AUSTRIAN JOURNAL OF STATISTICS
Volume 37 (2008), Number 2, 175–184

Model Selection Using Modified Akaike's Information Criterion: An Application to Maternal Morbidity Data

A.H.M. Mahbub Latif[1], M. Zakir Hossain[2], and M. Ataharul Islam[2]

[1] Institute of Statistical Research and Training, University of Dhaka, Bangladesh

[2] Department of Statistics, University of Dhaka, Bangladesh

Abstract: The most commonly used model selection criterion, Akaike's Information Criterion (AIC), cannot be used when the Generalized Estimating Equations (GEE) approach is considered for analyzing multivariate binary response. Recently, a modified version of AIC (mAIC) which is based on quasi-likelihood function is proposed as a model selection criterion. This model selection criterion can be used in the GEE setup. In this study, an application of mAIC is showed in selecting important covariates associated with pregnancy related complications of Bangladeshi women.

Zusammenfassung: Das am häufigsten verwendete Modellwahl Kriterium, das Akaike Informationskriterium (AIC), kann nicht verwendet werden, wenn der Ansatz der Generalisierten Schätzgleichungen (GEE) in Betracht gezogen wird um multivariate binäre Daten zu analysieren. Unlängst wurde eine modifizierte Version des AIC (mAIC) als Modellwahl Kriterium empfohlen, das auf die Quasi-Llikelihood Function basiert. Dieses Modellwahl Kriterium kann im GEE Umfeld verwendet werden. In dieser Studie wird eine Anwendung des mAIC gezeigt und damit wichtige Kovariablen ausgewählt, die mit schwangerschaftsbezogenen Komplikationen von Bangladeshi Frauen zusammenhängen.

Keywords: Multivariate Binary Response, Generalized Estimating Equations, Modified AIC.

1 Introduction

Millions of women in developing countries like Bangladesh experience life-threatening and other health related problems during pregnancy and post-partum period when they require professional care, but many of them either don't perceive the seriousness of their condition or they don't have favorable conditions to seek care. About 16000 maternal deaths occurred in Bangladesh due to pregnancy and delivery related complications in the year 2000.

Though experiencing complications during pregnancy and post-partum period is very common to Bangladeshi women, not many studies have been conducted in Bangladesh on this topic. Recently, Bangladesh Institute of Research for Promotion of Essential and Reproductive Health Technologies (BIRPERHT), a non-governmental organization, conducted a prospective survey on maternal morbidity in Bangladesh where the selected women were followed during the pregnancy and post-partum period. Among other important pregnancy-related variables, presence/absence of any complication during pregnancy is recorded over the follow-up period for each of the selected women.

Since several measurements are made from each woman over different time points, the responses are usually positively correlated and responses of this type are known as multivariate or correlated binary response.

The methods for analyzing multivariate binary responses can be classified into two broad classes of methods: likelihood based and estimating equation based methods. The likelihood based methods require complete specification of the joint distribution of the multivariate responses, whereas the estimating equation based methods can be employed when joint distribution is not fully specified. The most common likelihood based methods for multivariate binary data are multivariate probit and multivariate logit models which consider univariate normal and logistic distributions as univariate margins, respectively (see, Joe, 1997). On the other hand, the generalized estimating equations (GEE) methodology, an estimating equation based method which is proposed by Liang and Zeger (1986) (also see, Zeger and Liang, 1986), is widely used for analyzing multivariate binary response. GEE can be used for analyzing both continuous and discrete multivariate responses within the generalized linear model framework. This method can provide consistent estimators of the regression parameters if the specification of the marginal means is correct. They introduced the "working" correlation matrix in which a larger value of working correlation parameter is used if there is more dependence in the data.

