

Effect of Correlation Structure in Generalized Estimating Equations and Quasi Least Squares: An Application in Type 2 Diabetes Patients

Dilip C. Nath
Department of Statistics
Gauhati University
Guwahati-781014
India
dilipc.nath@gmail.com

Atanu Bhattacharjee
Department of Statistics
Gauhati University
Guwahati-781014
India
atanustat@gmail.com

Abstract

The Quasi-Least Squares (QLS) is useful for various correlation structures when using Generalized Estimating Equations (GEE). The purpose of this work is to compare the regression parameters in the presence of different correlation structures when using GEE and QLS method. The comparison of estimated regression parameters has been performed in clinical trial data set; studying the effect of drug treatment (metformin with pioglitazone) vs (gliclazide with pioglitazone) in type 2 diabetes patients. In the case of QLS, the correlation coefficient for post-parandinal blood glucose (PPBG) under a tridiagonal correlation structure is 0.008 using GEE. It has been found that the combination of metformin with pioglitazone is more effective when compared to the combination of gliclazide with pioglitazone.

Keywords: MCMC, AR(1), Exchangeable, Unstructured Correlation.

Introduction

The appropriate and early diagnosis of type 2 diabetes can be helpful to start early treatment and improve the likelihood of achieving adequate glucose control. Thus, appropriate diagnosis with different biochemical parameters helps to reduce the various complications from diabetes. Though fasting blood glucose (FBG) has been considered for many years to determine the level of diabetes, PPBG (postprandial blood glucose) also needs to be considered for better management of type 2 diabetes. There is a chance that 30% of patients being tested for diabetes may be miss-diagnosed if screened by FBG alone (Ronald, 2005). However, these two biochemical parameters (PPBG and FBG) do not reflect the previous status of blood glucose levels, whereas glycosylated haemoglobin (HBA1C) does. The Collaborative Analysis of Diagnostic Criteria in Europe (DECODE, 2001) trial shows that as many as one-third of diabetic patients can be missed by simply measuring the FBG level. Therefore, in this analysis the level of PPBG, HBA1C and the drug effect as the covariates over the response FBG have been considered.

In marginal modeling, Liang and Zeger (1986) have considered GEE to obtain consistent estimates of the regression parameters. Stuart et al. (2009) have applied the joint estimates in the marginal models by a single modified GEE approach in a children heart function comparison. Recently, Jianfeng et al. (2009), have developed a bivariate analytical method for genome-wide association study based on the modified extended GEE. Clarke et al. (2010) have applied the negative binomial regression in GEE to estimate and compare event rates, hospital utilization, and costs associated with major diabetes-related complications in under-developed, developing and developed countries.

The specification of existing correlation is important when working with GEE. QLS is also useful for this purpose.

Objective

The goal of this paper is to compare the drug treatment effect on biochemical parameters among the patients at three time points, i.e. baseline and two follow up visits. This paper looks at the existing covariance structure to explain its effect on the model.

Statistical Method

In this paper, the outcomes y_{ij} of i^{th} subject and the covariates

$x_{ij} = (x_{i1}, \dots, x_{im})$ 'have been composed for different observation times $j=1; \dots, n_i$; for each subjects $i = 1, \dots, m$ (e.g. Liang et al. (1986)).

The mean and variance of the outcome variable has been obtained

$$\text{as } E(y_{ij}) = g(x_{ij}) = \mu_{ij} \quad (1)$$

$$\text{and } \text{var}(y_{ij}) = \phi h(\mu_{ij}) \quad (2)$$

where ϕ and μ are the dispersion parameter and mean response respectively. The $g(\cdot)$ and $h(\cdot)$ are the link functions (e.g. Liang et al. (1986)). In contrast to GLM, both the QLS and GEE assume the covariance matrix of

$$\text{Cov}(y_{ij}) = \phi A_i^{1/2} R_i(\alpha) A_i^{1/2} \quad (3)$$

$$\text{where } A_i = \text{diag}(h(\mu_{i1}), \dots, h(\mu_{in_i})) \quad (4)$$

and $R_i(\alpha)$ is the working correlation matrix (e.g. Liang et al. (1986)). The $R_i(\alpha)$ explains the pattern of association between the repeated observations. The forms of estimating equations in GEE and QLS have been described by McCullagh and Nelder (1989).

