

# A Model for Assessing the Association in the Repeated Measures of Depression among the Elderly

M. Ataharul Islam<sup>a</sup>, Rafiqul. I. Chowdhury<sup>b</sup>, S. Bae<sup>c</sup>, Karan. P. Singh<sup>c</sup>

<sup>a</sup>Department of Applied Statistics, East West University, Aftabnagar, Dhaka-1212, Bangladesh

<sup>b</sup>Department of Epidemiology and Biostatistics, Western University, London, ON, Canada

<sup>c</sup>Division of Preventive Medicine and the Comprehensive Cancer Center, University of Alabama at Birmingham, Medical Towers Building, Suite 647A, 1717 11th Avenue South Birmingham, Alabama

35294-4410, USA

Email: [mataharul@yahoo.com](mailto:mataharul@yahoo.com)

**Abstract:** The dependence in the outcome variables is a major issue of concern in modeling the correlated data stemmed from the repeated observations. The marginal models such as GEE and the conditional models based on Markov chain have been employed for longitudinal data in the past. However, it has been evident that without addressing the underlying association parameters, the analysis of repeated outcome variables remains far from being resolved. In this paper, a method has been demonstrated to model such data using the underlying dependence in the outcome variables as well as dependence between outcome and explanatory variables. An extension of the regressive model is shown in this paper and a comparison is demonstrated between the existing model (reduced model) and the proposed model (extended model). The models are illustrated for depression among the elderly population in the USA using the Health and Retirement Study data from 1992 to 1998.

**Keywords:** *Bivariate binary outcomes, Conditional model, Joint model, Marginal model, Regressive model, Transition probability*

## 1 INTRODUCTION

The outcome variables generated from longitudinal studies are generally correlated. To resolve the problem of dependence in outcome variables over time, we can employ marginal model like GEE or conditional models based on the Markov chain or Markov process. There are limitations in those techniques due to either correlation structure or number of parameters to be estimated. In this paper, the regressive model approach is extended to address such problem.

The modeling of correlated binary outcomes has been discussed in Bonney (1987), Prentice (1988), Zeger and Qaqish (1988), Neuhaus et al. (1991), Liang et al. (1992), McDonald (1993), le Cessie and van Houwelingen (1994), and Solis-Trapala et al. (2007). Most of the previous works for different association measures of dependence were based on the marginal response probabilities. Muenz and Rubinstein (1985), Bonney (1986, 1987), Azzalini (1994), Islam and Chowdhury (2006, 2007, 2010) and Islam et al. (2009) employed the conditional regressive logistic models under the Markov assumptions. Edwards (2000) provided an approach of graphical modeling where the Simpson's paradox is illustrated and the log-linear model is used to take into account interactions among the factors for assessing dependence. We can generalize the works of Bonney (1986, 1987), Islam and Chowdhury (2006, 2007, 2010) and Islam et al. (2009) to include both binary outcomes in previous times as well as covariates in the conditional models. The joint mass function for two outcome variables  $Y_1$  and  $Y_2$  at follow-ups 1 and 2, respectively, in the presence of covariates  $X_1, \dots, X_p$ , and let  $\mathbf{X} = (1, X_1, \dots, X_p)$ , can be expressed as product of the conditional and marginal probability mass functions for given values of covariates as follows:

$$P_{Y_1, Y_2 | \mathbf{X}}(y_1, y_2 | \mathbf{x}) = P_{Y_1 | \mathbf{X}}(y_1 | \mathbf{x}) \times P_{Y_2 | Y_1, \mathbf{X}}(y_2 | y_1, \mathbf{x}) \quad (1)$$

where  $P_{Y_1, Y_2 | \mathbf{X}}(y_1, y_2 | \mathbf{x})$  is the joint mass function for  $Y_1$  and  $Y_2$ ,  $P_{Y_1 | \mathbf{X}}(y_1 | \mathbf{x})$  is the marginal mass function for  $Y_1$ ,  $P_{Y_2 | Y_1, \mathbf{X}}(y_2 | y_1, \mathbf{x})$  is the conditional probability for  $Y_2$  given  $Y_1 = y_1$ ,  $\mathbf{x} = (1, x_1, \dots, x_p)'$ ,  $i = 1, 2, \dots, n$ , is the vector of covariate values, and  $y_j, j = 1, 2$  is the value of the outcome variable at the  $j$ -th follow-up.