Model selection is an important part of data analysis which leads to a search "best" model. By "best" model, we mean selecting the best subset of the covariates from the available covariates in the data. Usually model selection is done by using a specific criterion. For likelihood-based methods, Akaike's Information Criterion (AIC) (Akaike, 1973) is widely used as a model selection criterion. But for non-likelihood-based methods, e.g., GEE, no such criterion is available for model selection. Recently, a modified Akaike's Information Criterion (mAIC), which is based on the quasi-likelihood function (McCullagh and Nelder, 1989), was proposed as a model selection criterion Pan (2001a). Among other non-likelihood based methods for model selection Pan (2001b) proposed the bootstrap smoothed cross-validation (BCV), a general model selection criterion that minimizes the expected predictive bias (EPB). Again Pan and Lee (2001) suggested the basic and bias-corrected bootstrap approaches to estimate the predictive mean squared error (PMSE) of a model and use the PMSE for model selection. Cantoni et al. (2005) suggested a generalized version of Mallows's C_p (GC_p) suitable for use with both parametric and non-parametric models, that provides an estimate of a measure of model's adequacy for prediction. Recently, Cantoni et al. (2008) also proposed a cross-validation Markov Chain Monte Carlo (MCMC) procedure as a general variable selection tool which avoids the need to visit all candidate models.

The main objective of this paper is to select best models from a given set of covariates when the response is multivariate binary. The generalized estimating equation (GEE) approach is considered for modeling multivariate binary response and appropriate information criterion is used to select the best subset of the available covariates. A procedure of selecting working correlation structure for the selected subset of covariates is also described with an example of maternal morbidity data. In Section 2, the method of generalized estimating equations and modified Akaike's Information Criterion are briefly described. In Section 3, a short description of the sampling procedure and estimates of different models considered in the analysis are given, and Section 4 contains the conclusion.

2 Methods

2.1 Generalized Estimating Equations

Let $\mathbf{y}_i = (y_{i1}, \ldots, y_{id_i})'$ be the response vector corresponding to the ith woman, $i = 1, \ldots, n$, where the binary response y_{ij} corresponds to the jth time-point of the ith woman, representing whether or not the woman suffers from specific complication. Let $\mathbf{x}_{ij} = (x_{ij1}, \ldots, x_{ijp})'$ be the vector of covariates corresponding to the jth response of the ith woman, where $x_{ij1} = 1$ for all i, j. Let us assume that y_{ij} follows a distribution from the exponential family and the dependence of the mean function $\mu_{ij} = \Pr(y_{ij} = 1)$ on the covariate set $\mathbf{x}_{ij}$ can be expressed by the link function $h(\cdot)$ as $\mu_{ij} = h^{-1}(\boldsymbol{\beta}'\mathbf{x}_{ij})$, where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ is the parameter vector of interest.

Liang and Zeger (1986) used a *working* correlation matrix $\mathbf{R}_i(\boldsymbol{\alpha})$, $i = 1, \ldots, n$, of order $d_i \times d_i$ to specify the within-subject dependence. The form of the working correlation matrix is assumed to be fully specified by the parameters $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_q)'$. The common correlation structures such as *independence* and the *exchangeable* correlation structure can be obtained by considering $\mathbf{R}_i(\boldsymbol{\alpha}) = \mathbf{I}_{d_i}$ and $\mathbf{R}_i(\boldsymbol{\alpha}) = (1-\rho)\mathbf{I}_{d_i} + \rho\mathbf{J}_{d_i}$, respectively, where $\rho = \operatorname{corr}(y_{ij}, y_{ik})$, $j, k = 1, \ldots, d_i$, $j \neq k$, where $\mathbf{I}_{d_i}$ is the identity matrix of order $d_i \times d_i$ and $\mathbf{J}_{d_i}$ is a $d_i \times d_i$ matrix with all elements equal to one.

