Supplementary Materials for Model Selection for Generalized Estimating Equations Accommodating Dropout Missingness

Chung-Wei Shen and Yi-Hau Chen*

Institute of Statistical Science, Academia Sinica, 11529 Taipei, Taiwan

yhchen@stat.sinica.edu.tw

1 Appendix

1.1 Derivation of Expressions (3) and (4) in the Paper

Let $\epsilon_i = Y_i - \mu_i^0$ and write

$$\begin{aligned}
(\hat{\mu}_i - \mu_i^0)^t(\hat{\mu}_i - \mu_i^0) &= (\hat{\mu}_i - Y_i + Y_i - \mu_i^0)^t(\hat{\mu}_i - Y_i + Y_i - \mu_i^0) \\
&= (\hat{\mu}_i - Y_i + \epsilon_i)^t(\hat{\mu}_i - Y_i + \epsilon_i) \\
&= (Y_i - \hat{\mu}_i)^t(Y_i - \hat{\mu}_i) - 2\epsilon_i^t(Y_i - \hat{\mu}_i) + \epsilon_i^t\epsilon_i \\
&= (Y_i - \hat{\mu}_i)^t(Y_i - \hat{\mu}_i) - 2\epsilon_i^t(\mu_i^0 - \hat{\mu}_i + \epsilon_i) + \epsilon_i^t\epsilon_i \\
&= (Y_i - \hat{\mu}_i)^t(Y_i - \hat{\mu}_i) + 2\epsilon_i^t(\hat{\mu}_i - \mu_i^0) - \epsilon_i^t\epsilon_i \\
&= (Y_i - \hat{\mu}_i)^t(Y_i - \hat{\mu}_i) + 2\mathrm{Tr}\{\epsilon_i(\hat{\mu}_i - \mu_i^*)^t\} + 2\epsilon_i^t(\mu_i^* - \mu_i^0) - \mathrm{Tr}(\epsilon_i\epsilon_i^t),
\end{aligned}$$

where μ_i^* is the limiting value of $\hat{\mu}_i$. Taking summation over $i = 1, \cdots, n$ and expectation on both sides of the above equation, we obtain expression (3) in the paper:

$$\begin{aligned}
&\mathrm{E}\Big\{\sum_{i=1}^n(\hat{\mu}_i - \mu_i^0)^t(\hat{\mu}_i - \mu_i^0)\Big\} \\
=\ &\mathrm{E}\Big\{\sum_{i=1}^n(Y_i - \hat{\mu}_i)^t(Y_i - \hat{\mu}_i)\Big\} + 2\sum_{i=1}^n \mathrm{Tr}\{\mathrm{cov}(Y_i, \hat{\mu}_i)\}) - \mathrm{Tr}(V_i^0)
\end{aligned}$$

by the fact that $\mathrm{E}\{\epsilon_i^t(\mu_i^* - \mu_i^0)\} = 0$.

Next we derive the covariance penalty term $\mathrm{cov}(Y_i, \hat{\mu}_i)$ for a candidate GEE model. Let $\hat{\beta}$ denote the regression parameter in this model estimated by GEE, and β^* the limiting value of $\hat{\beta}$. By Taylor expansion we have

$$\begin{aligned}
\hat{\mu}_i &= g^{-1}(X_i\hat{\beta}) = g^{-1}(X_i\beta^*) + \frac{\partial g^{-1}(X_i\beta)}{\partial \beta^t}(\hat{\beta} - \beta^*) + o_p(n^{-1/2}) \\
&= \mu_i^* + D_i(\hat{\beta} - \beta^*) + o_p(n^{-1/2}) \\
&= \mu_i^* + D_i\mathcal{H}_n^{-1}\sum_{k=1}^{n} S_k + o_p(n^{-1/2}), \qquad (1)
\end{aligned}$$

where $\mathcal{H}_n = \mathrm{E}\{\sum_{i=1}^n D_k^t V_k^{-1} D_k\}$, $S_k = D_k^t V_k^{-1}(Y_k - \mu_k^*)$. Hence

$$\begin{aligned}
\mathrm{cov}(Y_i, \hat{\mu}_i) &\approx \mathrm{cov}\{Y_i, D_i H_n^{-1}\sum_{k=1}^{n} S_k\} \\
&= \mathrm{E}\{(Y_i - \mu_i^0)S_i^t\}H_n^{-1}D_i^t \\
&= V_i^0 V_i^{-1} D_i H_n^{-1} D_i^t.
\end{aligned}$$