Generalized Estimating Equation

The covariance among the responses can be specified by

$$V_i = \phi A_i^{\frac{1}{2}} R_i(\alpha) A_i^{\frac{1}{2}} \quad (5)$$

$\Rightarrow A_i = \text{diag} V(\mu_{ij})$ diagonal matrix with diagonal elements $v(\mu_{ij})$

$\Rightarrow R_i(\alpha)$ is an $n_i \times n_i$ correlation matrix,

Liang and Zeger (1986) have described the $R_i(\alpha)$ as a correlation matrix. Where, α is the vector of parameter associated to the specified model of $\rho_{ist}(\alpha) = \rho_{ist} = \text{Corr}(Y_{is}, Y_{it}); \alpha, s \neq t$, and s, t are the times of observation. In GEE, the working covariance V_i can be denoted by $\text{Cov}(Y_i)$. The matrix $R_i(\alpha)$ is referred to as the "working correlation matrix" (e.g. Crowder (1995)). However, Wang (2003) has used the term "working covariance" for $R_i(\alpha)$. $R_i(\alpha) = I$, is useful to reduce the GEE in QLS with the assumption of independent repeated measurements. Prentice (1988) has proposed using the parameterization of within-subject association by the linear function of α (or, $\rho_{ist}(\alpha) = \alpha_{st}$). However, α or $\rho_{ist}(\alpha)$ could in principle be dependent on other covariates.

Normally, when V_i and α are not known, it is necessary to estimate $\rho_{ist}(\alpha) = \text{Corr}(Y_{is}, Y_{it})$.

Some types of correlation are (1) "exchangeable correlation" where $\rho_{ist} = \alpha$ for all $s < t$;

(2) "first-order autoregressive correlation" (AR (1)) for, $\rho_{ist} = \alpha^{|t-s|}$ for $0 < \alpha < 1$, where the correlation decreases as the time between measurements ($|t-s|$) increases; and (3) "unstructured correlation" when $\rho_{ist} = \alpha_{st}$ has no specified form.

Quasi Least Square

The specification of particular correlation is an important task in marginal modeling. Crowder (1995) has pointed out that the estimates of α could be wrong due to miss-specification of working correlation in GEE. In that case, the iterative method fails to converge for feasible estimate of α . The QLS can be useful in obtaining a feasible estimate of α . It is a two stage procedure with the computation of the correlation parameter in the composition of GEE. It is solvable for (Liang and Zeger, (1986)) β in estimating equations through the current value

$$\text{of } \alpha \text{ by two stages viz, (i) } \sum_{i=1}^m D_i^T A_i^{-\frac{1}{2}} R_i^{-1}(\alpha) A_i^{-\frac{1}{2}} (Y_i - U_i(\beta)) = 0 \quad (6)$$

where, $U_i(\beta) = E(Y_i)$ and $D_i = \frac{\partial U_i}{\partial \beta}$.

(ii) updating the correlation parameter α by minimizing generalized error sum of squares by $Q(\beta, R(\alpha)) = \sum_{i=1}^m z_i^T (\beta) R_i^{-1}(\alpha) Z_i(\beta)$ (7)

where $\alpha \in \Omega \subseteq R^s$, and $Z_i(\beta) = A_i^{-1/2}(Y_i - \mu_i) = (Z_{i1}, Z_{i2}, \dots, Z_{in})_{n \times 1}$ are known as the Pearson residuals (e.g. Jicgum et al (2009)) from the current value of α .

Chaganty (1997) has shown that the estimated value of α is inconsistent. To obtain a consistent estimate of α , the parameter β should be chosen from the two stage procedure by, α_{QLS} .

