*Journal of Statistical Research* *ISSN 0256 - 422 X*
*2012, Vol. 46, No. 2, pp. 203-217*

# TESTS FOR DEPENDENCE IN BINARY REPEATED MEASURES DATA

M ATAHARUL ISLAM

*Department of Statistics and OR, King Saud University, Saudi Arabia*

*Email: mataharul@yahoo.com*

RAFIQUL I CHOWDHURY

*Department of Epidemiology and Biostatistics, University of Western Ontario, London, ON, Canada*

*Email: mchowd23@uwo.ca*

ABDULHAMID ALZAID

*Department of Statistics and OR, King Saud University, Saudi Arabia*

*Email: alzaid@ksu.edu.sa*

SUMMARY

If we observe repeated binary outcomes over time then there may be dependence in outcomes and a test for dependence may be sought for such data. However, tests for dependence in models for repeated measures remain a challenge where covariates are associated with previous outcomes and both covariates and previous outcomes are included simultaneously in a model. This paper displays the nature of such problems (i.e. dependence among outcomes may depend on the association between covariates and previous outcomes) inherent in models for repeated binary outcomes that can distort the estimates and may produce misleading results. In the context of application of regressive models, this paper discusses conditions for which the regressive models can be safely employed. All these are shown on the basis of simple relationships between the conditional, marginal and joint probability mass functions for the bivariate binary outcomes which can be extended to the multivariate data stemmed from repeated measures. Some test procedures are suggested and applications are demonstrated using both simulations and real life data. Both the applications clearly indicate the utility of the proposed tests.

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

*AMS Classification:* 62H20, 62F03, 62F10, 62N03

# 1 Introduction

It is evident from the studies on repeated measures that there may be dependence in repeated outcomes over time and the problem becomes more complex if the dependence of outcomes depends on the explanatory variables as well. This problem still poses a formidable difficulty to both researchers and potential users in various disciplines. This paper proposes a simple way to understand the mechanism through which appropriate models can be specified. Although the relationship is displayed for the bivariate binary outcomes for the purpose of illustration, it can be generalized for a model of any order.

The modeling of correlated binary outcomes has been discussed in Bonney [1], Prentice [2], Zeger and Qaqish [3], Neuhaus et al. [4], Liang et al. [5], McDonald [6], le Cessie and van Houwelingen [7], and Solis-Trapala et al. [8]. Most of the previous works for different association measures of dependence were based on the marginal response probabilities. Muenz and Rubinstein [9], Bonney [1, 10], Azzalini [11], Islam and Chowdhury [12-14] and Islam et al. [15] employed the conditional regressive logistic models under the Markov assumptions. Edwards [16] 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. The works of Bonney [1, 10], Islam and Chowdhury [12-14] and Islam et al. [15] can be generalized 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 $X_1, \ldots, X_p$, and let $\mathbf{X} = (1, X_1, \ldots, X_p)'$, can be expressed as product of the conditional and marginal probability mass functions for given values of covariates as follows:

$$P_{Y_1,Y_2|\mathbf{X}}(y_1,y_2\,|\mathbf{x}) = P_{Y_1|\mathbf{X}}(y_1\,|\mathbf{x}) \times P_{Y_2|Y_1,\mathbf{X}}(y_2\,|y_1,\mathbf{x}), \tag{1.1}$$

where $P_{Y_1,Y_2|\mathbf{X}}(y_1, y_2\,|\mathbf{x})$ is the joint mass function for $\mathrm{Y}_1$ and $\mathrm{Y}_2$, $P_{Y_1|\mathbf{X}}(y_1\,|\mathbf{x})$is the marginal mass function for $\mathrm{Y}_1$, $P_{Y_2|Y_1,\mathbf{X}}(y_2\,|y_1,\mathbf{x})$ is the conditional probability for $Y_2$ given $Y_1 = y_1$, $\mathbf{x} = (1, x_1, \ldots, x_p)'$, $i = 1, 2, \ldots n$, 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 [1] proposed a regression model for the conditional probabilities as shown below:

$$P_{Y_2|Y_1,\mathbf{X}}(y_2 \mid y_1;\mathbf{x}) = \tfrac{e^{\theta y_2}}{1+e^{\theta}}, \text{ with } \theta = \gamma_0 + \gamma_1 x_1 + \cdots + \gamma_p x_p + \beta_1 y_1, \tag{1.2}$$

where $\gamma_0$ is the intercept, $\beta_1$ is the coefficient of the previous outcome, $Y_1$, and $\gamma_1, \ldots, \gamma_p$ are the coefficients of the covariates $X_1, \ldots, X_p$, respectively. Here $\theta$ is the logit defined as

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

# 2 Test for Dependence: An Extended Regressive Approach

Let us consider the following regressive model for the jth follow-up $(j = 1, \ldots, J)$ with two binary outcomes to test for the dependence in the outcome variables as well as between the covariates and

the outcome variables. For simplicity, let us consider that $J = 2$ which can be extended further but this paper focuses on data from bivariate Bernoulli outcomes from two consecutive follow-ups. 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:

$$P_{Y_2|Y_1,\mathbf{X},\mathbf{Z}}(Y_2 = s \mid y_1, \mathbf{x}, \mathbf{z}) = \frac{e^{(\boldsymbol{\gamma}'\mathbf{x}+\beta_1 y_1+\boldsymbol{\eta}'\mathbf{z})s}}{1 + e^{(\boldsymbol{\gamma}'\mathbf{x}+\beta_1 y_1+\boldsymbol{\eta}'\mathbf{z})}}, \quad s = 0, 1 \tag{2.1}$$

where $\mathbf{x} = (1, x_1, \ldots, x_p)'$, $\boldsymbol{\gamma}' = (\gamma_0, \gamma_1, \ldots, \gamma_p)$, $\beta_1$ is the parameter corresponding to $Y_1$, $\mathbf{z} = (z_{1,\ldots,z_p})'=(x_1 y_1, \ldots, x_p y_1)'$, $\boldsymbol{\eta}' = (\eta_1, \ldots, \eta_p)$. It may be noted here that this model is an extension of (1.2) as it contain interaction terms $\mathbf{Z} = (Z_1, \ldots, Z_p)'$.