Therefore, we have

$$\mathrm{Tr}\{\mathrm{cov}(Y_i, \hat{\mu}_i)\} = \mathrm{Tr}\{V_i^0 V_i^{-1} D_i H_n^{-1} D_i^t\} = \mathrm{Tr}\{H_n^{-1} D_i^t V_i^{-1} V_i^0 D_i\}.$$

1.2 Derivation of the MLIC Statistic

When the outcome Y is subject to dropout missingness and the estimated mean $\hat{\mu}_i$ is obtained by a candidate GEE model with weighted GEE estimation, we consider the expected weighted quadratic loss

$$\mathrm{E}\bigg\{\sum_{i=1}^{n}(\hat{\mu}_i - \mu_i)^t W_i(\hat{\mu}_i - \mu_i)\bigg\}.$$

Let $\epsilon_i = Y_i - \mu_i^0$, we have

$$\begin{aligned}
(\hat{\mu}_i - \mu_i^0)^t W_i (\hat{\mu}_i - \mu_i^0) &= (\hat{\mu}_i - Y_i + Y_i - \mu_i^0)^t W_i (\hat{\mu}_i - Y_i + Y_i - \mu_i^0) \\
&= (\hat{\mu}_i - Y_i + \epsilon_i)^t W_i (\hat{\mu}_i - Y_i + \epsilon_i) \\
&= (Y_i - \hat{\mu}_i)^t W_i (Y_i - \hat{\mu}_i) - 2\epsilon_i^t W_i (Y_i - \hat{\mu}_i) + \epsilon_i^t W_i \epsilon_i \\
&= (Y_i - \hat{\mu}_i)^t W_i (Y_i - \hat{\mu}_i) - 2\epsilon_i^t W_i (\mu_i^0 - \hat{\mu}_i + \epsilon_i) + \epsilon_i^t W_i \epsilon_i \\
&= (Y_i - \hat{\mu}_i)^t W_i (Y_i - \hat{\mu}_i) + 2(\hat{\mu}_i - \mu_i^0)^t W_i \epsilon_i - \epsilon_i^t W_i \epsilon_i \\
&= (Y_i - \hat{\mu}_i)^t W_i (Y_i - \hat{\mu}_i) + 2(\hat{\mu}_i - \mu_i^*)^t W_i \epsilon_i + 2(\mu_i^* - \mu_i^0)^t W_i \epsilon_i - \epsilon_i^t W_i \epsilon_i,
\end{aligned}$$

μ_i^* is the limiting value of $\hat{\mu}_i$. By the fact that $\mathrm{E}(W_i|Y_i, X_i) = I_T$, the identity matrix of size T,

$$\begin{aligned}
\mathrm{E}\{(\mu_i^* - \mu_i^0)^t W_i \epsilon_i\} &= \mathrm{E}\{(\mu_i^* - \mu_i^0)^t \mathrm{E}(W_i|Y_i, X_i) \epsilon_i\} \\
&= \mathrm{E}\{(\mu_i^* - \mu_i^0)^t \epsilon_i\} = 0,
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{E}(\epsilon_i^t W_i \epsilon_i) &= \mathrm{E}\{\epsilon_i^t \mathrm{E}(W_i|Y_i, X_i) \epsilon_i\} \\
&= \mathrm{E}(\epsilon_i^t \epsilon_i) = \mathrm{Tr}(V_i^0).
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\mathrm{E}\left\{\sum_{i=1}^n (\hat{\mu}_i - \mu_i^0)^t W_i (\hat{\mu}_i - \mu_i^0)\right\} \\
= \ &\mathrm{E}\left\{\sum_{i=1}^n (Y_i - \hat{\mu}_i)^t W_i (Y_i - \hat{\mu}_i)\right\} + 2\sum_{i=1}^n \mathrm{Tr}(\mathrm{E}\{W_i \epsilon_i (\hat{\mu}_i - \mu_i^*)^t\}) - \mathrm{Tr}(V_i^0) \\
= \ &\mathrm{E}\left\{\sum_{i=1}^n (Y_i - \hat{\mu}_i)^t W_i (Y_i - \hat{\mu}_i)\right\} + 2\sum_{i=1}^n \mathrm{Tr}\{\mathrm{cov}(\mathcal{E}_i, \hat{\mu}_i)\} - \mathrm{Tr}(V_i^0),
\end{aligned}$$

where $\mathcal{E}_i = W_i\epsilon_i = W_i(Y_i - \mu_i^0)$. Note that the last term above is constant over models and can be omitted for model selection.

