RESEARCH ARTICLE

COMPARISON OF POWER VALUES IN GENERALIZED LINEAR MIXED MODEL (GLMM) UNDER THE DIFFERENT ESTIMATION METHODS

[1]Md. Abu Manju* and [2]Soma Chowdhury Biswas

[1]Assistant Director, Statistics Department, Bangladesh Bank, Head Office, Dhaka, Bangladesh,

[2]Professor, Department of Statistics, University of Chittagong, Bangladesh

ABSTRACT

In longitudinal studies, the Generalized Estimating Equation (GEE) and Maximum Likelihood (ML) methods are commonly used to estimate the parameters of Generalized Linear Mixed Models (GLMM). Power and sample size estimation constitutes an important component for longitudinal study design and planning of modern clinical trials. In this paper, the power of the Wald Test for two different estimation methods of parameters in GLMM was compared. The data employed in this study were diabetes mellitus data with three consecutive follow ups. A pattern was found in which the power values were better under ML approach than GEE approach for the GLMM.

INTRODUCTION

Generalized linear mixed-effects models, more commonly known as generalized linear mixed models (GLMM), are widely commonly frequently used in longitudinal data analysis. The GLMM are natural combinations of two modeling strands, the linear mixed models and the generalized linear models. Linear mixed models (e.g., Hartville, [5]; Laird and Ware, [8] are linear regression models that include normally distributed random effects in addition to fixed effects. A natural application is to longitudinal data where the random effects vary between subjects and induce within-subject dependence among repeated measurements after conditioning on observed covariates. Generalized linear models Nelder and Wedderburn, [11]; Wedderburn, R. W. M. [20] unify regression models for different response types such as linear models for continuous responses, logistic models for binary responses, and log-linear models for counts. GLMMs are the generalized linear models that include multivariate normal random effects in the linear predictor. Early contributions explicitly discussing this idea are Wong and Mason [19] and West [22] in frequentist and Bayesian settings, respectively. The term "generalized linear mixed model" appears to have been coined by Gilmour, Anderson, and Rae [21], although it's widespread use is probably due to the highly cited paper by Breslow and Clayton [2]. In social statistics and other areas, generalized linear mixed models are also known as hierarchical, multilevel, or random-coefficient models.

The simplest cases of GLMMs are the random-intercept models with a single normally distributed random effect. Random-intercept models are sometimes referred to as variance components, error components, or random-effects models. Early applications of such models include the random-intercept probit models discussed by Lawley [9], in psychometrics. Similar models were later considered by Heckman and Willis [27] in econometrics. Pierce and Sands [13] and Anderson and Aitkin [1] discussed random-intercept logit models, and Hinde [6] and Brillinger and Preisler [28] discussed random-intercept Poisson models.

Introducing random effects with conjugate distributions into special cases of generalized linear models has a very long history, including the negative binomial model for counts Greenwood and Yule, [4] and the beta-binomial model for proportions Skellam, [18]. Although these models, which are strictly not GLMMs, a major limitation is that they cannot be extended to random-coefficient models where the effects of covariates vary randomly between subjects. Zeger, Liang, and Albert [23] discussed a generalized estimating equation (GEE) approach to fit the GLMM and both classes of models (Subject-Specific & Population Average) for discrete and continuous outcomes. The subject-specific parameters are assumed to follow a Gaussian distribution, simple relationships between the Population Averaged (PA) and Subject Specific (SS) parameters are available. Actually the SS model is the simplest example of GLMM. Sometimes we have to make a definite decision one way or another about a particular hypothesis in the GLMM; in this situation a test of hypothesis is appropriate. Although in science we never accept