Bonney (1987) proposed a regression model for the conditional probabilities as shown below:

$$P_{Y_2 | Y_1, \mathbf{X}}(y_2 | y_1; \mathbf{x}) = \frac{e^{\theta y_2}}{1 + e^{\theta}}, \quad (2)$$

where  $\theta = \gamma_0 + \gamma_1 x_1 + \dots + \gamma_p x_p + \beta_1 y_1$ ,  $\gamma_0$  is the intercept,  $\beta_0$  is the coefficient of the previous outcome,  $Y_1$ , and  $\gamma_1, \dots, \gamma_p$  are the coefficients of the covariates  $X_1, \dots, X_p$ , respectively. Here  $\theta$  is the logit defined as

$$\theta = \ln \frac{P_{Y_2 | Y_1, \mathbf{X}}(Y_2 = 1 | y_1, \mathbf{x})}{P_{Y_2 | Y_1, \mathbf{X}}(Y_2 = 0 | y_1, \mathbf{x})}.$$

## 2 TEST FOR DEPENDENCE: AN EXTENDED REGRESSIVE APPROACH

Let us consider the following regressive model for the  $j$ -th follow-up ( $j = 1, \dots, J$ ) with two binary outcomes to test for the dependence in the outcome variables as well as between the covariates and the outcome variables. Let us define  $Y_j = s, s = 0, 1$  at follow-up  $j = 1, 2$ . Then the model with prior outcome and  $p$  covariates is (Islam et al., 2012):

$$P_{Y_2 | Y_1, \mathbf{X}, \mathbf{Z}}(Y_2 = s | y_1, \mathbf{x}, \mathbf{z}) = \frac{e^{(\gamma' \mathbf{x} + \beta_1 y_1 + \eta' \mathbf{z})s}}{1 + e^{(\gamma' \mathbf{x} + \beta_1 y_1 + \eta' \mathbf{z})s}}, s = 0, 1 \quad (3)$$

where  $\mathbf{x} = (1, x_1, \dots, x_p)'$ ,  $\gamma' = (\gamma_0, \gamma_1, \dots, \gamma_p)$ ,  $\beta_1$  is the parameter corresponding to  $Y_1$ ,  $\mathbf{z} = (z_1, \dots, z_p)' = (x_1 y_1, \dots, x_p y_1)'$ ,  $\eta' = (\eta_1, \dots, \eta_p)$ . It may be noted here that this model is an extension of (2) as it contains interaction terms  $\mathbf{Z} = (Z_1, \dots, Z_p)'$ . We can extend this further considering  $Y_j = s, s = 0, 1$  at follow-up  $j = 1, 2, \dots, J$  as follows:

$$P_{Y_j | \mathbf{X}, \mathbf{Y}_{j-1}, \mathbf{V}_{j-1}, \mathbf{Z}}(Y_j = s | \mathbf{x}, \mathbf{y}_{j-1}, \mathbf{v}_{j-1}, \mathbf{z}_{j-1}) = \frac{e^{(\gamma' \mathbf{x} + \beta'_{j-1} \mathbf{y}_{j-1} + \rho'_{j-1} \mathbf{v}_{j-1} + \eta'_{j-1} \mathbf{z}_{j-1})s}}{1 + e^{(\gamma' \mathbf{x} + \beta'_{j-1} \mathbf{y}_{j-1} + \rho'_{j-1} \mathbf{v}_{j-1} + \eta'_{j-1} \mathbf{z}_{j-1})s}}, s = 0, 1 \quad (4)$$

where  $\mathbf{x} = (1, x_1, \dots, x_p)'$ ,  $\gamma' = (\gamma_0, \gamma_1, \dots, \gamma_p)$ ,  $\mathbf{y}_{j-1} = (y_1, \dots, y_{j-1})$ ,  $\beta_{j-1} = (\beta_1, \dots, \beta_{j-1})$  is the vector of parameters corresponding to  $\mathbf{Y}_{j-1} = (Y_1, \dots, Y_{j-1})$ ,  $\mathbf{v}_{j-1} = (v_{12}, \dots, v_{12 \dots j-1})' = (y_1 y_2, \dots, y_1 y_2 \dots y_{j-1})'$ ,

