SAS ユーザー総会 2016

欠測を含む順序カテゴリカル経時データの解析 −GEEプロシジャの有用性−

駒嵜弘[1]、藤原正和[2]
（[1]マルホ株式会社、[2]塩野義製薬株式会社）

Ordinal longitudinal data analysis with missing data

-Usefulness of Proc GEE-

Hiroshi Komazaki[1],Masakazu Fujiwara[2]
[1]Maruho Co, Ltd., [2]Shionogi & Co., Ltd.

2発表の構成

前発表

- カテゴリカル経時データの発生
- ↓
- 欠測データの発生
- ↓
- 多重補完（REG、MCMC、FCS）
- ↓

本発表

- 解析（CATMOD、GEE）

要旨：

順序カテゴリカル経時データの解析方法としてGEEプロシジャによる方法及びCATMODプロシジャによる方法をそれぞれ紹介するとともに、シミュレーションデータを用いて性能評価を行う。

キーワード： CATMODプロシジャ、GEEプロシジャ、Generalized Estimating Equations (GEE)、alternating logistic regression(ALR)

Outline

- ✓ 順序カテゴリカル経時データの解析仕様
  （ CATMODプロシジャ、GEEプロシジャ）
- ✓ シミュレーション結果
- ✓ 考察

Background 及びMotivation

順序カテゴリカル経時データに対する解析方法

PROC	特徴
CATMOD	重み付き最小二乗法（スライド11の式参照）
GENMOD	GEEによる推定 相関構造はINDしか使用できず （ただしロバスト分散の使用が可）
GEE	GEEによる推定 相関構造はEXCHが指定可能（スライド8、14の式参照）

SAS 9.4(stat 14.1)より、PROC GEEは応答変数が順序カテゴリカルデータに対しても実行できるようDIST=MULTINOMIALが、

さらに相関構造も指定できるようLOGOR= EXCHが追加された。

→PROC CATMODとPROC GEEを用いてカテゴリカル経時データの解析方法を紹介し、さらに性能評価を行う。

順序カテゴリカル経時データにおいて、
CATMODプロシジャでの解析方法とGEEプロシジャでの解析方法は
Liang and Zeger (1986)及びPrentice(1988)のGEE法に類似。

Liang and Zeger (1986)及びPrentice(1988)のGEE法との相違点

CATMODプロシジャ	GEEプロシジャ
重み付き最小二乗法 分散構造はsaturated correlation structure	相関パラメータαの推定方程式を改良

Liang and Zeger (1986)及びPrentice(1988)のGEE法を提示後、
それぞれのプロシジャでの推定方程式を紹介する。

＊順序カテゴリカルデータを表記する上での留意点

本発表にて、順序カテゴリカルデータは指示変数（indicator variable）を用いて表記

例）5カテゴリの応答変数の場合　$y_{pit} = I(O_{pit} \le c)$　$c \in [1,2,3,4]$

Yのカテゴリ	指示変数 $(y_{pit1} \quad y_{pit2} \quad y_{pit3} \quad y_{pit4})$
1	$(1 \quad 0 \quad 0 \quad 0)$
2	$(1 \quad 1 \quad 0 \quad 0)$
3	$(1 \quad 1 \quad 1 \quad 0)$
4	$(1 \quad 1 \quad 1 \quad 1)$
5	$(0 \quad 0 \quad 0 \quad 0)$

P	：群
i	：被験者
t	：時点

順序カテゴリカルデータのGEE(Liang and Zeger (1986)、Prentice(1988))
(主効果 β の推定方程式)

$$U(\beta)=\sum_{p=0}^{1} D'_p V(\alpha)_p^{-1}(Y_p-\hat{\pi}_p)=0$$

$\hat{\beta}_{0k}$:カテゴリkの切片

X :デザイン行列
(群、時点、群×時点)

$Y_p : y_{pitk}$ 各群(p=0,1)、各症例、各時点(t=1,2,3,4)の指示変数

例:症例iの指示変数 $y_{pit}=\{y_{pit1} \quad y_{pit2} \quad y_{pit3} \quad y_{pit4}\}$

$\hat{\pi}_p$ Y_p の期待値 $\left(\phi_{pitk}=L(\hat{\pi}_{pitk})=\ln\left[\dfrac{\hat{\pi}_{pit1}+\hat{\pi}_{pit2}+\cdots+\hat{\pi}_{pitk}}{\hat{\pi}_{pit(k+1)}+\cdots+\hat{\pi}_{pit5}}\right]=\hat{\beta}_{0k}+X'\beta\right)$