*Corresponding author: tba_man@yahoo.com
soma_chow@hotmail.com

a hypothesis outright, but rather continually modify our ideas and laws as new knowledge is obtained, in some situations, such as in clinical practice, we cannot afford this luxury. To understand the concepts of validity and power, it will be helpful if we consider the case in which a decision must be made, one way or the other, with the result that some wrong decisions will inevitably be made. Clearly, we wish to act in such a way that the probability of making a wrong decision is minimized. Repeated measurements arising from longitudinal studies occur frequently in applied research. Methods to calculate power in the context of repeated measures are available for experimental settings where the covariate of interest is a discrete treatment indicator. However, no closed form expression exists to calculate power for generalized linear models with non-zero within-cluster correlation that are common in epidemiological and observational studies in which the covariate of interest varies over time and is often measured on a continuous scale, and where the researchers control for several potential confounders. Gastan, McLaren & Delfinob [29] discussed a procedure to calculate power of the test for the normal linear mixed model, and the logistic regression model, both with repeated measurements and non-zero within-cluster correlation. Correlated binary data are common in biomedical studies. Such data can be analyzed using Liang and Zeger's generalized estimating equations (GEE) approach. An attractive point of the GEE approach is that one can use a misspecified working correlation matrix, such as the working independence model (i.e., the identity matrix), and draw (asymptotically) valid statistical inference by using the so-called robust or sandwich variance estimator. Wei Pan [24] derived some explicit formulas for sample size and power calculations under various common situations. Gwowen Shieh [15] described A direct extension of the approach described in Self, Mauritsen, and Ohara [17] for power and sample size calculations in generalized linear models is presented. The major feature of the proposed approach is that the modification accommodates both a finite and an infinite number of covariate configurations. Gwowen Shieh [30] wrote another paper in which A method is proposed for improving sample size calculations for logistic and Poisson regression models by incorporating the limiting value of the maximum likelihood estimates of nuisance parameters under the composite null hypothesis. The method modifies existing approaches of Whitteinore [25] & Signorini [31] provides explicit formulae for determining the sample size needed to test hypotheses about a single parameter at a specified significance level and power. Özdemir and Eyduran [12] described a study comparatively analyzed in point of power of test of Chi-Square and Likelihood Ratio Chi-Square statistics. Steven G. Self & Robert H. Mauritsen [16] proposed an approach for estimating power/sample size within the framework of generalized linear models. This approach is based on an asymptotic approximation to the power of the score test under contiguous alternatives and is applicable to tests of composite null hypotheses. The power properties of traditional repeated measures and hierarchical linear models have not been clearly determined in the balanced design for longitudinal studies. Hua Fang, Gordon P. Brooks & Maria L. Rizzo [7] proposed a procedure A Monte Carlo power analysis of traditional repeated measures and hierarchical multivariate linear models under three variance-covariance structures. In methodology we have discussed the estimation procedure of GLMM by ML & GEE approaches and the procedure of

conducting the power function for Wald Chisquare test. Actually we have tried to show whether there exist any differences in power values of the same test due to different methods of estimation. In results section we have illustrated our methodology for subject-specific logistic regression model.

METHODOLOGY

Let y_{it} denote the tth observation in the ith cluster, $t = 1, \cdots, T_i$. Let x_{it} denote a column vector of values of explanatory variables, for fixed effect model parameters β. Let u_i denote the vector of random effect values for cluster i. This is common to all observations in the cluster. Let z_{it} denote a column vector of their explanatory variables. Often, the random effect is univariate. Conditional on u_i , a GLMM resembles an ordinary GLM. Let $u_{it} = E(Y_{it}|u_i)$. The linear predictor for a GLMM has the form

$$g(\mu_{it}) = x'_{it}\beta + z'_{it}u_i \qquad (1.0)$$

for link function $g(\cdot)$. The random effect vector u_i is assumed to have a multivariate normal distribution $N(0, \sum)$. The covariance matrix $\sum$ depends on unknown variance components and possibly also correlation parameters. Denote $\text{var}(Y_{it}|u_i) = \phi v(\mu_{it})$, where the variance function $v(\cdot)$ describes how the (conditional) variance depends on the mean and ϕ is an unknown dispersion parameter. The variability among u_i induces a nonnegative association among the responses, for the marginal distribution averaged over the subjects. This is caused by the shared random effect u_i for each observation in a cluster. In (1.0), the random effect enters the model on the same scale as the predictor terms. This is convenient but also natural for many applications. For instance, random effects sometimes represent heterogeneity caused by omitting certain explanatory variables.

