Available at
www.ElsevierMathematics.com
POWERED BY SCIENCE DIRECT
Journal of Statistical Planning and Inference 128 (2005) 43–59
journal of statistical planning and inference
www.elsevier.com/locate/jspi

On power and sample size calculations for Wald tests in generalized linear models

Gwowen Shieh

Department of Management Science, National Chiao Tung University, Hsinchu 30050, Taiwan, ROC

Received 27 February 2003; accepted 6 September 2003

Abstract

A Wald test-based approach for power and sample size calculations has been presented recently for logistic and Poisson regression models using the asymptotic normal distribution of the maximum likelihood estimator, which is applicable to tests of a single parameter. Unlike the previous procedures involving the use of score and likelihood ratio statistics, there is no simple and direct extension of this approach for tests of more than a single parameter. In this article, we present a method for computing sample size and statistical power employing the discrepancy between the noncentral and central chi-square approximations to the distribution of the Wald statistic with unrestricted and restricted parameter estimates, respectively. The distinguishing features of the proposed approach are the accommodation of tests about multiple parameters, the flexibility of covariate configurations and the generality of overall response levels within the framework of generalized linear models. The general procedure is illustrated with some special situations that have motivated this research. Monte Carlo simulation studies are conducted to assess and compare its accuracy with existing approaches under several model specifications and covariate distributions.

Keywords: Information matrix; Likelihood ratio test; Logistic regression; Noncentral chi-square; Poisson regression; Score test

1. Introduction

Generalized linear models were first introduced by Nelder and Wedderburn (1972) and are broadly applicable in almost all scientific fields. A thorough development can be found in McCullagh and Nelder (1989). The class of generalized linear models is specified by assuming that independent scalar response variables Y_i, $i=1, \ldots, N$, follow

E-mail address: gwshieh@mail.nctu.edu.tw (G. Shieh).

0378-3758/$ - see front matter
doi:10.1016/j.jspi.2003.09.017

a probability distribution belonging to the exponential family of probability distributions with probability function of the form

$$\exp[\{Y\theta - b(\theta)\}/a(\phi) + c(Y, \phi)]. \tag{1}$$

The expected value $E(Y) = \mu$ is related to the canonical parameter θ by the function $\mu = b'(\theta)$, where b' denotes the first derivative of b. The link function g relates the linear predictors η to the mean response $\eta = g(\mu)$. The linear predictors can be written as

$$\eta = \mathbf{X}^{\mathrm{T}}\beta,$$

where $\mathbf{X}=(X_1, \ldots, X_k)^{\mathrm{T}}$ is a $k \times 1$ vector of covariates, and $\beta=(\beta_1, \ldots, \beta_k)^{\mathrm{T}}$ represents the corresponding $k \times 1$ vector of unknown regression coefficients. The scale parameter ϕ is assumed to be known. Assume $(y_i, \mathbf{x}_i)$ is a random sample from the joint distribution of $(Y, \mathbf{X})$ with probability function $f(Y, \mathbf{X}) = f(Y|\mathbf{X})f(\mathbf{X})$, where $f(Y|\mathbf{X})$ has the form defined in (1) and $f(\mathbf{X})$ is the probability function for $\mathbf{X}$. The form of $f(\mathbf{X})$ is assumed to depend on none of the unknown parameters β. The likelihood function associated with the data is

$$L(\beta) = \prod_{i=1}^{N} f(y_i, \mathbf{x}_i) = \prod_{i=1}^{N} f(y_i|\mathbf{x}_i)f(\mathbf{x}_i).$$

Let $\beta_1 = (\beta_1, \ldots, \beta_q)^{\mathrm{T}}$ and $\beta_2 = (\beta_{q+1}, \ldots, \beta_k)^{\mathrm{T}}$ represent the first q and the last p unknown regression coefficients of β, respectively ($k=q+p$, $q \geqslant 1$, $p \geqslant 1$). We wish to test the composite null hypothesis H_0: $\beta_2=\mathbf{0}$ against the alternative hypothesis H_1: $\beta_2 \neq \mathbf{0}$, while treating β_1 as nuisance parameters. It follows from the standard asymptotic theory that the maximum likelihood estimator $\hat{\beta} = (\hat{\beta}_1^{\mathrm{T}}, \hat{\beta}_2^{\mathrm{T}})^{\mathrm{T}}$ is asymptotically normally distributed with mean $\beta = (\beta_1^{\mathrm{T}}, \beta_2^{\mathrm{T}})^{\mathrm{T}}$ and with variance–covariance matrix given by the inverse of the $k \times k$ Fisher information matrix $\mathbf{I}(\beta_1, \beta_2)$, where the (i, j)th element of $\mathbf{I}$ is

$$I_{ij} = -E\left(\frac{\partial^2 \log L}{\partial \beta_i \partial \beta_j}\right), \quad i, j = 1, \ldots, k$$

and $E[\,\cdot\,]$ denotes the expectation taken with respect to the joint distribution of $(Y_1, \ldots, Y_N, \mathbf{X}_1, \ldots, \mathbf{X}_N)$. The Wald test statistic of the hypothesis is

$$W = \hat{\beta}_2^{\mathrm{T}}\hat{\mathbf{V}}^{-1}\hat{\beta}_2, \tag{2}$$

where $\hat{\mathbf{V}}$ is the lower-right $p \times p$ sub-matrix of $\mathbf{I}^{-1}(\hat{\beta}_1, \hat{\beta}_2)$. The actual test is performed by referring the statistic to its asymptotic distribution under the null hypothesis, which is a chi-square distribution with p degrees of freedom. In general, there is no simple closed-form expression for Fisher's information matrix except in some special cases.

For the purpose of power and sample size calculations, an approximate expression for Fisher's information matrix was provided in Whittemore (1981) for logistic regression model. The approximation employs the moment generating function of the covariates and is valid when the overall response probability is small. A formula for determining the sample size is developed from the resulting asymptotic variance of the maximum likelihood estimator of the parameters. Later, the technique was extended to the Poisson regression model in Signorini (1991). However, in this case, the expression of Fisher's information matrix is exact and there is no restriction of use in terms of the overall response level. Shieh (2001) has recently presented a direct modification of the sample size formulas in Whittemore (1981) and Signorini (1991). The major consideration of Shieh (2001) is that the value of the nuisance parameter under the restriction specified in the null model is different from that for the alternative model. In contrast, Whittemore (1981) and Signorini (1991) set the identical values for nuisance parameter under both the null and alternative models. According to the pedagogical arguments to be discussed in Section 2 and simulation results in Shieh (2001), the method of Shieh (2001) is more accurate than the approaches of Whittemore (1981) and Signorini (1991) for logistic and Poisson regression models, respectively. Therefore we will not discuss the latter two approaches further in this article. However, it is important to note that these Wald-type or Z test-based approaches are applicable only for tests of a single parameter ($p=1$).