Now let $\hat{\beta}$ denote the regression parameter in this model estimated by weighted GEE, and β^* the limiting value of $\hat{\beta}$. By Taylor expansion we have

$$\begin{aligned}
\hat{\mu}_i &= g^{-1}(X_i\hat{\beta}) = g^{-1}(X_i\beta^*) + \frac{\partial g^{-1}(X_i\beta)}{\partial\beta^t}(\hat{\beta} - \beta^*) + o_p(n^{-1/2}) \\
&= \mu_i^* + D_i(\hat{\beta} - \beta^*) + o_p(n^{-1/2}) \\
&= \mu_i^* + D_i H_n^{-1}\sum_{k=1}^{n}(U_k - G_k) + o_p(n^{-1/2}), \qquad (2)
\end{aligned}$$

where $H_n = \mathrm{E}\{\sum_{i=1}^n D_k^t V_k^{-1} W_k D_k\}$, $U_k = D_k^t V_k^{-1} W_k(Y_k - \mu_k^*)$,

$$G_k = (\sum_{m=1}^{n} U_m F_m^t)(\sum_{m=1}^{n} F_m F_m^t)^{-1} F_m,$$

and F_k the score component of the kth individual in the partial likelihood for the dropout model (Robins et al. 1995).

Note that $\mathrm{Tr}\{\mathrm{cov}(\mathcal{E}_i, \hat{\mu}_i)\} = \mathrm{Tr}\{\mathrm{cov}(\hat{\mu}_i, \mathcal{E}_i)\}$. By (2), we have

$$\begin{aligned}
\mathrm{cov}(\hat{\mu}_i, \mathcal{E}_i) &\approx \mathrm{cov}\{D_i H_n^{-1}\sum_{k=1}^{n}(U_k - G_k), \mathcal{E}_i\} \\
&= \mathrm{E}\{D_i H_n^{-1}(U_i - G_i)\mathcal{E}_i^t\} \\
&= D_i H_n^{-1}\{D_i^t V_i^{-1}\mathrm{E}(\mathcal{E}_i\mathcal{E}_i^t) - \mathrm{E}(G_i\mathcal{E}_i^t)\} = D_i H_n^{-1}(D_i^t V_i^{-1}\Sigma_i - \Lambda_i),
\end{aligned}$$

where $\Sigma_i = \mathrm{E}(\mathcal{E}_i\mathcal{E}_i^t)$ and $\Lambda_i = \mathrm{E}(G_i\mathcal{E}_i^t)$, and

$$\mathrm{Tr}\{D_i H_n^{-1}(D_i^t V_i^{-1}\Sigma_i - \Lambda_i)\} = \mathrm{Tr}\{H_n^{-1}(D_i^t V_i^{-1}\Sigma_i - \Lambda_i)D_i\}.$$

Replacing the population quantities with their empirical versions, we propose the Missing Longitudinal Information Criterion (MLIC), defined as

$$\mathrm{MLIC} = \sum_{i=1}^{n}(Y_i - \hat{\mu}_i)^t W_i(Y_i - \hat{\mu}_i) + 2\mathrm{Tr}(\hat{H}_n^{-1}J_n),$$

where

$$\hat{H}_n = \sum_{i=1}^{n} D_i^t V_i^{-1} W_i D_i,$$

$$J_n = \sum_{i=1}^{n} (D_i^t V_i^{-1} \mathcal{E}_i \mathcal{E}_i^t - G_i \mathcal{E}_i^t) D_i,$$

evaluated at $\hat{\beta}$ and $\hat{\alpha}$.