Estimation of GLMM by MLE

Maximum Likelihood Method: Model fitting is rather complex for GLMMs. The main difficulty is that the likelihood function does not have a closed form. Numerical methods for approximating it can be computationally intensive for models with multivariate random effects. In this section we outline the basic ideas of ML fitting of GLMMs. The GLMM is a two-stage model. At the first stage, conditional on the random effects, observations are assumed to follow a GLM. That is, observation y_{it} in cluster i has distribution in the exponential family with expected value u_{it} linked to a linear predictor,

$$g(\mu_{it}) = x'_{it}\beta + z'_{it}u_i$$

Then, $z'_{it}u_i$ is a known offset and observations in a cluster are independent. At the second stage, the random effects $\{u_i\}$ are

assumed independent from a $N(0,\Sigma)$ distribution. For example the simplest example of the GLMM is the Subject-Specific logistic regression model or logistic regression model with the random intercept and the form of the model

$$\log it[P(Y_{it}=1|u_i)]=x'_{it}\beta+u_i ....(1.1)$$

where $u_i \sim iid \ N(0,\sigma^2)$

For a discrete variable, denote the vector of all the observations by **y** and the vector of all the random effects by **u**. Let $f(y|u;\beta)$ denote the conditional mass function of **y**, given **u**. Let $f(u;\Sigma)$ denote the normal density function for **u**. The likelihood function $l(\beta,\Sigma;y)$ for a GLMM is the probability mass function $f(y;\beta,\Sigma)$ of **y**, viewed as a function of β and Σ. This mass function refers to the marginal distribution of **y** after integrating out the random effects,

$$l(\beta,\Sigma;y)=f(y;\beta,\Sigma)=\int f(y|u;\beta)f(u;\Sigma)du \quad …(1.2)$$

It is often called a marginal likelihood. For example, the likelihood functions $l(\beta,\sigma^2;y)$ for the logistic-normal model (1.1)

$$\prod_i (\int_{-\infty}^{\infty} \prod_t \left[\frac{\exp(x'_{it}\beta+u_i)}{1+\exp(x'_{it}\beta+u_i)}\right]^{y_{it}} *$$

$$\left[\frac{1}{1+\exp(x'_{it}\beta+u_i)}\right]^{1-y_{it}} f(u_i;\sigma^2)du_i)$$

The likelihood function is evaluated numerically and maximized as a function of β and Σ. Many methods have been developed to do this. We next discuss one of the popular methods.

Gauss–Hermite Quadrature Method

The integral determining the likelihood function has dimension that depends on the random effects structure. Gauss Hermite quadrature is a method for approximating the integral of a function $f(\cdot)$ multiplied by another function having the shape of a normal density. The approximation is a finite weighted sum that evaluates the function at certain points. In the univariate normal random effects case, the approximation has the form

$$\int_{-\infty}^{\infty} f(u)\exp(-u^2)du \approx \sum_{k=1}^{q} c_k f(s_k)$$

with weights $\{c_k\}$ and quadrature points $\{s_k\}$ that are tabulated. The approximation improves as q, the number of quadrature points, increases. The approximated likelihood can be maximized with standard algorithms such as Newton Raphson, yielding ML estimates $\hat{\beta}$ and $\hat{\Sigma}$. Inverting an approximation for the observed information matrix provides standard errors for the ML estimates. For complex models, second partial derivatives for the Hessian may be computed

numerically rather than analytically. Adequate approximation usually requires larger q for standard errors than for $\hat{\beta}$.