$\rho' = (\rho_{12}, \dots, \rho_{12\dots j-1})$ ,  $\mathbf{z}_{j-1} = (z_{11}, \dots, z_{1p}, \dots, z_{j-11}, \dots, z_{j-1p})' = (x_1 y_1, \dots, x_p y_1, \dots, x_1 y_{j-1}, \dots, x_p y_{j-1})'$ ,  $\eta' = (\eta_{11}, \dots, \eta_{1p}, \dots, \eta_{j-11}, \dots, \eta_{j-1p})$ . It may be noted here that this model is an extension of (3) as it contains interaction terms  $\mathbf{z}_{j-1} = (z_{11}, \dots, z_{1p}, \dots, z_{j-11}, \dots, z_{j-1p})'$ .

Let  $\lambda'_{j-1} = (\gamma', \beta'_{j-1}, \rho'_{j-1}, \eta'_{j-1})$ , and  $\mathbf{W}'_{j-1} = (\mathbf{X}', \mathbf{Y}'_{j-1}, \mathbf{V}'_{j-1}, \mathbf{Z}'_{j-1})$ . Then

$$P_{Y_j | \mathbf{W}_{j-1}} (Y_j = s | \mathbf{w}_{j-1}) = \frac{e^{(\lambda'_{j-1} \mathbf{w}_{j-1})s}}{1 + e^{(\lambda'_{j-1} \mathbf{w}_{j-1})s}}, s = 0, 1 \tag{5}$$

and if  $\theta_{j-1} = (\lambda'_{j-1} \mathbf{w}'_{j-1})$ , then

$$P_{Y_j | \mathbf{W}_{j-1}} (Y_j = s | \mathbf{w}_{j-1}) = \frac{e^{\theta_{j-1}s}}{1 + e^{\theta_{j-1}s}}, s = 0, 1. \tag{6}$$

The log likelihood function is

$$l = \sum_{i=1}^n [Y_{ji} \theta_{j-1,i} - \log(1 + e^{\theta_{j-1,i}})] \tag{7}$$

and the score function is

$$U = \sum_{i=1}^n [Y_{ji} - (1 + e^{\theta_{j-1,i}})] \left( \frac{\partial \theta_{j-1,i}}{\partial \lambda} \right) \tag{8}$$

The observed information can be obtained as follows:

$$-\frac{\partial^2 l(\lambda)}{\partial \lambda \partial \lambda'} = \sum_{i=1}^n \frac{e^{\theta_{j-1,i}(\lambda)}}{[1 + e^{\theta_{j-1,i}(\lambda)}]^2} \left( \frac{\partial \theta_{j-1,i}}{\partial \lambda} \right) \left( \frac{\partial \theta_{j-1,i}}{\partial \lambda} \right)' \tag{9}$$

From the models 3 and 4, it appears that if the interaction terms are ignored, the test for independence based on the previous outcomes alone may result in misleading conclusions. Hence, under independence, the model (4) can be defined as:

$$P_{Y_j | \mathbf{X}, \mathbf{Y}_{j-1}, \mathbf{V}_{j-1}, \mathbf{Z}} (Y_j = s | \mathbf{x}, \mathbf{y}_{j-1}, \mathbf{v}_{j-1}, \mathbf{z}_{j-1}) = \frac{e^{(\gamma' \mathbf{x})s}}{1 + e^{\gamma' \mathbf{x}}}, s = 0, 1. \tag{10}$$