累積ロジットモデル

D'_p $\partial\pi_p/\partial\beta$

$V(\alpha)_p$ 各群(p=0,1)の各時点、各カテゴリ間の分散共分散行列

ロジットモデル

$$\ln\left[\frac{\hat{\pi}_{pit}}{1-\hat{\pi}_{pit}}\right]=\beta_0+X'\beta$$

累積ロジットモデル
カテゴリ数:5の場合

$$\ln\left[\frac{\hat{\pi}_{pit1}+\hat{\pi}_{pit2}+\cdots+\hat{\pi}_{pitk}}{\hat{\pi}_{pit(k+1)}+\cdots+\hat{\pi}_{pit5}}\right]=\hat{\beta}_{0k}+X'\beta$$

$$\ln\left[\frac{\hat{\pi}_{pit1}}{\hat{\pi}_{pit2}+\hat{\pi}_{pit3}+\hat{\pi}_{pit4}+\hat{\pi}_{pit5}}\right] \quad \ln\left[\frac{\hat{\pi}_{pit1}+\hat{\pi}_{pit2}}{\hat{\pi}_{pit3}+\hat{\pi}_{pit4}+\hat{\pi}_{pit5}}\right]$$

各カテゴリにて、共変量βは同じと仮定
→比例オッズモデル

$$\ln\left[\frac{\hat{\pi}_{pit1}+\hat{\pi}_{pit2}+\hat{\pi}_{pit3}}{\hat{\pi}_{pit4}+\hat{\pi}_{pit5}}\right] \quad \ln\left[\frac{\hat{\pi}_{pit1}+\hat{\pi}_{pit2}+\hat{\pi}_{pit3}+\hat{\pi}_{pit4}}{\hat{\pi}_{pit5}}\right]$$

順序カテゴリカルデータのGEE
（相関パラメータ α の推定方程式）

$$U(\alpha)=\sum_{p=0}^{1}E_p'W_p^{-1}(Z_p-\hat{\eta}_p(\alpha))=0$$

Z_p　相関係数　$Z_{p(tk)(t'k')}=\dfrac{(y_{pitk}-\hat{\pi}_{pitk})(y_{pit'k'}-\hat{\pi}_{pit'k'})}{\left[\hat{\pi}_{pitk}(1-\hat{\pi}_{pitk})\hat{\pi}_{pit'k'}(1-\hat{\pi}_{pit'k'})\right]^{1/2}}$　Miller(1993)

$\hat{\eta}_p$　Z_pの期待値　$\eta_{pi(tk)(t'k')}(\alpha)=\dfrac{\exp\left(f_{pi(tk)(t'k')}\alpha\right)-1}{\exp\left(f_{pi(tk)(t'k')}\alpha\right)+1}$

E_p'　$\partial\eta_p(\alpha)/\partial\alpha$　（ $f_{pi(tk)(t'k')}$： X_{pit} と $X_{pit'}$ の関数）

W_p　$Z_{pi(tk)(t'k')}$ の分散共分散行列

順序カテゴリカルデータのGEE

推定量 $\hat{\beta}$

$$\hat{\beta}^{(t+1)} = \underline{\left[\sum_{p=0}^{1} X'_p G_p^t X_p\right]^{-1} \left[\sum_{p=0}^{1} X'_p G_p^t \right.}\left.\left(X_p \beta^t + \left(\frac{\partial \phi_p}{\partial \pi_p}\right)^t \left(p_p - \pi_p^{(t)}\right)\right)\right]$$

CATMODプロシジャの
推定量

$$\boxed{G_p^t = n_p \left(\frac{\partial \pi_p^t}{\partial \phi_p}\right)' \left(V(\alpha)_p^t\right)^{-1} \left(\frac{\partial \pi_p^t}{\partial \phi_p}\right)}$$

分散共分散構造がsaturated correlation structure*のとき、
CATMODの推定量 $\hat{\beta}$ は、上式の推定量の赤線部分に対応している(Miller,1993)

*saturated correlation structure
→群間、時点間、カテゴリ間で異なる相関パラメータαを指定

☆PROC GEEではHeagerty and Zeger(1996)の方法を使用

Liang and Zeger (1986)、Prentice(1988)のGEE

$\hat{\beta}$ の推定方程式(スライド8)