For estimating the regression parameters, Liang and Zeger (1986) proposed the following set of estimating equations

$$\mathbf{U}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left(\frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}}\right)' \mathbf{V}_i^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_i)\,, \tag{1}$$

where $\mathbf{V}_i$ is the *working* covariance matrix considered for the ith subject, which can be expressed as a function of the *working* correlation matrix as

$$\mathbf{V}_i = \mathbf{A}_i^{1/2}\mathbf{R}_i(\boldsymbol{\alpha})\mathbf{A}_i^{1/2}\,,$$

where $\mathbf{A}_i = \operatorname{diag}(\operatorname{var}(y_{i1}), \ldots, \operatorname{var}(y_{id_i}))$ and $\operatorname{var}(y_{ij}) = a(\phi)\mu_{ij}(1-\mu_{ij})$ is a function of the known mean function μ_{ij} and the dispersion parameter ϕ. Thus, the estimating equations (1) are functions of the regression parameters $\boldsymbol{\beta}$, the dispersion parameters $\boldsymbol{\alpha}$, and ϕ. If the regression parameters are of main interest, the estimating equations can be reduced as a function of $\boldsymbol{\beta}$ by replacing $\boldsymbol{\alpha}$ and ϕ by $\hat{\boldsymbol{\alpha}}(\mathbf{y}, \boldsymbol{\beta}, \phi)$ and $\hat{\phi}(\mathbf{y}, \boldsymbol{\beta})$, respectively. So the estimating equations can be written as

$$\begin{aligned}\mathbf{U}\left(\boldsymbol{\beta}, \hat{\boldsymbol{\alpha}}(\boldsymbol{\beta}, \hat{\phi}(\boldsymbol{\beta}))\right) &= \sum_{i=1}^{n} \left(\frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}}\right)' \mathbf{V}_i^{-1}(\boldsymbol{\beta}, \hat{\boldsymbol{\alpha}}(\boldsymbol{\beta}, \phi))(\mathbf{y}_i - \boldsymbol{\mu}_i) \\ &= \sum_{i=1}^{n} \mathbf{C}_i\mathbf{B}_i\mathbf{A}_i\,. \end{aligned} \tag{2}$$

According to Liang and Zeger (1986), given the estimators of $\boldsymbol{\alpha}$ and ϕ, the estimator of the regression parameters $\hat{\boldsymbol{\beta}}$, which is the solution of $\mathbf{U}(\boldsymbol{\beta}, \hat{\boldsymbol{\alpha}}(\boldsymbol{\beta}, \hat{\phi}(\boldsymbol{\beta}))) = \mathbf{0}$, is consistent and asymptotically multivariate normal with mean $\boldsymbol{\beta}$ and covariance matrix

$$\mathbf{V} = \left(\sum_{i=1}^{n} \mathbf{C}_i\mathbf{B}_i\mathbf{C}_i'\right)^{-1} \left(\sum_{i=1}^{n} \mathbf{C}_i\mathbf{B}_i\mathbf{A}_i\mathbf{A}_i'\mathbf{B}_i'\mathbf{C}_i'\right) \left(\sum_{i=1}^{n} \mathbf{C}_i\mathbf{B}_i'\mathbf{C}_i'\right)^{-1} . \tag{3}$$

One of the attractive property of GEE approach is that it provides consistent estimator of $\boldsymbol{\beta}$ even if the correlation matrix $\mathbf{R}$ is misspecified.

Though Liang and Zeger (1986) did not mention any connection between the GEE approach and the likelihood based approach. For multivariate binary responses it can be shown that the estimating equations (2) are score function derived from multivariate logistic distribution (see Molenberghs and Lesaffre, 1994, Joe and Latif, 2005) with constant third and fourth order moments. It can also be shown that the estimating equations (2) are equivalent to the score functions obtained from the quasi-likelihood function (McCullagh and Nelder, 1989) with *independent* correlation structure (Pan, 2001a). However, for a more general correlation structure there is no guarantee that a corresponding quasi-likelihood function exists unless certain conditions are satisfied.