We need to test for  $H_0 : \lambda^*_{j-1} = \mathbf{0}$  against  $H_1 : \lambda^*_{j-1} \neq \mathbf{0}$  for model (5) which is denoted as full model and model is the reduced model under the null hypothesis, where  $\lambda^*_{j-1} = (\beta_{j-1}, \rho_{j-1}, \eta_{j-1})'$  and the total number of parameters need to be tested is  $(2^{j-1} - 1)$  for  $\beta_{j-1}$  and  $\rho_{j-1}$  and  $(j - 1)p$  parameters for  $\eta_{j-1}$ . This test can be performed using the likelihood ratio and the test statistic follows chi-square asymptotically with  $(2^{j-1} - 1) + (j - 1)p$  degrees of freedom.

### 3 APPLICATION TO DEPRESSION DATA

For this study, the panel data from the Health and Retirement Study (HRS), sponsored by the National Institute of Aging (grant number NIA U01AG09740) and conducted by the University of Michigan (HRS, 2002), have been used. The respondents of this study are individuals over age 50 and their spouses. We have used the panel data from the four rounds or waves for 1992, 1994, 1996 and 1998. The depression index is computed using the scale proposed by the Center for Epidemiologic Studies on Depression (CESD). Severity of the emotional health can be measured from the CESD score. We have used 9761 respondents for analyzing depression among the elderly in the USA during 1992-1998. Many studies on clinical and non-clinical populations, CESD scale is employed to measure depressive symptoms (Steffick, 2000). Cheng et al. (2006) demonstrated the utility and validity of a short version of the CESD scale. The elderly population may suffer from repeated spells of depression resulting in fatality in some cases (Wasylenki, 1980). Depression is prevalent among the elderly and in many instances difficult to diagnose. Evans and Mottram (2000) observed that the spells of depression may change over time. A third of the minor depression patients may develop major depression over time, and a half of those with major depression may suffer from minor depression after recovery.

We considered the following dependent and explanatory variables: depression status (no depression (CESD score 0) = 0, depression (CESD score >0) = 1), we may denote  $Y_1$  = depression status at 1992,  $Y_2$  = depression

status at 1994,  $Y_3$ = depression status at 1996, and  $Y_4$ = depression status at 1998; gender (male=1, female=0), marital status (married/partnered =1, single/widowed/divorced=0), ethnic group (white=1, else 0; black=1, else 0; others= reference category), age, drinking habit (drink=1, does not drink=0), number of health conditions and total income.

Table 1 shows the fit of the full and reduced models. The full model includes: (i) the potential risk factors age, gender, marital status, white and black ethnic groups, drinking habit of alcohol, and number of conditions regarding previous health problems, (ii) previous outcomes, (iii) interaction between variables in (i) and (ii), interaction between variables in (ii). The reduced model takes into account the variables in (i) and (ii). In both the models, we observe that marital status, number of conditions and previous outcome CESD 2 (from wave 2) are significantly associated with depression status in wave 4 ( $Y_4$ ). In addition, the reduced model demonstrates association with gender, marital status, total income, CESD 1 and CESD 3 while the full model shows association with age and marital status. The full model shows that gender, marital status and white race, as compared to other races, are negatively associated with depression and the model is significant (p-value<0.001) in the reduced model but gender appears to be non-significant in the full model. In the full model, it is observed that some of the interaction terms such as CESD 1 and age (p-value<0.10), CESD 1 and white race (p-value=0.055), CESD 1 and black race (p-value < 0.01), CESD 3 and number of conditions (p-value=0.051), CESD 3 and total income (p-value <0.05) and CESD 1 and CESD 3 (p-value <0.01). This finding confirms our assertion that the reduced model which is used extensively fails to specify the model adequately.

The likelihood ratio test is performed for the hypothesis  $H_0 : \lambda^*_{j-1} = 0$  against  $H_1 : \lambda^*_{j-1} \neq 0$  for model (5) and the value of chi-square is 51.497 (p-value = 0.003). This indicates that the full model with interaction terms is more specified and without the interaction terms there is a possibility of coming up with misleading results for analyzing the dependence in outcome variables which depend on the association between previous outcomes and potential risk factors and among the previous outcomes as well. The results demonstrate that some of the interactions are statistically significant and hence the full model is preferred rather than the reduced model.