In this work the first stage has been considered as,

$$\sum_{i=1}^m \text{trace} \left\{ \frac{\Delta R_i^{-1}(\partial)}{\Delta \partial} R_i(\alpha) \right\}_{\alpha_{QLS} = \alpha} = 0 \quad (8)$$

In this analysis, the estimated value of α and the next step value of β have been obtained by solving the GEE's β value, evaluated at α_{GEE} (e.g Chaganty and Shults (1999)).

The advantage of QLS is due to the application of two stage procedures to point the coefficient parameters for the feasible and consistent results. The asymptotic distribution for β is the same regardless of the method (QLS or GEE)

Methodology

The secondary data in type 2 diabetes drug comparisons has been obtained from a clinical trial in Minakshi Mission Hospital, Tamil Nadu for two combination drug treatments viz

- (1) A combination of metformin with pioglitazone i.e. treatment 1 and
- (2) A combination of pioglitazone with gliclazide i.e. treatment 2.

A total of 100 patients were randomized to each treatment group and followed over a period of 12 months. The measurements of clinical parameters have been taken at baseline, three month and at the end of the twelve months. Participants in this study have been allocated to either of two diabetes drug (denoted treatment 1 and treatment 2). The effects of different covariates (PPBG and HBA1C) have been obtained under the specification of various correlation structures available in QLS and GEE. Prior to receiving the treatment, baseline data of the FBG are recorded for each patient. The FBG level have been categorized to (1) more than a normal and (2) less than the normal range in the binary form (e.g. Stokes et al. (2000)). In this work, a total of four covariates have been considered for analysis i.e. PPBG, Drug, HIBA1C, and ID, where PPBG and HBA1C are the value of postprandial and hemoglobin A1-C, at each visit, the type of treatment (labeled Drug) is coded as 1 for the (metformin with

pioglitazone) and 0 for the (gliclazide with pioglitazone). The ID is the identity number assigned to each patient. The programmes to perform QLS have been written by Ratcliffe et al. (2006) with the help of the Cholesky decomposition..

Statistical Analysis

During the treatment, blood sugar level as a representative of the treatment effect has been obtained for each of three visits. The variable Drug is in the binary format with two levels.

The explanatory variables Drug, PPBG and HBA1C have been specified as a vector of explanatory variables $x' = [x_{ij1}, x_{ij2}, \dots, x_{ij3}]'$.

The drug effect has been classified

$$\text{as: } \begin{cases} X_{ij} = 1(\text{drug 1})(\text{metformin with pioglitazone}) \\ 0(\text{drug 2})(\text{gliclazide with pioglitazone}) \end{cases} \quad (9)$$

Here, Y_{ij} is the FBG value for the i^{th} patient on the j^{th} visit, $j=1, \dots, 3$. and $\mu_{ij} = E(y_{ij})$ is the mean of FBG. The variance function of the binomial distribution is $v(\mu_{ij}) = \mu_{ij}(1-\mu_{ij})$, in the logit link function of $g(\mu_{ij}) = \log(\mu_{ij})/(1-\mu_{ij})$. The model for the mean response is $g(\mu_{ij}) = x_{ij}'\beta$. The coefficient β is the vector of regression parameters.

In the software R 2.9.1, the library geepack and function geeglm have been used to specify the “working covariance” and variance functions.

Results

The poisson assumption has been applied for the variance, $\text{Var}(Y_{ij}) = \phi \mu_{ij}$, where it is assumed that $\phi > 1$. The area of interest is which treatment of diabetes reduces the FBG level of the body in the follow-up period to the overall level of FBG. It has been expressed to the following marginal model for the expected value of FBG.