相関パラメータ $\hat{\alpha}$ の推定方程式(スライド10)

Heagerty and Zeger(1996)のGEE

$\hat{\beta}$ の推定方程式(スライド8)

相関パラメータ $\hat{\alpha}$ の推定方程式(スライド14)

$$\xi_{pi(t,t')(c_1,c_2)} = E\left(y_{itc_1} \middle| y_{it'c_2}\right)$$

GEEプロシジャの順序カテゴリカルデータのGEE法は、Heagerty and Zeger(1996)の手法に基づいている。

→相関パラメータ算出の際、alternating logistic regression(ALR)を使用

✓ 主効果 β の推定方程式はLiang and Zeger (1986) 、Prentice(1988)のGEEと同じ

✓ 相関パラメータ α は、時点間、カテゴリ間の条件付き期待値を利用して推定

→条件付き期待値の回帰パラメータが**対数オッズ比**となる(スライド17参照)

$$\xi_{pi(t,t')(c_1,c_2)} = E\left(y_{itc_1} \middle| y_{it'c_2}\right)$$

ALRによる相関パラメータ α の推定方程式

$$\boxed{U(\alpha)=\sum_{p=0}^{1}\left[\frac{\partial\xi_p}{\partial\alpha}\right]^t V_p^{-1}(Y_p^*-\xi_p)=0}$$

$$y_{pit}^* = \left(y_{p1}\otimes 1_4,\ldots,y_{p(n_i-1)}\otimes 1_4\right)$$

各群(p=0,1)の各時点(t=1,2,3,4)の指示変数

各データを、
（全時点-当該時点）個分、(カテゴリ数-1)個分、
重複生成している（→詳細は次ページ参照）

$$\xi_{p(t,t')(c_1,c_2)} = E\left(y_{tc_1}\middle| y_{t'c_2}\right)$$

時点t'でカテゴリc2が得られた条件のもとで、
時点tでカテゴリc1が得られる期待値（t'>t）

$$V_{pi} = diag\left[\xi_{pi}(1-\xi_{pi})\right]$$

例：説明の簡略化のため時点数t：**3**、カテゴリ数k:**3**を想定。

以下の2時点のデータが得られたとする（時点3のデータは説明変数で使用し、応答変数には用いないため省略）

	時点1 y_{i1}	**時点2** y_{i2}
カテゴリ	1	2
指示変数	(1,0)	(1,1)

n_i ：被験者iの時点数

各時点、各カテゴリごとの条件付き期待値を求めるにあたり、応答変数は以下の通りデータを重複発生する必要がある。

$$y_i^* = \Big(\overbrace{y_{i1}\otimes 1_{(k-1)},\ldots,y_{i1}\otimes 1_{(k-1)}}^{n_i-1},\overbrace{y_{i2}\otimes 1_{(k-1)},\ldots,y_{i2}\otimes 1_{(k-1)}}^{n_i-2},\ldots\ldots,\overbrace{y_{i(n_i-1)}\otimes 1_{(k-1)}}^{1}\Big)'$$

$$y_i^* = \left[y_{i1}\otimes 1_2 \quad y_{i1}\otimes 1_2 \quad y_{i2}\otimes 1_2\right]$$
$$= \left[(1,1,0,0) \quad (1,1,0,0) \quad (1,1,1,1)\right]$$

例 : 各時点、各カテゴリごとの条件付き期待値の構造は以下の通り

$$\xi_i = \begin{bmatrix} \xi_{12} & \xi_{13} & \xi_{23} \end{bmatrix}$$
$$= \begin{bmatrix} \underline{\xi_{12(11)}} & \xi_{12(12)} & \xi_{12(21)} & \xi_{12(22)} & \xi_{13(11)} & \xi_{13(12)} & \xi_{13(21)} & \xi_{13(22)} & \xi_{23(11)} & \xi_{23(12)} & \xi_{23(21)} & \xi_{23(22)} \end{bmatrix}$$

時点2で1が得られた条件のもと、時点1で1が得られる期待値

従って本例の場合、残差 $y_p^* - \xi_p$ は12個のデータで構成される

$$y_i^* = \begin{bmatrix} y_{i1} \otimes 1_2 & y_{i1} \otimes 1_2 & y_{i2} \otimes 1_2 \end{bmatrix}$$
$$= \begin{bmatrix} (1,1,0,0) & (1,1,0,0) & (1,1,1,1) \end{bmatrix}$$