2.2 Akaike's Information Criterion

Akaike's information criterion (Akaike, 1973) was introduced as an approximately unbiased estimator of the expected Kullback-Leibler information of the fitted model. Let $\mathcal{D} = \{(\mathbf{y}_i, \mathbf{x}_{ij})\}$ be the data at hand, where $\mathbf{y}_i$ is the response vector and $\mathbf{x}_{ij}$ is a set of covariates as defined in the previous section. Also let M and $M^\star$ be a candidate and the true model, respectively. Further let $L(\boldsymbol{\beta}; \mathcal{D})$ and $L(\boldsymbol{\beta}^\star; \mathcal{D})$ be the log-likelihood functions corresponding to the models M and $M^\star$, respectively, where $\boldsymbol{\beta}$ and $\boldsymbol{\beta}^\star$ are the corresponding regression parameters. The Kullback-Leibler information, also known as cross entropy, between the models M and $M^\star$ is

$$\Delta(\boldsymbol{\beta}, \boldsymbol{\beta}^\star) = \mathrm{E}_{M^\star}\left(-2L(\boldsymbol{\beta}; \mathcal{D})\right) ,$$

where the expectation $\mathrm{E}_{M^\star}$ is taken under the true model $M^\star$. For a given set of competing models, we choose that model as the best model for which the Kullback-Leibler information $\Delta(\boldsymbol{\beta}, \boldsymbol{\beta}^\star)$ is the smallest. In practice, both $\boldsymbol{\beta}$ and $\boldsymbol{\beta}^\star$ are unknown, as an asymptotically unbiased estimator of $\mathrm{E}_{M^\star}(\Delta(\hat{\boldsymbol{\beta}}, \boldsymbol{\beta}^\star))$ which is actually the AIC can be used as a model selection criterion, where $\hat{\boldsymbol{\beta}}$ is the maximum likelihood estimator of $\boldsymbol{\beta}$ under any competing model. Notationally, the AIC can be written as

$$\mathrm{AIC} = -2L(\hat{\boldsymbol{\beta}}; \mathcal{D}) + 2p ,$$

where p is the order of the vector $\boldsymbol{\beta}$. A model which minimizes the AIC is considered to be the "best" model. This definition implies that when there are several models whose values of maximum likelihood are about the same level, we should choose the one with the smallest number of free parameters. A more detailed discussion on AIC can be found in Linhart and Zucchini (1986) and a review of model selection can be found in Zucchini (2000).

2.3 Modified Akaike's Information Criterion

The AIC is one of the most widely used model selection criterion when the likelihood function can be fully specified. But on the other hand, when the likelihood function cannot be fully specified, e.g., as in the GEE setup, the AIC cannot be used for model selection purposes. In such a situation, the modified Akaike's Information Criterion (mAIC) which is based on the quasi-likelihood function (McCullagh and Nelder, 1989, p. 325), can be used instead. Under the working independence correlation structure, the quasi-likelihood function based new discrepancy measure can be defined as

$$\Delta_m(\boldsymbol{\beta}, \boldsymbol{\beta}^\star, \mathbf{I}) = \mathrm{E}_{M^\star}\left(-2Q(\boldsymbol{\beta}; \mathbf{I}, \mathcal{D})\right) ,$$

where $\mathbf{I}$ is for the independence working correlation structure. Then the modified Akaike's Information Criterion can be defined for a general working correlation structure $\mathbf{R}$ as

$$\mathrm{mAIC}(\mathbf{R}) = -2Q(\hat{\boldsymbol{\beta}}(\mathbf{R}); \mathbf{I}, \mathcal{D}) + 2\mathrm{trace}(\hat{\boldsymbol{\Omega}}_{\mathbf{I}}\hat{\mathbf{V}}_{\mathbf{R}})\,, \tag{4}$$

where $\hat{\boldsymbol{\beta}}(\mathbf{R})$ is a solution of the estimating equations defined in (1) under the working correlation structure $\mathbf{R}$, $\hat{\mathbf{V}}_{\mathbf{R}}$ is the estimated robust variance-covariance matrix under the general working correlation structure $\mathbf{R}$ which is defined in (3), and

$$\hat{\boldsymbol{\Omega}}_{\mathbf{I}} = -\left.\frac{\partial^2 Q(\boldsymbol{\beta}; \mathbf{I}, \mathcal{D})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}\right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}(\mathbf{R})}$$

is a consistent estimator of

$$\boldsymbol{\Omega}_{\mathbf{I}} = \mathrm{E}_{M^\star}\left(-\left.\frac{\partial^2 Q(\boldsymbol{\beta}; \mathbf{I}, \mathcal{D})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}\right|_{\boldsymbol{\beta}=\boldsymbol{\beta}^\star}\right) = \sum_{i=1}^{n} \mathbf{C}_i\mathbf{B}_i\mathbf{C}_i'\,.$$