Estimation of GLMM by GEE

GEE Approach: Zeger, Liang and Albert (1988) discussed a procedure to estimate GLMM by Generalized estimating Equation (GEE) approach. To use the GEE approach for the GLMM, we calculate the marginal moments, μ_i and V_i, from the conditional moments and the random effects distribution, F. We then solve the equation

$$U(\beta^*) = \sum_{i=1}^{K} \frac{\partial \mu'_i}{\partial \beta^*} V_i^{-1}(\alpha)(Y_i-\mu_i)=0$$

................(1.3)

discussed by Liang and Zeger (1986). Given the conditional moments in

$$g(\mu_{it})=x'_{it}\beta+z'_{it}u_i$$

and a distribution, F, for the random effects, the marginal expectation, μ_i has the form of

$$\mu_{it}=E(y_{it})=E[E(y_{it}|u_i)]=$$
$$\int h^{-1}(x'_{it}\beta+z'_{it}u_i)dF(u_i)$$

The marginal covariance matrix is

$$V_i = cov[E(Y_i|u_i)]+E[cov(Y_i|u_i)]$$

with s, t element

$$[V_i]_{s,t} = \int (u_{it}-\mu_{it})(u_{is}-\mu_{is})dF(u_i)+$$
$$\phi I(s=t)\int g(u_{it})dF(u_i)$$

where $u_{it}=h^{-1}(x'_{it}\beta+z'_{it}u_i)$ and I(s=t) is the indicator function with value 1 if s = t and 0 otherwise. Having evaluated μ_i and V_i for each subject/cluster, we solve the GEE given in (1.3) for β. Note that the GEE is a function of F, which is assumed to be known at a given iteration. For example If we consider the Subject-Specific logistic regression model/Logistic normal regression model and assuming $u_i \sim iid \ N(0,\sigma^2)$, the expression for the marginal mean simplifies by Zeger et al. (1988) and given by

$$\mu_{it} = E(Y_{it}) \approx \left[\exp(cx'_{it}\beta) \Big/ [1+\exp(cx'_{it}\beta)]\right]$$

where $c=\left[1+0.6\sigma^2\right]^{-1/2}$

Expanding the link function in a Taylor series about $u_i=0$ gives the approximation

$$cov(Y_i) \approx cov\left[h^{-1}(x'_{it}\beta)+\frac{\partial h^{-1}}{\partial u_i}(x'_{it}\beta)u_i\right]+$$

$$\phi E\left\{g\left[h^{-1}(x'_{it}\beta)+\frac{\partial h^{-1}}{\partial u_i}(x'_{it}\beta)u_i\right]\right\}$$

$$\approx L_i Z_i \sum Z'_i L_i + \phi A_i = \hat{V}_i$$

Where

$$L_i = \text{diag}\left\{ \frac{\partial h^{-1}(u)}{\partial u}, u = x'_{it}\beta, t = 1,...,n_i \right\}$$

Using $\widetilde{V}_i$, as an approximation to V_i, we have that conditional on F, $\sqrt{K}(\beta - \hat{\beta})$ is asymptotically multivariate Gaussian with mean 0 (if an exact expression for μ_{it} is used and small bias otherwise) and with variance that can be consistently estimated by

$$\hat{V}_{\hat{\beta}} = \left(\sum_{i=1}^{K} \frac{\partial \hat{\mu}'_i}{\partial \beta} \hat{\widetilde{V}}_i^{-1} \frac{\partial \hat{\mu}_i}{\partial \beta} \right)^{-1} *$$

$$\left(\sum_{i=1}^{K} \frac{\partial \hat{\mu}'_i}{\partial \beta} \hat{\widetilde{V}}_i^{-1} (Y_i - \hat{\mu}_i)(Y_i - \hat{\mu}_i)' \hat{\widetilde{V}}_i^{-1} \frac{\partial \hat{\mu}_i}{\partial \beta} \right) *$$

$$\left(\sum_{i=1}^{K} \frac{\partial \hat{\mu}'_i}{\partial \beta} \hat{\widetilde{V}}_i^{-1} \frac{\partial \hat{\mu}_i}{\partial \beta} \right)^{-1}$$

The approximate estimates of variance covariance matrix of the random effects is

$$\hat{\Sigma} \approx (Z'_i Z_i)^{-1} Z'_i L_i (V_i - \phi A_i) L_i^{-1} Z_i (Z'_i Z_i)^{-1}$$

and the estimates of the scale parameter is

$$\hat{\phi} = \sum_{i=1}^{K} \sum_{t=1}^{n_i} \frac{(y_{it} - \hat{\mu}_{it})^2 - (\hat{L}_i)^2 z'_{it} \hat{D} z_{it}}{g(\hat{\mu}_{it})}$$