In the regressive model (1.2), the dependence between $Y_1$ and $Y_2$ is examined on the basis of the test for $H_0 : \beta_1 = 0$ against $H_1 : \beta_1 \neq 0$. Lack of evidence against the null hypothesis may indicate a possible independence between these variables. Here, it is assumed that for given values of $Y_1 = 0$ or $Y_1 = 1$ the relationship between $\mathbf{X}$ and $Y_2$ remain unchanged. However, it is clearly evident from the modified model (equation 2.1) that there is dependence through $\mathbf{Z}$ as well. Hence, without a proper investigation about the underlying relationships may lead to misleading conclusions in many instances if we employ model (1.2). Therefore, we need to test for the hypothesis that for given values of $Y_1 = 0$ or $Y_1 = 1$ the relationship between the covariates, $X$ and $Y_2$, remains unchanged. If this does not hold, the test for dependence may provide misleading result due to model misspecification. The following sections illustrate the underlying relationships and their consequences based on the extended model proposed in (2.1).

# 3 Dependence in Three Binary Variables

Let us consider two binary outcome variables $Y_1$ and $Y_2$ and one covariate, X, which is also considered as binary here for the purpose of illustration. The joint probability can be expressed as follows:

$$P_{Y_1,Y_2|X}(y_2, y_1 \mid x) = P_{Y_2|Y_1,X}(y_2 \mid y_1, x) \times P_{Y_1|X}(y_1 \mid x).$$

The conditional probability can be expressed as follows for the underlying conditional independence of the variables:

(i) If $Y_1$ and $Y_2$ are conditionally independent for given $X = x$ then $P_{Y_2|Y_1,X}(y_2 \mid y_1, x) = P_{Y_2|X}(y_2 \mid x)$, for all values of $x$, and

(ii) If $Y_2$ and $X$ are conditionally independent for given $Y_1 = y_1$ then $P_{Y_2|Y_1,X}(y_2 \mid y_1, x) = P_{Y_2|Y_1}(y_2 \mid y_1)$, for all values of $y_1$.

The conditional model (2.1) can be expressed as follows:

$$P_{Y_2|Y_1,X,Z_1}(Y_2 = s \mid y_1, x, z_1) = \frac{\exp\{(\gamma_0 + \gamma_1 x + \beta_1 y_1 + \eta_1 z_1)s\}}{1 + \exp\{(\gamma_0 + \gamma_1 x + \beta_1 y_1 + \eta_1 z_1)\}}, \quad s = 0, 1 \tag{3.1}$$

and the conditional model for $X = 0$ can be shown as:

$$P_{Y_2|Y_1,X}(Y_2 = s \mid y_1, X = 0) = \frac{\exp\{(\beta_{00} + \beta_{01}y_1)s\}}{1 + \exp\{(\beta_{00} + \beta_{01}y_1)\}}, \;\; s = 0, 1, \tag{3.2}$$

where $\beta_{00} = \gamma_0$ and $\beta_{01} = \beta_1$. Similarly, the conditional model for $X = 1$ is

$$P_{Y_2|Y_1,X}(Y_2 = s \mid y_1, X = 1) = \frac{\exp\{(\beta_{10} + \beta_{11}y_1)s\}}{1 + \exp\{(\beta_{10} + \beta_{11}y_1)\}}, \;\; s = 0, 1, \tag{3.3}$$

where $\beta_{10} = \gamma_0$ and $\beta_{11} = (\beta_1 + \eta_1)$. It is noteworthy that $\beta_{00} = \beta_{10} = \gamma_0$ in both the conditional models (3.2) and (3.3). Hence, in case of conditional independence of $Y_1$ and $Y_2$ for given $X = x$, we need to test the null hypotheses $H_0 : \beta_{01} = \beta_{11} = 0$.

Similarly, if $Y_2$ and $X$ are conditionally independent for given $Y_1 = y_1$ then $P_{Y_2|Y_1,X}(y_2 \mid y_1, x) = P_{Y_2|Y_1(y_2|y_1)}$, for all values of $y_1$. The model (3.1) can be expressed as follows for the values of $Y_1 = 0$ and $Y_1 = 1$:

$$P_{Y_2|Y_1,X}(Y_2 = s \mid Y_1 = 0, x) = \frac{\exp\{(\gamma_{00} + \gamma_{01}x)s\}}{1 + \exp\{(\gamma_{00} + \gamma_{01}x)\}}, \; s = 0, 1, \tag{3.4}$$

where $\gamma_{00} = \gamma_0$ and $\gamma_{01} = \gamma_1$, and

$$P_{Y_2|Y_1,X}(Y_2 = s \mid Y_1 = 1, x) = \frac{\exp\{(\gamma_{10} + \gamma_{11}x)s\}}{1 + \exp\{(\gamma_{10} + \gamma_{11}x)\}}, \; s = 0, 1, \tag{3.5}$$

where $\gamma_{10} = \gamma_0$ and $\gamma_{11} = (\gamma_1 + \eta_1)$. Hence, in case of conditional independence of $Y_2$ and $X$ for given $Y_1 = y_1$, we need to test the null hypothesis $H_0 : \gamma_{01} = \gamma_{11} = 0$.

The joint independence of $Y_1$ and $Y_2$ for given $X = x$ in model (1.1) can be shown from the following relationship:

$$P_{Y_1,Y_2|X}(y_2, y_1 \mid x) = P_{Y_2|Y_1,X}(y_2 \mid y_1, x) \times P_{Y_1|X}(y_1 \mid x).$$

If we employ the conditional independence tests in the above relationship for simultaneous tests in the joint probability function of $Y_1$ and $Y_2$ for given $X = x$, it can be shown for the model (3.1) that the following conditions need to be satisfied: (i) for the test for independence of $Y_2$ and $Y_1$, the condition to be satisfied for equations (3.4) and (3.5) is that $\gamma_{01} = \gamma_{11} = \gamma$, and (ii) for the test for independence of $Y_2$ and $X$, the condition to be satisfied for equations (3.2) and (3.3) is that $\beta_{01} = \beta_{11} = \beta$. These are illustrated in sections 4 and 5.

# 4 Test for Conditional Independence of $Y_1$ and $Y_2$

Let us consider the conditional probabilities for dependence between $Y_1$ and $Y_2$ for given $\mathbf{X}$, where $Y_1$ and $Y_2$ are outcomes in the follow-ups 1 and 2 and $\mathbf{X} = (1, X_1, \ldots, X_p)$ is the vector of $p$ covariates as shown in section 1 (see [12, 15]) for further details.