The right hand side of equation (4) is approximately equal to $\mathrm{E}_{M^\star}(\Delta_m(\hat{\boldsymbol{\beta}}, \boldsymbol{\beta}^\star, \mathbf{I}))$, which ignores a term that is difficult to estimate (Pan, 2001a). However, this term converges to 0 if the model is correctly specified.

For analyzing regression models with dependent responses using a GEE approach, a minimum mAIC strategy can be used to find the best model from a set of competing models. The mAIC can be helpful not only to select the best set of covariates but also to select the best working correlation structure. Among all competing models, the best model is the one that has the smallest mAIC value. The difference between two mAIC values may not be meaningful. One of the limitations of the mAIC as a model selection criterion is that no probability distribution is associated with it, so the difference between two mAIC values cannot be compared using any statistical hypothesis testing procedure. Another limitation of the mAIC is its weak consistency in the sense that its consistency is assured only if the model is correctly specified. For details about the mAIC see Pan (2001a).

3 Results

3.1 Data and Variables

This paper is based on the data from the survey on maternal morbidity in Bangladesh conducted by the Bangladesh Institute of Research for Promotion of Essential and Reproductive Health Technologies (BIRPERHT) during the period of November 1992 to December 1993. There have been a number of papers published using this data set, e.g., Islam et al. (2004), Gulshan et al. (2005), and Chakraborty et al. (2003).

A multistage sampling design was used in the survey where in the first stage the districts are randomly selected in such a way that exactly one district was chosen from each division. In the second stage, one thana was randomly selected from each of the chosen districts and in the third stage, two unions were randomly selected from each of the selected thanas. All the pregnant women of duration at most six months of the selected unions comprised the sample. All the selected women were followed till 90 days after delivery. A total of 1020 pregnant women were interviewed in the survey. Information on socio-economic and demographic characteristics, pregnancy related care and practice, morbidity during the period of follow-up as well as in the past, information concerning complications at the time of delivery and during the postpartum period, etc., were also collected for all the selected pregnant women.

One of the objectives of the BIRPHERT survey was to identify important factors associated with pregnancy related complications. The major life-threatening antenatal complications are hemorrhage, oedema, excessive vomiting, and fits or convulsion. In this study the response variable is considered as binary taking the value 1 if at least one of the complications was present. Notationally, it can be defined as

$$y = \begin{cases} 1, & \text{if the woman suffers at least one of the major complications,} \\ 0, & \text{otherwise.} \end{cases}$$

Among the available covariates, only five important covariates are considered in this study, which are: educational level of the respondents (EDU), age at marriage (AgeM), economic status (ECON), desired the index pregnancy (DIP), and food supplement (FS). All these covariates are coded as binary with the reference categories, never attended school for educational level, 15 years or less for age at marriage, less than average for economic status, and no for desired index pregnancy, and food supplement, respectively.

3.2 Selection of Best Models

One of the main objectives of this paper is to show applications of the mAIC to select the best model within this GEE setup. All possible models that can be considered from the selected five covariates are examined and the best models with different number of covariates are shown in Table 1 with different correlation structures.

