum_{i=1}^n \boldsymbol{\mu}_i' \mathbf{A}_i^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i) \\ \vdots \\ \sum_{i=1}^n \boldsymbol{\mu}_i' \mathbf{A}_i^{-1/2} \mathbf{M}_k \mathbf{A}_i^{-1/2} (\mathbf{y}_i - \boldsymbol{\mu}_i) \end{pmatrix}. \quad (3)$$

Then, the GEE (1) is expressed as $n\mathbf{a}'(\boldsymbol{\rho})\bar{\mathbf{g}}_n(\boldsymbol{\beta})$, where $\mathbf{a}(\boldsymbol{\rho}) = (a_0(\boldsymbol{\rho}), a_1(\boldsymbol{\rho}), \dots, a_k(\boldsymbol{\rho}))'$ is the vector of the coefficients in expansion (2). Because there are more equations than unknown parameters in (3), the generalized method of moments proposed by Hansen [11] is then applied to minimize the following quadratic inference function:

$$Q_n(\boldsymbol{\beta}) = n\bar{\mathbf{g}}_n'(\boldsymbol{\beta})\mathbf{C}_n^{-1}(\boldsymbol{\beta})\bar{\mathbf{g}}_n(\boldsymbol{\beta}), \quad (4)$$

where $\mathbf{C}_n(\boldsymbol{\beta}) = n^{-1} \sum_{i=1}^n \mathbf{g}_i(\boldsymbol{\beta})\mathbf{g}_i'(\boldsymbol{\beta})$ is the sample covariance matrix of $\bar{\mathbf{g}}_n$. It is noted that the objective function given in (4) only contains $\boldsymbol{\beta}$, and only the basis matrices from the working correlation structure are used to formulate this function. This implies that the QIF estimator, $\hat{\boldsymbol{\beta}} = \arg \min_{\boldsymbol{\beta}} Q_n(\boldsymbol{\beta})$, is obtained without estimating the nuisance parameter $\boldsymbol{\rho}$. This property delivers substantial ease in the design setting because the QIF sample size determination does not require knowledge of parameter $\boldsymbol{\rho}$.

According to Qu *et al.* [7], the QIF estimator $\hat{\boldsymbol{\beta}}$ has the usual large sample properties; for example, it is \sqrt{n} -consistent and asymptotically normal. The asymptotic variance is given by the inverse of the Godambe information matrix (or the sandwich estimator) with consistent estimate $\mathbf{J}_n(\hat{\boldsymbol{\beta}})^{-1}$, with $\mathbf{J}_n(\hat{\boldsymbol{\beta}}) = \dot{\bar{\mathbf{g}}}_n(\hat{\boldsymbol{\beta}})' \mathbf{C}_n^{-1}(\hat{\boldsymbol{\beta}}) \dot{\bar{\mathbf{g}}}_n(\hat{\boldsymbol{\beta}})$. The asymptotic normality allows us to establish the Wald test for the hypothesis with respect to the regression coefficients. Again, we note that this asymptotic covariance matrix \mathbf{J}_n is not dependent on the parameter $\boldsymbol{\rho}$. Similarly, the Wald test also applies to the GEE method, in which the asymptotic covariance matrix actually depends on the correlation parameter $\boldsymbol{\rho}$ explicitly. Therefore, our following derivation and comparison of sample size and power will focus on the Wald test that is available in both QIF and GEE.

We have coded this QIF method both in SAS (SAS Institute, Cary, NC, USA) and R language. We have made both packages available at <http://www-personal.umich.edu/pxsong/>.

3. Sample size determination

In this section, we derive sample size n under a fixed m . We begin by describing our model and then present steps related to the derivation of sample size in both QIF and GEE. We discuss in detail two examples of designs based on the CS and AR-1 working correlation structures.

3.1. Longitudinal logistic model

We consider the following logistic model with longitudinal dichotomous outcomes:

$$\text{logit}(\mu_{ij}) = \beta_0 + \beta_1 d_i + \beta_2 t_j + \beta_3 d_i t_j, \quad (5)$$

where $\mu_{ij} = P(y_{ij} = 1 | d_i, t_j)$ is the probability of a favorable clinical outcome ($y_{ij} = 1$) at visit time t_j for subject i , covariate d_i is the indicator of treatment group, defined as

$$d_i = \begin{cases} 1, & \text{if subject } i \text{ is in test treatment (Rx) group,} \\ 0, & \text{if subject } i \text{ is in controlled treatment group,} \end{cases}$$

and covariate t_j is the time of the j th visit for subject i . It follows from model (5) that the design matrix is $\mathbf{X}_i = (\mathbf{1}, d_i \mathbf{1}, \mathbf{t}, d_i \mathbf{t})'$, where $\mathbf{1}$ is an m -element vector of all ones, $\mathbf{t} = (t_1, t_2, \dots, t_m)'$, and the vector of regression coefficients is $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)'$. Also, in the logistic model, the variance function is $V_{ij} = V(\mu_{ij}) = \mu_{ij}(1 - \mu_{ij})$ and the dispersion parameter $\phi = 1$. Note that under the balanced design with homogeneous visit times, there are only two versions of design matrices and so are their induced matrices, corresponding to the two respective treatment arms. In all the subsequent expressions, a sub-index 1 (or 0) denotes terms from the test (or controlled) treatment arm. For example, the design matrix for the test drug arm is \mathbf{X}_1 and the counterpart for the controlled drug arm is \mathbf{X}_0 .

We are interested in testing for one of these three hypotheses: total effect, $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$; main Rx effect, $H_0 : \beta_1 = 0$; and joint Rx effect, $H_0 : \beta_1 = \beta_3 = 0$. In general, we may express the three scenarios in a unified form as $H_0 : \mathbf{H}\boldsymbol{\beta} = \mathbf{0}$ versus $H_1 : \mathbf{H}\boldsymbol{\beta} = \mathbf{h}_0 \neq \mathbf{0}$, where \mathbf{H} is a suitable matrix determined by the null hypothesis.

Let $\mathbf{J}_n(\boldsymbol{\beta})$ be the asymptotic covariance matrix of either the QIF estimator or the GEE estimator and let λ_n be the noncentrality parameter given by $\lambda_n = \mathbf{h}'_0 \{\mathbf{H}\mathbf{J}_n(\hat{\boldsymbol{\beta}})^{-1}\mathbf{H}'\}^{-1} \mathbf{h}_0$. Then, the Wald test statistic is $(\mathbf{H}\hat{\boldsymbol{\beta}})' \{\mathbf{H}\mathbf{J}_n(\hat{\boldsymbol{\beta}})^{-1}\mathbf{H}'\}^{-1} (\mathbf{H}\hat{\boldsymbol{\beta}})$, and the power is given by $1 - \eta = \int_{\chi^2_{rk(\mathbf{H})}(1-\alpha)}^{\infty} f(x, rk(\mathbf{H}), \lambda_n) dx$, where α is the type I error, η is the type II error, and $f(x, rk(\mathbf{H}), \lambda_n)$ is the noncentral chi-square density function with degrees of freedom $rk(\mathbf{H})$. Ultimately, we need to find the smallest n such that $1 - \eta \leq \int_{\chi^2_{rk(\mathbf{H})}(1-\alpha)}^{\infty} f(x, rk(\mathbf{H}), \lambda_n) dx$. Numerically, it is obtained by the forward search algorithm. Notably, the sample size calculation we provided is based on asymptotic properties of the two methods. Teerenstra *et al.* [12] have considered a small sample setting for GEE sample size. Similar setting for QIF needs separate investigation.