The conditional probabilities for $Y_2$ given $Y_1 = 0$, $Y_1 = 1$ and $\mathbf{X} = \mathbf{x}$ are:

$$P_{Y_2|Y_1,X}(Y_2 = 1 \mid Y_1 = 0, \mathbf{x}) = \frac{\exp(\boldsymbol{\gamma}'_{01}\mathbf{x})}{1 + \exp(\boldsymbol{\gamma}'_{01}\mathbf{x})}, \tag{4.1}$$

where $\mathbf{x} = (1, x_1, \ldots, x_p)'$, $\boldsymbol{\gamma}'_{01} = (\gamma_{010}, \gamma_{011}, \ldots, \gamma_{01p})$, and

$$P_{Y_2|Y_1,X}(Y_2 = 1 \mid Y_1 = 1, \mathbf{x}) = \frac{\exp({\gamma'}_{11}\mathbf{x})}{1 + \exp({\gamma'}_{11}\mathbf{x})}, \tag{4.2}$$

where $\mathbf{x} = (1, x_1, \ldots, x_p)'$, $\boldsymbol{\gamma}'_{11} = (\gamma_{110}, \gamma_{111}, \ldots, \gamma_{11p})$. It may be noted here that

$$P_{Y_2|Y_1,\mathbf{X}}(Y_2 = 0 \mid Y_1 = 0, \mathbf{x}) + P_{Y_2|Y_1,\mathbf{X}}(Y_2 = 1 \mid Y_1 = 0, \mathbf{x}) = 1 \text{ and}$$
$$P_{Y_2|Y_1,\mathbf{X}}(Y_2 = 0 \mid Y_1 = 1, \mathbf{x}) + P_{Y_2|Y_1,\mathbf{X}}(Y_2 = 1 \mid Y_1 = 1, \mathbf{x}) = 1.$$

For a single covariate, $X$, models (4.1) and (4.2) will reduce to (3.4) and (3.5), respectively, for $s = 1$. The marginal probabilities for $Y_1$ are defined as follows:

$$P_{Y_1|\mathbf{X}}(Y_1 = 1 \mid \mathbf{x}) = \frac{\exp(\boldsymbol{\gamma}'_1\mathbf{x})}{1 + \exp({\gamma'}_1\mathbf{x})}, \tag{4.3}$$

where $\mathbf{x} = (1, x_1, \ldots, x_p)'$, ${\boldsymbol{\gamma}'}_1 = (\gamma_{10}, \gamma_{11}, \ldots, \gamma_{1p})$ and

$$P_{Y_1|\mathbf{X}}(Y_1 = 0 \mid \mathbf{x}) = \frac{1}{1 + \exp(\boldsymbol{\gamma}'_1\mathbf{x})}. \tag{4.4}$$

We obtain the following joint probabilities using the conditional and marginal probabilities as follows:

$$P_{Y_1,Y_2|\mathbf{X}}(Y_1 = 0, Y_2 = 1 \mid \mathbf{x}) = \frac{\exp(\boldsymbol{\gamma}'_{01}\mathbf{x})}{(1 + \exp(\boldsymbol{\gamma}'_{01}\mathbf{x}))(1 + \exp(\boldsymbol{\gamma}'_1\mathbf{x}))} \tag{4.5}$$

$$P_{Y_1,Y_2|\mathbf{X}}(Y_1 = 0, Y_2 = 0 \mid \mathbf{x}) = \frac{1}{(1 + \exp(\boldsymbol{\gamma}'_{01}\mathbf{x}))(1 + \exp(\boldsymbol{\gamma}'_1\mathbf{x}))} \tag{4.6}$$

$$P_{Y_1,Y_2|\mathbf{X}}(Y_1 = 1, Y_2 = 1 \mid \mathbf{x}) = \frac{\exp(\boldsymbol{\gamma}'_{11}\mathbf{x})\exp(\boldsymbol{\gamma}'_1\mathbf{x})}{(1 + \exp(\boldsymbol{\gamma}'_{11}\mathbf{x}))(1 + \exp(\boldsymbol{\gamma}'_1\mathbf{x}))} \tag{4.7}$$

$$P_{Y_1,Y_2|\mathbf{X}}(Y_1 = 1, Y_2 = 0 \mid \mathbf{x}) = \frac{\exp(\boldsymbol{\gamma}'_1\mathbf{x})}{(1 + \exp(\boldsymbol{\gamma}'_{11}\mathbf{x}))(1 + \exp(\boldsymbol{\gamma}'_1\mathbf{x}))} \tag{4.8}$$

The marginal probabilities for $Y_1$ are obtained by summing (4.5) and (4.6) and (4.7) and (4.8) which are (4.4) and (4.3), respectively. Similarly, the marginal probabilities for $\mathrm{Y}_2$ by summing (4.5) and (4.7) and (4.6) and (4.8) respectively and under the $H_0 : \gamma_{01} = \gamma_{11} = \gamma$ the probabilities are:

$$P_{Y_2|\mathbf{X}}(Y_2 = 1 \mid \mathbf{x}) = \frac{\exp(\boldsymbol{\gamma}'\mathbf{x})}{1 + \exp(\boldsymbol{\gamma}'\mathbf{x})} \text{ and } P_{Y_2|\mathbf{X}}(Y_2 = 0 \mid \mathbf{x}) = \frac{1}{1 + \exp(\boldsymbol{\gamma}'\mathbf{x})} \tag{4.9}$$

These probabilities are same as the marginal probabilities of $Y_1$ under the $H_0 : \gamma_1 = \gamma$.

If $Y_1$ and $Y_2$ are conditionally independent for any value of $\mathbf{X} = \mathbf{x}$ then (4.1) and (4.2) are equal. In other words, the conditional independence is true if

$$P_{Y_2|Y_1,\mathbf{X}}(Y_2 = 1 \mid Y_1 = 0, \mathbf{x}) = P_{Y_2|Y_1,\mathbf{X}}(Y_2 = 1 \mid Y_1 = 1, \mathbf{x}) = \frac{e^{{\boldsymbol{\gamma}'}_{01}\mathbf{x}}}{1 + e^{{\boldsymbol{\gamma}'}_{01}\mathbf{x}}} = \frac{e^{\boldsymbol{\gamma}'_{11}\mathbf{x}}}{1 + e^{\boldsymbol{\gamma}'_{11}\mathbf{x}}} = \frac{e^{\boldsymbol{\gamma}'\mathbf{x}}}{1 + e^{\boldsymbol{\gamma}'\mathbf{x}}}.$$

It shows that the conditional independence is achieved only if $\gamma_{01} = \gamma_{11} = \gamma$. This is equivalent to the test for $H_0 : \beta = 0$ in the regressive model (2.1).