2 Extended Simulations

2.1 Selection of the Mean Model with Multivariate Normal Data

We examine the performance of the MLIC method by performing a further simulation under multivariate normal distribution. In the simulations, the outcome variable Y_i is multivariate normal with marginal mean $\mu_{i,j}$ given by

$$\mu_{i,j} = \beta_0 + \beta_1 X_{1,i,j} + \beta_2 X_{2,i,j}, i = 1, \cdots, n, j = 1, \cdots, T,$$

where the covariate x_1 and x_2 are cluster-level covariates following a Bernoulli distribution with success probability 0.5 and a uniform(0,2) distribution, respectively. The number of visits $T = 3$ or 5 for each subject, and the number of subjects $n = 50$ or 100. The Pearson correlation coefficients between $Y_{i,j}$ and $Y_{i,k}$ for $1 \leq j, k \leq T$ are a constant $\rho = 0.5$; that is, the true correlation structure is "compound symmetry" or "exchangeable". The parameter values in the true model are $\beta_0 = -2$, $\beta_1 = 1$, $\beta_2 = -1$. Given Y_i and $R_{i,j-1} = 1$, the variable $R_{i,j}$ indicating whether subject i is observed ($R_{i,j} = 1$) or missing ($R_{i,j} = 1$) at time j is generated from the logistic model

$$\log\left(\frac{\lambda_{i,j}}{1-\lambda_{i,j}}\right) = \alpha_0 + \alpha_1 Y_{i,j-1} + \alpha_2 Y_{i,j-2} I(j > 2), j = 2, \cdots, T, \quad (3)$$

where $\lambda_{i,j} = \text{pr}(R_{i,j} = 1 | R_{i,j-1} = 1, Y_i, X_i)$, $I(j > 2) = 1$ if $j > 2$ and 0 otherwise, $\alpha_1 = -1.5$, $\alpha_2 = -1$, and α_0 is set to -2.2 or -3.3 when $T = 3$, and set to -1.5 or -2.5 when $T = 5$, so that the overall proportion of missing observations, i.e. the number of missing observations over nT, is 20% or 35%.

In addition to the true covariates x_1 and x_2 given above, we consider 4 redundant variables $x_3, \ldots, x_6$, where x_3 is a time-dependent covariate with $x_{3,j} = j - 1$ for $j = 1, \ldots, T$, and x_4 to x_6 are two-way interactions among (x_1, x_2, x_3) with $x_{4,j} = x_{1,j}x_{2,j}$, $x_{5,j} = x_{1,j}x_{3,j}$, and $x_{6,j} = x_{2,j}x_{3,j}$, $j = 1, \ldots, T$. The full model we considered is the linear model include all the 6 covariates.

We apply the proposed MLIC method, based respectively on the weighted GEE analyses under "independent" (Ind) and "compound symmetry" (CS) working correlation matrices. The weight matrix W_i in the weighted GEE analysis is evaluated with the observing probability $w_{i,t}$ estimated by the maximum likelihood method mentioned in Section 2 of the paper, where the likelihood is based on the true dropout model (3). Besides, we also apply the "naive" QIC method based on the naive GEE analysis under working independence, which ignores missing observations and treats data as if they were complete. Over 1000 simulated datasets, we then calculate for each method the percentage of choosing a model among a set of candidate models; recall that a model is chosen by the MLIC/QIC method when it has the smallest MLIC/QIC value among the candidate models. For illustration, we consider 10 non-nested candidate models including various subsets of covariates, as listed in Tables S1 and S2.

The results shown in Table S1 ($T = 3$) and Table S2 ($T = 5$) reveal that, the MLIC and QIC methods perform equally well when data are fully observed. When data are subject to

Table S1. Results for variable selection by MLIC: Percentage of selecting a candidate marginal linear model over 1000 replications with $T = 3$. The true model is model $(3) = \{X_1, X_2\}$