The basic idea of Shieh (2001) can be easily demonstrated by considering the determination of sample sizes needed to detect a difference between two proportions p_1 and p_2. Assume an asymptotic normality of the transformed binomial proportions, the widely used formula for calculating the required sample sizes for the equal group size designs is

$$N = 2\{Z_{\alpha/2}[2\bar{p}(1-\bar{p})]^{1/2} + Z_\gamma[p_1(1-p_1)+p_2(1-p_2)]^{1/2}\}^2/(p_1-p_2)^2,$$

where $\bar{p}=(p_1+p_2)/2$ and Z_α represents the $100(1-\alpha)$th percentile of a standard normal distribution. Similar formula was given by Fleiss (1981, Eq. (3.14)) and Sahai and Khurshid (1996, Eq. (7)). More importantly, the formula can be derived from the general result in Shieh (2001) under the formulation of simple logistic regression. Note that this formula takes into account the different variance structures associated with the null hypothesis H_0: $p_1=p_2$ and alternative hypothesis H_1: $p_1 \neq p_2$. In contrast, an alternative simple approximation to the formula is obtained by replacing $2\bar{p}(1-\bar{p})$ with $p_1(1-p_1)+p_2(1-p_2)$:

$$N = 2[p_1(1-p_1)+p_2(1-p_2)](Z_{\alpha/2}+Z_\gamma)^2/(p_1-p_2)^2,$$

see Sahai and Khurshid (1996, Eq. (13)). Although these two formulas give similar results for balanced designs, they are fundamentally different with respect to the specification of variance under the null hypothesis. Unfortunately, this notion is overlooked and is rarely addressed for analogous adaptation under unbalanced designs and other more complex models. The discrepancy between these two approaches could be substantial for unequal allocation as shown later in the simulation study.

Along the same line of power and sample size calculations within the generalized linear models framework, two other major formulas have been proposed. They are the score and likelihood ratio test approaches developed by Self and Mauritsen (1988) and Self et al. (1992), respectively. The likelihood ratio test approach of Self et al. (1992) is easier to implement and more accurate over a much wider range of model specifications than the score test method of Self and Mauritsen (1988). Nevertheless these two approaches are limited to the cases that the number of covariate configurations is finite. This assumption was relaxed for the likelihood ratio test approach in Shieh

(2000) to accommodate covariate with an infinite number of configurations. Note that the probability function $f(\mathbf{X})$ for covariate $\mathbf{X}$ defined above covers implicitly discrete and/or continuous cases.

Unlike the score and likelihood ratio test procedures mentioned above, there is no simple and direct extension of the Wald test approach of Shieh (2001) for testing more than a single parameter ($p > 1$). Although there is some undesired property with the Wald test reported in Hauck and Donner (1977), this statistic is calculated routinely by standard statistical packages and continues to be an important alternative in the context of statistical hypotheses testing for a wide range of complicated situations. Furthermore, the Wald test is quite appealing for constructing the confidence intervals and regions of parameters without much extra effort. Hence, it should be extremely useful to generalize the approach of Shieh (2001) such that one can compute the power and sample size for tests involving arbitrary number of parameters ($p \geqslant 1$). In that case, the Wald test can be employed for the whole line of analyses throughout the study, including the accurate sample size calculations at the planning stage of research, without resorting to other test procedures. This article aims to provide power and sample size formula for the Wald test within the framework of generalized linear models defined in (1). The distinguishing features of the proposed approach are the accommodation of tests about multiple parameters, the flexibility of covariate configurations and the generality of overall response levels.

In Section 2, the proposed methodology is described. In Section 3, the procedures are illustrated with examples. Simulation studies are performed and results are presented in Section 4. Section 5 contains some concluding remarks.

2. The proposed method

In order to perform power analyses and sample size calculations, we need to examine the asymptotic mean and variance of $\hat{\beta}_2$ under both the alternative and null hypotheses. It can be shown that $\mathbf{I}(\beta_1, \beta_2) = N \cdot \boldsymbol{\Xi}$ under the alternative model with unrestricted parameter estimators, where

$$\boldsymbol{\Xi} = E_{\mathbf{X}}\left[a^{-1}(\phi)b''(\theta)\left(\frac{\partial\theta}{\partial\eta}\right)^2 \mathbf{X}\mathbf{X}^{\mathrm{T}}\right].$$

$E_{\mathbf{X}}[\cdot]$ denotes the expectation taken with respect to the distribution of $\mathbf{X}$ and b'' denotes the second derivative of b. Therefore, the asymptotic distribution of $\hat{\beta}_2$ is normal with mean β_2 and variance $\mathbf{V} = \boldsymbol{\Sigma}/N$, where $\boldsymbol{\Sigma}$ is the lower-right $p \times p$ sub-matrix of $\boldsymbol{\Xi}^{-1}$. Furthermore, we approximate the distribution of the Wald statistic W defined in (2) by a noncentral chi-square distribution with p degrees of freedom and noncentrality parameter $N\delta$ with

$$\delta = \beta_2^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\beta_2 \tag{3}$$

denoted by $\chi_p^2(N\delta)$.

For the distribution of $\hat{\beta}_2$ under the restriction $\beta_2 = \mathbf{0}$ of the null model, our formulation is analogous to that of Shieh (2001). Let $\beta_1^* = (\beta_1^*, \ldots, \beta_q^*)$ denote the solution of the equation $\lim_{N\to\infty} N^{-1}E[S_N(\beta_1,\mathbf{0})]=\mathbf{0}$, where S_N represents derivatives of the log-likelihood function with respect to $\beta_1=(\beta_1,\beta_2,\ldots,\beta_q)$ by setting $\beta_2=\mathbf{0}$, and the expectation is taken with respect to the true value of $\beta=(\beta_1,\beta_2)$. With the restricted parameter estimates β_1^*, we approximate the distribution of $\hat{\beta}_2$ by a normal distribution with mean $\mathbf{0}$ and variance–covariance matrix $\mathbf{V}^*=\mathbf{\Sigma}^*/N$, where $\mathbf{\Sigma}^*$ is the proper component of $\mathbf{\Xi}^{*-1}$ as that of $\mathbf{\Sigma}$ to $\mathbf{\Xi}^{-1}$ mentioned above, and

$$\mathbf{\Xi}^*=E_{\mathbf{X}}\left[a^{-1}(\phi)b''(\theta)\left(\frac{\partial\theta}{\partial\eta}\right)^2\mathbf{X}\mathbf{X}^{\mathrm{T}}\bigg|_{(\beta_1^*,\mathbf{0})}\right]$$

$$=E_{\mathbf{X}}\left[a^{-1}(\phi)b''(\theta^*)\left(\frac{\partial\theta^*}{\partial\eta^*}\right)^2\mathbf{X}\mathbf{X}^{\mathrm{T}}\right].$$

θ^* and η^* denote the canonical parameter θ and linear predictor η evaluated at $\beta=(\beta_1^*,\mathbf{0})$, respectively. In general, $\mathbf{\Sigma}$ and $\mathbf{\Sigma}^*$ are not equivalent and neither are $\mathbf{V}$ and $\mathbf{V}^*$. This distinction of variance–covariance matrix under null and alternative models plays an important role in sample size determination shown next.