Hypothesis Tests

After estimating t

he vector of regression coefficients $\hat{\beta}$, it may be of interest to test hypotheses concerning the elements of β. Consider hypotheses of the form $H_0 : C\beta = d$

where C is a $c \times p$ matrix of constants imposing c linearly independent constraints on the elements of β and d is a $p \times 1$ vector of constants. There exists a established property that the ML estimates are approximately normally distributed with respective mean and variance and Liang and Zeger (1988) shown that the GEE estimates also normally distributed with respective mean and variance. Again since $\hat{\beta}$ is asymptotically normal, so the Wald test statistic

$$Q_C = (C\hat{\beta} - d)' \left[C \hat{Var}(\hat{\beta}) C' \right]^{-1} (C\hat{\beta} - d)$$

has an asymptotic χ_c^2 distribution if H_0 is true.

Power of the Wald Test

Now we are interested to test the null hypothesis $H_0 : C\beta = d$ against alternative hypothesis $H_1 : C\beta = b \neq d$, where C is a $c \times p$ matrix of constants imposing c linearly independent constraints on the elements of β and d is a $p \times 1$ vector of constants. To construct the power function of the Wald test, at first we have to know the

distribution of Wald test under alternative hypothesis H_1. PHILLIPS, P. C. B. [14] discussed the exact distribution of the Wald test under alternative hypothesis. According to PHILLIPS, P. C. B. [14], the probability density function of the Wald statistic (w) under H_1 is given by

$$pdf(w) = \frac{\exp(-m'm/2)w^{\frac{q}{2}-1}}{N^{q/2}\Gamma(q/2)} \sum_{j,k} \frac{(-1)^f (1/2)^k (w/N)^f}{j!k!(1/2)_k C_{(f)}(I_q)}$$

$$\cdot \left[\det(L(\partial X \otimes I)L')^{1/2} \cdot C_f^{j,k}(L(\partial X \otimes I)L', (L(\partial X \otimes I)L',)mm') \det(I - X)^{-N/2} \right] X=0$$

$$= \chi_q^2(\delta^2)$$

with the non centrality parameter is

$$\delta^2 = m'm = b'V^{-1}b = b'(C\sum_{\beta}C')^{-1}b$$

where $L = V^{-1/2}C^{1/2}V^{1/2}C'^{-1/2}$

According to Shieh, G. [15], to calculate the power function, at first we have to find the $100(1-\alpha)$ th percentile of a central chi-square distribution with p d.f., denoted by $\chi^2_{c,1-\alpha}$ and then find the non centrality parameter δ^2 of the non central chi-square distribution with c d.f such that the 100β th percentile, denoted by $\chi^2_{c,\beta}(\delta^2)$ is equal to $\chi^2_{c,1-\alpha}$, mathematically it can be written as

$$\psi(\delta^2, \alpha) = 1 - F_{\chi_c^2(\delta^2)}(p_{1-\alpha})$$

where $p_{1-\alpha}$ denotes the $(1-\alpha)th$ percentile of the central chi-square $\chi^2_{c,1-\alpha}$ and $F_{\chi_c^2(\delta^2)}(p_{1-\alpha})$ is the cumulative density of the noncentral chisquare distribution with c degrees of freedom and noncentrality parameter δ^2 at $p_{1-\alpha}$.