In other words, for testing $H_0 : \beta = 0$, in a regressive model, is equivalent to testing the null hypothesis $H_0 : \gamma_{01} = \gamma_{11}$ in conditional models (4.1) and (4.2).

The test statistic is

$$\chi^2 = (\hat{\gamma}_{01} - \hat{\gamma}_{11})' \big[\widehat{\text{Var}}(\hat{\gamma}_{01} - \hat{\gamma}_{11})\big]^{-1} (\hat{\gamma}_{01} - \hat{\gamma}_{11}) \tag{4.10}$$

which is chi-square with $p$ degrees of freedom.

An alternative test can also be performed under the equality of parameters due to the fact that the marginal probabilities of $Y_1$ and $Y_2$ are same as shown in (4.1–4.2) and (4.9) respectively. Hence, the alternative test can be performed for

$$H_0 : \gamma_{01} = \gamma \quad \text{vs} \quad H_1 : \gamma_{01} \neq \gamma$$

The test statistic is

$$\chi^2 = (\hat{\gamma}_{01} - \hat{\gamma})' \big[\widehat{\text{Var}}(\hat{\gamma}_{01} - \hat{\gamma})\big]^{-1} (\hat{\gamma}_{01} - \hat{\gamma}) \tag{4.11}$$

which asymptotically follows chi-square with $p$ degrees of freedom. In addition to test hypothesis $H_0 : \gamma_{11} = \gamma \quad \text{vs} \quad H_1 : \gamma_{11} \neq \gamma$, We can use the test statistic

$$\chi^2 = (\hat{\gamma}_{11} - \hat{\gamma})' \big[\widehat{\text{Var}}(\hat{\gamma}_{11} - \hat{\gamma})\big]^{-1} (\hat{\gamma}_{11} - \hat{\gamma}), \tag{4.12}$$

which is also chi-square with $p$ degrees of freedom. For single covariate, these hypotheses can be performed using the asymptotically normal tests.

# 5 Test for Conditional Independence of $\mathbf{Y}_2$ and X for Given $\mathbf{Y}_1$

Let us consider that $Y_1$ and $Y_2$ are the outcomes in the follow-ups 1 and 2 and X is a covariate as shown in section 3 instead of a vector of covariates shown in section 4. This is considered to keep the illustration simple. Without any loss of generality, this can be extended to any p-covariate vector producing any set of covariate patterns.

We know that the conditional probabilities of $Y_2$ given $X = x$ and $Y_1 = y_1$ can be shown as follows:

$$P_{Y_2|Y_1,X}\,(Y_2{=}\,1|y_1, X = 0) = \frac{e^{\beta_{01} y_1}}{1 + e^{\beta_{01} y_1}} \tag{5.1}$$

and

$$P_{Y_2|Y_1,X}\,(Y_2{=}\,1|y_1, X = 1) = \frac{e^{\beta_{11} y_1}}{1 + e^{\beta_{11} y_1}}. \tag{5.2}$$

By definition, $Y_2$ and X are conditionally independent for given $Y_1 = y_1$ if

$$P_{Y_2|Y_1,X}(y_2 \mid y_1, X = 0) = P_{Y_2|Y_1,X}(y_2 \mid y_1, X = 1).$$

It can be shown from (5.1) and (5.2) that $\beta_{01} = \beta_{11} = \beta$ satisfies this condition. Now the joint probability function

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

where the regressive model (1.2) represents the conditional probability $P_{Y_2|Y_1,X}(y_2 \mid y_1, x)$ which reduces to only $P_{Y_2|Y_1}(y_2 \mid y_1)$ if $\gamma = 0$. Alternatively, an equivalent test can be considered for $\beta_{01} = \beta_{11} = \beta$ if conditional models (5.1) and (5.2) are considered. We can use the Wald chi-square test with 1 degree of freedom as shown below:

$$\chi^2 = \frac{(\hat{\beta}_{01} - \hat{\beta}_{11})^2}{\widehat{\text{Var}}(\hat{\beta}_{01} - \hat{\beta}_{11})}. \tag{5.3}$$

# 6 Simulation

In order to examine the proposed method, we have employed both simulation and application to real life data. For the purpose of generating correlated binary data for simulations, a software package, bindata, developed by Leisch et al. [17] has been used. Based on this method, data can be generated from multivariate Bernoulli distributions. The joint distribution of $Y_1$ and $Y_2$ are fully specified by the marginal and conditional probabilities. We have considered two binary outcome variables, $Y_1$ and $Y_2$, and one binary covariate, $X$. For these variables, data are generated from a trivariate Bernoulli distribution. These three binary variables are not necessarily independent. For generating data, we have used various combinations of the following probabilities: $P_{Y_1}(Y_1 = 1)$, $P_{Y_2}(Y_2 = 1)$, $P_X(X = 1)$, $P_{Y_1,Y_2}(Y_1 = 1, Y_2 = 1)$, $P_{Y_1,X}(Y_1 = 1, X = 1)$ and $P_{Y_2,X}(Y_2 = 1, X = 1)$. We have used different combinations of the joint probabilities which are denoted as models based on relationship between $Y_1$ and $Y_2$ and relationship between $Y_2$ and $X$ for given values of $Y_1$. Eight models are considered each for 500 simulations with samples of size 350 and 1000, respectively. In defining the models, we have considered the following to obtain the variations in the relationships in the bivariate binary outcome variables and a single covariate: (i) independence or near independence of $Y_1$ and $Y_2$, (ii) dependence in $Y_1$ and $Y_2$, (iii) close conditional probability of $Y_2$ and $X$ for given $Y_1$ =0 and $Y_1$ =1, and (iv) difference in the conditional probabilities of $Y_2$ and $X$ for given $Y_1 = y_1$. Based on the simulated data, we estimated/fitted the following: (i) the odds ratios, (ii) the conditional correlations between $Y_2$ and $X$ for given $Y_1 = y_1$, (iii) the conditional models for $Y_2$ for given $Y_1 = y_1$ and $X = x$ (equation 4.1 and 4.2), (iv) the conditional models for $Y_2$ for given $Y_1 = y_1$ and $X = x$ (equations 5.1 and 5.2), (v) the regressive models (equation 1.2), and (vi) the proposed tests shown in (4.10) and (5.3).