#### 4 CONCLUDING REMARKS

In a regressive models both the covariates and the previous outcomes are included. In this paper, an extended regressive model is proposed to reveal the underlying relationships between outcome variables, relationship between previous outcomes and potential risk factors and the association between pairs of previous outcomes. The proposed and reduced models are fitted employing the longitudinal depression data using the proposed full model and the existing reduced model. The results indicate that there are statistically significant interaction terms which reveal the dependence not only among the outcome variables but also between previous outcomes and also between previous outcomes and the potential risk factors. The proposed model provides more specified model and the results are more meaningful to the policymakers.

#### ACKNOWLEDGEMENT

The authors acknowledge gratefully to the HRS (Health and Retirement Study) which is sponsored by the National Institute of Aging (grant number NIA U01AG09740) and conducted by the University of Michigan.

**Table 1.** Estimates of parameters of the proposed and the reduced models for the status on depression among elderly population

Parameter	Reduced Model			Full Model		
	Estimate	Standard Error	p-value	Estimate	Standard Error	p-value
<b>Intercept</b>	-0.889	0.399	0.026	-1.446	0.622	0.020
<b>Age (wave I)</b>	0.008	0.006	0.228	0.023	0.010	0.022
<b>Sex</b>	-0.105	0.056	0.061	-0.107	0.084	0.202
<b>Marital status (wave I)</b>	-0.222	0.065	0.001	-0.246	0.104	0.018
<b>White</b>	-0.101	0.157	0.519	-0.336	0.256	0.190
<b>Black</b>	0.082	0.170	0.628	-0.278	0.281	0.322
<b>Drink (wave I)</b>	-0.065	0.058	0.260	-0.081	0.087	0.348
<b># of cond. (wave I)</b>	0.143	0.027	0.0001	0.132	0.044	0.003
<b>Total income (wave I)</b>	0.000002	0.000001	0.0001	-0.000001	0.000001	0.178
<b>CESD1 (wave I)</b>	0.567	0.062	0.0001	1.099	0.857	0.200
<b>CESD2 (wave II)</b>	0.924	0.059	0.0001	2.174	0.843	0.010
<b>CESD3 (wave III)</b>	1.137	0.058	0.0001	0.473	0.833	0.571
<b>CESD1*Age</b>				-0.026	0.014	0.060
<b>CESD2*Age</b>				-0.019	0.014	0.158
<b>CESD3*Age</b>				0.007	0.014	0.601
<b>CESD1*Sex</b>				-0.053	0.127	0.677
<b>CESD2*Sex</b>				-0.055	0.120	0.647
<b>CESD3*Sex</b>				0.109	0.118	0.360
<b>CESD1*Marital status</b>				0.186	0.145	0.198
<b>CESD2*Marital status</b>				-0.132	0.138	0.341
<b>CESD3*Marital status</b>				0.052	0.136	0.704
<b>CESD1*White</b>				0.631	0.329	0.055
<b>CESD2*White</b>				-0.025	0.326	0.938
<b>CESD3*White</b>				-0.046	0.324	0.888
<b>CESD1*Black</b>				0.959	0.359	0.008
<b>CESD2*Black</b>				-0.027	0.354	0.939
<b>CESD3*Black</b>				-0.038	0.351	0.914
<b>CESD1*Drink</b>				0.109	0.130	0.402
<b>CESD2*Drink</b>				-0.161	0.123	0.191
<b>CESD3*Drink</b>				0.121	0.123	0.325
<b>CESD1*# of cond.</b>				-0.036	0.057	0.526
<b>CESD2*# of cond.</b>				-0.039	0.057	0.494
<b>CESD3*# of cond.</b>				0.112	0.057	0.051
<b>CESD1* Total income</b>				0.000	0.000	0.154
<b>CESD2* Total income</b>				0.000	0.000	0.280
<b>CESD3* Total income</b>				0.000	0.000	0.034
<b>CESD1* CESD2</b>				0.109	0.128	0.394
<b>CESD1* CESD3</b>				0.408	0.129	0.002
<b>CESD2* CESD3</b>				0.045	0.119	0.706
<b>-2 Log L</b>		8057.884			8006.387	
<b>Likelihood Ratio (DF, p-value)</b>		1720.1421			1771.64	
		(11,			(38,p<0.01)	
		p<0.01)				