$$\log E(Y_{ij}) = \log \mu_{ij} = \beta_1 + \beta_2 ppbs + \beta_3 drug_i + \beta_4 hba1c \quad (10)$$

where Y_{ij} is the value of FBG for the i^{th} patient in the j^{th} period of observation ($j=1,2,3$). In the variable of Drug, $drug_1$ and $drug_2$ are indicator for treatment 1 (metformin with pioglitazone) and treatment 2 (gliclazide with pioglitazone) respectively. Because, patients have been allocated to one of the two treatments, the model does not include main effects of treatment. To complete the specification of the model, the variances of the counts and the within-subject association in the repeated counts has been obtained. To understand the degree of overdispersion in data, we assume that the variance of Y_{ij} is given by $\text{Var}(Y_{ij}) = \phi \mu_{ij}$ where ϕ is the over dispersion parameter. Finally, the with-in subject association is accounted by assuming a common correlation, $\text{Corr}(Y_{i1}, Y_{i2}) = \alpha$. In this marginal model, the covariates are binary and the log-

linear regression parameter has been used to interpret the term of (log)rate ratio. We summaries the interpretation of β in term of the log expected count in the three observations i.e. at baseline and two treatment follow-ups.

The mean pattern changes for the response variables are shown in the Table 3. The mean PPBG and HBA1C are reduced in the both treatment groups. The expected count of FBG at baseline in the group of drug is e^{β_1} . The rate ratio of FBG, in the follow-up periods is $e^{\{\beta_1+\beta_2\}}$. Thus, $e^{\{\beta_2\}}$ is the rate ratio of FBG, comparing the follow-up period to baseline, in the both drug groups; (metformin with pioglitazone) and (gliclazide with pioglitazone). The changes in the expected FBG are expressed by β coefficients between the groups. The estimated values of the regression coefficients are shown in Table 1 and Table 2. The induced model has been fit by,

$$\log E(Y_{ij}) = \log \mu_{ij} = \beta_1 + \beta_2 DRUG + \beta_3 PPBS + \beta_4 HBA1C + HBA1C * PPBS + PPBS * DRUG + DRUG * HBA1C + DRUG * HBA1C * PPBS \quad (11)$$

Here the variable drug is an indicator variable for type of treatment. The estimated regression coefficients through binomial link function are given in Table 4. In case of GEE and QLS, the estimated values of PPBG coefficient are 0.01 under the unstructured correlation structure. The common estimate of the log ratio, comparing post-treatment rates of PPBG in the diabetes group (drug 1), to group (drug 2), is 0.01. Thus, the rate ratio is 1.01 (or $e^{\{0.01\}}$), with 95% confidence interval(1.00 to 1.02) indicates that treatment 1 (metformin with pioglitazone) significantly reduce the averages' number of FBG compared to treatment 2 (gliclazide with pioglitazone). The estimate of the pair wise correlation ($\hat{\alpha}$) is low (approximately 0), indicating that there may be considerable homogeneity between the patients iwith a specific disease severity.

The estimated scale parameter ($\hat{\phi}$) has been approximated as 0.505 showing the over dispersion in the features of the data.

The estimated value of parameters in presence of different correlation structure by the both GEE and QLS cases are obtained and given in Table 1 and Table 2 respectively. In this study, the FBG observations have not been taken in equal spaced time. In case of unequal space time repeated observations with Markov correlation structure becomes useful for data analysis. However, in this analysis the estimates of the different regression parameters in the presence of AR(1), equi-correlated and tri-diagonal correlation structures have been explored and compared. In QLS, the estimated correlation for PPBG in Markov structure is 0.007. The p value with respect to the treatment effect is 0.06. In both the cases QLS and GEE, with a presence of exchangeable correlation structure the estimated coefficient for PPBG is 0.005 with non-significant p-value 0.12 that concludes the non-significant relation of PPBG on the response observation. In the presence of AR(1) structure the estimated coefficient of PPBG for QLS and GEE are 0.007 and 0.005 respectively. In case of QLS the p-value for the PPBG is 0.06, i.e. significant but for the GEE it is 0.12 with non-significant effect with

FBG. In fam and tri-diagonal correlation structure under the QLS, the PPBG regression parameter comes with 0.009 and 0.008 is in highly significant p-value. It again shows that PPBG is closely related to the FBG value. It can be concludes that the PPBG observation is needed to screen at the time of diabetes diagnosis. In tri-diagonal correlation structure the regression parameter of PPBG in the presence of QLS is 0.008 that has failed to obtain through GEE. The p-value of PPBG is 0.03, represents the significant relation between PPBG and FBG.