RESULTS AND DISCUSSION

In this section, we have illustrated the general approach with an example. For illustration purposes, we fitted Subject-Specific logistic regression model which is considered as the simplest case of GLMM for the repeated measures and the parameters have estimated by MLE and GEE approaches. To carry out the analysis we had used the repeated measures data of diabetes mellitus. Here the follow up data on 965 patients registered Bangladesh Institute of Research and Rehabilitation in Diabetes, Endocrine and Metabolic Disorders' (BIRDEM) in 2002 was used to identify the risk factors responsible for the transitions from controlled diabetic state as well as confirm diabetic to controlled stage of diabetes. Here we had taken into account the 3 consecutive visits of the patients from the registration. The response variable was defined in terms of the observed glucose level two hours of 75gms-glucose load follow up visit. The cut-off point for the blood glucose level is 11.1 mmol/liter. If the observed response was less than 11.1, then the patient was defined as non diabetic (categorized as 0) if the response was greater than or equal to 11.1 then the patients was said to be diabetic (categorized as 1). In our study we considered two covariates age and sex respectively. The variable age represents the age of the respondents at each visit, which is used as a continuous variable directly in the analysis.

Table 1. Fixed Effect Estimates of Subject-Specific Logistic Regression Model Obtained by Maximum Likelihood (ML) Methods with Associated Wald Test

Variables	Estimated Coefficients	Standard Error	(Coefficients /S.E)^2	Odds- Ratio	P-value
Intercept (Fixed)	-0.129146	0.171574	0.567009	-------	0.45162
Age	-0.001502	0.003171	0.224676	0.998499	0.63560
Sex	-0.248922	0.076806	10.504081	0.779640	0.00119*

Table 2. Random Effect Estimates of Subject-Specific Logistic Regression Model Obtained by Maximum Likelihood (ML) Methods

Variables	Estimated Variance	Std. Dev.
Intercept (Random)	0.018059	0.13438

Table 3. Variance Covariance Matrix of the Estimated Fixed Effect Parameters of Subject-Specific Logistic Regression Model Obtained by Maximum Likelihood (ML) Methods

	Constant	Age	Sex
Constant	0.0294376090	-0.0004494745	-0.0022280906
Age	-0.0004494745	1.005209e-05	-0.0000244092
Sex	-0.0022280906	-0.0000244092	0.0058992143

Table 4. Fixed Effect Estimates of Subject-Specific Logistic Regression Model Obtained by Generalized Estimating Equation (GEE) Approach with Associated Wald Test

Variables	Estimated Coefficients	Standard Error	(Coefficients /S.E)^2	Odds- Ratio	P-value
Intercept (Fixed)	-0.146814	0.197544	0.552341	-------	0.457362
Age	-0.001503	0.004138	0.131994	0.998497	0.716373
Sex	-0.247109	0.094552	6.830247	0.781055	0.008962

Table 5. Random Effect Estimates of Subject-Specific Logistic Regression Model Obtained by Generalized Estimating Equation (GEE) Approach

Variables	Estimated Variance	Std. Dev.
Intercept (Random)	0.015135	0.12302

Table 6. Variance Covariance Matrix of the Estimated Fixed Effect Parameters of Subject-Specific Logistic Regression Model Obtained by Generalized Estimating Equation (GEE) Approach

	Constant	Age	Sex
Constant	0.039152492	-0.000766086	-0.002768442
Age	-0.000766086	1.709722e-05	-4.592932e-05
Sex	-0.002768442	-4.592932e-05	8.966033e-03

Table 7. Comparison of Power for Testing the Null Hypothesis $H_0 : \beta_2 = 0$ Against the Alternative Hypothesis $H_1 : \beta_2 \neq 0$ by the Wald Chisquare Test When the Parameters are Estimated by Maximum Likelihood (ML) & Generalized Estimating Equation (GEE) Approaches Respectively

Parameter Values	-0.4	-0.3	-0.2	-0.1	0.1	0.2	0.3	0.4
Power Values under ML Method	0.999	0.974	0.740	0.255	0.255	0.740	0.974	0.999
Power Values under GEE Method	0.987	0.885	0.558	0.183	0.183	0.558	0.885	0.987
Power Differences	0.012	0.089	0.182	0.072	0.072	0.182	0.089	0.012

The variable sex is dichotomous, where 0 stands for female and 1 stands for male.