The average odds ratios (Table 1) indicate that the Models 1, 2, 3 and 5 show no evidence of association between $Y_1$ and $Y_2$ and the odds ratios are close to 1. However, the Models 1 and 5 indicate that there are close but non-zero conditional correlations between $Y_2$ and $X$ for given $Y_1$ but one is positive (Model 1) and the other is negative (Model 5). This means that we expect no association between $Y_1$ and $Y_2$ but association between $Y_2$ and $X$ and due to close (approximately) but non-zero correlations, the conditional models and the regressive model are expected to give

similar estimates. The average estimated coefficients for X variable from sample size 350 and 1000 are 1.394 and 1.386 for the conditional model (4.1) and 1.700 and 1.349 for the conditional model (4.2), respectively. Similar finding is observed for Model 5. The significant models (at 5% level) were found 496 and 498 times out of 500 samples of size 350 and 500 out of 500 samples of size 1000. It is observed that with the increase in sample size the results become more consistent. The equality of parameters for both the variables $X$ and $Y_1$ indicate that the regressive model can be employed without any distortion. The significant results were obtained mostly in less than 5% of the cases for both sample size 350 of Model 1 and 1000 of Model 5, respectively.

The Model 2 shows no association between $Y_1$ and $Y_2$ but different correlations between X and the outcome variables. In that case, we observe marked variations in the estimated parameters for the regressive model. The proposed tests for the equality of parameters corresponding to X shows significant results in 344 and 477 out of 500 times for samples of size 350 and 1000, respectively. This is evident more sharply for the Model 3 where the correlations are taken in opposite directions for the conditional models (4.1) and (4.2), respectively. It clearly reveals that under situations for the Models 2 and 3, the regressive models are not good choices. The proposed tests show significant results in almost all the cases of simulations. In the Models 1 and 5, the odds ratio between $Y_1$ and $Y_2$ appear to be near 1 implying independence and correlations between $X$ and $Y_1$ and $X$ and $Y_2$ are close but non-zero. In both the cases, the proposed tests for the equality of parameters of the conditional models indicate the acceptance of null hypothesis for $Y_2$ given $Y_1 = y_1$ as well as for $X$ and $Y_2$ for given $Y_1 = y_1$. However, the Model 4 displays association between $Y_1$ and $Y_2$ and equal but non-zero correlation between $X$ and outcome variables $Y_2$ for given $Y_1 = y_1$. The proposed tests indicate the hypothesis of equality of parameters for the conditional models for both given $X = x$ and given $Y_1 = y_1$ can be accepted. In that case, the conditional models and the regressive models provide similar estimates. The estimated parameters for the conditional models and the regressive models appear to be similar for the Model 4 and the proposed tests show that the equality of parameters for the conditional models failed in only less than 4 percent cases. Hence, the regressive models can be employed without any difficulty. Model 6, on the other hand, shows the association between $Y_1$ and $Y_2$ and different correlations between $X$ and $Y_2$ for conditional models (4.1) and (4.2). The conditional models show that the null hypothesis of equality of parameters can be rejected in this model which is reflected from the proposed tests. This shows clearly that the regressive model estimates are quite different and the proposed tests of the equality of parameters reveals significant results in 480 and 500 out of 500 times for samples of size 350 and 1000 respectively. Under this circumstance, the regressive model fails to provide any reasonable estimate of the relationship with the variables of interest.

# 7 Application to Depression Data

For this study, an application is displayed in this section from the Health and Retirement Study (HRS) data [18]. The HRS is sponsored by the National Institute of Aging (grant number NIA U01AG09740) and conducted by the University of Michigan (2002). This study is conducted nationwide for individuals over age 50 and their spouses. We have used the panel data from the two

Table 1: Results averaged from 500 simulations for different dependence patterns for samples of size 350 and 1000