Let $\chi^2_{p,\alpha}$ denote the $100(1-\alpha)$th percentile of a central chi-square distribution with p degrees of freedom, and $\chi^2_{p,1-\gamma}(\lambda^*)$, represent the $100\cdot\gamma$th percentile of a noncentral chi-square distribution with p degrees of freedom and noncentrality parameter λ^*. Also, let random variable $\mathbf{Z}$ have a p-dimensional normal density with mean $\mathbf{0}$ and variance–covariance matrix $\mathbf{\Sigma}$, denoted by $N_p(\mathbf{0},\mathbf{\Sigma})$. For a generalized linear model with specified parameter values $\beta=(\beta_1,\beta_2)$ and chosen covariate distribution $f(\mathbf{X})$, the sample size needed to test hypothesis H_0: $\beta_2=\mathbf{0}$ with specified significance level α and power $1-\gamma$ against the alternative hypothesis H_1: $\beta_2\neq\mathbf{0}$ is computed as follows. First, compute the adjusted significance level

$$\alpha^*=P(\mathbf{Z}^{\mathrm{T}}\mathbf{\Sigma}^{*-1}\mathbf{Z}>\chi^2_{p,\alpha}). \tag{4}$$

Next, find the noncentrality parameter λ^* of a noncentral chi-square distribution with p degrees of freedom such that $\chi^2_{p,1-\gamma}(\lambda^*)=\chi^2_{p,\alpha^*}$. Then the sample size estimate is computed as

$$N_{\mathrm{P}}=\lambda^*/\delta, \tag{5}$$

where δ is defined in (3). Note that this procedure can be reversed to calculate the statistical power. Given parameter values $\beta=(\beta_1,\beta_2)$, chosen covariate distribution $f(\mathbf{X})$, and sample size N, the statistical power achieved for testing hypothesis H_0: $\beta_2=\mathbf{0}$ with specified significance level α against the alternative H_1: $\beta_2\neq\mathbf{0}$ is the probability

$$P\{\chi^2_p(N\delta)>\chi^2_{p,\alpha^*}\}.$$

The notion of adjusted significance level α^* defined by (4) is motivated by knowledge that, when $p=1$, similar adjustment is naturally incorporated in the sample size determination. It follows immediately from (4) with $p=1$ that N_{P} in (5) is the required sample size such that

$$P\{\chi^2_1(N_{\mathrm{P}}\delta)>\chi^2_{1,\alpha^*}\}=1-\gamma, \quad \text{where } \delta=\beta_k^2/\Sigma, \tag{6}$$

where Σ is the univariate form of $\mathbf{\Sigma}$. With the relationship between the square of a standard normal variable and a chi-square variable of 1 degree of freedom, it can be demonstrated that $\chi^2_{1,\alpha^*} = \chi^2_{1,\alpha^*}(\Sigma^*/\Sigma)$, where Σ^* is the univariate version of $\mathbf{\Sigma}^*$. Such simplification leads essentially to the formula proposed in Shieh (2001)

$$N_{\mathrm{P}} = [Z_{\alpha/2} \cdot (\Sigma^*)^{1/2} + Z_{\gamma} \cdot (\Sigma)^{1/2}]^2/\beta_k^2. \tag{7}$$

Hence, the proposed approach subsumes the method of Shieh (2001) as a special case.

Furthermore, the restriction of small response rate for logistic regressions in Shieh (2001) is relaxed here. It appears, however, that there is no analytic form for the adjusted significance level except in the special case of $p = 1$ presented above. A useful and analogous expression for the adjusted significance level in (4) is

$$\alpha^* = P(Q > \chi^2_{p,\alpha}),$$

where $Q = \sum_{l=1}^{p} \lambda_l W_l$ with λ_l, $l = 1, \ldots, p$, are the eigenvalues of $\mathbf{\Sigma}^{1/2}\mathbf{\Sigma}^{*-1}\mathbf{\Sigma}^{1/2}$ and W_l, $l = 1, \ldots, p$, are independent central chi-square χ^2_1 random variables. Note that both $\mathbf{\Sigma}$ and $\mathbf{\Sigma}^*$ are positive definite, therefore, $\lambda_l > 0$, $l = 1, \ldots, p$. Consequently, Q is a positive linear combination of central chi-square random variables of 1 degree of freedom. It was shown in Wood (1989) that the distribution of Q can be adequately approximated by a three-parameter F distribution. Accordingly, we propose to approximate the adjusted significance level by

$$\alpha^* \cong P\left(F > \frac{a_2 t_2}{a_1 t_1} \chi^2_{p,\alpha}\right),$$

where F has an F distribution with degrees of freedom $2a_1$ and $2a_2$, $a_1 = 2k_1(k_3 k_1 + k_1^2 k_2 - k_2^2)/t_1$, $a_2 = 3 + 2k_2(k_2 + k_1^2)/t_2$, $t_1 = 4k_2^2 k_1 + k_3(k_2 - k_1^2)$, $t_2 = k_3 k_1 - 2k_2^2$, and

$$k_r = 2^{r-1}(r-1)! \sum_{l=1}^{p} \lambda_l^r \quad \text{for } r = 1, 2 \text{ and } 3.$$

This approach is implemented in the numerical assessments shown later. While more involved iterative computing algorithms can be employed to provide exact calculations of α^*, it is still of great interest to have good and simple approximation. With respect to the trade-off between numerical accuracy and computation-wise complexity, the advantage of simplicity becomes more prominent as p grows larger. More importantly, according to our finding, the non-iterative three-parameter F approximation appears to provide satisfactory results. For the general developments regarding the derivation of distributions of linear combinations of independent chi-square variables, see Johnson et al. (1994, Section 18.8) for a comprehensive discussion.

Instead of considering the adjustment of significance level, an alternative direct method of obtaining sample size, with the same general setup and definition of δ just presented, is given by

$$N_{\mathrm{D}} = \lambda/\delta, \tag{8}$$

where λ is the noncentrality parameter of a noncentral chi-square distribution with p degrees of freedom such that $\chi^2_{p,1-\gamma}(\lambda) = \chi^2_{p,\alpha}$. Actually, this formula has been implemented in the commercial software POWER AND PRECISION (Biostat, 2000). At the first sight, this seems to be a questionable approach because the distribution property of $\hat{\beta}_2$ under the restriction prescribed in the null hypothesis is not addressed. Furthermore, a close examination of this procedure, as in the previous discussion of the proposed approach for tests of a single parameter, shows that this formula sets the variance of $\hat{\beta}_k$ under the restriction of null model exactly the same as that under the unrestricted alternative model as follows:

$$N_{\mathrm{D}} = [Z_{\alpha/2} + Z_\gamma]^2 \Sigma/\beta_k^2. \tag{9}$$

However, since the Fisher information matrix is a function of the parameters β, accordingly the variance–covariance matrices given by the inverse of the Fisher information matrix generally disagree for different parameter settings. Therefore, formula (8) and the proposed approach (5) are fundamentally different with respect to the specification of variance–covariance approximation of $\hat{\beta}_2$ under the composite null hypothesis. This phenomenon is illustrated in the following examples.