n	missing proportion	method	model (1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
50	0%	MLIC(CS)	0.1	0.1	67.7	0.0	0.0	12.9	13.4	2.2	1.9	1.7
		MLIC(Ind)	0.1	0.1	67.7	0.0	0.0	12.9	13.4	2.2	1.9	1.7
		QIC(Ind)	0.1	0.1	64.7	0.0	0.0	16.1	12.2	2.9	1.8	2.1
	20%	MLIC(CS)	0.2	1.4	49.2	0.2	0.4	14.8	16.3	4.0	6.6	6.9
		MLIC(Ind)	0.2	1.4	49.5	0.2	0.2	14.8	16.1	4.1	6.6	6.9
		QIC(Ind)	0.1	0.5	38.9	0.2	0.0	11.2	27.8	7.9	6.4	7.0
	35%	MLIC(CS)	0.2	3.6	33.8	0.7	0.6	13.6	14.7	5.9	11.0	15.9
		MLIC(Ind)	0.2	3.8	33.9	0.3	0.5	13.4	14.7	6.1	11.2	15.9
		QIC(Ind)	0.0	0.7	27.5	0.3	0.0	9.3	31.1	8.3	8.7	14.1
100	0%	MLIC(CS)	0.0	0.0	70.2	0.0	0.0	11.3	13.9	2.2	0.9	1.5
		MLIC(Ind)	0.0	0.0	70.2	0.0	0.0	11.3	13.9	2.2	0.9	1.5
		QIC(Ind)	0.0	0.0	69.1	0.0	0.0	12.6	13.2	2.5	0.9	1.7
	20%	MLIC(CS)	0.0	0.0	54.1	0.2	0.2	13.7	18.6	2.1	4.9	6.2
		MLIC(Ind)	0.0	0.1	54.5	0.0	0.0	13.8	18.5	2.1	4.8	6.2
		QIC(Ind)	0.0	0.0	30.1	0.0	0.0	5.1	44.6	7.8	6.1	6.3
	35%	MLIC(CS)	0.1	0.4	41.3	0.4	0.5	13.3	18.1	5.2	7.2	13.5
		MLIC(Ind)	0.1	0.4	42.2	0.0	0.0	13.4	17.8	5.1	7.4	13.6
		QIC(Ind)	0.0	0.0	19.8	0.0	0.0	4.2	49.9	9.9	6.7	9.5

Note: model (1)=$\{X_1\}$, (2)=$\{X_2\}$, (3)=$\{X_1, X_2\}$, (4)=$\{X_1, X_3\}$, (5)=$\{X_1, X_3, X_5\}$, (6)=$\{X_1, X_2, X_4\}$, (7)=$\{X_1, X_2, X_3\}$, (8)=$\{X_1, X_2, X_3, X_4\}$, (9)=$\{X_1, X_2, X_3, X_4, X_5\}$, (10)=$\{X_1, X_2, X_3, X_4, X_5, X_6\}$

dropout missingness, the MLIC method using either the independent (wrong) or compound symmetry (correct) correlation matrix still performs reasonably well; over the candidate models, MLIC selects the true model most frequently. On the contrary, the naive application of QIC method tends to select the wrong model containing $\{x_1, x_2, x_3\}$ (model (7)). It is also noticed that the performance of MLIC is virtually independent of the chosen working correlation.

Table S2. Results for variable selection by MLIC: Percentage of selecting a candidate marginal linear model over 1000 replications with $T = 5$. The true model is model $(3) = \{X_1, X_2\}$

n	missing proportion	method	model (1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
50	0%	MLIC(CS)	0.0	0.2	65.9	0.0	0.0	13.4	15.2	2.8	1.3	1.2
		MLIC(Ind)	0.0	0.2	65.9	0.0	0.0	13.4	15.2	2.8	1.3	1.2
		QIC(Ind)	0.0	0.1	63.6	0.0	0.0	16.2	13.6	3.3	1.6	1.6
	20%	MLIC(CS)	0.2	0.7	50.9	0.1	0.5	13.8	18.0	4.5	4.2	7.1
		MLIC(Ind)	0.2	0.8	51.3	0.1	0.1	13.9	17.6	4.4	4.2	7.4
		QIC(Ind)	0.1	0.4	39.1	0.0	0.1	10.8	30.8	5.5	5.0	8.2
	35%	MLIC(CS)	0.6	1.5	38.2	0.1	1.2	15.4	17.4	6.2	8.1	11.3
		MLIC(Ind)	0.6	1.7	38.2	0.1	0.6	15.1	17.8	6.4	8.2	11.3
		QIC(Ind)	0.3	0.5	27.2	0.2	0.1	9.7	34.0	9.7	0.7	11.3
100	0%	MLIC(CS)	0.0	0.0	68.0	0.0	0.0	13.2	13.2	2.4	1.5	1.7
		MLIC(Ind)	0.0	0.0	68.0	0.0	0.0	13.2	13.2	2.4	1.5	1.7
		QIC(Ind)	0.0	0.0	66.4	0.0	0.0	14.6	13.1	2.3	1.6	2.0
	20%	MLIC(CS)	0.0	0.0	51.9	0.1	0.0	11.9	20.6	2.8	4.5	8.2
		MLIC(Ind)	0.0	0.0	52.2	0.0	0.0	11.8	20.5	3.0	4.7	7.8
		QIC(Ind)	0.0	0.0	25.6	0.0	0.0	5.6	48.9	5.3	6.1	8.5
	35%	MLIC(CS)	0.1	0.0	43.7	0.0	0.3	11.3	19.8	5.6	6.6	12.6
		MLIC(Ind)	0.1	0.0	44.0	0.0	0.0	11.6	19.9	5.4	6.6	12.5
		QIC(Ind)	0.0	0.0	14.8	0.0	0.0	2.6	51.8	9.0	8.2	13.6