The fitted regression line for FBG under nine different correlation structure models can be specified by,

(I)In QLS

AR1:-

$$FBG = INTERCEPT (0.51) + 0.00 * PPBG - 0.01 * HBAIC + 0.22 * DRUG$$

EXCHANGEABLE:

$$FBG = INTERCEPT (0.49) + 0.00 * PPBG - 0.01 * HBAIC + 0.21 * DRUG$$

FAM:-

$$FBG = INTERCEPT(0.49) + 0.00 * PPBG - 0.02 * HBAIC + 0.22 * DRUG$$

MARKOV:-

$$FBG = INTERCEPT (0.51) + 0.00 * PPBG - 0.01 * HBAIC + 0.22 * DRUG.$$

TRIDIAGONAL:-

$$FBG = INTERCEPT (0.51) + 0.00 * PPBG - 0.02 * HBAIC + 0.23 * DRUG.$$

(II)In GEE

AR1:-

$$FBG = INTERCEPT(0.51) + 0.00 * PPBG - 0.01 * HBAIC + 0.22 * DRUG$$

EXCHANGEABLE:-

$$FBG = INTERCEPT (0.49) + 0.00 * PPBG - 0.01 * HBAIC + 0.21 * DRUG$$

INDEPENDENT:-

$$FBG = INTERCEPT(0.49) + 0.00 * PPBG - 0.02 * HBAIC + 0.22 * DRUG.$$

UNSTRUCTURED:-

$$FBG = INTERCEPT (0.49) + 0.00 * PPBG - 0.01 * HBAIC + 0.22 * DRUG.$$

Discussion

This study examined the long-term impact of drug effect on type 2 diabetes patients under supervision over a one year study period. The drug effect comparison has been performed with FBG responses. It is not possible to consider all kinds of correlation structure by the both GEE and QLS. So the challenge has been faced by recognizing that some correlation structures do not work with GEE and some of them do not work with QLS. In the case of GEE, the unstructured correlation has produced the PPBG correlation coefficient that could not be produced by QLS. In the AR(1) correlation structure the estimated regression parameter for both GEE and QLS are same for HBA1C. We conclude that treatment of metformin with pioglitazone does a better job of reducing FBG compared to gliclazide with pioglitazone. The results indicate that PPBG level is significantly associated with the FBG level in the type 2 diabetes bloods. Although HBA1C is associated with FBG level among the patients over follow-up periods, the association is significant. The results of the current study lead to a series of proposal including the requirement to take measurement of HBA1C and PPBG with the FBG in the type 2 diabetes patients. It is, therefore, possible that the narrowed differences between the two groups with regard to treatments and in the resultant HbA1c levels may be partly obscure differences in long-term outcomes over time. Meng et al (2005) have concluded that the patients with pioglitazone completed study with higher HBA1c level in comparison to gliclazide group. Bailey et al (1996) have concluded that metformin lowers the fasting blood glucose levels. Turner et al (1999) have described that when monotherapy fails, treatment is need to change in combined drug, or insulin therapy. Recently, Charbonnel et al (2005) have shown that the combination of pioglitazone with metformin is effective alternative in comparison to monotheray. In the present study, patients have been followed for a year to complete the clinical trial. The goal of the trial is to compare the drug combination of metformin with pioglitazone over a combination of pioglitazone with gliclazide to reduce the diabetes parameter like FBG for assess the treatment effect. In any kind of longitudinal data set there can be the infinite number of models in presence of infinite number of correlation structure to analyse it. The principle of model comparison is not to determine a 'accurate' model but to infer from the model, given a set of reasonable choices, is most 'useful' i.e. stand for an optimal equilibrium between accuracy and complexity. In other words, statistical inference has nothing to say about 'accurate' models. All that it grants an inference about which is more to be expected in a given data set. This computation has been performed in R. In this context , it can be concluded that the attachment of QLS gives more scope to point out different correlation structures in addition to the GEE approach. These, finding indicate that patients in high FBG seem to be no effect on PPBG level. On the other hand at the end of the study the high FBG level has been found in the drug therapy group pioglitazone with gliclazide in comparison to metformin with pioglitazone.