The logistic regression model is given by

$$\log it(Y_i = 1 \mid age_i, sex_i, v_i) = \log it(\mu_i^{SS}) = \beta_0^{SS} + \beta_1^{SS} age_i + \beta_2^{SS} sex_i + v_i$$

such that the population-averaged odds ratio for sex given by

$$OR_{sex}^{SS} = \frac{P(Y_i = 1 \mid sex_i = 1, age, v_i) \big/ P(Y_i = 0 \mid sex_i = 1, age, v_i)}{P(Y_i = 1 \mid sex_i = 0, age, v_i) \big/ P(Y_i = 0 \mid sex_i = 0, age, v_i)} = \exp(\beta_2^{SS})$$

represents the ratio of the odds of a given patient having with confirmed diabetic if the patient is a male compared to the odds of the same confirmed diabetic if the patient is a female. Alternatively, ratio of the odds among male patients with confirmed diabetic having the same value of unobserved random effect to the ratio of the odds of those confirmed diabetic patients with the same value of unobserved random effect who are female. Let $\hat{\beta}_2$ be the estimated coefficient of the covariate sex under ML or GEE approach. We are interested to test the null hypothesis $H_0 : \beta_2 = 0$ against the alternative hypothesis $H_1 : \beta_2 \neq 0$ by the Wald test.

We have discussed the issue of power calculations in GLMM with correlated data. We have presented relevant formulas for several important special cases. In this paper, we have developed a systematic approach to power analysis for the most popular longitudinal data approaches, the GEE and MLE. By extending existing methods to accommodate more practical considerations, this unified approach improves the limitations of these methods and provides a quite general approach for power and sample size estimation in GLMM for repeated measures. The approach presented here allowed us to compute power for a relatively complex study for which no exact methods are available. This study compares the performance of different methods of estimation for GLMM using the power values.

According to the diabetes data considered in the study, we may conclude that if we are interested to calculate power values of the Wald test, we should use the ML method of estimation because the Wald test gives greater power values under this method of estimation than the method GEE. Under both the method of estimations the fixed effect covariate sex shows the significance contribution to the diabetes status by the Wald Chisquare test and gives the approximate estimates -0.25.

Table 7 shows that power differences increases if the parameter values tend to the magnitude 0.25 which means that the proposed Wald Chisquare test detects highly the effect size under the ML methods of estimation. Most longitudinal studies, consider varying follow up times for the subjects which produces the missing data. In the article, the GEE and ML methodologies with non missing data was addressed. Hence, a common, constant number of follow ups for each subject was considered.
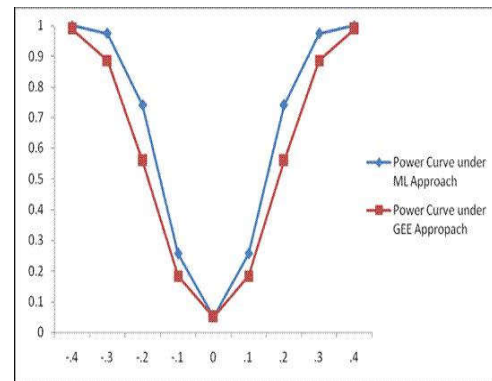
Power Curves Comparison for Testing the Null Hypothesis $H_0 : \beta_2 = 0$ Against the Alternative Hypothesis $H_1 : \beta_2 \neq 0$ by the Wald Chisquare Test When the Parameters are Estimated by Maximum Likelihood (ML) and Generalized Estimating Equation (GEE) Approaches Respectively