……

$$\xi_i = \begin{bmatrix} \xi_{12} & \xi_{13} & \xi_{23} \end{bmatrix}$$
$$= \begin{bmatrix} \xi_{12(11)} & \xi_{12(12)} & \xi_{12(21)} & \xi_{12(22)} & \xi_{13(11)} & \xi_{13(12)} & \xi_{13(21)} & \xi_{13(22)} & \xi_{23(11)} & \xi_{23(12)} & \xi_{23(21)} & \xi_{23(22)} \end{bmatrix}$$

$\xi_{p(t,t')(c_1,c_2)} = E\left(y_{tc_1} \middle| y_{t'c_2}\right)$ のモデル

対数オッズ比　　二項目は切片

$$\log it\left[E\left(Y_{itc_1} \mid Y_{it'c_2}\right)\right] = \hat{\gamma}_{i(tt')(c_1c_2)} Y_{it'c_2} + \log\left[\frac{\mu_{itc_1} - E\left(Y_{itc_1} Y_{it'c_2}\right)}{1 - \mu_{itc_1} - \mu_{it'c_2} + E\left(Y_{itc_1} Y_{it'c_2}\right)}\right]$$

$$\gamma_{i(tt')(c_1c_2)} = \log\left(OR\left(Y_{itc_1}, Y_{it'c_2}\right)\right) = Z^t_{i(tt')(c_1c_2)}\alpha = \log\left(\frac{1+\rho_{i(tt')(c_1c_2)}}{1-\rho_{i(tt')(c_1c_2)}}\right)$$

時点間、カテゴリ間の対数オッズ比より相関パラメータ α を推定

→相関係数 $\rho_{i(tt')(c_1c_2)}$ が求まる（$Z_i^t = 1$ のとき相関構造はexchangeable）

数式の詳細は、Heagerty and Zeger(1996)、Lipsitz(1991)、SAS HELP参照

シミュレーションの流れ

カテゴリカル経時データの発生

↓

欠測データの発生

↓

多重補完（REG、MCMC、FCS）

↓

解析（CATMOD、GEE）

シミュレーションデータ

Donneau(2015)及びIbrahim and Suliadi(2011)を参考にシミュレーションデータを作成

$$\log it\left[\Pr\left(Y_{ij} \le k \middle| x_i, t\right)\right] = \beta_{0k} + \beta_x x_i + \beta_t I(t=j) + \beta_{tx} x_i I(t=j)$$

i :被験者　j :時点(1,2,3,4)　k :応答変数のカテゴリ(1,2,3,4)

y :応答変数

x :投与群(0:プラセボ、1:実薬)

t :時点(1,2,3,4)

パラメータ真値	
β_{01}	-1.39
β_{02}	-0.41
β_{03}	0.41
β_{04}	1.39
β_x	0.5
β_t	-0.4、-0.2、-0.1、0
β_{xt}	-0.4、-0.2、-0.1、0

シミュレーション回数:1000回

シミュレーションデータ
各時点、各カテゴリの周辺分布

プラセボ群

%

K	VISIT 1	VISIT 2	VISIT 3	VISIT 4
1	14	17	18	20
2	16	18	19	20
3	19	20	20	20
4	23	21	21	20
5	27	23	22	20

実薬群

%

K	VISIT 1	VISIT 2	VISIT 3	VISIT 4
1	16	22	25	29
2	17	21	22	23
3	20	20	20	19
4	22	19	17	16
5	25	18	16	13

シミュレーションデータ

観測確率モデル

$$\mathrm{logit}(p_t) = 2 - 0.1 \cdot Y_{t-1}$$

- ✓ 1時点前のデータが欠測の有無に影響
- ✓ 各群、最終時点までに40%脱落

CATMODプロシジャの指定方法

データセットの構造

時点1～時点4のカテゴリのパターン

sim	Imputation Number	arm	_1	_2	_3	_4	count
1	1	0	1	1	1	1	0.01
1	1	0	1	1	1	2	1
1	1	0	1	1	1	3	0.01
1	1	0	1	1	1	4	0.01
1	1	0	1	1	1	5	
:	:	:	:	:	:	:	:
1000	20	1	5	5	5	5	0.01

各時点の応答変数パターンを度数で格納
度数0のパターンは代わりに0.01を格納（値の妥当性は未検証）
→分散共分散行列のランク落ち、累積ロジスティックモデルの準完全分離を防ぐため。
ERROR: Logits cannot be computed. One of the marginal probabilities is equal to zero.
ERROR: The covariance matrix of the response functions is singular in population 1.