3.2. Quadratic inference functions sample size under known true correlation

Let us begin with an ideal scenario where the true correlation structure is known. We will remove this rather idealized condition in the next section. We focus on two important cases of designs, respectively under the CS and AR-1 working correlation structures, which are widely used in practice. For both working correlations, we present the details of calculating the asymptotic covariance matrix \mathbf{J}_n for the QIF method and leave the calculation of this matrix for the GEE to Appendix B.

Design under CS working correlation. As shown in Section 2, the CS structure leads to two basis matrices. Under model (5), we explicitly derive the elements of the extended score vector in (3) to be

$$\mathbf{g}_i = (\mathbf{1}', d_i \mathbf{1}', \mathbf{t}'_i, d_i \mathbf{t}'_i, \boldsymbol{\omega}'_i \mathbf{A}_i^{-\frac{1}{2}}, d_i \boldsymbol{\omega}'_i \mathbf{A}_i^{-\frac{1}{2}}, \tilde{\boldsymbol{\omega}}'_i \mathbf{A}_i^{-\frac{1}{2}}, d_i \tilde{\boldsymbol{\omega}}'_i \mathbf{A}_i^{-\frac{1}{2}})' \mathbf{e}_i \stackrel{\text{def}}{=} \mathbf{B}_i \mathbf{e}_i, \quad (6)$$

where $\mathbf{e}_i = (y_{i1} - \mu_{i1}, y_{i2} - \mu_{i2}, \dots, y_{im} - \mu_{im})'$, $\mathbf{A}_i = \text{diag}(V_{i1}, \dots, V_{im})$ and

$$\begin{aligned} \boldsymbol{\omega}'_i &= \sum_{j=1}^m V(\mu_{ij})^{\frac{1}{2}} \mathbf{1}' - [V(\mu_{i1})^{\frac{1}{2}}, \dots, V(\mu_{im})^{\frac{1}{2}}], \\ \tilde{\boldsymbol{\omega}}'_i &= \sum_{j=1}^m t_j V(\mu_{ij})^{\frac{1}{2}} \mathbf{1}' - [t_1 V(\mu_{i1})^{\frac{1}{2}}, \dots, t_m V(\mu_{im})^{\frac{1}{2}}]. \end{aligned}$$

Note that there are only two versions of \mathbf{A}_i 's, $\boldsymbol{\omega}_i$'s, $\tilde{\boldsymbol{\omega}}_i$'s as explained in Section 3.1. Now, denote \bar{d} as the proportion of subjects assigned to the test treatment arm. Then, there are $\bar{d}n = \sum_{i=1}^n d_i$ and $(1 - \bar{d})n$ subjects in the test and controlled treatment arms, respectively. Denoting the true correlation matrix as \mathbf{R}_T , we obtain

$$\mathbf{C}_n(\boldsymbol{\beta}) = \bar{d} \mathbf{B}_1 \mathbf{A}_1^{\frac{1}{2}} \mathbf{R}_T \mathbf{A}_1^{\frac{1}{2}} \mathbf{B}_1 + (1 - \bar{d}) \mathbf{B}_0 \mathbf{A}_0^{\frac{1}{2}} \mathbf{R}_T \mathbf{A}_0^{\frac{1}{2}} \mathbf{B}_0, \quad (7)$$

$$\dot{\mathbf{g}}_n = - \{ \bar{d} \mathbf{B}_1 \mathbf{A}_1 \mathbf{x}'_1 + (1 - \bar{d}) \mathbf{B}_0 \mathbf{A}_0 \mathbf{x}'_0 \}, \quad (8)$$

which are then used to calculate Godambe information matrix $\mathbf{J}_n(\boldsymbol{\beta}) = n \dot{\mathbf{g}}_n'(\boldsymbol{\beta}) \mathbf{C}_n^{-1}(\boldsymbol{\beta}) \dot{\mathbf{g}}_n(\boldsymbol{\beta})$.

Design under AR-1 working correlation. For the AR-1 structure, there are three basis matrices, as stated in Section 2. According to Qu *et al.* [7], the third matrix \mathbf{M}_2 makes little contribution to the QIF and hence is omitted in the formulation of QIF for convenience. Consequently, we obtain

$$\tilde{\mathbf{g}}_i = (\mathbf{1}, d_i \mathbf{1}, \mathbf{t}_i, d_i \mathbf{t}_i, \mathbf{X}_i \mathbf{A}_i^{\frac{1}{2}} \mathbf{M}_1 \mathbf{A}_i^{-\frac{1}{2}})' \mathbf{e}_i \stackrel{\text{def}}{=} \tilde{\mathbf{B}}_i \mathbf{e}_i. \quad (9)$$

The Godambe information matrix is obtained in the same manner as in the first case of CS correlation, in which \mathbf{B}_i is replaced by $\tilde{\mathbf{B}}_i$ in (7) and (8).

In both the aforementioned cases, we also derived the sensitivity matrix $\mathbf{S}_n(\boldsymbol{\beta})$ and the variability matrix $\mathbf{W}_n(\boldsymbol{\beta})$ in the GEE context, and hence the Godambe information matrix of the GEE estimator is given by $\mathbf{J}_n(\boldsymbol{\beta}) = \mathbf{S}_n(\boldsymbol{\beta})' \mathbf{W}_n(\boldsymbol{\beta})^{-1} \mathbf{S}_n(\boldsymbol{\beta})$. See Appendix B for details regarding the \mathbf{S}_n and \mathbf{W}_n matrices.

It is worth pointing out that for model (5) with either CS or AR-1 correlation structure, the weight matrix \mathbf{C}_n in the QIF method is singular. Hence, we delete linearly dependent elements in the extended score vector. Because such elements do not present new information, it is sensible to not use them. In fact, when model (5) only includes the treatment effect and the working correlation is CS, QIF degenerates to QIF, exactly as obtained by reducing these dependent elements. We can find related details in Appendix A. Also, we again emphasize that in QIF, the working correlation structure contributes the two basis matrices, not the value of parameter ρ , to the calculation of matrix \mathbf{J}_n . This is not the case for the GEE, where the matrix \mathbf{J}_n depends on the entire \mathbf{R}_W matrix, including the actual value of ρ .

