

**ON OPTIMAL GQL ESTIMATION APPROACH FOR LONGITUDINAL
DYNAMIC BINARY MODEL: APPLICATION TO THE MATERNAL
MORBIDITY DATA OF BANGLADESH**

Farzana Afroz* and Wasimul Bari

*Department of Statistics, Biostatistics & Informatics, University of Dhaka,
Dhaka-1000, Bangladesh*

Abstract

A longitudinal non-linear binary dynamic (NLBD) model which provides a full range for the longitudinal correlation parameter was attempted to analyze longitudinal morbidity data for determining the potential determinants of antenatal care (ANC) seeking behavior of pregnant women during their pregnancies. The regression and dependence parameters were estimated by optimal generalized quaslikelihood (OGQL) estimation approach. The estimate of dependence parameter was found to be significantly large. Age at marriage, economic status, husband's occupation, distance from health facility and numbers of previous pregnancies were found to have significant impact on the ANC seeking behavior.

Key words: Antenatal care, dependence parameter, longitudinal binary data, NLBD model, OGQL estimation approach

Introduction

Longitudinal binary and count data are of often interest in medical and public health sciences. Longitudinal studies play an important role in enhancing the understanding of the development and persistence of disease in health sciences. The distinguishing feature of longitudinal studies is that the study participants are measured repeatedly throughout the duration of the study. Hence, the direct assessment of changes in the response variable over time is possible. Suppose that there are T repeated discrete responses $y_{i1}, \dots, y_{it}, \dots, y_{iT}$ and a p dimensional vector of time dependent/time independent covariates $x_{it} = (x_{it1}, \dots, x_{itw}, \dots, x_{itp})'$ are observed from the i^{th} individual ($i = 1, \dots, K$) at time t ($t = 1, \dots, T$). In a longitudinal set up, usually $K \rightarrow \infty$ and T is small. Suppose that $\beta = (\beta_1, \dots, \beta_w, \dots, \beta_p)'$ is the $p \times 1$ vector of regression effects of x_{it} on y_{it} . Since repeated responses are collected from each individual, the responses are likely to be correlated. Therefore, one needs to take into account this correlation to estimate the regression parameter β . The likelihood estimation approach is cumbersome for discrete longitudinal data since the multivariate distribution of $y_i = (y_{i1}, \dots, y_{it}, \dots, y_{iT})'$ is complicated. To overcome this difficulty, Liang and Zeger (1986) proposed a 'working' correlation based generalized estimating equations (GEE) for the regression parameters. The GEE approach has many pitfalls as discussed by Crowder (1995) and Sutradhar and Das (1999). Crowder (1995) argued that the 'working' correlation parameter may not have any solution in many situations. Even if it exists, it may not

*Corresponding author: <farzana_bd87@yahoo.com>.

converge to the true correlation parameter. Moreover, Sutradhar and Das (1999) showed that this 'working' correlation based GEE approach may produce inefficient estimates as compared to the estimates obtained assuming that the repeated responses are uncorrelated. As a remedy, Sutradhar (2003) suggested to use a generalized autocorrelation matrix instead of 'working' correlation matrix in the GEE.