Among the five models with one covariate, the model with FS as the only covariate (Model I) is found to be the best one because the corresponding mAIC value is the smallest. Model I is found to be the best model for all three correlation structures that have been considered in this analysis. Among the 10 models with two covariates, Model II, which includes AgeM and FS as covariates, is the best choice. For three covariates, the model with the covariates AgeM, ECON, and FS is found to be the best one, we denote this model as Model III. The best model with four covariates (Model IV) includes the covariate EDU in addition to the covariates of Model III. The only model with five covariates is denoted as Model V which contains all the covariates that are considered in this study. Among all the five models (Model I up to Model V), Model III can be considered

Table 1: Best models with different number of covariates for different correlation structures

	Covariates					Correlation Structure		
Model	EDU	AgeM	ECON	DIP	FS	Indep	Exchang	Unstruc
I					✓	692.63	692.63	709.36
II		✓			✓	690.17	690.17	705.76
III		✓	✓		✓	689.90	689.90	705.36
IV	✓	✓	✓		✓	690.57	690.57	706.14
V	✓	✓	✓	✓	✓	695.32	695.32	710.42

as the best model because the corresponding mAIC value is the smallest and this is true for all three correlations structures. For all cases, the selected best models are found to be the best model for all three correlation structures.

3.3 Analysis of Morbidity Data using Different Correlation Structures

Table 2 shows the estimates of the parameters of the best model (Model III) under different correlation structures, namely, independence, exchangeable, and unstructured. It is found that only covariate FS is significant irrespective of the choice of the correlation structure. The analysis shows that taking special food during the pregnancy period reduces the number of complications. More specifically, women who do not take food supplements during pregnancy experience about twice as more pregnancy related complications compared to women taking food supplements. Age at marriage is found to be significant (at a 10 percent level) only if the unstructured correlation structure is assumed for the model. The analysis shows that women who got married before their fifteenth birthday experience more pregnancy related complications than women who married later. The other variable of the best model (Model III), economic status, is found to have a non-significant effect on pregnancy related complications.

Table 2: Estimates of the parameters of Model III (with p-values in parenthesis)

Correlation Structure	Intercept	AgeM	ECON	FS
Independence	−0.023	−0.129	−0.023	−0.598
	(0.787)	(0.251)	(0.841)	(0.000)
Exchangeable	−0.029	−0.129	−0.023	−0.598
	(0.779)	(0.250)	(0.841)	(0.000)
Unstructured	0.093	−0.183	−0.007	−0.514
	(0.347)	(0.089)	(0.952)	(0.000)

4 Conclusion

Health related problems during pregnancy and the post-partum period are very common to Bangladeshi women. Not many studies have been considered in order to identify the important covariates associated with such pregnancy related problems. Recently, BIRPHERT has conducted a survey on pregnant women in Bangladesh to identify such factors associated with pregnancy related problems. In this study, the BIRPHERT data is used to show an application of recently proposed modified Akaike's Information Criterion (mAIC) as a model selection criterion. The mAIC is very useful in situations when the response is multivariate non-normal and a fully specified likelihood function is not available.

Among the five covariates we have considered in this analysis, age at marriage, the economic status, and taking food supplements are found to be the best subset of the covariates among all possible subsets of covariates. The selection of the best model does not depend on the choice of the correlation structure.

The analysis of the best model shows that taking food supplements during the pregnancy period significantly reduces complications during pregnancy period. This means that the probability of developing some major complications during pregnancy is smaller for women who took special food during pregnancy than for those who did not. In a study conducted in the late nineties in Gambia, Ceesay et al. (1997) also found that food supplements have a significant effect on increasing weight gain during pregnancy and also on increasing birth weight.

The variable age at marriage is also found to have a significant effect on pregnancy related complications if only the unstructured correlation structure is considered for the model. Age at marriage is an important covariate for pregnancy related studies in a developing country like Bangladesh where more than 50% women married at the age 18. Akhter et al. (1996) also found a significant effect of age at marriage on pregnancy related complications. Recent studies show that female education plays a vital role in reducing maternal mortality, more specifically, a low incidence of maternal morbidities was found among the educated females (Choolani and Ratnam, 1995). Chowdhury et al. (2007) examined the trends in maternal mortality in Matlab, Bangladesh over 30 years and revealed female education and poverty reduction are two important variables in reducing the maternal mortality. In our analysis the variable female education has not been selected in the best model.