3.3. Optimal sample size with unknown true correlation

Let R_T denote the true correlation matrix. So, the actual sample size would be $n(R_T)$ if the R_T were known. To overcome the difficulty of an unknown true correlation structure in the sample size determination, we propose to vary the underlying correlation matrix among a class of possible candidates according to a Wishart prior distribution, instead of fixing it to be a prechosen single correlation matrix. This will allow us to reduce the subjectivity related to the choice of the true correlation matrix in the design. To be precise, let correlation matrix $\mathbf{R} \sim \text{Wishart}(\cdot|\mathbf{R}_0)$, where \mathbf{R}_0 is a prespecified correlation matrix. Then, for each sampled correlation matrix \mathbf{R} , the sample size $n = n(\mathbf{R})$ is obtained by the sample size determination procedure in Section 3.2. We aim to choose an optimal sample size that minimizes the following average risk:

$$E\{n(\mathbf{R}) - n(\mathbf{R}_T)\}^2 = \int_{\mathbf{R} \in \mathcal{R}} \{n(\mathbf{R}) - n(\mathbf{R}_T)\}^2 g(\mathbf{R}) d\mathbf{R},$$

where $g(\mathbf{R})$ is the density of Wishart distribution and \mathcal{R} is the space of correlation matrices. It is easy to see that the optimal sample size is $n^* = E_{\mathbf{R}} n(\mathbf{R}) = \int_{\mathbf{R} \in \mathcal{R}} n(\mathbf{R}) g(\mathbf{R}) d\mathbf{R}$. Practically, this integral can be evaluated by the Monte Carlo simulation method. In effect, the Monte Carlo simulation will bring in variation of correlation matrices, and as a result, we can obtain not only the mean sample size n^* but also a sample size distribution, which in practice, may be more valuable as it provides more options to practitioners. Moreover, this procedure gives us a venue where we can conduct a fair comparison between QIF and GEE analysis because, in this case, the comparison will not depend on specific choices of true correlation structures.

3.4. Numerical illustration

In this section, we provide three examples to illustrate the proposed QIF and GEE sample size formulas and the comparison of these two methods on the basis of known and unknown true correlation structures, respectively.

Example 1

Consider several different combinations of true (\mathbf{R}_T) and working (\mathbf{R}_W) correlation structures, as listed in Table I, where we set effect size $\beta = (\beta_0, \beta_1, \beta_2, \beta_3)$ as $(1, 0.5, 0.4, 0.1)$, $\rho = 0.5$, type I error rate at 0.05, and visit times at 0, 2, 4, and 6. Under a design of complete randomization, we choose a 50:50 assignment of subjects into two treatment arms. In Figure 1, we see that the sample size increases as power increases and the QIF sample size is smaller than (or equal to) that of the GEE when the working correlation is the same as (or different than) the true one.

Example 2

We fix the power to be 80% and use the following true unstructured (UN) correlation matrix:

$$\mathbf{R}_T = \begin{pmatrix} 1 & 0.4 & 0.3 & 0.2 \\ & 1 & 0.5 & 0.4 \\ & & 1 & 0.6 \\ & & & 1 \end{pmatrix}. \tag{10}$$

The 1-dependence structure means that only the measurements at two nearby time points are correlated. Mathematically, it is characterized by a width-three-banded matrix—with elements 1's along

Table I. A comparison of the sample size requirements by GEE and QIF analysis with QIF-calculated sample size and percentage of reduction in parenthesis.

True	Null	Working correlation	
		CS	AR-1
Independent	$\beta_1 = 0$	665 (642, 3.5)	690 (642, 7.0)
	$\beta_1 = \beta_3 = 0$	321 (281, 12.5)	285 (281, 1.4)
CS	$\beta_1 = 0$	634 (634, 0)	702 (635, 9.5)
	$\beta_1 = \beta_3 = 0$	643 (643, 0)	682 (644, 5.6)
AR-1	$\beta_1 = 0$	811 (739, 8.9)	736 (736, 0)
	$\beta_1 = \beta_3 = 0$	603 (565, 6.3)	565 (565, 0)
1-dep	$\beta_1 = 0$	899 (773, 14.0)	757 (749, 1.1)
	$\beta_1 = \beta_3 = 0$	557 (479, 14.0)	483 (479, 0.8)
UN	$\beta_1 = 0$	605 (600, 0.8)	693 (616, 11.1)
	$\beta_1 = \beta_3 = 0$	527 (520, 1.3)	601 (535, 11.0)

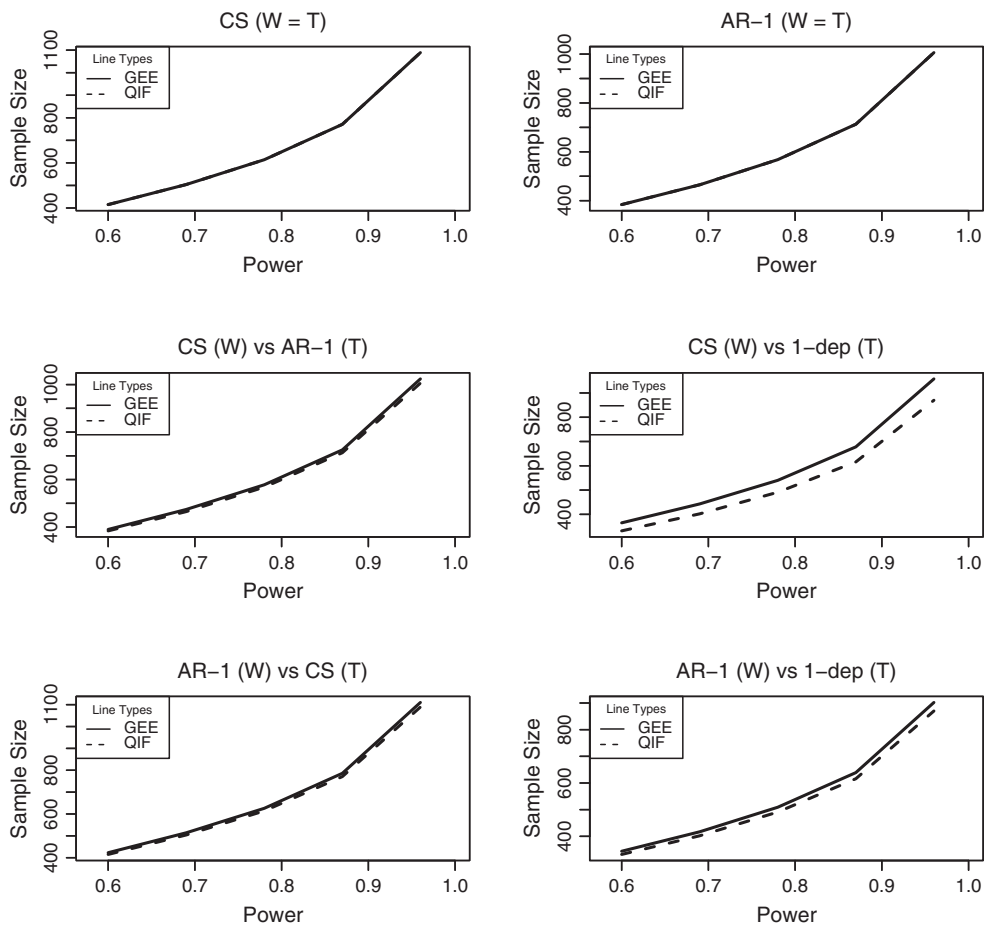


Figure 1. Comparison of sample sizes for testing joint treatment effect between QIF and GEE by power. Effect size, (1, 0.5, 0.4, 0.1); correlation, 0.5; type I error rate, 0.05; and time points, 0, 2, 4, 6.

the main diagonal and ρ at the first upper and lower diagonals. When the correlation matrix is correctly specified ($\mathbf{R}_T = \mathbf{R}_W$), the GEE and QIF analysis require the same sample size, as indicated by the bolded numbers in Table I. This table also clearly indicates the sample size saving of the QIF analysis in all cases with misspecified correlation structures, and the amount of saving by the QIF varies from 1.9% to 10.1%, depending on how severely the working correlation matrix deviates from the true correlation

structure. Clearly, the QIF-based design is advantageous over the GEE-based design in terms of study cost saving, when the working correlation structure used in the design is different from the true one. Misspecification is indeed the case in practice.