One way of avoiding the use of correlation parameter in the estimating equations of the regression parameters under the longitudinal set up is to define a dynamic model for repeated data. In the dynamic model, correlation among the repeated responses of an individual is explained by the dependence parameter γ . Note that this dependence parameter γ is considered as a regression parameter similar to β in the model. It appears as the regression parameter associated with y_{it-1} in the conditional expectation of Y_{it} for given $Y_{it-1} = y_{it-1}$, $t = 2, \dots, T$ as well as in the covariance between Y_{iu} and Y_{iv} , $u, v = 1, \dots, T$. Amemiya (1985) and Manski (1987) used non-linear binary dynamic (NLBD) model in their study with application to the econometric data. Qaqish (2003) introduced a general conditional linear binary dynamic (CLBD) model to analyze the binary dependent data. The limitation of this model is that the correlation parameter considered in the model does not cover the full range from -1 to 1. Sutradhar and Farrel (2007) extended the NLBD model for longitudinal data and proposed generalized quaslikelihood (GQL) and optimal generalized quaslikelihood (OGQL) estimation approaches for estimation of regression parameter as well as the dependence parameter. Because y_{it-1} is a different regression variable from the fixed effect covariate vector x_{it} , considering $y_i = (y_{i1}, \dots, y_{it}, \dots, y_{iT})'$ as a basic statistic to construct the GQL estimating equation does not appear to exploit sufficient information for the estimation of the model parameters, specially for γ . Thus to improve the GQL estimates of the parameters β and γ , Sutradhar and Farrel (2007) construct a new GQL estimating equation by using all possible distinct pair wise products and the first-order responses, instead of using only the first-order responses. This estimating approach is known as the optimal generalized quaslikelihood (OGQL) estimating approach. In this paper, the NLBD model for longitudinal data and the OGQL estimation approach are illustrated to the longitudinal data on maternal morbidity of Bangladesh collected by Bangladesh Institute of Research for Promotion of Essential and Reproductive Health and Technologies (BIRPERHT) (Akhter *et al.* 1996) during the period of November 1992 to December 1993. For the purpose of analysis, antenatal care (ANC) seeking behavior of a pregnant mother was considered as a response variable. Although the data set was collected during 1992-93, we are considering this because of two reasons: (a) this is the only follow-up data in Bangladesh on maternal morbidity and (b) though the ANC seeking behavior has increased, it is still low among the developing countries.

According to WHO, antenatal care includes recording medical history, assessment of individual needs, advice and guidance on pregnancy and delivery, screening tests, education on self-care during pregnancy, identification of conditions detrimental to health during pregnancy, first-line management and referral if necessary. Again WHO measures 'antenatal care' as the percentage of women who utilized antenatal care provided by skilled birth attendants for reasons

related to pregnancy at least once during pregnancy among all women who gave birth to a live child in a given time period. It is well established that ANC reduces remarkably the maternal morbidity and mortality during pregnancy as well as the infant mortality (WHO 1999). Bangladesh is one of the poorest countries where maternal mortality ratio is 194 per 100000 live births (BMMS 2010) and the infant mortality rate is 52 per 1000 live births (BDHS 2009). In the developing countries, 68% of the women receive at least one ANC visits from medically trained provider (WHO 2003), whereas, it is only 52% in Bangladesh (BDHS 2009). One of the challenges of Bangladesh to attain the millennium development goals (MDGs) is to improve the rate of receiving ANC substantially. For this reason, one may need to identify vulnerable groups and to provide them proper health care services. This paper is organized as follows. The NLBD model for longitudinal data and the OGQL estimation approach for the regression parameter β and the dependence parameter γ are briefly discussed in section 2. The model and estimation approach are illustrated to longitudinal morbidity data collected by BIRPERHT in section 3. In section 4, the conclusion and recommendation are given.

Non-linear binary dynamic (NLBD) model and OGQL estimation approach

Non-linear binary dynamic (NLBD) model:

The non-linear binary dynamic (NLBD) model for the longitudinal data can be written as follows

$$P(Y_{i1} = 1) = \mu_{i1}^*, \text{ with } \mu_{i1}^* = \frac{\exp(x'_{i1}\beta)}{1 + \exp(x'_{i1}\beta)}, \text{ and}$$

$$\lambda_{(i|t,t-1)}^* = P(Y_{it} = 1 | Y_{i,t-1}), \text{ with } \lambda_{(i|t,t-1)}^* = \frac{\exp(x'_{it}\beta + \gamma Y_{i,t-1})}{1 + \exp(x'_{it}\beta + \gamma Y_{i,t-1})}, t = 2, \dots, T;$$

where γ is the dynamic dependence parameter. This model produces the correlations with full ranges (-1 to 1). Under this set up, the mean of Y_{it} can be obtained as

$$\mu_{it} = E(Y_{it}) = Pr(Y_{it} = 1) = \mu_{it}^* + \mu_{i,t-1}(\bar{\mu}_{it} - \mu_{it}^*), \quad t = 2, \dots, T,$$