CATMODプロシジャの指定方法

```
proc catmod data=IN_DATA ;
  response    clogits;
  population   arm;
  weight        count ;

  model  _1*_2*_3*_4    = (1   0   0   0    0    1  0  0    0  0  0,
                            0   1   0   0    0    1  0  0    0  0  0,
                            0   0   1   0    0    1  0  0    0  0  0,
                            0   0   0   1    0    1  0  0    0  0  0,
                            ~~~~~~~~~~~~~~~~~~~~~~~~
                            ~~~~~~~~~~~~~~~~~~~~~~~~
                            1   0   0   0    0    0  0  0    0  0  0,
                            0   1   0   0    0    0  0  0    0  0  0,
                            0   0   1   0    0    0  0  0    0  0  0,
                            0   0   0   1    0    0  0  0    0  0  0,
                            1   0   0   0    1    1  0  0    1  0  0,
                            0   1   0   0    1    1  0  0    1  0  0,
                            0   0   1   0    1    1  0  0    1  0  0,
                            0   0   0   1    1    1  0  0    1  0  0,
                            ~~~~~~~~~~~~~~~~~~~~~~~~
                            ~~~~~~~~~~~~~~~~~~~~~~~~
                            1   0   0   0    1    0  0  0    0  0  0,
                            0   1   0   0    1    0  0  0    0  0  0,
                            0   0   1   0    1    0  0  0    0  0  0,
                            0   0   0   1    1    0  0  0    0  0  0 )
                       (1 2 3 4 = 'intercept',  5='arm' ,6 7 8='VISIT', 9 10 11='arm*VISIT') ;
run;
```

← 累積ロジットモデルの指定

← 投与群の指定

← 度数変数の指定

列1~4　：(応答変数のカテゴリ数-1)個の切片

列5　　：投与群

列6~8　：時点

列9~11　：投与群×時点の交互作用

GEEプロシジャの指定方法

データセットの構造

sim	Imputation Number	arm	ID	VISIT	AVAL
1	1	0	1	1	1
1	1	0	1	2	2
1	1	0	1	3	2
1	1	0	1	4	.
:	:	:	:	:	:
1000	20	1	100	4	4

応答変数の
順序カテゴリカルデータ
(AVAL:1,2,3,4,5)

被験者ごと、VISITごとに応答変数データを格納

GEEプロシジャの指定方法

```
proc gee data=IN_DATA;
  class    arm ID;
  model AVAL_M =visit  arm  visit*arm  / dist=multinomial;
  repeated subject=ID /  logor=exch;
 run;
```

累積ロジットモデルの指定

相関構造にexchangeableを指定

arm	:投与群(0 or 1)
visit	:時点(1,2,3,4)
Visit*arm	:投与群×時点の交互作用

結果1 : モデルを正しく特定ver

CATMOD

	真値	推定値(MSE)	BIAS	RB	αエラー
MCMC	0.5	0.418 (0.078)	0.082	83.6	2.80
REG	0.5	0.436 (0.090)	0.064	87.1	2.50
FCS	0.5	0.433 (0.091)	0.067	86.6	2.50
補完なし	0.5	0.420 (0.092)	0.08	84.0	2.70
完全データ	0.5	0.450 (0.059)	0.05	90.1	2.50

$$\frac{1}{1000}\sum_{i}^{1000}(\hat{\mu}-\mu) \qquad \frac{1}{1000}\sum_{i}^{1000}\frac{\hat{\mu}}{\mu}$$

GEE

	真値	推定値(MSE)	BIAS	RB	αエラー
MCMC	0.5	0.460 (0.086)	0.040	92.0	2.20
REG	0.5	0.479 (0.095)	0.021	95.8	2.60
FCS	0.5	0.477 (0.095)	0.023	95.4	2.00
補完なし	0.5	0.490 (0.094)	0.010	98.0	2.40
完全データ	0.5	0.497 (0.059)	0.003	99.3	2.30