Note: model (1)=$\{X_1\}$, (2)=$\{X_2\}$, (3)=$\{X_1, X_2\}$, (4)=$\{X_1, X_3\}$, (5)=$\{X_1, X_3, X_5\}$, (6)=$\{X_1, X_2, X_4\}$, (7)=$\{X_1, X_2, X_3\}$, (8)=$\{X_1, X_2, X_3, X_4\}$, (9)=$\{X_1, X_2, X_3, X_4, X_5\}$, (10)=$\{X_1, X_2, X_3, X_4, X_5, X_6\}$

Table S3. Results for selecting a correlation structure by MLICC: Percentage of selecting a correlation matrix over 1000 replications. The true correlation matrix is AR-1 (first-order autoregressive).

	missing proportion	method	$n = 50$ Ind	CS	AR-1	$n = 100$ Ind	CS	AR-1
$T = 3$	0%	MLICC	15.6	19.2	66.2	12.6	12.4	75.0
		QIC	19.4	18.2	63.0	21.8	11.0	67.2
	20%	MLICC	15.8	32.8	51.4	17.6	24.6	57.8
		QIC	84.2	5.0	10.8	97.2	1.4	1.4
	35%	MLICC	16.8	33.4	49.8	16.1	31.5	52.4
		QIC	86.8	8.4	4.8	97.5	1.0	1.5
$T = 5$	0%	MLICC	13.4	16.2	70.6	11.6	11.0	77.4
		QIC	19.0	15.4	65.8	17.4	10.8	71.8
	20%	MLICC	19.4	28.6	52.0	19.6	23.4	57.0
		QIC	84.6	4.0	11.4	98.0	0.0	2.0
	35%	MLICC	21.0	36.8	42.2	19.4	34.8	45.8
		QIC	92.4	3.4	4.2	98.8	0.2	1.0

2.2 Selection of Correlation under AR-1

To examine the performance of the proposed MLICC method for selection of the correlation structure under dropout missingness, we perform a further simulation study under the same settings as those for Table 3 in the paper, except that now the true correlation structure is AR-1. The results are shown in Table S3.

We can see from Table S3 that, when data are fully observed, the MLICC and QIC methods have similar performance and choose the correct correlation structure with reasonably high probability. When there exists dropout missingness, MLICC still performs satisfactorily, while the naive QIC procedure treating incomplete data as complete has a very low chance to select the true correlation structure. These conclusions are the same as those drawn from Table 3 in the paper.

Table S4. Results for variable selection by MLIC: Percentage of selecting a candidate marginal logistic model over 1000 replications with $T = 3$. The true mean model is $\{1,2\} = \{X_1, X_2\}$ and the dropout model is misspecified.