3. Examples

Due to the complex nature of the proposed approach, it is useful to examine the general formula described above in important special situations. We now consider the two-sample problem in terms of simple linear predictor of the form $\eta = \beta_1 + X\beta_2$, where X is a Bernoulli random variable with $P(X = 1) = \pi$ and $P(X = 0) = 1 - \pi$. For canonical link $\eta = \theta$, it can be shown that the asymptotic variance approximate of $\hat{\beta}_2$ is $V = \Sigma/N$, where $\Sigma = 1/[(1 - \pi)b''(\theta_0)] + 1/[\pi \cdot b''(\theta_1)]$, $\theta_0 = \eta_0 = \beta_1$ and $\theta_1 = \eta_1 = \beta_1 + \beta_2$. For a hypothesis testing of group difference with $\mathrm{H}_0\colon \beta_2 = 0$ versus $\mathrm{H}_1\colon \beta_2 \neq 0$, the restricted estimate is $\beta_1^* = g(\mu^*)$, where $\mu^* = (1 - \pi)b'(\theta_0) + \pi \cdot b'(\theta_1)$. Thus, the proposed variance of $\hat{\beta}_2$, under the parameter restriction $\beta_2 = 0$, is $V^* = \Sigma^*/N$, where $\Sigma^* = 1/[\pi(1 - \pi)b''(\theta^*)]$, and $\theta^* = \eta^* = \beta_1^*$. In the following, we restrict our attention to the logistic and Poisson regression models.

Example 1: simple logistic regression.

For binary outcomes, the probability of response is given by $\mu = b'(\theta) = e^\theta/(1 + e^\theta)$ and the variance function $\sigma^2 = b''(\theta) = e^\theta/(1 + e^\theta)^2$. In this case, $a(\phi) = 1$ and $g(\mu) = \log\{\mu/(1 - \mu)\}$. Straightforward substitution into the general formula above gives

$$\Sigma = \frac{1}{(1 - \pi) \cdot \mu_1(1 - \mu_1)} + \frac{1}{\pi \cdot \mu_2(1 - \mu_2)}$$

and

$$\Sigma^* = \frac{1}{\pi(1 - \pi)\mu^*(1 - \mu^*)},$$

where $\mu_1 = \exp(\beta_1)/[1 + \exp(\beta_1)]$, $\mu_2 = \exp(\beta_1 + \beta_2)/[1 + \exp(\beta_1 + \beta_2)]$ and $\mu^* = (1 - \pi)\mu_1 + \pi \cdot \mu_2$.

Essentially, the test of log odds ratio or covariate coefficient $\mathrm{H}_0\colon \beta_2 = 0$ versus $\mathrm{H}_1\colon \beta_2 \neq 0$ is equivalent to the test of attributable risk or equivalence of two binomial proportions $\mathrm{H}_0\colon \mu_1 = \mu_2$ versus $\mathrm{H}_1\colon \mu_1 \neq \mu_2$. It follows from the delta method that the proposed large-sample distributions for $\hat{\beta}_2$ conform to the usual asymptotic normal approximations for the difference of two independent sample proportions, under both the respective alternative and null hypotheses. For sample size determinations, the proposed formula (6) or more exactly the simplified version (7) is in line with the asymptotic normal method of Sahai and Khurshid (1996, Eq. (20)) in testing the equality of two binomial proportions. In our notation, it gives

$$N = \left\{ Z_{\alpha/2} \left[\frac{\mu^*(1-\mu^*)}{\pi(1-\pi)} \right]^{1/2} + Z_\gamma \left[\frac{\mu_1(1-\mu_1)}{1-\pi} + \frac{\mu_2(1-\mu_2)}{\pi} \right]^{1/2} \right\}^2 \bigg/ (\mu_1 - \mu_2)^2.$$

This method is also given in Fleiss (1981), Machin and Campbell (1987), and Rosner (1994). In a similar fashion as the proposed procedure, expression (9) corresponds to the simple normal method assuming heterogeneity proposed in Sahai and Khurshid (1996, Eq. 26) as follows:

$$N = [Z_{\alpha/2} + Z_\gamma]^2 \left[\frac{\mu_1(1-\mu_1)}{1-\pi} + \frac{\mu_2(1-\mu_2)}{\pi} \right] \bigg/ (\mu_1 - \mu_2)^2.$$

From the summary of Sahai and Khurshid (1996), it appears that this procedure for unequal group numbers has never been closely examined. However, the special case of equal group numbers ($\pi=0.5$) in their Eq. (13) was also documented in Pocock (1982), Machin and Campbell (1987), and Snedecor and Cochran (1989). Note that all earlier results on sample size and power calculations for covariate $\mathbf{X}$ with discrete probability function of finite support are still applicable even though $\mathbf{X}$ is nonrandom with a finite number of configurations. The major modification occurs in the interpretation of sample allocation. For the current two-sample example, therefore, the actual sample size is being fixed as $N(1-\pi)$ and $N\cdot\pi$ for groups 1 and 2, respectively, in the setting of Sahai and Khurshid (1996). In our formulation, however $(1-\pi)$ and π represent the *expected* weights of samples in each group.

Example 2: simple Poisson regression.

The Poisson regression models outcomes that are counts, with the variance function $\sigma^2 = b''(\theta)$ is the same as the mean response $\mu = b'(\theta) = e^\theta$. Moreover, $a(\phi) = 1$ and $g(\mu) = \log(\mu)$. In this particular case,

$$\Sigma = \frac{1}{(1-\pi)\mu_1} + \frac{1}{\pi \cdot \mu_2}$$

and

$$\Sigma^* = \frac{1}{\pi(1-\pi)\mu^*},$$

where $\mu_1 = \exp(\beta_1)$, $\mu_2 = \exp(\beta_1 + \beta_2)$ and $\mu^* = (1-\pi)\mu_1 + \pi \cdot \mu_2$. With Σ and Σ^*, the proposed sample size is given by (7), whereas the direct method yields the sample size in (9). Likewise, the connection between the test of group difference in terms of $\mathrm{H}_0\colon \beta_2 = 0$ and $\mathrm{H}_0\colon \mu_1 = \mu_2$ can be carried out in a similar fashion as the simple logistic regression described above. With the simplified asymptotic normal approximations without or with the heterogeneity assumption, the resulting formulas are

$$N = \left\{ Z_{\alpha/2} \left[\frac{\mu^*}{\pi(1-\pi)} \right]^{1/2} + Z_\gamma \left[\frac{\mu_1}{1-\pi} + \frac{\mu_2}{\pi} \right]^{1/2} \right\}^2 \Big/ (\mu_1 - \mu_2)^2$$

and

$$N = [Z_{\alpha/2} + Z_\gamma]^2 \left[\frac{\mu_1}{1-\pi} + \frac{\mu_2}{\pi} \right] \Big/ (\mu_1 - \mu_2)^2,$$

respectively.