Acknowledgements

The authors are grateful to an anonymous referee for all valuable suggestions.

References

Akaike, H. (1973). *Information theory and an extension of the maximum likelihood principle.* Budapest: Akademiai Kiado.

Akhter, H. H., Chowdhury, M. E., and Sen, A. (1996). *A Cross-Sectional Study on Maternal Morbidity in Bangladesh*. Dhaka: Bangladesh Institute of Research for Promotion of Essential and Reproductive Health and Technologies.

Cantoni, E., Flemming, J. M., and Ronchetti, E. (2005). Variable selection for marginal longitudinal generalized linear models. *Biometrics*, *61*, 507-514.

Cantoni, E., Flemming, J. M., and Ronchetti, E. (2008). Longitudinal variable selection by cross-validation in the case of many covariates. *Statistics in Medicine*. (in press)

Ceesay, S. M., Prentice, A. M., Cole, T. J., Foord, F., Poskitt, E., Weaver, L. T., et al. (1997). Effects on birth weight and perinatal mortality of maternal dietary supplements in rural Gambia: 5-year randomized controlled trial. *British Medical Journal*, *315*, 786-790.

Chakraborty, N., Islam, M. A., Chowdhury, R. I., and Bari, W. (2003). Analysis of ante-partum maternal morbidity in rural Bangladesh. *Australian Journal of Rural Health*, *11*, 22-27.

Choolani, M., and Ratnam, S. S. (1995). Maternal morbidity: a global overview. *Journal of the Indian Medical Association*, *93*, 36-40.

Chowdhury, M. E., Botlero, R., Koblinsky, M., Saha, S. K., Dieltiens, G., and Ronsmans, C. (2007). Determinants of reduction in maternal mortality in Matlab, Bangladesh: a 30-year cohort study. *Lancet*, *370*, 1320-1328.

Gulshan, J., Chowdhury, R. I., Islam, M. A., and Akhter, H. H. (2005). GEE models for maternal morbidity in rural Bangladesh. *Austrian Journal of Statistics*, *34*, 295-304.

Islam, M. A., Chowdhury, R. I., Chakraborty, N., and Bari, W. (2004). A multistage model for maternal morbidity during antenatal, delivery and postpartum periods. *Statistics in Medicine*, *23*, 137-158.

Joe, H. (1997). *Multivariate Dependence Concept*. London: Chapman & Hall.

Joe, H., and Latif, A. H. M. M. (2005). Computations for familial analysis of binary traits. *Computational Statistics*, *20*, 439-448.

Liang, K.-Y., and Zeger, S. L. (1986). Longitudinal data analysis with generalized linear models. *Biometrika*, *73*, 13-22.

Linhart, L., and Zucchini, W. (1986). *Model Selection*. New York: John Wiley and Sons.

McCullagh, P., and Nelder, J. A. (1989). *Generalized Linear Models* (2nd ed.). London: Chapman & Hall.

Molenberghs, G., and Lesaffre, E. (1994). Marginal modeling of correlated ordinal data using a multivariate Plackett distribution. *Journal of the American Statistical Association*, *89*, 633-44.

Pan, W. (2001a). Akaike's information criterion in generalized estimating equations. *Biometrics*, *57*, 120-125.

Pan, W. (2001b). Model selection in estimating equations. *Biometrics*, *57*, 529-534.

Pan, W., and Lee, C. T. (2001). Bootstrap model selection in generalized linear models. *Journal of Agricultural, Biological & Environmental Statistics*, *6*, 49-61.

Zeger, S. L., and Liang, K.-Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, *42*, 121-130.

Zucchini, W. (2000). An introduction to model selection. *Journal of Mathematical Psychology*, *44*, 41-61.

Corresponding Author's current Address

A.H.M. Mahbub Latif
School of Mathematical Sciences
University of London, Queen Mary
Mile End Road
London E1 4NS
United Kingdom

E-Mail: mlatif@isrt.ac.bd