A Comparison of Three Models in Multivariate Binary Longitudinal Data

Hissah Alzahrani

Longitudinal Data Analysis

Longitudinal data structure

$$
\begin{array}{cccc}
ID & Y_{it} & time & \begin{matrix} X_{it1} & X_{it2} & \cdots & X_{itC} \end{matrix} \\
\begin{pmatrix} 1 \\ 1 \\ \vdots \\ 2 \\ 2 \\ \vdots \\ \vdots \\ N \\ N \\ \vdots \\ N \end{pmatrix} &
\begin{pmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{21} \\ y_{22} \\ \vdots \\ \vdots \\ y_{N1} \\ y_{N2} \\ \vdots \\ y_{NT} \end{pmatrix} &
\begin{pmatrix} 1 \\ 2 \\ \vdots \\ 1 \\ 2 \\ \vdots \\ \vdots \\ 1 \\ 2 \\ \vdots \\ T \end{pmatrix} &
\begin{pmatrix}
x_{111} & x_{112} & \cdots & x_{11C} \\
x_{121} & x_{122} & \cdots & x_{12C} \\
\vdots & \vdots & \ddots & \vdots \\
x_{211} & x_{212} & \cdots & x_{21C} \\
x_{221} & x_{222} & \cdots & x_{22C} \\
\vdots & \vdots & \ddots & \vdots \\
\vdots & \vdots & \ddots & \vdots \\
x_{N11} & x_{N12} & \cdots & x_{N1C} \\
x_{N21} & x_{N22} & \cdots & x_{N2C} \\
\vdots & \vdots & \ddots & \vdots \\
x_{NT1} & x_{NT2} & \cdots & x_{NTC}
\end{pmatrix}
\end{array}
$$

Multivariate longitudinal data structure

$$
\begin{array}{c}
ID \\
\begin{pmatrix} 1 \\ 1 \\ \vdots \\ 2 \\ 2 \\ \vdots \\ \vdots \\ N \\ N \\ \vdots \\ N \end{pmatrix}
\end{array}
\quad
\begin{array}{c}
\begin{array}{cccc} Y_{it1} & Y_{it2} & \cdots & Y_{itK} \end{array} \\
\begin{pmatrix}
y_{111} & y_{112} & \cdots & y_{11K} \\
y_{121} & y_{122} & \cdots & y_{12K} \\
\vdots & \vdots & \ddots & \vdots \\
y_{211} & y_{212} & \cdots & y_{21K} \\
y_{221} & y_{222} & \cdots & y_{22K} \\
\vdots & \vdots & \ddots & \vdots \\
\vdots & \vdots & \ddots & \vdots \\
y_{N11} & y_{N12} & \cdots & y_{N1K} \\
y_{N21} & y_{N22} & \cdots & y_{N2K} \\
\vdots & \vdots & \ddots & \vdots \\
y_{NT1} & y_{NT2} & \cdots & y_{NTK}
\end{pmatrix}
\end{array}
\quad
\begin{array}{c}
time \\
\begin{pmatrix} 1 \\ 2 \\ \vdots \\ 1 \\ 2 \\ \vdots \\ \vdots \\ 1 \\ 2 \\ \vdots \\ T \end{pmatrix}
\end{array}
\quad
\begin{array}{c}
\begin{array}{cccc} X_{it1} & X_{it2} & \cdots & X_{itC} \end{array} \\
\begin{pmatrix}
x_{111} & x_{112} & \cdots & x_{11C} \\
x_{121} & x_{122} & \cdots & x_{12C} \\
\vdots & \vdots & \ddots & \vdots \\
x_{211} & x_{212} & \cdots & x_{21C} \\
x_{221} & x_{222} & \cdots & x_{22C} \\
\vdots & \vdots & \ddots & \vdots \\
\vdots & \vdots & \ddots & \vdots \\
x_{N11} & x_{N12} & \cdots & x_{N1C} \\
x_{N21} & x_{N22} & \cdots & x_{N2C} \\
\vdots & \vdots & \ddots & \vdots \\
x_{NT1} & x_{NT2} & \cdots & x_{NTC}
\end{pmatrix}
\end{array}
$$