With all the sample size formulas in both Examples 1 and 2, it is obvious that the expressions depend explicitly on the two design components: parameter configurations and covariate distributions. Consequently, the adequacy of these approaches is affected by these two factors and more exactly by their interrelationship. For the simple logistic and Poisson regressions, the proposed approach and the direct method with respect to their fundamental discrepancy in conjunction with covariate distributions will be further investigated in the following simulation study.

4. Simulation studies

The finite-sample adequacy of our formula was assessed through simulations, in which we also compared the proposed method with the direct method and the likelihood ratio test-based approach of Shieh (2000). For illustrative purposes, we concentrate on the two most prominent models in the class of generalized linear models: logistic and Poisson regressions. For both regression models, two linear predictors are examined. First, we investigated the simple linear predictor described in the previous section, namely $\eta = \beta_1 + X\beta_2$, where X is a Bernoulli random variable with $P(X = 1) = \pi$ and $P(X = 0) = 1 - \pi$. The parameter of interest β_2 is taken to be log(2). The intercept parameter β_1 is chosen to satisfy the overall response $\mu^* = 0.2$ with respect to the Bernoulli distribution of X with $\pi = 0.1$, 0.3, 0.5, 0.7 and 0.9, where μ^* is defined in Examples 1 and 2 for logistic and Poisson regression models, respectively.

The second predictor is of the form $\eta = X_1\beta_1 + X_2\beta_2 + X_3\beta_3 + X_4\beta_4$ with $X_1 \equiv 1$. The joint distribution of (X_2, X_3) is assumed to be multinomial with probabilities π_1, π_2, π_3 and π_4, corresponding to (x_2, x_3) values of $(0,0)$, $(0,1)$, $(1,0)$ and $(1,1)$, respectively. Three sets of $(\pi_1, \pi_2, \pi_3, \pi_4)$ are studied to represent different distributional shapes, namely (0.76, 0.19, 0.01, 0.04), (0.4, 0.1, 0.1, 0.4), and (0.04, 0.01, 0.19, 0.76). The covariate X_4 has a standard normal distribution and is independent of (X_2, X_3). The parameters $(\beta_2, \beta_3, \beta_4)$ are set as (log(1.5), log(2), 0.1). The intercept parameter β_1 is chosen to satisfy the overall response $\bar{\mu} = 0.1$, where $\bar{\mu} = E_{\mathbf{X}}[\exp(\eta)/\{1 + \exp(\eta)\}]$ and $\bar{\mu} = E_{\mathbf{X}}\{\exp(\eta)\}$ for the logistic and Poisson regression models, respectively, and $E_{\mathbf{X}}[\,\cdot\,]$ denotes the expectation taken with respect to the joint distribution of (X_2, X_3, X_4). This model mimics a design with two dichotomous main effects and a continuous confounder. We are interested in the tests of H_0: $\beta_2 = \beta_3 = 0$ for treatment effects $(p = 2)$ and H_0: $\beta_2 = \beta_3 = \beta_4 = 0$ for overall effects $(p = 3)$.

For a given model, covariate distribution, parameter values, and overall response level, the estimates of sample sizes required for testing the specified hypothesis with significance level 0.05 and power (0.90, 0.95) are calculated. The resulting sample sizes correspond to the direct method (8), the proposed approach (5), and Shieh (2000) are denoted by N_{D}, N_{P}, and N_{S}, respectively. These estimates of sample size allow comparison of relative efficiencies of the approaches. However, the magnitude of the sample size affects the accuracy of the asymptotic distribution and the resulting formula, a fair comparison among these approaches must adjust for this factor. Hence, we unify the sample sizes in the simulations by choosing the sample size N_{D} as the benchmark to recalculate the nominal powers for all competing approaches.

Estimates of the true power associated with given sample size and model configuration are then computed through Monte Carlo simulation of 10,000 independent data sets. For each replicate, N_{D} covariate values are generated from the selected distribution. These covariate values determine the incidence rates for generating N_{D} Bernoulli or Poisson outcomes. Then the test statistic is computed and the estimated power is the proportion of the 10,000 replicates whose test statistic values exceed the critical value $\chi^2_{p,\alpha}$. Note that both the Wald statistic and the likelihood ratio test have the identical asymptotic chi-square distribution under the null hypothesis. The adequacy of the sample size formula is determined by the difference between the estimated power and nominal power specified above. All calculations are performed using programs written with SAS/IML (SAS Institute, 1989).

The results of the simulation studies are presented in Tables 1–4. Tables 1 and 2 contain results for the simple linear predictor, while Tables 3 and 4 contain results for the multiple linear predictor. The adjusted significance levels described in (4) for the proposed method are presented in the footnote of the tables in accordance to the sequence of listed models. For a concise visualization of the results, the errors associated with larger sample size (which gives power very close to 0.95 for the direct method) are plotted with different covariate distributions for the three procedures in Figs. 1–4 corresponding to the four models in Tables 1–4, respectively.

In general, the errors are larger for power 0.90 than power 0.95 for all competing methods. Furthermore, the skewed covariate distribution appears to degrade the accuracy of sample size calculations. The results suggest that the direct method is extremely vulnerable to the unbalanced allocation of samples and gives mostly the largest errors among the three formulas for both logistic and Poisson regression models. However, the proposed method and the approach of Shieh (2000) generally maintain a close agreement between the estimated power and nominal power. The only exceptions are with the extremely unbalanced Bernoulli covariate of $\pi = 0.1$ for the simple regressions and with the exceedingly skewed multinomial covariate distribution $(0.72, 0.18, 0.02, 0.08)$ for the multiple regressions. Obviously, in these cases, there is essentially limited or inadequate information for discriminating the alternative hypothesis against the null hypothesis because these designs are composed of a comparatively small treatment group and a large control group. For the rest of model configurations, the proposed method performs well and is comparable to Shieh's (2000) approach.

Table 1
Calculated sample sizes and estimates of actual power at specified sample size for simple logistic regression with Bernoulli covariate