with $\bar{\mu}_{it} = \frac{\exp(x'_{it}\beta + \gamma)}{1 + \exp(x'_{it}\beta + \gamma)}$ and $\mu_{it}^* = \frac{\exp(x'_{it}\beta)}{1 + \exp(x'_{it}\beta)}$. Note that $\mu_{i1} = \mu_{i1}^*$. Then, the variance of Y_{it} is $\sigma_{it} = var(Y_{it}) = \mu_{it}(1 - \mu_{it})$. Since the covariance between Y_{it} and Y_{iu} , $u < t$, is given by

$$Cov(Y_{iu}, Y_{it}) = \sigma_{it} = \mu_{it}(1 - \mu_{it}) \prod_{j=u+1}^t (\bar{\mu}_{ij} - \mu_{ij}^*)$$

OGQL Estimation Approach

Following Sutradhar and Farrel⁸, optimal generalized quasi-likelihood (OGQL) estimating equations for $\theta = (\beta', \gamma)'$ can be written as

$$\sum_{i=1}^k \frac{\partial \xi_i'}{\partial \theta} \Omega_i^{-1} (f_i - \xi_i) = 0.$$

Here, $f_i = (y_i', s_i')$ with $y_i = (y_{i1}, \dots, y_{it}, \dots, y_{iT})'$ being a $T \times 1$ vector and $s_i = (y_{i1}y_{i2}, \dots, y_{it}y_{it+1}, \dots, y_{i(T-1)}y_{iT})'$ being a $\frac{T \times (T-1)}{2} \times 1$ vector. Then, $\xi_i = E[F_i] = (\mu_i', \rho_i')$, with $\mu_i = (\mu_{i1}, \dots, \mu_{it}, \dots, \mu_{iT})'$ and $\rho_i = (\rho_{i12}, \dots, \rho_{iuv}, \dots, \rho_{i(T-1)T})'$.

Note that $\rho_{iuv} = E(Y_{iu}Y_{iv}) = \sigma_{iuv} + \mu_{iu}\mu_{iv}$, $u < v$. The $\frac{T \times (T+1)}{2} \times \frac{T \times (T+1)}{2}$ variance-covariance matrix $\Omega_i = \text{Cov}(F_i)$ is defined as

$$\Omega_i = \begin{bmatrix} \Sigma_i & \psi_i \\ \psi_i' & \Delta_i \end{bmatrix},$$

with $\Sigma_i = \text{Cov}(Y_i)$, $\psi_i = \text{Cov}(Y_i, S_i)$ and $\Delta_i = \text{Cov}(S_i)$. The components of ψ_i and Δ_i can be computed by using the $\text{Cov}(Y_{iu}, Y_{iv}, Y_{it}) = E(Y_{iu}Y_{iv}Y_{it}) - E(Y_{iu})E(Y_{iv})E(Y_{it})$ and $\text{Cov}(Y_{iu}Y_{iv}, Y_{it}Y_{it}) = E(Y_{iu}Y_{iv}Y_{it}Y_{it}) - E(Y_{iu}Y_{iv})E(Y_{it}Y_{it})$ respectively.

Note that $E(Y_{iu}Y_{iv}Y_{it}) = P(Y_{iu} = 1, Y_{iv} = 1, Y_{it} = 1)$, $E(Y_{iv}Y_{it}) = P(Y_{iv} = 1, Y_{it} = 1)$ and $E(Y_{iu}Y_{iv}Y_{it}Y_{it}) = P(Y_{iu} = 1, Y_{iv} = 1, Y_{it} = 1, Y_{it} = 1)$.