Example 3

Arguably, knowing the true correlation matrix of the data that are not yet collected is not possible. Figure 2 displays several examples in which the minimal average risk strategy is applied to the optimal sample size. All the panels on the left column are the sample size distributions when the working CS structure is used in the design, whereas those on the right column are based on the use of the working AR-1 structure in the design. As seen in all the cases, the mean sample size from the QIF study design appears consistently smaller than that of the GEE study design. The top two panels are the sample size distributions when the matrix parameter \mathbf{R}_0 in the Wishart distribution is specified as the chosen working correlation. The middle and bottom panels are based on unstructured \mathbf{R}_0 given in (10) and 1-dependence, respectively. Clearly, at the same type I error and power level, the QIF sample size is on average smaller than the GEE sample size, even if the correlation matrices are sampled from a Wishart distribution centered at the working correlation matrix (as in the top panels). When the correlation matrix is simulated from a Wishart distribution with the \mathbf{R}_0 being a 1-dependence matrix, the percentage of QIF sample size saving relative to the GEE is about 6%. When the prior correlation matrix \mathbf{R}_0 is unstructured, the two designs require very similar sample sizes.

To demonstrate the relation of n and m , in Figure 3, we set power as 0.8, type I error as 0.05, and $\rho = 0.5$. When $m = 2$, it is easy to show that the sample size for the two designs is effectively the same. We can find an analytic explanation for this equivalency in Appendix A. Again, as shown by the top two panels, the QIF sample size requirement is identical to that of the GEE when no misspecification

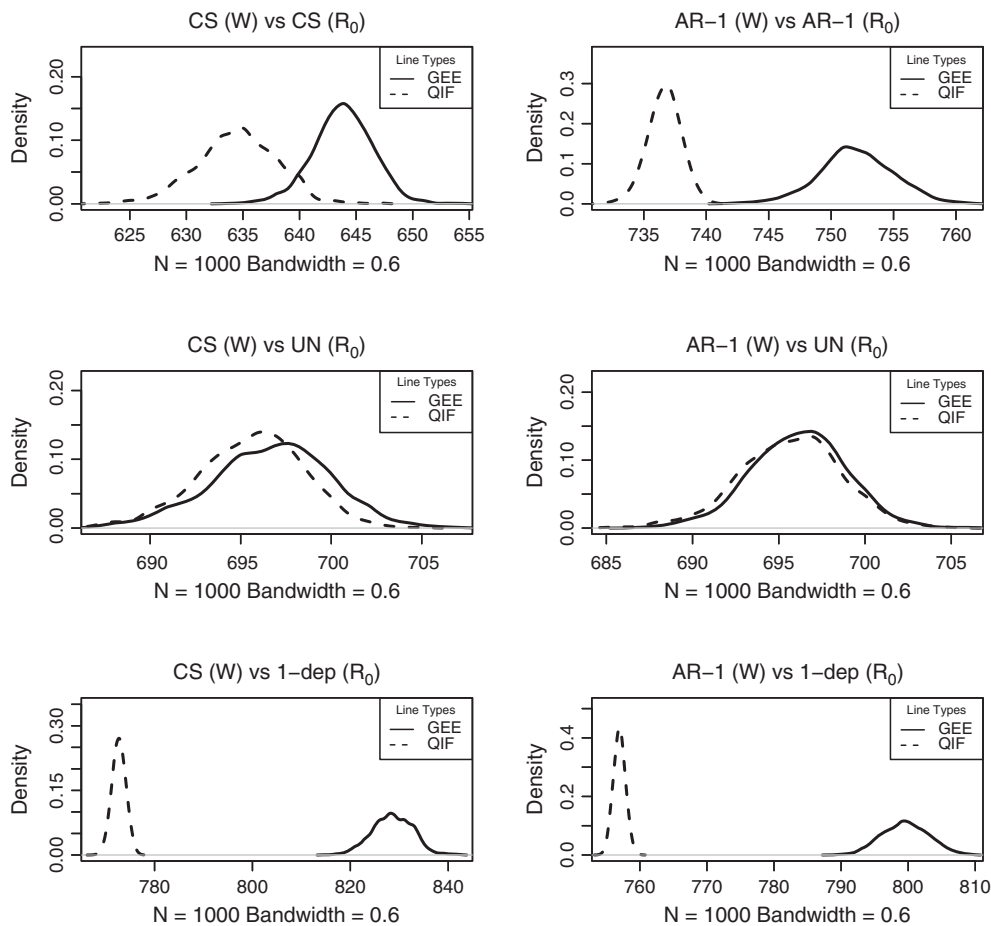


Figure 2. Distributions of the sample size required by the GEE and QIF analysis. \mathbf{R}_0 denotes the prior correlation structure in the Wishart distribution. The top two panels correspond to using the working correlation as the matrix parameter for the Wishart distribution.