	The direct method		The proposed method[a]		Shieh (2000)	
Power	0.90	0.95	0.90	0.95	0.90	0.95
Bernoulli (0.1)						
Sample size[b] (N_D, N_P, N_S)	1173	1451	1377	1677	1261	1559
Nominal power[c] at N_D	0.9001	0.9501	0.8441	0.9155	0.8785	0.9355
Estimated power	0.8762	0.9294	0.8762	0.9294	0.8635	0.9214
Error	−0.0239	−0.0207	0.0321	0.0139	−0.0150	−0.0141
Bernoulli (0.3)						
Sample size[b] (N_D, N_P, N_S)	587	726	626	769	602	744
Nominal power[c] at N_D	0.9001	0.9501	0.8806	0.9384	0.8931	0.9455
Estimated power	0.8924	0.9431	0.8924	0.9431	0.8887	0.9403
Error	−0.0077	−0.0070	0.0118	0.0047	−0.0044	−0.0052
Bernoulli (0.5)						
Sample size[b] (N_D, N_P, N_S)	583	720	561	696	565	698
Nominal power[c] at N_D	0.9004	0.9500	0.9106	0.9560	0.9092	0.9557
Estimated power	0.9097	0.9560	0.9097	0.9560	0.9106	0.9572
Error	0.0093	0.0060	−0.0009	0.0000	0.0014	0.0015
Bernoulli (0.7)						
Sample size[b] (N_D, N_P, N_S)	822	1016	716	899	751	928
Nominal power[c] at N_D	0.9003	0.9501	0.9330	0.9686	0.9241	0.9650
Estimated power	0.9274	0.9673	0.9274	0.9673	0.9323	0.9699
Error	0.0271	0.0172	−0.0056	−0.0013	0.0082	0.0049
Bernoulli (0.9)						
Sample size[b] (N_D, N_P, N_S)	2267	2803	1797	2278	1953	2415
Nominal power[c] at N_D	0.9001	0.9500	0.9492	0.9773	0.9374	0.9728
Estimated power	0.9442	0.9777	0.9442	0.9777	0.9505	0.9802
Error	0.0441	0.0277	−0.0050	0.0004	0.0131	0.0074

[a]The adjusted significance levels are 0.0257, 0.0390, 0.0575, 0.0810 and 0.1086.
[b]Sample sizes needed to achieve power 0.9 and 0.95, respectively.
[c]Nominal powers at calculated sample sizes of the direct method in[b].

5. Concluding remarks

In this article we have extended the sample size and power methodology for Wald statistics in generalized linear models to handle test of hypothesis with any number of parameters. Unlike other approaches, the proposed method is applicable under general conditions that there are no particular limitations in the overall response rate and covariate distribution. According to the general formulation, the proposed approach naturally encompasses both fixed and random covariate configurations.

The notion of adjusted significance level is introduced to improve the determinations of power and sample size by taking into account the discrepancy between the

Table 2
Calculated sample sizes and estimates of actual power at specified sample size for simple Poisson regression with Bernoulli covariate

	The direct method		The proposed method[a]		Shieh (2000)	
Power	0.90	0.95	0.90	0.95	0.90	0.95
Bernoulli (0.1)						
Sample size[b] (N_D, N_P, N_S)	736	910	1011	1214	856	1058
Nominal power[c] at N_D	0.9004	0.9502	0.7654	0.8614	0.8525	0.9168
Estimated power	0.8449	0.9051	0.8449	0.9051	0.8169	0.8890
Error	−0.0555	−0.0451	0.0795	0.0437	−0.0356	−0.0278
Bernoulli (0.3)						
Sample size[b] (N_D, N_P, N_S)	440	545	488	598	457	565
Nominal power[c] at N_D	0.9000	0.9503	0.8664	0.9300	0.8893	0.9432
Estimated power	0.8797	0.9378	0.8797	0.9378	0.8728	0.9322
Error	−0.0203	−0.0125	0.0133	0.0078	−0.0165	−0.0110
Bernoulli (0.5)						
Sample size[b] (N_D, N_P, N_S)	493	609	459	572	464	574
Nominal power[c] at N_D	0.9005	0.9501	0.9187	0.9606	0.9165	0.9603
Estimated power	0.9124	0.9611	0.9124	0.9611	0.9153	0.9623
Error	0.0119	0.0110	−0.0063	0.0005	−0.0012	0.0020
Bernoulli (0.7)						
Sample size[b] (N_D, N_P, N_S)	753	931	608	769	654	809
Nominal power[c] at N_D	0.9002	0.9501	0.9465	0.9759	0.9357	0.9719
Estimated power	0.9380	0.9752	0.9380	0.9752	0.9455	0.9781
Error	0.0378	0.0251	−0.0085	−0.0007	0.0098	0.0062
Bernoulli (0.9)						
Sample size[b] (N_D, N_P, N_S)	2194	2713	1568	2011	1774	2194
Nominal power[c] at N_D	0.9001	0.9501	0.9627	0.9841	0.9501	0.9798
Estimated power	0.9540	0.9840	0.9540	0.9840	0.9629	0.9874
Error	0.0539	0.0339	−0.0087	−0.0001	0.0128	0.0076

[a]The adjusted significance levels are 0.0117, 0.0330, 0.0646, 0.1030 and 0.1446.
[b]Sample sizes needed to achieve power 0.9 and 0.95, respectively.
[c]Nominal powers at calculated sample sizes of the direct method in[b].

asymptotic distribution approximations of maximum likelihood estimators in both mean and variance–covariance under the alternative and null hypotheses. Using heuristic arguments and computer simulations, it is shown that the proposed approach performed well over most of the range of conditions we considered here. The exceptions occurred only when the model especially consists of an asymmetric allocation scheme that is disproportionate for the purpose of addressing specific scientific hypotheses and confirming credible treatment effects. Nevertheless, neither the naive Wald test-based direct method nor the likelihood ratio test-based method of Shieh (2000) is immune to unbalanced designs. It appears that the direct method tends to provide inappropriate

Table 3
Calculated sample sizes and estimates of actual power at specified sample size for multiple logistic regression with multinomial and standard normal covariates

	The direct method		The proposed method[a]		Shieh (2000)	
Power	0.90	0.95	0.90	0.95	0.90	0.95
Test 1: treatment effects ($p=2$)						
1. Multinomial (0.72, 0.18, 0.02, 0.08)						
Sample size[b] (N_D, N_P, N_S)	913	1115	1107	1326	1014	1237
Nominal power[c] at N_D	0.9000	0.9501	0.8241	0.9026	0.8656	0.9266
Estimated power	0.8691	0.9256	0.8691	0.9256	0.8447	0.9096
Error	−0.0309	−0.0245	0.0450	0.0230	−0.0209	−0.0170
2. Multinomial (0.40, 0.10, 0.10, 0.40)						
Sample size[b] (N_D, N_P, N_S)	676	824	619	763	622	759
Nominal power[c] at N_D	0.9004	0.9500	0.9237	0.9634	0.9236	0.9646
Estimated power	0.9240	0.9688	0.9240	0.9688	0.9306	0.9720
Error	0.0236	0.0188	0.0003	0.0054	0.0070	0.0074
3. Multinomial (0.08, 0.02, 0.18, 0.72)						
Sample size[b] (N_D, N_P, N_S)	2323	2835	1653	2095	1899	2318
Nominal power[c] at N_D	0.9000	0.9500	0.9655	0.9854	0.9505	0.9799
Estimated power	0.9509	0.9834	0.9509	0.9834	0.9671	0.9898
Error	0.0509	0.0334	−0.0146	−0.0020	0.0166	0.0099
Test 2: overall effects ($p=3$)						
1. Multinomial (0.72, 0.18, 0.02, 0.08)						
Sample size[b] (N_D, N_P, N_S)	966	1170	1094	1309	1061	1286
Nominal power[c] at N_D	0.9003	0.9501	0.8534	0.9214	0.8680	0.9282
Estimated power	0.8753	0.9296	0.8753	0.9296	0.8544	0.9149
Error	−0.0250	−0.0205	0.0219	0.0082	−0.0136	−0.0133
2. Multinomial (0.40, 0.10, 0.10, 0.40)						
Sample size[b] (N_D, N_P, N_S)	725	878	674	822	668	810
Nominal power[c] at N_D	0.9002	0.9500	0.9210	0.9620	0.9240	0.9649
Estimated power	0.9179	0.9612	0.9179	0.9612	0.9260	0.9646
Error	0.0177	0.0112	−0.0031	−0.0008	0.0020	−0.0003
3. Multinomial (0.08, 0.02, 0.18, 0.72)						
Sample size[b] (N_D, N_P, N_S)	2241	2715	1685	2106	1877	2274
Nominal power[c] at N_D	0.9000	0.9500	0.9604	0.9828	0.9470	0.9781
Estimated power	0.9395	0.9770	0.9395	0.9770	0.9597	0.9847
Error	0.0395	0.0270	−0.0209	−0.0058	0.0127	0.0066