The $(p+1) \times \frac{T \times (T+1)}{2}$ derivative matrix $\frac{\partial \xi_i'}{\partial \theta}$ is defined as $\frac{\partial \xi_i'}{\partial \theta} = \begin{bmatrix} \delta \mu_i' & \delta \rho_i' \end{bmatrix}$, where

$$\frac{\delta \rho_{iut}}{\delta \beta} = \sum_{Y_{iu}Y_{iv} \in S^*} \left[g_{i1}^* \prod_{j=2}^T g_{(i,j|j-1)}^* \right] \left\{ \sum_{k=2}^T (y_{ik} - \lambda_{(i,k|k-1)}) y_{ik} \left[+ (y_{i1} - \mu_{i1}) y_{i1} \right] \right\}_{Y_{iu}=1, Y_{iv}=1}$$

$$\frac{\delta \rho_{iut}}{\delta \gamma} = \sum_{Y_{iu}Y_{iv} \in S^*} \left[g_{i1}^* \prod_{j=2}^T g_{(i,j|j-1)}^* \right] \left\{ \sum_{k=2}^T (y_{ik} - \lambda_{(i,k|k-1)}) y_{i,k-1} \right\}_{Y_{iu}=1, Y_{iv}=1}$$

with $g_{i1}^* = \mu_{i1}^{y_{i1}} [1 - \mu_{i1}]^{1-y_{i1}}$ and $g_{(i,j|j-1)}^* = (\lambda_{(i,j|j-1)}^{y_{ij}}) (1 - \lambda_{(i,j|j-1, \dots, 1)}^{y_{ij}})^{1-y_{ij}}$.

\sum indicates the summation over all $y_{ij} = 0, 1$ for $j \neq u, t$.

An illustration: Longitudinal maternal morbidity data

For the purpose illustration of non-linear binary dynamic (NLBD) model for the longitudinal data, the prospective morbidity data set collected by BIRPERHT was used. In the baseline, 1020

pregnant women (pregnant for < 6 months) were selected and prospective subjects were followed-up with an interval on average of once a month, through full-term pregnancy, delivery and until 90 days postpartum or 90 days after any other pregnancy outcome. Information was collected on socioeconomic status, background, pregnancy-related care and practice, extent of morbidity during the index pregnancy, delivery, and postpartum period or abortions. In this paper, for computational simplicity, 719 pregnant women out of 1020 who provided responses in first three follow ups were selected. We considered the ANC seeking behavior (whether seeking or not) at each follow-up as dependent variable and considered age at marriage (<15 years, \geq 15 years), number of previous pregnancies (0, 1 - 4, \geq 5), husband's occupation (farmer, businessman, day laborer, other), economic status (poor, average, good), distance from health facility (< 1 mile, \geq 1 mile), and wanted pregnancy (yeas, no) as categorical covariates in our analysis as these variables were found as the potential factors for health service seeking behavior in previous studies (WHO 2003, Chakraborty *et al.* 2003)

At first, we conducted a bivariate analysis to examine the association between the selected covariates with the ANC seeking behavior by using the information obtained at the first follow-up. Results of this analysis are given in Table 1. It is clear from Table 1 that mothers who got married before 15 years of age took more antenatal care than the mothers who got married after 15 years of age. Again women who are becoming mother for the first time took more antenatal care than the

Table 1. Examining the association between antenatal care and selected variable.

Explanatory variables	Category	Antenatal care		Total
		Yes (%)	No (%)	
*** Age at marriage	< 15	128 (27.8)	332 (72.2)	460
	\geq 15	127 (23.8)	406 (76.2)	533
* Number of previous pregnancies	0	82 (30.5)	187 (69.5)	269
	1 - 4	137 (24.2)	430 (75.8)	567
	\geq 5	36 (22.9)	121 (77.1)	157
Mother's education	Uneducated	131 (24.1)	413 (75.9)	544
	Educated	124 (27.6)	325 (72.4)	449
** Husband's occupation	Farmer	100 (29.7)	237 (70.3)	337
	Businessman	101 (25.3)	298 (74.7)	399
	Day Laborer	35 (18.7)	152 (81.3)	187
	Other	19 (27.1)	51 (72.9)	70
*** Economic status	Poor	43 (17.9)	197 (82.1)	240
	Average	136 (24.6)	417 (75.4)	553
	Good	76 (38.0)	124 (62.0)	200
* Distance from health facility	<1 mile	117 (27.9)	303 (72.1)	420
	\geq 1 mile	138 (24.1)	435 (75.9)	573
** Wanted pregnancy	No	69 (22.0)	245 (78.0)	314
	Yes	186 (27.4)	493 (72.6)	679
Total		255	738	993