n	missing proportion	method	$\{1\}$	$\{3\}$	$\{1,2\}$	$\{1,3\}$	$\{2,3\}$	$\{1,2,3\}$	$\{1,2,3,4\}$
100	20%	MLIC(CS)	30.5	23.3	20.8	3.4	13.7	2.0	6.3
		MLIC(Ind)	30.6	23.2	20.5	3.4	13.9	2.0	6.4
		QIC(Ind)	39.2	26.5	12.9	4.2	9.9	0.9	6.4
	35%	MLIC(CS)	34.6	25.0	17.5	4.0	11.4	1.7	5.8
		MLIC(Ind)	34.6	25.0	17.5	4.0	11.4	1.7	5.8
		QIC(Ind)	43.2	29.1	10.7	4.6	6.8	0.7	4.9
200	20%	MLIC(CS)	22.5	15.2	31.1	2.5	18.5	2.0	8.2
		MLIC(Ind)	22.5	15.2	31.1	2.5	18.5	2.0	8.2
		QIC(Ind)	36.5	21.2	20.2	2.7	10.9	1.4	7.1
	35%	MLIC(CS)	32.7	19.6	22.7	3.8	13.6	2.0	5.6
		MLIC(Ind)	32.7	19.6	22.6	3.8	13.6	2.0	5.7
		QIC(Ind)	42.0	24.6	15.1	4.4	7.2	1.8	4.9

2.3 Misspecified Dropout Model

We perform a simulation study where data are generated as in Tables 1-3 of the paper, while the dropout model $\lambda_{i,j}$ in equation (11) of the paper is misspecified by omitting its dependence on $Y_{i,j-2}$ when performing weighted GEE analysis and calculating the MLIC/MLICC. Results of this study are provided in Tables S4-S6. We can see that, except when the number of visits T for each subject is small, the MLIC method still has nice performance for variable selection under a misspecified dropout model. Moreover, the performance of MLICC for selecting the correlation structure is affected only slightly by misspecification of the dropout model. In both cases, the proposed MLIC/MLICC performs better than the naive QIC procedure, even though the dropout model is misspecified.

Table S5. Results for variable selection by MLIC: Percentage of selecting a candidate marginal logistic model over 1000 replications with $T = 5$. The true mean model is $\{1, 2\} = \{X_1, X_2\}$ and the dropout model is misspecified.

n	missing proportion	method	model $\{1\}$	$\{3\}$	$\{1,2\}$	$\{1,3\}$	$\{2,3\}$	$\{1,2,3\}$	$\{1,2,3,4\}$
100	20%	MLIC(CS)	3.4	3.8	48.3	0.8	32.3	3.6	7.8
		MLIC(Ind)	3.4	3.8	48.5	0.8	32.5	3.5	7.5
		QIC(Ind)	9.2	7.2	43.9	0.9	28.8	3.3	6.7
	35%	MLIC(CS)	11.3	8.5	40.1	1.0	29.6	2.4	7.1
		MLIC(Ind)	11.5	8.5	39.9	1.0	29.6	2.4	7.1
		QIC(Ind)	19.5	17.2	31.7	1.4	20.2	2.3	7.7
200	20%	MLIC(CS)	0.2	0.2	60.7	0.0	24.2	5.7	9.0
		MLIC(Ind)	0.2	0.2	60.8	0.0	24.3	5.7	8.8
		QIC(Ind)	1.7	0.9	59.5	0.2	22.8	5.9	9.0
	35%	MLIC(CS)	2.4	1.6	54.8	0.4	28.5	6.0	6.3
		MLIC(Ind)	2.5	1.6	54.9	0.4	28.7	5.9	6.7
		QIC(Ind)	8.2	3.9	48.9	1.8	24.0	6.3	6.9

Table S6. Results for selecting a correlation structure by MLICC: Percentage of selecting a correlation matrix over 1000 replications. The true correlation matrix is CS (compound symmetry) and the dropout model is misspecified.

	missing proportion	method	$n = 50$			$n = 100$		
			Ind	CS	AR-1	Ind	CS	AR-1
$T = 3$	20%	MLICC	23.3	50.6	26.1	32.5	48.6	18.9
		QIC	86.1	5.0	8.9	98.1	0.3	1.6
	35%	MLICC	24.2	45.0	30.8	28.7	49.6	21.7
		QIC	85.0	6.3	8.7	96.9	0.3	2.8
$T = 5$	20%	MLICC	28.2	54.6	17.2	30.2	59.9	9.9
		QIC	95.7	0.3	4.0	99.6	0.0	0.4
	35%	MLICC	30.9	51.2	17.9	30.6	58.4	11.0
		QIC	94.1	0.2	5.7	99.9	0.0	0.1