[a]The adjusted significance levels are 0.0206, 0.0700, 0.1547, 0.0280, 0.0673 and 0.1372.
[b]Sample sizes needed to achieve power 0.9 and 0.95, respectively.
[c]Nominal powers at calculated sample sizes of the direct method in[b].

sample sizes and is not recommended. In fact, the findings in the sensitivity to the distribution of the covariates are consistent with those of Self and Mauritsen (1988), Self et al. (1992) and Shieh (2001) which are obtained from the score, likelihood

Table 4
Calculated sample sizes and estimates of actual power at specified sample size for multiple Poisson regression with multinomial and standard normal covariates

	The direct method		The proposed method[a]		Shieh (2000)	
Power	0.90	0.95	0.90	0.95	0.90	0.95
Test 1: treatment effects ($p=2$)						
1. Multinomial (0.72, 0.18, 0.02, 0.08)						
Sample size[b] (N_D, N_P, N_S)	709	865	900	1074	826	1008
Nominal power[c] at N_D	0.9001	0.9500	0.8003	0.8862	0.8480	0.9136
Estimated power	0.8507	0.9083	0.8507	0.9083	0.8121	0.8866
Error	−0.0494	−0.0417	0.0504	0.0221	−0.0359	−0.0270
2. Multinomial (0.40, 0.10, 0.10, 0.40)						
Sample size[b] (N_D, N_P, N_S)	637	777	567	700	573	699
Nominal power[c] at N_D	0.9003	0.9501	0.9304	0.9672	0.9294	0.9682
Estimated power	0.9238	0.9666	0.9238	0.9666	0.9307	0.9699
Error	0.0235	0.0165	−0.0066	−0.0006	0.0013	0.0017
3. Multinomial (0.08, 0.02, 0.18, 0.72)						
Sample size[b] (N_D, N_P, N_S)	2288	2792	1529	1953	1818	2219
Nominal power[c] at N_D	0.9001	0.9500	0.9717	0.9883	0.9558	0.9827
Estimated power	0.9537	0.9843	0.9537	0.9843	0.9723	0.9893
Error	0.0536	0.0343	−0.0180	−0.0040	0.0165	0.0066
Test 2: overall effects ($p=3$)						
1. Multinomial (0.72, 0.18, 0.02, 0.08)						
Sample size[b] (N_D, N_P, N_S)	752	911	882	1052	869	1052
Nominal power[c] at N_D	0.9001	0.9501	0.8370	0.9109	0.8492	0.9146
Estimated power	0.8566	0.9083	0.8566	0.9083	0.8205	0.8841
Error	−0.0435	−0.0418	0.0196	−0.0026	−0.0287	−0.0305
2. Multinomial (0.40, 0.10, 0.10, 0.40)						
Sample size[b] (N_D, N_P, N_S)	679	823	620	758	614	744
Nominal power[c] at N_D	0.9002	0.9502	0.9254	0.9645	0.9291	0.9681
Estimated power	0.9193	0.9672	0.9193	0.9672	0.9286	0.9711
Error	0.0191	0.0170	−0.0061	0.0027	−0.0005	0.0030
3. Multinomial (0.08, 0.02, 0.18, 0.72)						
Sample size[b] (N_D, N_P, N_S)	2170	2629	1544	1943	1781	2157
Nominal power[c] at N_D	0.9001	0.9500	0.9669	0.9860	0.9512	0.9803
Estimated power	0.9467	0.9821	0.9467	0.9821	0.9653	0.9885
Error	0.0466	0.0321	−0.0202	−0.0039	0.0141	0.0082

[a]The adjusted significance levels are 0.0161, 0.0780, 0.1813, 0.0234, 0.0721, and 0.1598.
[b]Sample sizes needed to achieve power 0.9 and 0.95, respectively.
[c]Nominal powers at calculated sample sizes of the direct method in[b].

ratio and Wald tests, respectively. Conceivably, the allocation schemes or the covariate distributions play an important role in the accuracy of the existing power and sample size methodology within the framework generalized linear models.

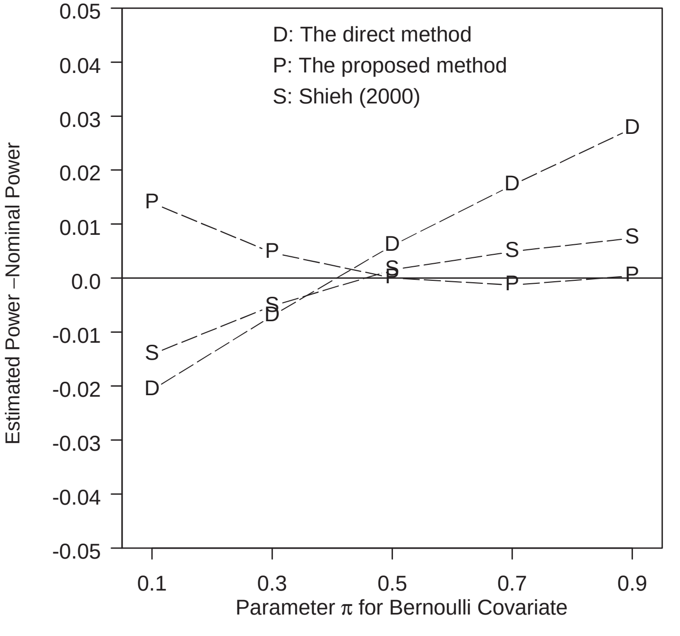

Fig. 1. The estimated errors for simple logistic regression.