## REFERENCES

- Azzalini, A. (1994). Logistic regression for autocorrelated data with application to repeated measures. *Biometrika* 81, 767–775.
- Bonney, G. E. (1986). Regressive logistic models for familial disease and other binary trials. *Biometrics* 42, 611–625.
- Bonney, G. E. (1987). Logistic regression for dependent binary observations. *Biometrics* 43, 951–973.
- Cheng, S. T., A. C. M. Chan, and H. H. Fung (2006). Factorial structure of a short version of the center for epidemiologic studies depression scale. *International Journal of Geriatric Psychiatry* 21, 333–336.
- Edwards, D. (2000). *Introduction to Graphical Modelling (2nd Edition)*. Springer, New York.
- Evans, M. and P. Mottram (2000). Diagnosis of depression in elderly patients. *Advances in Psychiatric Treatment* 6, 49–56.
- HRS (2002). *Health And Retirement Study, (Wave [14]/Year[1992-1998]) public use dataset. Produced and distributed by the University of Michigan with funding from the National Institute on Aging (grant number NIAU01AG09740)*. Ann Arbor, MI.
- Islam, M. A. and R. I. Chowdhury (2006). A higher-order Markov model for analyzing covariate dependence. *Applied Mathematical Modelling* 30, 477–488.
- Islam, M. A. and R. I. Chowdhury (2007). *First and higher order transition models with covariate dependence. In Progress in Applied Mathematical Modeling, F. Yang (ed), pp. 153-198*. Nova Science, New York.
- Islam, M. A. and R. I. Chowdhury (2010). Prediction of disease status: A regressive model approach for repeated measures. *Statistical Methodology* 7, 520–540.
- Islam, M. A., R. I. Chowdhury, and A. Alzaid (2012). Tests for dependence in binary repeated measures data. *Journal of Statistical Research* 46, 203–217.
- Islam, M. A., R. I. Chowdhury, and S. Huda (2009). *Markov Models with Covariate Dependence for Repeated Measures*. Nova Science, New York.
- le Cessie, S. and J. C. van Houwelingen (1994). Logistic regression for correlated binary data. *Applied Statistics* 43, 95–108.
- Liang, K. Y., S. L. Zeger, and B. Qaqish (1992). Multivariate regression-analyses for categorical data. *Journal of the Royal Statistical Society, Series B* 54, 3–40.
- McDonald, W. B. (1993). Estimating logistic regression parameters for bivariate binary data. *Journal of the Royal Statistical Society, Series B* 55, 391–397.
- Muenz, L. R. and L. V. Rubinstein (1985). Markov models for covariate dependence of binary sequences. *Biometrics* 41, 91–101.
- Neuhaus, J. M., J. D. Ktalbfleisch, and W. A. Hauck (1991). A comparison of duster-specific and population-averaged approaches for analyzing correlated binary data. *International Statistical Review* 59, 25–35.
- Prentice, R. L. (1988). Correlated binary regression with covariate specific to each binary observation. *Biometrics* 44, 1033–1048.
- Solis-Trapala, I. L., J. Carthey, V. T. Farewell, and M. R. L. de Leval (2007). Dynamic modelling in a study of surgical error management. *Statistics in Medicine* 26, 5189–5202.
- Steffick, D. E. (2000). *Documentation of Affective Functioning Measures in the Health and Retirement Study*. Survey Research Center University of Michigan Ann Arbor, MI.
- Wasylenki, D. (1980). Depressions in the elderly. *CMA Journal* 122, 525–532.
- Zeger, S. L. and B. Qaqish (1988). Markov regression models for time series: a quasi-likelihood approach. *Biometrics* 44, 1019–1031.