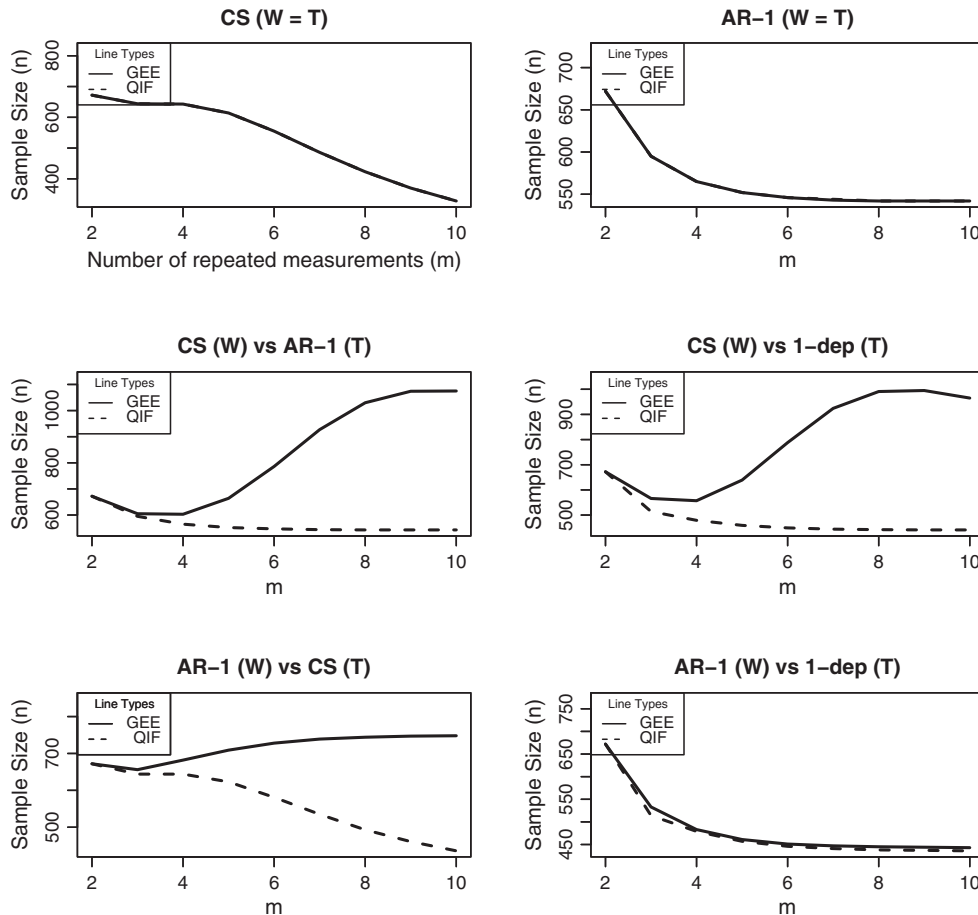


Figure 3. Sample size comparison between QIF and GEE for testing joint treatment effect with varying number of repeated measurements for each individual when misspecification occurs. The top panels consider cases with no misspecification; the middle panels use CS working correlation; and the bottom panels take AR-1 working correlation for both designs.

is present. With misspecification, all the other four panels indicate in Figure 3 that QIF requires smaller sample sizes. To be specific, the percent of sample size saving by QIF is up to 47.6%, 54.2%, 42.7% and 3.9%, respectively. Also, under misspecification, the GEE analysis tends to require a larger sample size with a greater m . As we illustrate in Section 4, this is probably because the degree of misspecification increases as m grows and thus more subjects are needed to fight against the difference between working and true correlation matrices. This phenomenon is also obvious in Figure 4, where we test for the total and main Rx effect. We should notice that this trend is substantially reduced for the QIF method, possibly because QIF uses only basis matrices and hence is more robust to misspecification.

As requested by a reviewer, we have used the setting of Example 1 to perform a sanity check on our sample size formulas. For simplicity, we considered $m = 3$ (i.e., visits at 0, 2, and 4) and simulated 1000 sets of correlated binary outcomes under $\beta_1 = 0$ and 0.5. With model (5), we calculated empirical type I error ($\beta_1 = 0$) and power ($\beta = 0.5$) in the scenarios with and without correlation misspecification. We summarize the results for two cases of no correlation misspecification as follows: (i) the true and working correlation are both CS, the power for both GEE and QIF is 0.8 and both type I error rates are 0.048; and (ii) the true and working correlation are AR-1, both power maintains as 0.8 and both type I errors are 0.052. Moreover, we summarize the results for two cases of misspecified correlations as follows: (i) under a case of the true as CS and the working as AR-1, the power is 0.81 for both methods and the type I errors are 0.052 and 0.054, respectively, for GEE and QIF; and (ii) when the true is AR-1 and the working is CS, the power is 0.8 for GEE and 0.83 for QIF, whereas the corresponding type I errors are 0.058 for GEE and 0.053 for QIF. All these cases have confirmed the desirable control of type I error and power using the given sample sizes obtained from our formulas.

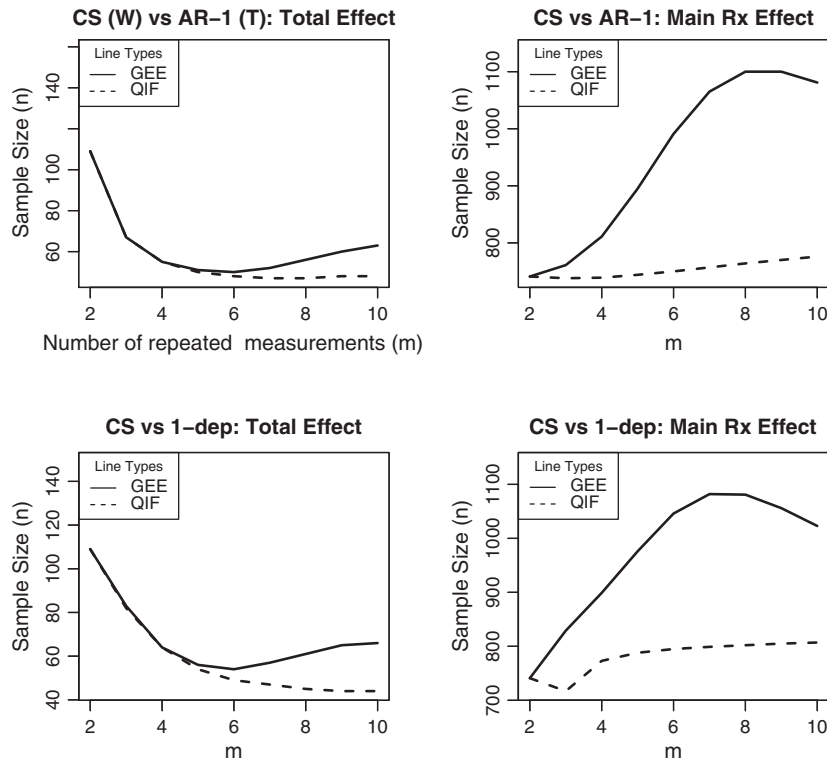


Figure 4. Sample size comparison between QIF and GEE for testing total effect and main Rx effect with varying number of repeated measurements for each individual when R_T is CS.

4. Relation to number of repeated measurements

A special operating characteristic of a longitudinal study is the number of repeated measurements (m). Under the ideal and simple scenarios [1,4], increasing m results in smaller n . Many clinical practitioners take this property for granted and use it as an effective remedy to the insufficiency in n for their studies. In fact, the relationship between n and m is far more complex than the monotonic inverse proportionality in the simple GEE settings.

When there is no misspecification, increasing m inherently produces more data, which is equivalent to increasing sample size. This means that if we can elongate the study period, we can recruit less patients. The top panels in Figure 2 verifies this statement. However, if there is misspecification and we are using a working correlation to approximate the true one, we then are introducing error in our modeling process. As Overall and Tonidandel [13] demonstrated, the degree of misspecification would affect the power of GEE analysis. If the gained information from increasing m is not sufficient to account for the error, we certainly will need to increase our sample size to maintain the power of the study.

To facilitate the discussion, let us first look at some numerical evidence shown in Figure 3 where the null hypothesis is $H_0 : \beta_1 = \beta_3 = 0$ (joint treatment effect). Using the setup of Example 1 in Section 3.4, we take a snapshot fixed at power 0.8, type I error 0.05, and $\rho = 0.5$. It is easy to visualize that (i) when $m = 2$, the sample size for the two designs is the same; (ii) when the working correlation is specified the same as the true correlation, as indicated by the plots in the top two panels, the QIF sample size requirement is identical to that of the GEE and the sample size decreases as the number of follow-up visits increases; (iii) three of the remaining plots in the middle and bottom panels, corresponding to the mismatched working correlation structures to the true ones, clearly indicate opposite behaviors: not only does the QIF method require smaller sample sizes but it also appears rather robust to the increased number of follow-up visits. However, the GEE design is sensitive to varying numbers of follow-up visits. Figure 3 confirms the evidence drawn from Figure 2 under the other two null hypotheses: $H_0 : \beta_1 = \beta_2 = \beta_3$ (total effect) and $H_0 : \beta_1 = 0$ (main treatment effect).