Longitudinal Data Analysis

Mixed effect models

Marginal models

Transition models

Multivariate longitudinal data structure

The sources of the correlation within each subject:

- Within outcome over time
- Within time among outcomes
- Across outcomes and times

$$
\begin{array}{c}
ID \\
\begin{pmatrix} 1 \\ 1 \\ \vdots \\ 2 \\ 2 \\ \vdots \\ \vdots \\ N \\ N \\ \vdots \\ N \end{pmatrix}
\end{array}
\quad
\begin{array}{c}
\begin{array}{cccc} Y_{it1} & Y_{it2} & \cdots & Y_{itK} \end{array} \\
\begin{pmatrix}
y_{111} & y_{112} & \cdots & y_{11K} \\
y_{121} & y_{122} & \cdots & y_{12K} \\
\vdots & \vdots & \ddots & \vdots \\
y_{211} & y_{212} & \cdots & y_{21K} \\
y_{221} & y_{222} & \cdots & y_{22K} \\
\vdots & \vdots & \ddots & \vdots \\
\vdots & \vdots & \ddots & \vdots \\
y_{N11} & y_{N12} & \cdots & y_{N1K} \\
y_{N21} & y_{N22} & \cdots & y_{N2K} \\
\vdots & \vdots & \ddots & \vdots \\
y_{NT1} & y_{NT2} & \cdots & y_{NTK}
\end{pmatrix}
\end{array}
$$

Research Aims

1) Reducing many outcomes into one summary outcome which requires a unique set of regression coefficients → (O'Brien and Fitzmaurice, 2004)
2) Analyzing each outcome separately which requires a set of regression coefficients for each outcome → (Shelton et al., 2004)
3) Analyzing the outcomes jointly which requires a set of regression coefficients for each outcome and the joint between them → The proposed method

Research Aims

1) Reducing many outcomes into one summary outcome which requires a unique set of regression coefficients ➡ $g(E(Y_{itk})) = X_{it}\,\beta$
2) Analyzing each outcome separately which requires a set of regression coefficients for each outcome ➡ $g(E(Y_{itk})) = \sum_{k=1}^{K} X_{itk}\,\beta_k$
3) Analyzing each outcome jointly which requires a set of regression coefficients for each outcome and the joint between them ➡ The proposed method

The Proposed Method

Let $Y_1 \sim$ Bernolli(π_1), $Y_2 \sim$ Bernolli(π_2), …. $Y_K \sim$ Bernolli(π_K)

$$\begin{matrix} Y_{it1} & Y_{it2} & \cdots & Y_{itK} \end{matrix}$$

$$\begin{pmatrix} y_{111} & y_{112} & \cdots & y_{11K} \\ y_{121} & y_{122} & \cdots & y_{12K} \\ \vdots & \vdots & \ddots & \vdots \\ y_{211} & y_{212} & \cdots & y_{21K} \\ y_{221} & y_{222} & \cdots & y_{22K} \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ y_{N11} & y_{N12} & \cdots & y_{N1K} \\ y_{N21} & y_{N22} & \cdots & y_{N2K} \\ \vdots & \vdots & \ddots & \vdots \\ y_{NT1} & y_{NT2} & \cdots & y_{NTK} \end{pmatrix}$$

Encoding

$$Z_{it}$$

$$\begin{pmatrix} z_{11} \\ z_{12} \\ \vdots \\ z_{21} \\ z_{22} \\ \vdots \\ \vdots \\ z_{11} \\ z_{21} \\ \vdots \\ z_{NT} \end{pmatrix}$$