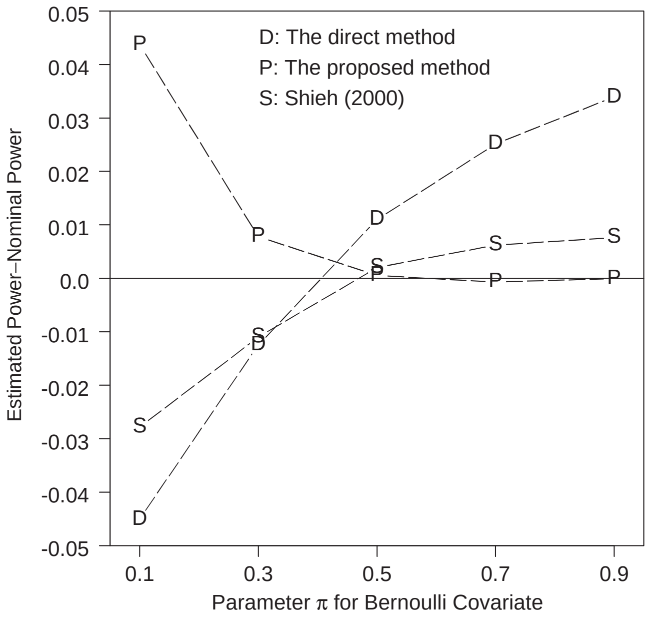

Fig. 2. The estimated errors for simple Poisson regression.

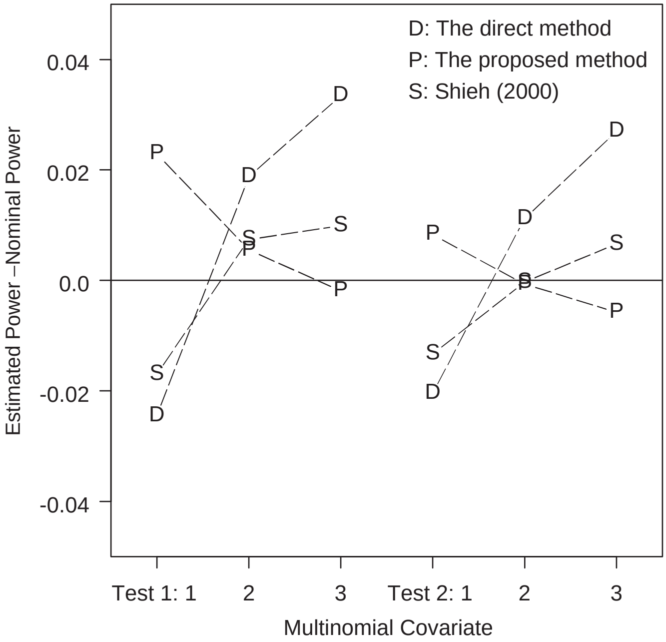

Fig. 3. The estimated errors for multiple logistic regression.

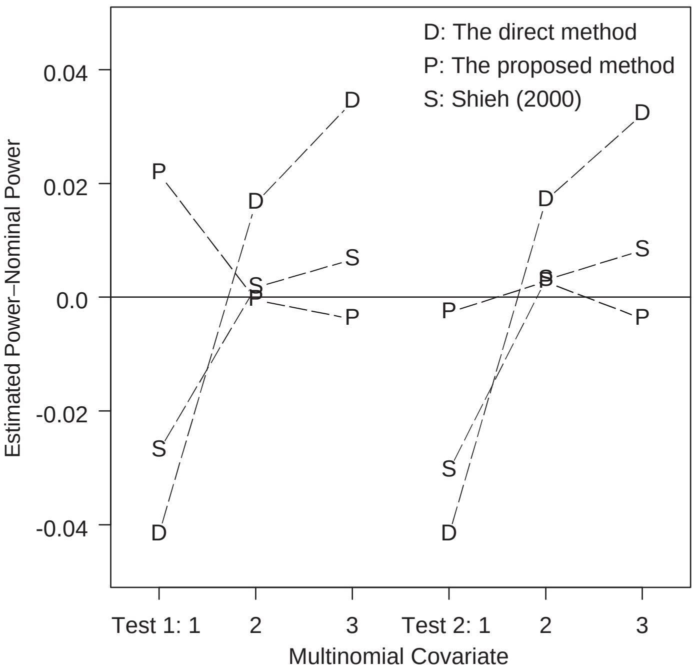

Fig. 4. The estimated errors for multiple Poisson regression.

Finally, we note that the generalized estimating equations approach was introduced by Liang and Zeger (1986) as a way of handling correlated data that, except for the correlation between responses, can be modeled with generalized linear models. The proposed approach could be extended for use in the context of generalized estimating equations. We are currently developing this extension and the results will be reported elsewhere.

Acknowledgements

This work was partially supported by the National Science Council. The author is grateful to the referee for helpful comments that have improved the exposition.

References

Biostat, Inc., 2000. Power and Precision, Version 2, Englewood,NJ.
Fleiss, J.L., 1981. Statistical Methods for Rates and Proportions, 2nd Edition. Wiley, New York.
Hauck, W.H., Donner, A., 1977. Wald's test as applied to hypotheses in logit analysis. J. Amer. Statist. Assoc. 72, 851–853.
Johnson, N.L., Kotz, S., Balakrishnan, N., 1994. Continuous Univariate Distributions, 2nd Edition. Wiley, New York.
Liang, K.-Y., Zeger, S.L., 1986. Longitudinal data analysis using generalized linear models. Biometrika 73, 13–22.
Machin, D., Campbell, M.J., 1987. Statistical Tables for the Design of Clinical Trials. Blackwell, Oxford.
McCullagh, P., Nelder, J.A., 1989. Generalized Linear Models, 2nd Edition. Chapman & Hall, London.
Nelder, J.A., Wedderburn, R.W.M., 1972. Generalized linear models. J. Roy. Statist. Soc. Ser. A 135, 370–384.
Pocock, S.J., 1982. Statistical aspects of clinical trials design. The Statistician 31, 1–18.
Rosner, B., 1994. Fundamentals of Biostatistics, 4th Edition. Duxbury, Belmont, CA.
Sahai, H., Khurshid, A., 1996. Formulae and tables for the determination of sample sizes and power in clinical trials for testing differences in proportions for the two-sample design: a review. Statist. Med. 15, 1–21.
SAS Institute, 1989. SAS/IML Software: Usage and Reference, Version 6. SAS Institute, Carey, NC.
Self, S.G., Mauritsen, R.H., 1988. Power/sample size calculations for generalized linear models. Biometrics 44, 79–86.
Self, S.G., Mauritsen, R.H., Ohara, J., 1992. Power calculations for likelihood ratio tests in generalized linear models. Biometrics 48, 31–39.
Shieh, G., 2000. On power and sample size calculations for likelihood ratio tests in generalized linear models. Biometrics 56, 1192–1196.
Shieh, G., 2001. Sample size calculations for logistic and Poisson regression models. Biometrika 88, 1193–1199.
Signorini, D.F., 1991. Sample size for Poisson regression. Biometrika 78, 446–450.
Snedecor, G.W., Cochran, W.G., 1989. Statistical Methods, 8th Edition. Iowa State University Press, Ames, IA.
Whittemore, A.S., 1981. Sample size for logistic regression with small response probability. J. Amer. Statist. Assoc. 76, 27–32.
Wood, T.A., 1989. An F approximation to the distribution of a linear combination of Chi-squared variables. Comm. Statist. Simulation Comput. 18, 1439–1456.