* $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

other groups. Educated mothers were more likely to receive antenatal care than uneducated mother. Antenatal taking behavior was the highest for the farmer's wives and it was the lowest for the day laborer's wives. Mothers from the families of good economic condition took more

antenatal care than others groups. Mothers who lived more than one mile far from the health care center (HCC) were less likely to receive ANC compared with those who lived within one mile distance. The tendency of taking antenatal care was higher for the women who wanted the index pregnancy. Note that all variables except the mother's education were found to have significant effects on ANC taking behavior.

For the purpose of regression analysis, the longitudinal non-linear binary dynamic (NLBD) model was fitted and the optimal generalized quaslikelihood (OGQL) estimation approach to estimate the parameters involved in the models. In NLBD model, variables which have been found to be significant in bivariate analysis were only included. The results are shown in Table 2.

Table 2. OGQL estimates of the NLBD model parameters for longitudinal morbidity data.

Variable	$\hat{\beta}$	S.E.	Wald test	p-value
Age at marriage (Ref < 15)				
≥15	-0.6936***	0.1063	-6.5255	0.0000
Number of previous pregnancies (Ref = 0)				
≥5	-1.0488***	0.1730	-6.0633	0.0000
1-4	-0.5256***	0.1080	-4.8686	0.0000
Husband's occupation (Ref = Farmer)				
businessman	-0.5713***	0.1138	-5.0189	0.0000
Day laborer	-0.1886	0.1385	-1.3621	0.1732
Other	-0.5529**	0.2315	-2.3881	0.0169
Economic status (Ref = Poor)				
Good	0.3872**	0.1539	2.5168	0.0118
Average	-0.0988	0.1204	-0.8208	0.4118
Distance from health facility (Ref < mile)				
≥ 1 mile	-0.1850*	0.1020	-1.8144	0.0696
Wanted pregnancy (Ref = No)				
Yes	-0.0208	0.1072	-0.1938	0.8463
Dependence parameter (γ)	2.2224***	0.1227	18.1058	0.0000

* p < 0.10; ** p < 0.05; *** p < 0.01.

As far as estimation of dependence parameter is concerned, we found that $\hat{\gamma} = 2.2224$, which was highly significant (p value = 0.0). This indicates that there exists dependence in the data. Therefore, one requires taking the dependence into account for the estimation of the regression parameters. It is clear from Table 2 that receiving antenatal care was more likely for the women who got married early. This result was found statistically significant (p value = 0.0). This is not surprising because young age at marriage leads to various complications during pregnancy. Women with increasing previous pregnancy experience were less likely to seek antenatal care. Note that, the results were found to be statistically significant. Husband's occupation seemed to be an important determinant of antenatal care receive, indicating that wives of businessman and others were less likely to seek antenatal care compared to the wives of farmers. However, the

wives of businessman were the most sufferers and this result was also statistically significant (p value = 0.0). Women from families of good economic condition were more likely to receive antenatal care from doctor or nurse than those whose economic condition was poor. The positive impact of good economic status on health care use was statistically significant (p value = 0.01). Women whose house were far (≥ 1 mile) from health facility center are less likely to receive antenatal care with compared to the women who were nearer to the health facility center (< 1 mile). However, this result is found to be significant at 10% level of significance. There was a negative association between wanted pregnancy and receiving ANC and this result was found statistically insignificant.

Conclusion and Recommendations

An attempt has been made to employ the longitudinal non-linear binary dynamic (NLBD) model for analyzing the prospective maternal morbidity data collected by BIRPERHT. To estimate the regression and dependence parameter involved in the model, the optimal generalized quaslikelihood (OGQL) was used as it provides consistent and efficient estimators which statistically sufficient. To be specific, potential determinants of the ANC seeking behavior during the pregnancy period were determined through the longitudinal NLBD model and OGQL estimation approach. The estimate of dependence parameter was found to be significantly large. It indicates that the prospective data is correlated and it is necessary to take this dependency into account for the estimation of regression parameters. We found that age at marriage, numbers of previous pregnancies, husband's occupation, economic status and distance from health facility have significant impact on the ANC seeking behavior.