It is a transformation from binary coding system to decimal coding system 2^K

The Proposed Method

$$\begin{array}{ccc} Y_{it1} & Y_{it2} & Y_{itk} \end{array}$$
$$\begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 0 & 0 & 1 \\ 1 & 1 & 1 \\ \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots \\ 1 & 1 & 0 \\ 0 & 1 & 1 \\ \vdots & \vdots & \vdots \\ 1 & 0 & 1 \end{pmatrix}$$

Using 2^3 transformation →

$$Z_{it}$$
$$\begin{pmatrix} 3 \\ 4 \\ . \\ 1 \\ 7 \\ . \\ . \\ 6 \\ 3 \\ . \\ 5 \end{pmatrix}$$

The Proposed Method

$$Z_{it}\quad \begin{pmatrix} z_{11} \\ z_{12} \\ \vdots \\ z_{21} \\ z_{22} \\ \vdots \\ \vdots \\ z_{11} \\ z_{21} \\ \vdots \\ z_{NT} \end{pmatrix}$$

Decoding →

$$\begin{matrix} Y_{it1} & Y_{it2} & \cdots & Y_{itK} \end{matrix}$$

$$\begin{pmatrix} y_{111} & y_{112} & \cdots & y_{11K} \\ y_{121} & y_{122} & \cdots & y_{12K} \\ \vdots & \vdots & \ddots & \vdots \\ y_{211} & y_{212} & \cdots & y_{21K} \\ y_{221} & y_{222} & \cdots & y_{22K} \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ y_{N11} & y_{N12} & \cdots & y_{N1K} \\ y_{N21} & y_{N22} & \cdots & y_{N2K} \\ \vdots & \vdots & \ddots & \vdots \\ y_{NT1} & y_{NT2} & \cdots & y_{NTK} \end{pmatrix}$$

The Comparison

Encoding Step:

The new variable $Z_{it} = 2^0 \times Y_{i1} + 2^1 \times Y_{i2} = 2Y_{i2} + Y_{i1}$. The new variable $Z \in 0, 1, 2, 3$

$$Z_{it} = \begin{cases} 0 & \text{when} \quad (Y_{i1} = 0) \,\&\, (Y_{i2} = 0) \\ 1 & \text{when} \quad (Y_{i1} = 0) \,\&\, (Y_{i2} = 1) \\ 2 & \text{when} \quad (Y_{i1} = 1) \,\&\, (Y_{i2} = 0) \\ 3 & \text{when} \quad (Y_{i1} = 1) \,\&\, (Y_{i2} = 1) \end{cases}$$

Application

- The Florida Dental Care Study (FDCS)
- The sample size is 873 subjects
- There are three binary outcomes that were measured over four time intervals (0-6, 6-12, 12-18, 18-24 months)

Y_1= Problem oriented visit , Y_2=Check up, Y_3=Dental cleaning

Application

Covariate	Definition
IRA	(1) The subject goes to a dentist regularly or occasionally whether or not has a problem, (0) The subject did go to a dental check-up once a year or more often in the previous 5 years.
Gender	(1) Female, (0) Male.
Cavit	(1) The subject reported having cavities (tooth decay) in the previous 6 months, (0) if not.
Loose	(l) The subject had a loose tooth, (0)had not
Able	(1) The subject able to pay unexpected US$ 500 dental bill, but with difficulty, (0) not able to pay