Acknowledgments

The authors would like to thank Rakesh Kumar Saroj for providing the data set for analysis from M.M.H.R.S, Madurai.

Table 1: QLS estimation of a parameter through different correlation structure

Correlation Structure	VARIABLE (Parameter)	ESTIMATION	S.E.	WALS	p-VALUE
FAM	INTERCEPT(β_0)	0.497449*	0.208903	5.67	0.0173
	PPBG(β_1)	0.00921*	0.000408	5.09	0.0240
	HBAIC(β_2)	-0.023046	0.026991	0.73	0.3932
	DRUG(β_3)	0.224971*	0.076581	8.63	0.0033
AR1	INTERCEPT(β_0)	0.5110138**	0.1789051	8.159	0.00429
	PPBG(β_1)	0.007076*	0.0003873	3.338	0.06769
	HBAIC(β_2)	-0.0194338	0.02457	72 0.625	0.42910
	DRUG(β_3)	0.2263851**	0.07642	64 8.774	0.00306
EXCHANGEABLE	INTERCEPT(β_0)	0.491916**	0.179951	7.47	0.0063
	PPBG(β_1)	0.00596	0.000390	2.33	0.1266
	HBAIC(β_2)	-0.013917	0.024952	0.31	0.5770
	DRUG(β_3)	0.218781**	0.077327	8.00	0.0047
MARKOV	INTERCEPT(β_0)	0.511014**	0.178905	8.16	0.0043
	PPBG(β_1)	0.00708*	0.000387	3.34	0.0677
	HBAIC(β_2)	-0.019434	0.024577	0.63	0.4291
	DRUG(β_3)	0.226385**	0.076426	8.77	0.0031
TRIDIAGONAL	INTERCEPT(β_0)	0.517281**	0.189560	7.45	0.0064
	PPBG(β_1)	0.00835*	0.000400	4.37	0.0367
	HBAIC(β_2)	-0.023865	0.025274	0.89	0.3450
	DRUG(β_3)	0.233804**	0.076890	9.25	0.0024

*p-value <0.1, **p-value <0.01, ***p-value <0.001,

Table 2: GEE estimation of parameter through different correlation structure

Correlation Structure	VARIABLE (Parameter)	ESTIMATION	S.E.	WALS	p-VALUE
AR1	INTERCEPT(β_0)	0.511236**	0.178691	1 8.19	0.0042
	PPBG(β_1)	0.00698 *	0.000390	3.20	0.0735
	HBAIC(β_2)	-0.019233	0.024651	0.61	0.4353
	DRUG(β_3)	0.226470**	0.076454	8.77	0.0031
EXCHANGEABLE	INTERCEPT(β_0)	0.491729**	0.179984	7.46	0.0063
	PPBG(β_1)	0.00594	0.000390	0 2.32	0.1281
	HBAIC(β_2)	-0.013848	0.024967	0.31	0.5791
	DRUG(β_3)	0.218751**	0.07733	2 8.00	0.0047
INDEPENDENT	INTERCEPT(β_0)	0.497449*	0.208903	5.67	0.0173
	PPBG(β_1)	0.00921*	0.000408	5.09	0.0240
	HBAIC(β_2)	-0.023046	0.026991	1 0.73	0.3932
	DRUG(β_3)	0.224971**	0.076581	1 8.63	0.0033
UNSTRUCTURED	INTERCEPT(β_0)	0.495471**	0.179553	7.61	0.0058
	PPBG(β_1)	0.00684*	0.000376	3.32	0.0684
	HBAIC(β_2)	-0.017269	0.024696	0.49	0.4844
	DRUG(β_3)	0.228661**	0.076399	8.96	0.0028