It is found that age at marriage of women plays an important role in ANC taking behavior. The rate of ANC taking behavior increases as the age of marriage of women increases. Though, in Bangladesh, the minimum legal age at marriage for girls is 18 years, the rate of marriage before the age of 18 is still 66% and over one third of girls are married before the age of 15. It is now essential to implement this law to increase the rate of ANC seeking behavior. An important key finding of this study is that the economic status of women is a major factor in their health care seeking behavior for taking ANC. Women from families of good economic status are more likely to receive antenatal care than those whose economic condition is poor. In families with low income level, pregnant women cannot afford to consult qualified doctors for their physical problems and illness. If women are financially solvent, it is expected that they will be able to make decisions regarding their healthcare services. Income-generating work opportunities need to be created for women so that they can be financially independent to make decisions and to seek better health care services. Again free treatment can be provided to every pregnant woman.

Distance from health facility is found to be an influential factor for taking ANC. In this regard, basic healthcare facilities need to be available in rural areas. Government and non-government organizations should take initiative in this purpose. The number of previous pregnancies is also an important factor for women's antenatal taking behavior. In this regard, education and awareness

programs have positive impacts on the motivation for fertility control and healthcare seeking behavior.

In conclusion, it is recommended that an integrated approach to increase women's educational level, create more income-generating opportunities for them, and provide basic healthcare facilities in rural areas of Bangladesh should be taken. A proper government monitoring system can make present situation better.

References

- Liang, K.Y. and S.L. Zeger. 1986. Longitudinal data analysis using generalized linear models. *Biometrika* 13-22.
- Crowder, M. 1995. On the use of a working correlation matrix in using generalized linear models for repeated measures. *Biometrika* **82**: 407-410.
- Sutradhar, B.C. and K. Das. 1999. On the efficiency of regression estimators in generalized linear models for longitudinal data. *Biometrika* 459-65.
- Sutradhar, B.C. 2003. An overview of regression models for discrete longitudinal responses. *Statistical Science* 377-93.
- Amemiya, T. 1985. *Advanced Econometrics*, Harvard University Press, Cambridge, UK.
- Manski, C.F. 1987. Semi-parametric analysis of random effects linear models from binary data. *Econometrica* 357-362.
- Qaqish, B.F. 2003. A family of multivariate binary distributions for simulating correlated binary variables with specified marginal means and correlations. *Biometrika* 455-63.
- Sutradhar, B.C. and P.J. Farrel. 2007. On optimal lag 1 dependence estimation for dynamic binary models with application to asthma data. *The Indian Journal of Statistics* 448-47.
- Akhter, H.H., M.E.E.K. Chowdhury and A. Sen. 1996. A cross sectional study on maternal morbidity in Bangladesh. BIRPERHT, Dhaka, Bangladesh.
- World Health Organization, 1999. Reduction of maternal mortality: a joint WHO/UNFPA/ UNICEF/ World Bank statement. Geneva.
- Bangladesh Maternal Mortality and Health Care Survey (BMMS). 2010. Summary of key findings and implications.
- Bangladesh Demographic and Health Survey, 2007, 2009. National Institute of Population Research and Training (NIPORT). Calverton, Maryland and Dhaka, NIPORT, Mitra and Associates and Macro International Inc.
- World Health Organization. 2003. Antenatal care in developing countries promises, achievements and missed opportunities: an analysis of trends, levels and differentials, 1990-2001.
- Chakraborty, N., M.A. Islam, R.I. Chowdhury, W. Bari and H.H. Akhter. 2003. Determinants of the use of maternal health services in rural Bangladesh. *Health Promotion International* **18**: 327-337.

(Manuscript received on 12 December, 2013; revised on 27 December, 2013)