Application

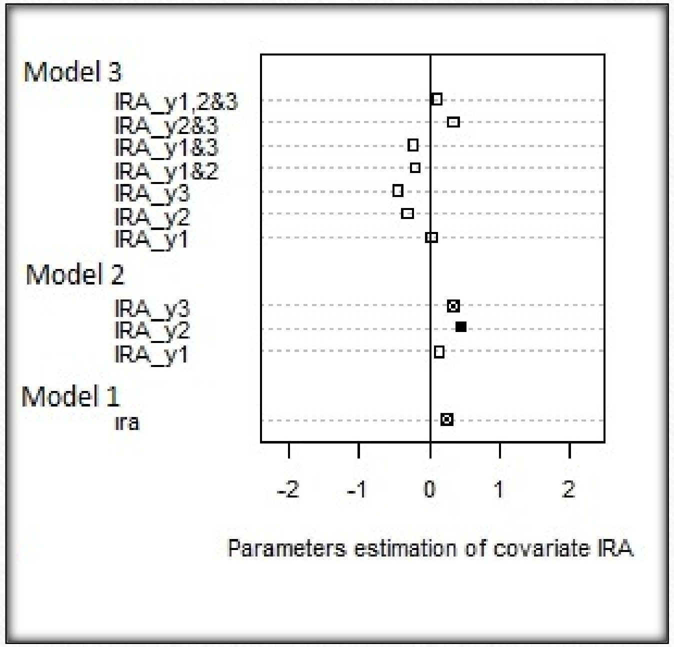
Model 1: $log(E(Y_{itk})) = \beta_0 + \beta_1 X_{IRA} + \beta_2 X_{gender} + \beta_3 X_{cavit} + \beta_4 X_{loose} + \beta_5 X_{able}$

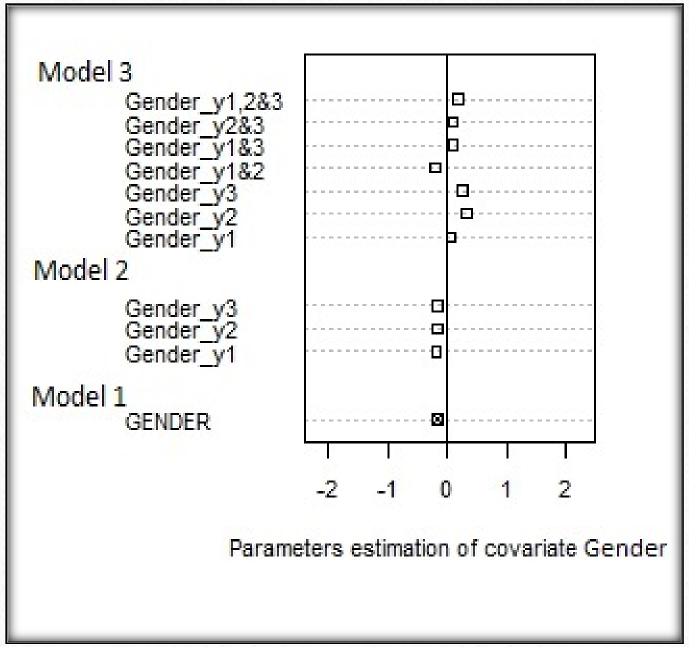
Model 2 : $log(E(Y_{itk})) = \sum_{k=1}^{K} \beta_{k0} + \beta_{k1} X_{IRA} + \beta_{k2} X_{gender} + \beta_{k3} X_{cavit} + \beta_{k4} X_{loose} + \beta_{k5} X_{able}$

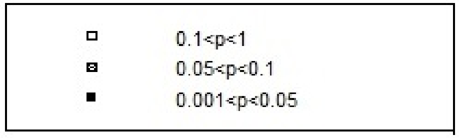
Model 3: $logit(E(Z_{it})) = \sum_{j=1}^{J} \beta_{j0} + \beta_{j1} X_{IRA} + \beta_{j2} X_{gender} + \beta_{j3} X_{cavit} + \beta_{j4} X_{loose} + \beta_{j5} X_{able}$