*p-value <0.1, **p-value <0.01, ***p-value <0.001,

Table 3: Description of Diabetes Data set

Variable (Parameter)	Observation	Mean(sd) for Drug-1	Mean(sd) for Drug-2
PPBG	Obs=1 st	267.3(52.88)	285.94(74.46)
	Obs=2 nd	200.55(26.75)	238.94(55.98)
	Obs=3 rd	151.91(0.15)	202.56(48.08)
HBA1C	Obs=1 st	9.62(1.09)	9.61(1.66)
	Obs=2 nd	8.41(0.8941)	8.72(1.49)
	Obs=3 rd	7.62(32.43)	8.13(1.34)

Table 4: Analysis of GEE parameter of binomial link for FBG under unstructured correlation structure

Variable	Parameter	Estimates	S.E.	P value
Intercept	β_1	-2.613903	1.1756	0.1184
DRUG	β_2	3.248042**	0.3689	0.000
PPBG	β_3	0.011731*	0.0139	0.013
HBA1C	β_4	0.180221	0.030	0.111
DRUG*HBA1C	β_5	-0.148303*	0.31	0.02
PPBG*HBA1C	β_6	-0.000794*	0.63	.036
DRUG*PPBG	β_7	-0.009301*	0.45	0.04
DRUG*PPBG*HBA1C	β_8	0.000522*	0.69	0.06
Scale Parameter	φ	0.505		

*p-value <0.1, **p-value <0.01, ***p-value <0.001,

Reference

1. Clarke PM, Glasziou P, Patel A, Chalmers J, Woodward M, et al. (2010) Event Rates, Hospital Utilization, and Costs Associated with Major Complications of Diabetes: A Multi-country Comparative Analysis. *PLoS Med* 7(2): e1000236.
2. Crowder, M. (1995). On the use of a working correlation matrix in using generalized linear models for repeated measurements. *Biometrika*, 82, 407–410.
3. Chaganty N. R. (1997). An alternative approach to the analysis of longitudinal data via generalized estimating equations. *Journal of Statistical Planning and Inference* 63, 39-54.
4. Jianfeng Liu, Yufang Pei, Chris J. Papasian and Hong-Wen Deng. (2009) Bivariate Association Analyses for the Mixture of Continuous and Binary Traits with the Use of Extended Generalized Estimating Equations. *Genet Epidemiol.* 33(3): 217–227.
5. Jichun Xie, Justine Shults and Peet Jon (2009). Quasi Least Square with Mixed Linear Correlation Structure. biostats.bepress.com/Upennbiostat/papers.art3.
6. McCullagh, P. and J. A. Nelder, (1989). *Generalised Linear Models*, 2nd ed. Chapman and Hall, New York.
7. Prentice, R.L. (1988) Correlated binary regression with covariates specific to each binary observation. *Biometrics*, 44, 1033–1048.
8. Ratcliffe S, Shults J. (2006). "GEEQBOX": A MATLAB toolbox for implementation of quasi-least squares and generalized estimating equations." *Journal of Statistical Software*, 25(14), 1-13.
9. Ronald A. Codario (2005). *Type 2 Diabetes, Pre-Diabetes, and the Metabolic Syndrome*, Humana Press 20.
10. Stokes M. E., Davis C. S., Koch G. G. (2000). *Categorical data analysis using the SAS system*, 2nd edition, Cary, NC: SAS Institute Inc.
11. Stuart R. Lipsitz, Garrett M. Fitzmaurice, Joseph G. Ibrahim, Debajyoti Sinha, Michael Parzen and Steven Lipschutz. (2009). Joint generalized estimating equations for multivariate longitudinal binary outcomes with missing data: An application to AIDS data. *J R Stat Soc Ser A Stat Soc.* 172(1): 3–20.
12. The European Diabetes Epidemiology Group (DECODE). (2001) Glucose tolerance and cardiovascular mortality: comparison of fasting and 2-hour diagnostic criteria. *Arch Intern Med* 161: 397–405.
13. Liang, K. Y., Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* 73: 13–22.
14. Wang, Y. G. and Carey, V. J. (2003). Working correlation structure misspecifying, estimating and covariates design: Implication for generalized estimating equations performance. *Biometrika* 90, 29-41.