We have attempted to provide some analytic insights as how the sample size n and the cluster size m would behave. However, it is difficult to provide a general theory. In Section 4.1, we can analytically

prove that when the \mathbf{R}_W used in the design is the same as the true \mathbf{R}_T , the monotonic inverse proportionality between n and m still holds. For the case of misspecification, we provide in Section 4.2 analytic arguments in an example to disprove the relationship.

4.1. Correctly specified correlation structure

When there is no correlation misspecification, the sensitivity matrix $\mathbf{S}_n(\boldsymbol{\beta})$ is the same as the variability matrix $\mathbf{W}_n(\boldsymbol{\beta})$ for GEE. Hence, the Godambe information matrix reduces to the sensitivity matrix $\mathbf{S}_n(\boldsymbol{\beta})$, given by (14) in Appendix B. Further, we partition two treatment-arm specific sensitivity matrices, $\mathbf{S}_{1,n}(\boldsymbol{\beta})$ and $\mathbf{S}_{0,n}(\boldsymbol{\beta})$, into 4×4 matrices, with elements denoted as $Q_{ij}^{(1)}$ and $Q_{ij}^{(0)}$, $i, j = 1, \dots, 4$.

In the following discussion, we focus on two cases where the CS and AR-1 structures are used in the design. We show that at a given power, the monotonicity relationship between n and m remains. Our argument shows that as m increases, all of $Q_{11}^{(1)}$, $Q_{33}^{(1)}$, $Q_{11}^{(0)}$, and $Q_{33}^{(0)}$ increase, and hence the resulting standard errors decrease, which leads to smaller sample sizes.

It is easy to show that for the CS working correlation,

$$Q_{11}^{(1)} = -\frac{(m-2)\rho + 1}{(\rho-1)[(m-1)\rho + 1]} \sum_j V_{1j} + \frac{\rho}{(\rho-1)[(m-1)\rho + 1]} \sum_{j \neq k} V_{0j}^{\frac{1}{2}} V_{1k}^{\frac{1}{2}}.$$

When $\sum_{j \neq k} V_{0j}^{\frac{1}{2}} V_{1k}^{\frac{1}{2}} = O(m)^{\ddagger}$, we have

$$Q_{11}^{(1)} = \frac{\rho}{1-\rho} \sum_{j=1}^m V_{1j} + o(1), \quad \text{as } m \rightarrow \infty.$$

Because $V_{1j} > 0$, then $\sum_{j=1}^m V_{1j}$ increases as m increases.

Similarly, for the case of AR-1 correlation, we have

$$Q_{11}^{(1)}(m) = \frac{1+\rho^2}{1-\rho^2} \sum_{j=1}^m V_{1j} - \frac{2\rho}{1-\rho^2} \sum_{j=1}^{m-1} V_{1j}^{\frac{1}{2}} V_{1,j+1}^{\frac{1}{2}} - \frac{\rho^2}{1-\rho^2} V_{11} - \frac{\rho^2}{1-\rho^2} V_{1m}$$

which leads to the observation

$$Q_{11}^{(1)}(m+1) - Q_{11}^{(1)}(m) = \frac{1}{1-\rho^2} (V_{1,m+1}^{\frac{1}{2}} - \rho V_{1m}^{\frac{1}{2}})^2 \geq 0,$$

indicating that $Q_{11}^{(1)}(m)$ increases as m increases.

4.2. Misspecified correlation structure

When the correlation structure is misspecified, the inverse monotonicity relationship between m and n no longer holds. It is interesting to note that the magnitude of deviation between variability matrix $\mathbf{W}(\boldsymbol{\beta})$ and sensitivity matrix $\mathbf{S}(\boldsymbol{\beta})$ affects the properties of this relationship. Similar to the measurement of goodness of fit by Pan [14] and White [15], we adopt the following notation d_R to quantify the deviation:

$$d_R = \|\mathbf{R}_W^{-1} \mathbf{R}_T \mathbf{W}_S^{-1} - \mathbf{R}_W^{-1}\|_2 = \|\mathbf{R}_W^{-1} (\mathbf{R}_T - \mathbf{R}_W) \mathbf{R}_W^{-1}\|_2, \quad (11)$$

with $\|\mathbf{A}\|_2$ denoting the L_2 -norm $\|\mathbf{A}\|_2 = \max_i \sqrt{\sum_j a_{ij}^2}$.

The discrepancy measure d_R in (11) does not have a closed analytical form except in some simple cases but can be easily obtained numerically. In Figure 3, we have seen in the middle left panel that the sample size required by GEE increases as m increases. This panel corresponds to the combination of CS as working and AR-1 as true correlation. To further explore this panel in terms of the relationship of sample size to the deviation d_R , we provide four additional plots in Figure 5 with $\rho = 0.1$ and $\rho = 0.9$, respectively. This figure indicates clearly that the inverse monotonicity relationship between n and m is

$\ddagger o(x)$ means increase in a smaller rate than x .

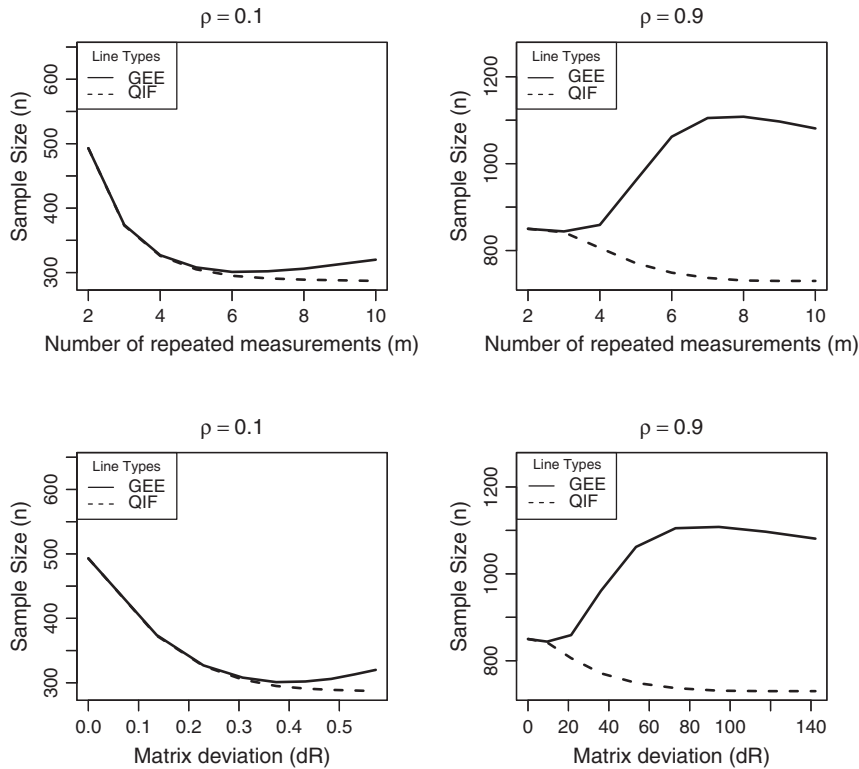


Figure 5. A graphic demonstration of the trend of GEE requiring more sample sizes as the number of repeated measurements and the deviation of true and working correlation matrices increase.

distorted in the GEE analysis, but interestingly, the QIF analysis is barely affected. What is also striking is that the severity of such distortion worsens for larger d_R (and ρ). We have tried other pairs of R_T and R_W scenarios and found very similar properties. One possible explanation is that the GEE sample size is critically dependent on the chosen working correlation matrix and how severely it deviates from the true correlation matrix.