REFERENCES

1. Anderson DA, Aitkin M. Variance components models with binary response interviewer variability. J.R. Statistical Soc (1985); Series B, 47, 203-210.
2. Breslow N, Clayton G. Approximate inference in generalized linear mixed models. J. Amer. Statist. Assoc (1983); 88: 9-25.
3. Chamberlain G. Analysis of covariance with qualitative data. Re®. Econ. Stud (1980); 47; 225-238.
4. Greenwood M, Yule GU. An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease or of repeated accidents. J. Roy. Statist. Soc (1920). Ser A 83; 255-279.
5. Hartville DA. Maximum likelihood approaches to variance component estimation and to related problems. Journal of American Statistical Association (1977). 72; 320-388.
6. Hinde J. Compound Poisson regression models. Pp. 109-121 in GLIM 82: Proc. International Conference on Generalised Linear Models (1982), ed. R. Gilchrist. New York: Springer-Verlag.
7. Fang H, Brooks GP, Rizzo, ML. A Monte Carlo power analysis of traditional repeated measures and hierarchical multivariate linear models in longitudinal data analysis, Journal of Modern Applied Statistical Methods (2008), Vol. 7, No. 1; 101-119.
8. Laird NM, Ware FH. Random effects models for longitudinal data. Biometrics (1982).38; 963-974.
9. Lawley DN. A general method for approximating the distribution of likelihood ratio criteria. Biometrika (1956). 43; 295-303.
10. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. Biometrika (1986). 73; 13-22.
11. Nelder JA, Weddervurn RWM. Generalized linear models. J. R. Statistical Soc (1972). Series A, 135, part 3, 370-384.
12. Özdemir, Eyduran. Comparison of chi-square and likelihood ratio chi-Square tests. Journal of Applied

Sciences Research 1(2): 242-244, 2005, © 2005, INSInet Publication

13. Pierce DA, Sands BR. Extra-Bernoulli variation in regression of binary data. Technical Report 46, Statistics Deptartment (1975), Oregon State University, Cornwallis, OR.

14. Phillips PCB. The exact distribution of the Wald statistic. Econometrica (1986). Vol. 54, No. 4, 881-895.

15. Shieh G. On power and sample Size calculations for likelihood ratio tests in generalized linear models. Biometrics (2000). 56, 1192-1196.

16. Self SG, Mauritsen RH. Power/sample size calculations for generalized linear models, Biometrics (1988). 44, 79-86

17. Self SG, Mauritsen RH, Ohara J. Power calculations for likelihood ratio tests in generalized linear models. Biornetrics (1992). 48, 31-39.

18. Skellam JG. A probability distribution derived from the binomial distribution by regarding the probability of success as variable between the sets of trials. J. Roy. Statist. Soc (1948). Ser. B 10: 257-261.

19. Wong GY, Mason WM. The hierarchical logistic regression model for multilevel analysis. J. Amer. Statist. Assoc (1985). 80: 513-524.

20. Wedderburn RWM. Quasi-likelihood functions, generalized linear models, and the gauss-newton method. Biometrika (1974). 61: 439-447.

21. Gilmour AR, Anderson RD, Rae AL. The analysis of binomial data by a generalized linear mixed model. Biometrika (1985). 72, 3, 596-599.

22. *West* L. A review of J. Kiefer's work on conditional *frequentist* statistics. INDIANA 47907-1399. PHILADELPHIA, PENNSYLVANIA (1985). 19104-6302.

23. Zeger SP, Liang KY, Albert PS. Models for longitudinal data: A generalized estimating equation Approach. Biometrics (1988). 44, 1049-1060.

24. Pan W. Sample size and power calculations with correlated binary data, Controlled Clinical Trials 22:211–227, Elsevier Science Inc (2001).

25. Whittemore AS. Sample size for logistic regression with small response probability. Journal of the American Statistical Association (1981).76, 27-32.

26. Zeger SL. The analysis of discrete longitudinal data: commentary. Statistics in Medicine (1988). 7, 161-168.

27. Heckman JJ, Willis RJ. "Estimation of a Stochastic Model of Reproduction An Econometnc Approach, 99-146 National Bureau of Economic Research (1976), Inc

28. Brillinger DR, Preisler HK. Maximum likelihood estimation in a latent variable problem. Studies in Econometrics, Time Series and Multivariate Statistics (1983). Academic Press: New York; 31–65.

29. Gastañaga VM, McLaren CE, Delfino RJ. Power calculations for generalized linear models in observational longitudinal studies. 27-33 (2006)

30. Shieh G. A simple approach to power and sample size calculations in logistic regression and Cox regression models. Statistics in Medicine (2004) Volume: 23, Issue: 11, Pages: 1781-1792

31. Signorini. Sample size for Poisson regression, Biometrika (1991). 78, 446-50

****