| | Sample Size = 350 | | | | | | | | Sample Size = 1000 | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Models | | | | | | | | Models | | | | | | | |
| Column No. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| $P_{Y_1,Y_2}(Y_1=0,Y_2=0)$ | 0.40 | 0.25 | 0.25 | 0.20 | 0.40 | 0.19 | 0.45 | 0.25 | 0.40 | 0.25 | 0.25 | 0.20 | 0.40 | 0.19 | 0.45 | 0.25 |
| $P_{Y_1,Y_2}(Y_1=0,Y_2=1)$ | 0.40 | 0.25 | 0.25 | 0.30 | 0.40 | 0.31 | 0.35 | 0.25 | 0.40 | 0.25 | 0.25 | 0.30 | 0.40 | 0.31 | 0.35 | 0.25 |
| $P_{Y_1,Y_2}(Y_1=1,Y_2=0)$ | 0.10 | 0.25 | 0.25 | 0.30 | 0.10 | 0.09 | 0.05 | 0.09 | 0.10 | 0.25 | 0.25 | 0.30 | 0.10 | 0.09 | 0.05 | 0.09 |
| $P_{Y_1,Y_2}(Y_1=1,Y_2=1)$ | 0.10 | 0.25 | 0.25 | 0.20 | 0.10 | 0.41 | 0.15 | 0.41 | 0.10 | 0.25 | 0.25 | 0.20 | 0.10 | 0.41 | 0.15 | 0.41 |
| Odds Ratio ($Y_1$,$Y_2$) | 1.03 | 1.018 | 1.039 | 0.45 | 1.022 | 2.772 | 4.131 | 4.605 | 1.006 | 1.014 | 1.008 | 0.447 | 1.021 | 2.667 | 3.951 | 4.446 |
| $r_{xy_2}^{01}$ | 0.245 | 0.255 | 0.241 | 0.612 | -0.244 | 0.909 | 0.117 | 0.238 | 0.248 | 0.244 | 0.246 | 0.618 | -0.247 | 0.908 | 0.116 | 0.249 |
| $r_{xy_2}^{11}$ | 0.269 | 0.005 | -0.268 | 0.61 | -0.271 | 0.055 | 0.113 | 0.064 | 0.268 | 0.004 | -0.269 | 0.616 | -0.273 | 0.056 | 0.118 | 0.058 |
| Conditional Model (based on $Y_1$=0 and $Y_1$=1) | | | | | | | | | | | | | | | | |
| $\gamma_{00}$ | -1.156 | -1.276 | -1.164 | -0.848 | 1.155 | -3.328 | -0.318 | -1.138 | -1.143 | -1.133 | -1.141 | -0.858 | 1.145 | -2.871 | -0.312 | -1.165 |
| p-values $< 0.05$ | 486 | 421 | 419 | 481 | 478 | 490 | 371 | 396 | 500 | 499 | 500 | 500 | 500 | 500 | 494 | 498 |
| $\gamma_{01}$ | 1.394 | 1.526 | 1.399 | 3.112 | -1.384 | 7.149 | 1.063 | 1.368 | 1.386 | 1.373 | 1.381 | 3.122 | -1.39 | 6.266 | 0.998 | 1.406 |
| p-values $< 0.05$ | 494 | 462 | 450 | 500 | 493 | 480 | 238 | 443 | 500 | 500 | 500 | 500 | 500 | 500 | 454 | 500 |
| $\gamma_{10}$ | -1.386 | -0.034 | 1.056 | -2.248 | 1.325 | 1.27 | 0.739 | 1.251 | -1.04 | -0.014 | 1.02 | -2.232 | 1.054 | 1.212 | 0.685 | 1.206 |
| p-values $< 0.05$ | 231 | 52 | 416 | 500 | 228 | 435 | 128 | 432 | 456 | 124 | 497 | 500 | 464 | 500 | 317 | 500 |
| $\gamma_{11}$ | 1.7 | 0.025 | -1.367 | 3.099 | -1.655 | 0.295 | 0.55 | 0.329 | 1.349 | 0.019 | -1.341 | 3.096 | -1.373 | 0.332 | 0.59 | 0.346 |
| p-values $< 0.05$ | 297 | 50 | 462 | 500 | 294 | 64 | 79 | 79 | 488 | 114 | 499 | 500 | 486 | 131 | 190 | 134 |
| # Sig. models | 496 | 434 | 496 | 500 | 498 | 500 | 254 | 428 | 500 | 500 | 500 | 500 | 500 | 500 | 460 | 500 |
| Conditional Model (based on $X_1$=0 and $X_1$=1) | | | | | | | | | | | | | | | | |
| $\beta_{00}$ | 1.156 | -1.278 | -1.164 | -0.848 | 1.155 | -3.326 | -0.318 | -1.138 | -1.143 | -1.133 | -1.141 | -0.858 | 1.145 | -2.871 | -0.312 | -1.165 |
| p-values $< 0.05$ | 486 | 421 | 419 | 481 | 478 | 490 | 371 | 396 | 500 | 499 | 500 | 500 | 500 | 500 | 494 | 498 |
| $\beta_{01}$ | -0.23 | 1.244 | 2.219 | -1.4 | 0.174 | 4.597 | 1.057 | 2.391 | 0.103 | 1.119 | 2.162 | -1.374 | -0.09 | 4.083 | 0.997 | 2.371 |
| p-values $< 0.05$ | 20 | 300 | 492 | 463 | 21 | 490 | 245 | 490 | 33 | 453 | 500 | 500 | 35 | 500 | 445 | 500 |
| # Sig. models | 31 | 311 | 492 | 467 | 36 | 500 | 260 | 492 | 33 | 456 | 500 | 500 | 32 | 500 | 449 | 500 |

**Note:** $r_{xy_2}^{01} = corr(X, Y_2|Y_1 = 0)$ **and** $r_{xy_2}^{11} = corr(X, Y_2|Y_1 = 1)$

Table 1. continued... results averaged from 500 simulations for different dependence patterns for samples of size 350 and 1000

| | Sample Size = 350 | | | | | | | | Sample Size = 1000 | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Models | | | | | | | | Models | | | | | | | |
| Column No. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| Conditional Model (based on $X_1$=0 and $X_1$=1) | | | | | | | | | | | | | | | | |
| $\beta_{10}$ | 0.238 | 0.25 | 0.235 | 2.264 | -0.229 | 3.787 | 0.75 | 0.23 | 0.244 | 0.24 | 0.24 | 2.264 | -0.246 | 3.395 | 0.686 | 0.241 |
| p-values $< 0.05$ | 229 | 171 | 158 | 500 | 191 | 490 | 121 | 135 | 430 | 313 | 314 | 500 | 436 | 500 | 308 | 322 |
| $\beta_{11}$ | 0.075 | -0.259 | -0.547 | -1.414 | -0.101 | -2.221 | 0.54 | 1.35 | 0.066 | -0.235 | -0.56 | -1.4 | -0.073 | -1.851 | 0.589 | 1.312 |
| p-values $< 0.05$ | 25 | 111 | 300 | 479 | 27 | 476 | 83 | 499 | 34 | 200 | 457 | 500 | 31 | 500 | 199 | 500 |
| # Sig. models | 26 | 112 | 301 | 479 | 30 | 488 | 85 | 499 | 35 | 200 | 457 | 500 | 33 | 500 | 197 | 500 |
| Regressive Model | | | | | | | | | | | | | | | | |
| $\gamma_0$ | -1.141 | -0.506 | 0.111 | -0.843 | 1.15 | -1.128 | -0.307 | -0.742 | -1.13 | -0.475 | 0.101 | -0.853 | 1.136 | -1.107 | -0.302 | -0.747 |
| p-values $< 0.05$ | 496 | 233 | 41 | 492 | 494 | 500 | 364 | 354 | 500 | 391 | 76 | 500 | 500 | 500 | 493 | 498 |
| $\gamma_1$ | 1.378 | 0.622 | -0.133 | 3.086 | -1.379 | 2.887 | 0.824 | 0.903 | 1.371 | 0.587 | -0.121 | 3.101 | -1.381 | 2.843 | 0.823 | 0.911 |
| p-values $< 0.05$ | 499 | 315 | 47 | 500 | 499 | 500 | 283 | 426 | 500 | 451 | 81 | 500 | 500 | 500 | 470 | 500 |
| $\beta_1$ | 0.08 | 0.02 | -0.008 | -1.39 | -0.1 | 0.634 | 0.846 | 1.542 | 0.075 | 0.03 | -0.016 | -1.381 | -0.078 | 0.62 | 0.824 | 1.525 |
| p-values $< 0.05$ | 31 | 53 | 43 | 497 | 29 | 271 | 289 | 500 | 42 | 99 | 77 | 500 | 44 | 451 | 467 | 500 |
| # Sig. models | 496 | 281 | 53 | 500 | 498 | 500 | 499 | 500 | 500 | 427 | 100 | 500 | 500 | 500 | 500 | 500 |
| Proposed Tests | | | | | | | | | | | | | | | | |
| $H_0 : \gamma_{01} = \gamma_{11}$ | | | | | | | | | | | | | | | | |
| p-values $< 0.05$ | 11 | 344 | 499 | 18 | 22 | 480 | 37 | 144 | 23 | 477 | 500 | 20 | 30 | 500 | 55 | 379 |
| $H_0 : \beta_{01} = \beta_{11}$ | | | | | | | | | | | | | | | | |
| p-values $< 0.05$ | 19 | 303 | 498 | 2 | 20 | 480 | 11 | 111 | 29 | 463 | 500 | 3 | 22 | 500 | 35 | 345 |