5. Concluding remarks

In this paper, we have developed the sample size and power calculation for both QIF and GEE analysis for dichotomous outcomes and made detailed comparisons between these two designs. We showed that the QIF approach enjoys a sample size saving over the GEE approach; on some occasions, the saving is substantial. We anticipate that such a benefit of sample size saving remains in other types of outcome variables when correlation structure is misspecified.

We proposed an optimal sample size determination in terms of minimal average risk in the scenario where the true correlation structure is unknown. Our strategy is to vary the sample size among possible correlation structures simulated from a prior Wishart distribution and then take the averaged sample size to be used in an actual design. We argue that this strategy is optimal in terms of minimal average risk.

In addition, we demonstrate the robust behavior of the QIF sample size in response to an increased number of follow-up visits, in contrast to GEE, which requires more subjects to follow-up patients over more visits. We regard this as an important property and a clear advantage of the QIF analysis over the GEE analysis as it can reduce the burden of subject recruitment and hence cost of study.

We detail an R package used to determine GEE and QIF sample sizes in Appendix C and make it available at <http://www-personal.umich.edu/pxsong>.

Appendices

In Appendix A, we provide the details concerning the $C_n(\beta)$ matrix. Appendix B presents the detail of the GEE sample size calculation for the logistic model (5), and Appendix C includes some examples of the R package QIFSAMS.

APPENDIX A. C_n matrix in quadratic inference functions

It is easy to see that exchanges between rows or column on $\bar{\mathbf{g}}_n$ will not change the matrix $C_n(\beta)$. We use symbol $\check{\mathbf{X}}$ to denote the post-exchange variable \mathbf{X} and hence have

$$\check{\mathbf{g}}_i = (\mathbf{1}', \mathbf{t}'_i, \omega'_i \mathbf{A}_i^{-\frac{1}{2}}, \tilde{\omega}'_i \mathbf{A}_i^{-\frac{1}{2}}, d_i \mathbf{1}, d_i \mathbf{t}'_i, d_i \omega'_i \mathbf{A}_i^{-\frac{1}{2}}, d_i \tilde{\omega}'_i \mathbf{A}_i^{-\frac{1}{2}})' \mathbf{e}_i \stackrel{\text{def}}{=} (\mathbf{Z}_i, d_i \mathbf{Z}_i)' \mathbf{e}_i \stackrel{\text{def}}{=} \check{\mathbf{B}}_i \mathbf{e}_i, \tag{12}$$

where $\mathbf{Z}_i = (\mathbf{1}', \mathbf{t}'_i, \omega'_i \mathbf{A}_i^{-\frac{1}{2}}, \tilde{\omega}'_i \mathbf{A}_i^{-\frac{1}{2}})'$. We then have

$$\begin{aligned} \check{C}(\beta) &= \bar{d} \check{\mathbf{B}}_1 \mathbf{A}_1^{\frac{1}{2}} \mathbf{R}_T \mathbf{A}_1^{\frac{1}{2}} \check{\mathbf{B}}_1 + (1 - \bar{d}) \check{\mathbf{B}}_0 \mathbf{A}_0^{\frac{1}{2}} \mathbf{R}_T \mathbf{A}_0^{\frac{1}{2}} \check{\mathbf{B}}_0 \\ &= \begin{pmatrix} \mathbf{G} & \mathbf{F} \\ \mathbf{F} & \mathbf{F} \end{pmatrix}, \end{aligned}$$

where

$$\begin{aligned} \mathbf{G} &= \bar{d} \mathbf{Z}_1 \mathbf{A}_1^{\frac{1}{2}} \mathbf{R}_T \mathbf{A}_1^{\frac{1}{2}} \mathbf{Z}'_1 + (1 - \bar{d}) \mathbf{Z}_0 \mathbf{A}_0^{\frac{1}{2}} \mathbf{R}_T \mathbf{A}_0^{\frac{1}{2}} \mathbf{Z}'_0, \\ \mathbf{F} &= \bar{d} \mathbf{Z}_1 \mathbf{A}_1^{\frac{1}{2}} \mathbf{R}_T \mathbf{A}_1^{\frac{1}{2}} \mathbf{Z}'_1. \end{aligned}$$

For $m \geq 2$, matrix \mathbf{F} is singular because the fourth row in \mathbf{Z}_i is a linear combination of the other three. This leads to the singularity of matrix $\check{C}_n(\beta)$, which can be overcome by removing the fourth row. Then, an inference function is given by

$$\hat{\mathbf{g}}_i = (\mathbf{1}' \mathbf{e}_i, \mathbf{t}'_i \mathbf{e}_i, \omega'_i \mathbf{A}_i^{-\frac{1}{2}} \mathbf{e}_i, d_i \mathbf{1} e_i, d_i \mathbf{t}'_i \mathbf{e}_i, d_i \omega'_i \mathbf{A}_i^{-\frac{1}{2}} \mathbf{e}_i)',$$

which was used throughout this paper under the CS correlation when $m \geq 3$. We also removed $\omega'_i \mathbf{A}_i^{-\frac{1}{2}}$ when $m = 2$. It should be aware that for this case, the inference functions of QIF and GEE are the same, and hence, their calculated sample sizes are the same.

We took a similar procedure for the case of AR-1 working correlation.

APPENDIX B. Godambe information matrix for generalized estimating equations

To obtain the GEE Godambe information matrix for the logistic model (5), we find that the sensitivity matrix is

$$\mathbf{S}_n(\beta) = - \sum_{i=1}^n \mathbf{D}'_i \mathbf{A}_i \Sigma_i^{-1} \mathbf{A}_i \mathbf{D}_i \tag{13}$$

$$= -\bar{d} n \mathbf{X}_1 \mathbf{A}_1^{\frac{1}{2}} \mathbf{R}_W^{-1} \mathbf{A}_1^{\frac{1}{2}} \mathbf{X}'_1 - (1 - \bar{d}) n \mathbf{X}_0 \mathbf{A}_0^{\frac{1}{2}} \mathbf{R}_W^{-1} \mathbf{A}_0^{\frac{1}{2}} \mathbf{X}'_0. \tag{14}$$