結果2：相関構造誤特定ver

真の相関構造 $\begin{pmatrix} 1 & 0.6 & 0.4 & 0.2 \\ 0.6 & 1 & 0.6 & 0.4 \\ 0.4 & 0.6 & 1 & 0.6 \\ 0.2 & 0.4 & 0.6 & 1 \end{pmatrix}$

CATMOD

	真値	推定値(MSE)	BIAS	RB	αエラー
MCMC	0.5	0.448 (0.080)	-0.052	89.6	2.10
REG	0.5	0.462 (0.085)	-0.038	92.5	1.50
FCS	0.5	0.451 (0.086)	-0.049	90.2	1.80
補完なし	0.5	0.452 (0.090)	-0.048	90.3	1.80
完全データ	0.5	0.471 (0.058)	-0.029	94.3	3.00

GEE

	真値	推定値(MSE)	BIAS	RB	αエラー
MCMC	0.5	0.493 (0.094)	0.007	98.5	2.60
REG	0.5	0.509 (0.098)	0.009	101.8	1.70
FCS	0.5	0.497 (0.097)	0.003	99.3	1.40
補完なし	0.5	0.524 (0.108)	0.024	104.7	2.60
完全データ	0.5	0.518 (0.065)	0.018	103.6	2.70

結果3:欠測メカニズムがMNAR ver

CATMOD

	真値	推定値(MSE)	BIAS	RB	αエラー
MCMC	0.5	0.415 (0.120)	-0.085	82.9	2.00
REG	0.5	0.426 (0.146)	-0.074	85.1	2.60
FCS	0.5	0.422 (0.147)	-0.078	84.5	2.50
補完なし	0.5	0.396 (0.137)	-0.104	79.3	2.50

GEE

	真値	推定値(MSE)	BIAS	RB	αエラー
MCMC	0.5	0.463 (0.137)	-0.037	92.6	2.60
REG	0.5	0.475 (0.158)	-0.025	94.9	3.00
FCS	0.5	0.473 (0.157)	-0.027	94.6	2.30
補完なし	0.5	0.498 (0.158)	-0.002	99.6	2.50

結果のまとめ

- モデルを正しく特定できたとき
  - ✓ 性能：　GEE ≧ CATMOD
  - ✓ GEEでは、補完によりバイアス増加
- 相関構造を誤特定したとき
  - ✓ 性能：　GEE > CATMOD　→GEEの頑健性を確認できた。
  - ✓ GEEでは、補完によりバイアス減少
- 欠測メカニズムを誤特定したとき
  - ✓ 性能はGEE> CATMOD　→相関構造の誤特定よりはバイアス小
MSEはいずれも増加。
  - ✓ GEEでは、補完によりバイアス増加
- αエラーの増大(>>2.5)は特に確認できなかった。

発表のまとめ

- ✓ カテゴリカル経時データの解析方法の紹介（CATMOD、GEE）
- ✓ GEEプロシジャでは相関パラメータαを対数オッズ比より推定
- ✓ 性能：　GEE＞CATMOD
- ✓ GEEでは補完した方がバイアスが増大することもあった。
- ✓ 累積ロジットモデルのとき、Weighted-GEEによる解析はSASのPROC　GEEでは実装されていないため、今後の導入が期待される。

参考文献

Donneau, A. F., Mauer, M., Molenberghs, G., & Albert, A. (2015). A simulation study comparing multiple imputation methods for incomplete longitudinal ordinal data. *Communications in Statistics-Simulation and Computation*, *44*(5), 1311-1338.

Heagerty, P. J., & Zeger, S. L. (1996). Marginal regression models for clustered ordinal measurements. *Journal of the American Statistical Association*, *91*(435), 1024-1036.

Ibrahim, N. A., & Suliadi, S. (2011). Generating correlated discrete ordinal data using R and SAS IML. *Computer methods and programs in biomedicine*, *104*(3), e122-e132.

Liang, K. Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, *73*(1), 13-22.

Lipsitz, S. R., Laird, N. M., & Harrington, D. P. (1991). Generalized estimating equations for correlated binary data: using the odds ratio as a measure of association. *Biometrika*, *78*(1), 153-160.

Miller, M. E., Davis, C. S., & Landis, J. R. (1993). The analysis of longitudinal polytomous data: Generalized estimating equations and connections with weighted least squares. *Biometrics*, 1033-1044.

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