rounds of the study conducted on individuals over age 50 years in 1992 and 1994 and documented by RAND. We have used the panel data on depression for the period, 1992-1994. The depression index is based on the score on the basis of the scale proposed by the Center for Epidemiologic Studies Depression (CESD). As indicated in the documentation of the RAND, the CESD score is computed on the basis of eight indicators attributing depression problem. The indicators of depression problem are based on six negative (all or most of the time: depressed, everything is an effort, sleep is restless, felt alone, felt sad, and could not get going) and two positive indicators (felt happy, enjoyed life). These indicators are yes/no responses of the respondent's feelings much of the time over the week prior to the interview. The CESD score is the sum of six negative indicators minus two positive indicators. Hence, severity of the emotional health can be measured from the CESD score. From the panels of data, we have used 9761 respondents for analyzing depression among the elderly in the USA during 1992-2002. Steffick [19] indicated that many studies have used the CESD scale to measure depressive symptoms in a wide range of both clinical and non-clinical populations. Cheng, Chan and Fung [20] showed the validity of a short version of the CESD scale. The study of depression is important among the elderly because repeated spells of depression may lead to fatality such as suicide [21]. It is further noted that the depression is very common among the elderly and may be difficult to diagnose. Evans and Mottram [22] observed that there is movement along the spectrum of depression 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 $\leqslant 0$) $= 0$, depression (CESD score>0)=1), we may denote $Y_1 =$ depression status at 1992, and $Y_2 =$ depression status at 1994; 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).

Table 2 shows the fit of the marginal model for the outcome variable $Y_1$ as well as fit of the regressive model considering the previous outcome as a covariate. It 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 regressive model but gender appears to be non-significant in the marginal model. Table 3 displays the transition count and transition probabilities for states no depression and depression. It is observed that the probabilities of remaining in no depression and depression states are 0.650 and 0.715, respectively. The probability of making a transition from no depression to depression during a two year period is 0.350 and the probability of a recovery from depression during consecutive follow-ups in two-years apart is 0.285. In other words, more people move to depression than that of the recovery.

Then we need to test for the equality of the parameters of the conditional models for the new cases of depression (transition from 0 to 1) and the old cases (transition type 1 to 1). The estimates of the parameters (Table 4) indicate that gender, marital status and white race are negatively associated with depression for both the models and the estimates appear to be visibly different for gender and marital status. The chi-square test value (as shown in (4.10)) is 838.49 for the equality of the parameters for the conditional models reflect the large variation and the null hypothesis of equality of parameters is rejected (p-value$<$0.001). In other words, in this example, we have demonstrated that

Table 2: Logistic Regression for Wave I and the regressive model (dependent variable= CESD score)

| Variables | Coefficient | Standard error | Wald Chi-square | p-value |
|---|---|---|---|---|
| **Marginal Model $Y_1$** | | | | |
| Constant | 0.4428 | 0.1205 | 13.492 | 0.0002 |
| Gender | -0.0479 | 0.0460 | 1.084 | 0.2977 |
| Marital Status | -0.6045 | 0.0522 | 133.873 | 0.0001 |
| White | -0.5912 | 0.1169 | 25.574 | 0.0001 |
| Black | -0.0368 | 0.1257 | 0.086 | 0.7700 |
| Likelihood Ratio ($\beta = 0$) (p-value) | | 307.826 (0.0001) | | |
| -2 Log L | | 11231.366 | | |
| **Regressive Model** | | | | |
| Constant | 0.3680 | 0.1428 | 6.639 | 0.0100 |
| Gender | -0.2378 | 0.0487 | 23.809 | 0.0001 |
| Marital Status | -0.4598 | 0.0568 | 65.500 | 0.0001 |
| White | -0.5871 | 0.1368 | 18.431 | 0.0001 |
| Black | -0.2086 | 0.1469 | 2.016 | 0.1557 |
| $Y_1$ | 1.4375 | 0.0503 | 817.756 | 0.0001 |
| Likelihood Ratio ($\beta = 0$) (p-value) | | 1211.616 (0.000) | | |
| -2 Log L | | 10121.338 | | |

**Note:** Gender: male=1, female=0; Marital Status: married/partnered, 0=single/divorced/separated; White: yes=1, no=0; Black: yes=1, no=0.

the conditional models need to be fitted and we also concluded that the independence of the outcome variables is rejected. Finally, a comparison of the conditional and regressive models indicates that the estimates of the parameters for the risk factors shows significant variation in the conditional models for depression depending on given previous outcome. Hence, a regressive model is not justified for such analysis.