$$j = \begin{cases} 1 & \text{when} \quad (Y_{i1} = 0)\ \&\ (Y_{i2} = 0)\ \&\ (Y_{i3} = 1) \\ 2 & \text{when} \quad (Y_{i1} = 0)\ \&\ (Y_{i2} = 1)\ \&\ (Y_{i3} = 0) \\ 3 & \text{when} \quad (Y_{i1} = 0)\ \&\ (Y_{i2} = 1)\ \&\ (Y_{i3} = 1) \\ 4 & \text{when} \quad (Y_{i1} = 1)\ \&\ (Y_{i2} = 0)\ \&\ (Y_{i3} = 0) \\ 5 & \text{when} \quad (Y_{i1} = 1)\ \&\ (Y_{i2} = 0)\ \&\ (Y_{i3} = 1) \\ 6 & \text{when} \quad (Y_{i1} = 1)\ \&\ (Y_{i2} = 1)\ \&\ (Y_{i3} = 0) \\ 7 & \text{when} \quad (Y_{i1} = 1)\ \&\ (Y_{i2} = 1)\ \&\ (Y_{i3} = 1) \end{cases}$$

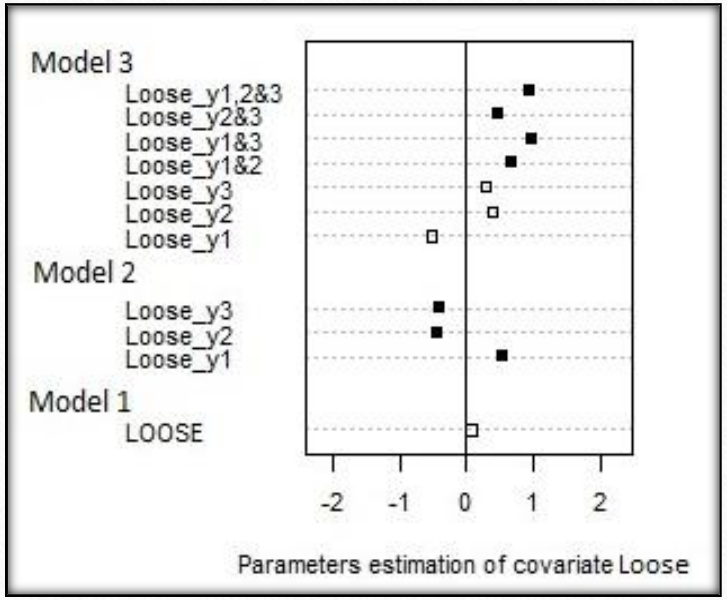
Application

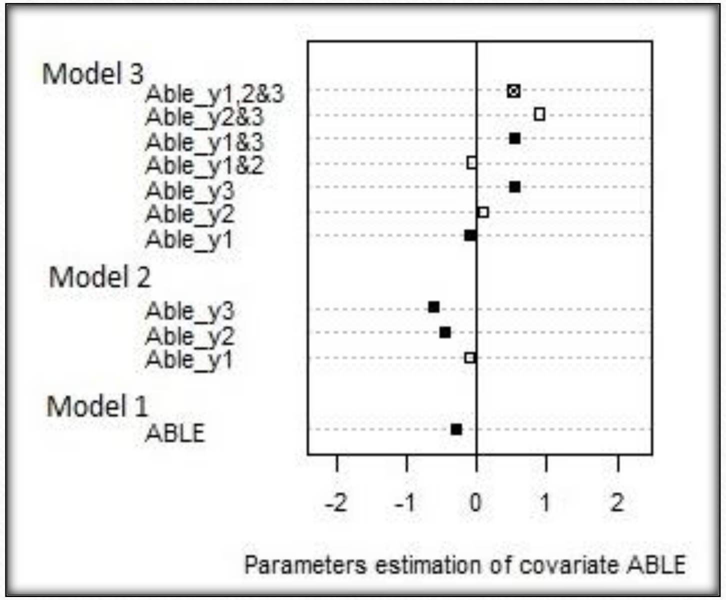


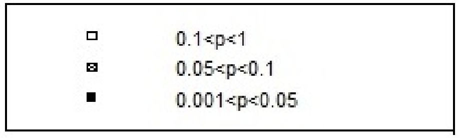




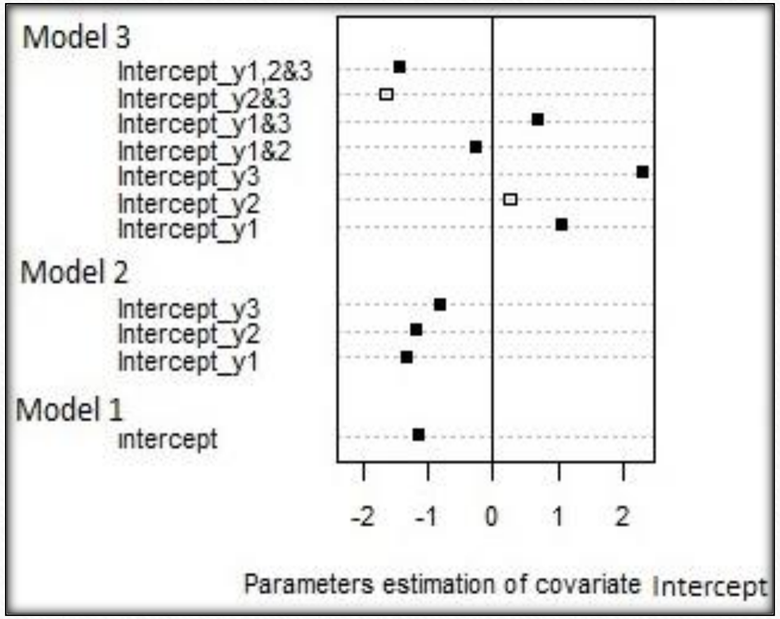
Application

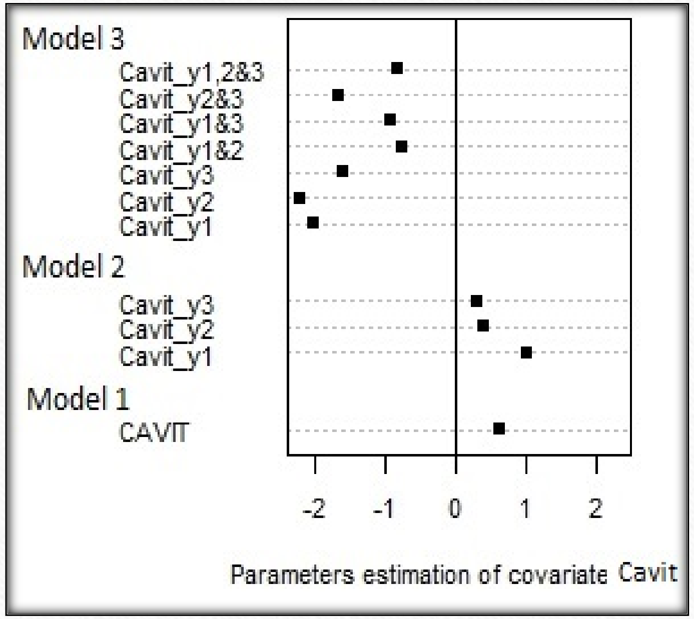


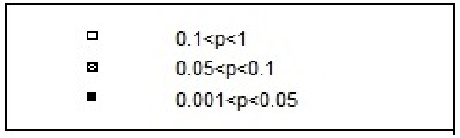




Application







Application

- The effect of each covariates on the responses varies over the three models.
- The most significant covariates over the three models are Cavity, Loose tooth, and the ability to pay an unexpected US$500 dental bill, but with difficulty.

Conclusion

Model 1: The estimated covariates effects are one for all outcomes.

➡ less parameters while accounting for the multivariate structure.

Model 2: The covariates effects are estimated for each outcomes.

➡ analyze the effects of the covariates on the outcomes separately

Model 3: Estimating the effects of the covariates for each joint case of all the outcomes.

➡ deeper analysis for the joint distribution of the multivariate longitudinal data.

Conclusion

Method 3 Cons:

- Limited to a small number of outcomes
- Estimation a lot of parameters especially when there are a lot of covariates.

Thank you