The variability matrix is

$$\begin{aligned} \mathbf{W}_n(\beta) &= \sum_{i=1}^n \mathbf{D}'_i \mathbf{A}_i \Sigma_i^{-1} \text{Var}(\mathbf{r}_i) \Sigma_i^{-1} \mathbf{A}_i \mathbf{D}_i \\ &= \bar{d} n \mathbf{X}_1 \mathbf{A}_1^{\frac{1}{2}} \mathbf{R}_W^{-1} \mathbf{R}_T \mathbf{R}_W^{-1} \mathbf{A}_1^{\frac{1}{2}} \mathbf{X}'_1 + (1 - \bar{d}) n \mathbf{X}_0 \mathbf{A}_0^{\frac{1}{2}} \mathbf{R}_W^{-1} \mathbf{R}_T \mathbf{R}_W^{-1} \mathbf{A}_0^{\frac{1}{2}} \mathbf{X}'_0. \end{aligned}$$

Then, the Godambe information matrix is calculated by $\mathbf{J}_n(\beta) = \mathbf{S}_n(\beta)' \mathbf{W}_n(\beta)^{-1} \mathbf{S}_n(\beta)$. The aforementioned sensitivity can be partitioned in the following fashion. When there is no model misspecification, denoting $R_W^{-1} = (R_{jk})_{mm}$, we find

$$\mathbf{S}_{1,n}(\beta) = \begin{pmatrix} Q_{11}^{(1)} & Q_{11}^{(1)} & Q_{13}^{(1)} & Q_{13}^{(1)} \\ Q_{11}^{(1)} & Q_{11}^{(1)} & Q_{13}^{(1)} & Q_{13}^{(1)} \\ Q_{13}^{(1)} & Q_{13}^{(1)} & Q_{33}^{(1)} & Q_{33}^{(1)} \\ Q_{13}^{(1)} & Q_{13}^{(1)} & Q_{33}^{(1)} & Q_{33}^{(1)} \end{pmatrix}, \quad \mathbf{S}_{0,n}(\beta) = \begin{pmatrix} Q_{11}^{(0)} & 0 & Q_{13}^{(0)} & 0 \\ 0 & 0 & 0 & 0 \\ Q_{13}^{(0)} & 0 & 0 & 0 \\ 0 & 0 & 0 & Q_{33}^{(0)} \end{pmatrix},$$

where

$$\begin{aligned}
 Q_{11}^{(1)} &= \sum_{j,k} R_{jk} V_{1j}^{\frac{1}{2}} V_{1k}^{\frac{1}{2}}, & Q_{11}^{(0)} &= \sum_{j,k} R_{jk} V_{0j}^{\frac{1}{2}} V_{0k}^{\frac{1}{2}} \\
 Q_{13}^{(1)} &= \sum_{j,k} R_{jk} t_k V_{1j}^{\frac{1}{2}} V_{1k}^{\frac{1}{2}}, & Q_{13}^{(0)} &= \sum_{j,k} R_{jk} t_k V_{0j}^{\frac{1}{2}} V_{0k}^{\frac{1}{2}} \\
 Q_{33}^{(1)} &= \sum_{j,k} R_{jk} t_j V_{1j}^{\frac{1}{2}} t_k V_{1k}^{\frac{1}{2}}, & Q_{33}^{(0)} &= \sum_{j,k} R_{jk} t_j V_{0j}^{\frac{1}{2}} t_k V_{0k}^{\frac{1}{2}}.
 \end{aligned}$$

APPENDIX C. R package: QIFSAMS

The R package QIFSAMS includes two functions that calculate sample size for the QIF and GEE analysis, respectively. We can download this package and the user's manual from webpage <http://www.umich.edu/~pxsong>.

To obtain the sample size, the user needs to specify effect sizes and correlation matrices. QIFSAMS takes effect sizes of zero time and/or interaction effect as well. QIFSAMS requires $m \geq 2$ to reflect a longitudinal study design. For example, we used the following commands to produce Table I.

```

GEE.n <-GEE_Size(typeIerror=0.05, power=0.8, coeff=c(1,0.5,0.4,0.1),
  num_repeated = 4, ratioTRT = 0.5,timeUnit = 2, visits = c(0,2,4),
  corr_T = 0.5, corr_W = 0.5, R_wk_ID = 1, R_true_ID = 1,R_un =
  corr.matrix.un, test.ID = 1)
QIF.n <-QIF_Size(typeIerror=0.05, power=0.8, coeff=c(1,0.5,0.4,0.1),
  num_repeated = 4, ratioTRT = 0.5,timeUnit = 2, visits = c(0,2,4),
  corr=0.5 ,R_wk_ID=1, R_true_ID=1, R_un=corr.matrix.un, test.ID=1).

```

The arguments type I error, power, effect size, number of repeated measurements, proportion of subjects in the test treatment group, time points, and correlation parameters need to be specified in the functions. Note that in GEE, we need to specify both true and working correlation parameters, but in QIF, we only need the true correlation parameters. For example, when both true and working correlation are CS, as shown previously, we set both `R_wk_ID` and `R_true_ID` to be 1 and obtain the corresponding two sample sizes. For other cases in Table I, we vary these two parameters.

Both functions combines `R_true_ID = 4` and `R_un` to specify a UN correlation, which can be sampled from a Wischart distribution. In Figure 2, we generated 1000 such samples, choose a working correlation, and then obtain the sample size distributions.

We can find more details from the help file from the package.

Acknowledgement

An NSF grant (DMS 0904177) to the second author partly supported this research.

References

1. Diggle P, Heagerty P, Liang K-Y, Zeger S. *Analysis of Longitudinal Data*. Oxford University Press: New York, 2002.
2. Song PXX. *Correlated Data Analysis*. Springer: New York, 2007.
3. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrics* 1986; **73**:13–22.
4. Pan W. Sample size and power calculation with correlated binary data. *Controlled Clinical Trials* 2001; **22**:211–227.
5. Jung S-H, Ahn C. Sample size estimation for GEE method for comparing slopes in repeated measurement data. *Statistics in Medicine* 2003; **22**:1305–1315.
6. Rochon J. Application of GEE procedures for sample size calculations in repeated measures experiments. *Statistics in Medicine* 1998; **98**(E2):3247–3259.
7. Qu A, Lindsay BG, Li B. Improving generalised estimating equations using quadratic inference function. *Biometrika* 2000; **87**(4):823–836.
8. Song PXX, Jiang Z, Park E, Qu A. Quadratic inference functions in marginal models for longitudinal data. *Statistics in medicine* 2009; **28**(29):3683–3696.
9. Demidenko E. Sample size determination for logistic regression revisited. *Statistics in Medicine* 2006; **26**(18).
10. Hsieh FY, Bloch DA, Larsen MD. A simple method of sample size calculation for linear and logistic regression. *Statistics in Medicine* July 1998; **17**(14):1623–34.
11. Hansen L. Large sample properties of generalized method of moments estimators. *Econometrica* 1982; **50**:1029–54.

12. Teerenstra S, Lu B, Preisser JS, van Achterberg T, Borm GF. Sample size considerations for GEE analyses of three-level cluster randomized trials. *Biometrics* 2010.
13. Overall JE, Tonidandel S. Robustness of generalize estimation equation (GEE) test of significance against misspecification of the error structure model. *Biometrics* 2004.
14. Pan W. Akaike's information criterion in generalized estimating equations. *Biometrics* 2001; **57**:120–125.
15. White H. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 1980; **48**:817–838.