# 8 Concluding Remarks

In regressive models both the covariates and the previous outcomes are included. In this paper, the possible relationships in the outcome variables, covariates and previous outcomes are demonstrated and tests are proposed based on the conditional, marginal and joint models for bivariate binary outcomes. It is clearly shown in this paper that under certain circumstances separate conditional models are preferred as compared to the regressive models where both covariates and previous outcomes

Table 3: Transition count and probability based on Consecutive Follow-ups I and II

| WAVE I | WAVE II | | | | | |
|---|---|---|---|---|---|---|
| | Transition Count | | | Transition Probability | | |
| | 0 | 1 | Total | 0 | 1 | Total |
| 0 | 3296 | 1773 | 5069 | 0.650 | 0.350 | 1.000 |
| 1 | 868 | 2179 | 3047 | 0.285 | 0.715 | 1.000 |

Table 4: Logistic Regression for Wave I and the regressive model (dependent variable= CESD score)

| Variables | Coefficient | Standard error | Wald Chi-square | p-value |
|---|---|---|---|---|
| Model $0 \rightarrow > 1$ | | | | |
| Constant | 0.2493 | 0.1832 | 1.8517 | 0.1736 |
| Gender | -0.2787 | 0.0608 | 21.0197 | 0.0001 |
| Marital Status | -0.2990 | 0.0747 | 16.0170 | 0.0001 |
| White | -0.5906 | 0.1771 | 11.1271 | 0.0009 |
| Black | -0.1246 | 0.1908 | 0.4267 | 0.5136 |
| Model $1 \rightarrow > 1$ | | | | |
| Constant | 1.8606 | 0.2249 | 68.4430 | 0.0001 |
| Gender | -0.1792 | 0.0826 | 4.7077 | 0.0300 |
| Marital Status | -0.5245 | 0.0932 | 31.6490 | 0.0001 |
| White | -0.5868 | 0.2171 | 7.3038 | 0.0069 |
| Black | -0.2998 | 0.2315 | 1.6763 | 0.1954 |
| Likelihood Ratio (p-value) | | 1210.549 (0.000) | | |

**Note:** Gender: male=1, female=0; Marital Status: married/partnered, 0=single/divorced/separated; White: yes=1, no=0; Black: yes=1, no=0.

are specified in order to examine the dependence between the current and the previous outcomes. The relationships between current outcome, previous outcome and covariate can have the following types: (i) both previous outcome and covariate are conditionally independent, (ii) one of the previous outcome or covariate is conditionally independent, and (iii) none of the two is conditionally independent. The regressive models can be applied for cases (i) and (ii) but fails to provide any feasible model for case (iii). Both simulations and real life applications clearly indicate the utility of the proposed tests.

# 9 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. We would like to express our deep gratitude to Professor Rahul Mukerjee for his comments on a previous version of the paper.

# References

[1] Bonney, G.E (1987). Logistic regression for dependent binary observations. *Biometrics,* **43***, 951-973.*

[2] Prentice, R.L (1988). Correlated binary regression with covariates specific to each binary observation. *Biometrics,* **44***, 1033-1048.*

[3] Zeger, S.L. Qaqish, B (1988). Markov regression models for time series: a quasi-likelihood approach. *Biometrics,* **44***, 1019-1031.*

[4] Neuhaus, J.M. Kalbfleisch, J.D. Hauck, W.W (1991). A comparison of duster-specific and population avenged approaches for analyzing correlated binary data. *International Statistical Review,* **59***, 25-35.*

[5] K.Y. Liang, S.L. Zeger, Qaqish, B (1992). Multivariate regression-analyses for categorical data. *Journal of Royal Statistical Society Series B. Statistical Methodology,* **54***, 3-40.*

[6] McDonald, W.B (1993). Estimating logistic regression parameters for bivariate binary Data. *Journal of Royal Statistical Society Series B. Statistical Methodology,* **55***, 391-397.*

[7] Le Cessie, S. van Houwelingen, J.C (1994). Logistic regression for correlated binary data. *Journal of Royal Statistical Society Series C. Applied Statistics,* **43***, 95-108.*

[8] Solis-Trapala, I.L. Carthey, J. Farewell, V.T. de Leval, M.R (2007). Dynamic modelling in a study of surgical error management. *Statistics in Medicine,* **26***, 5189–5202.*

[9] Muenz, L.R. Rubinstein, L.V (1985). Markov models for covariate dependence of binary sequences. *Biometrics,* **41***, 91-101.*

[10] Bonney, G.E (1986). Regressive logistic models for familial disease and other binary trials. *Biometrics,* **42***, 611-625.*

[11] Azzalini, A (1994). Logistic regression for autocorrelated data with application to repeated measures. *Biometrika,* **81***, 767-775.*

[12] Islam, M.A. Chowdhury, R.I (2006). A higher-order Markov model for analyzing covariate dependence. *Applied Mathematical Modelling,* **30***, 477-488.*

[13] Islam, M.A. Chowdhury, R.I (2007). *First and higher order transition models with covariate dependence*. In progress in applied mathematical modeling, F. Yang (ed), Nova Science Publishers Inc., New York, pp. 153-198.

[14] Islam, M.A. Chowdhury, R.I (2010). Prediction of disease status: a regressive model approach for repeated measures. *Statistical Methodology,* **7** *520-540.*

[15] Islam, M.A. Chowdhury, R.I. Huda, S. *Markov models with covariate dependence for repeated measures*. Nova Science Publishers, Inc., New York, 2009.

[16] Edwards, D. *Introduction to graphical modelling ($2^{nd}$ Edition)*. Springer, New York, 2000.

[17] Leisch, F. Weingessel, A. Hornik, K. *On the generation off correlated artificial binary data. working paper series*. Working paper No. 13, Vienna University of Economics and Business Administration, August 2-6, 1090 Wien, Austria, 1998.

[18] *Health and Retirement Study (HRS), (Wave [1-2]/Year [1992-1994]).* Public use dataset. Produced and distributed by the University of Michigan with funding from the National Institute on Aging (grant number NIA U01AG09740). Ann Arbor, MI, 2002.

[19] Steffick, D.E. *Documentation of Affective Functioning Measures in the Health and Retirement Study*. Survey Research Center University of Michigan Ann Arbor, MI, 2000.

[20] Cheng, S.T. Chan, A.C.M. Fung, H.H (2006). Factorial structure of a short version of the Center for Epidemiologic Studies Depression Scale. *International Journal of Geriatric Psychiatry,* **21***, 333-336*.

[21] Wasylenki, D (1980). Depressions in the elderly. *Canadian Medical Association Journal,* **122** *525-532*.

[22] Evans, M. Mottram, P (2000). Diagnosis of depression in elderly patients. *Advances in Psychiatric Treatment,* **6***, 49-56*.