Supporting Information for empirical-likelihood-based criteria for model selection on marginal analysis of longitudinal data with dropout missingness by Chixiang Chen, Biyi Shen, Lijun Zhang, Yuan Xue and Ming Wang

Appendix A. Proofs

In this section, we denote $\boldsymbol{\gamma} = (\boldsymbol{\beta}', \boldsymbol{\rho}^{c\prime})'$ and assume the over-dispersion parameter ϕ is known without loss of generality. Before presenting the proofs of Theorems 2.1 and 2.2, we give the following lemma, which is the key to the derivation of asymptotic behavior of our proposal.

Lemma (1): *Denote the true values of the parameters* $(\boldsymbol{\beta}', \boldsymbol{\rho}', \boldsymbol{\theta}')'$ *as* $(\boldsymbol{\beta}_0', \boldsymbol{\rho}_0', \boldsymbol{\theta}_0')'$*, and* $\widehat{\rho}_{ijk}(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0)$ *as* $e_{ij}(\boldsymbol{\beta}_0)e_{ik}(\boldsymbol{\beta}_0)R_{ik}/\omega_{ik}(\boldsymbol{\theta}_0)$ *with* $\rho_{0jk} = E\{\widehat{\rho}_{ijk}(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0)\}$ *for* $1 \leqslant j < k \leqslant T$*, we have* $E\{\widehat{\rho}_{ijk}(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0) - \rho_{0jk}\phi(1-p/n)\} = O(n^{-1})$ *for longitudinal data with dropout missingness under MAR, provided that the mean structure and the dropout model are correctly specified.*

Proof. According to Robins et al. (1995), $R_{ik}/\omega_{ik} = 1 + \sum_{t=2}^{k} \left(R_{it} - \lambda_{it} R_{i(t-1)}\right) \omega_{it}^{-1}, 2 \leqslant k \leqslant T$. Also, denote $M_i(t) = \sum_{j=2}^{T} \left(R_{ij} - \lambda_{ij} R_{i(j-1)}\right) \omega_{ij}^{-1}$, thus under MAR, $M_i(t)$ is realized to be a mean zero martingale process with respect to the filtration process which is defined by $\mathscr{F}_i(t) = \sigma\{R_{i1}, \ldots, R_{i(t-1)}, \mathbf{Y}_i, \mathbf{X}_i, \mathbf{H}_i\}$. It indicates that $e_{ij}(\boldsymbol{\beta}_0)e_{ik}(\boldsymbol{\beta}_0)(R_{ik}/\omega_{ik} - 1)$ is also a mean zero martingale with respect to $\mathscr{F}_i(t)$. Hence, for $i = 1, \ldots, n$,

$$
\begin{aligned}
&E\{\widehat{\rho}_{ijk}(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0) - \rho_{0jk}\phi(1 - \frac{p}{n})\} \\
=&E\{e_{ij}(\boldsymbol{\beta}_0)e_{ik}(\boldsymbol{\beta}_0)\} - \rho_{0jk}\phi(1 - \frac{p}{n}) + E\{M_i(k)e_{ij}(\boldsymbol{\beta}_0)e_{ik}(\boldsymbol{\beta}_0)\} \\
=&O(\mathbf{n}^{-1}).
\end{aligned}
$$

Under Lemma (1), we can easily get $E\{\mathbf{g}(\mathbf{X}_i, \mathbf{Y}_i, \boldsymbol{\beta}_0, \boldsymbol{\rho}_0; \boldsymbol{\theta}_0)\} = O(\mathbf{n}^{-1})$ for WGEE model and $E\{\mathbf{G}_F(\mathbf{X}_{Fi}, \mathbf{Y}_i, \widetilde{\boldsymbol{\beta}}_0, \boldsymbol{\rho}_0^c; \boldsymbol{\theta}_0)\} = O(\mathbf{n}^{-1})$ in our proposed empirical likelihood ratio criteria.

A.1 *Notations and Conditions*

For simplicity, we ignore the subscript of i for the following proof, and the estimating equations below are subject-level if no further clarification. Without loss of generality, we set $\mathbf{X}_F = (\mathbf{X}, \mathbf{Z})$ with corresponding parameters $\widetilde{\boldsymbol{\beta}} = (\boldsymbol{\beta}', \mathbf{0}')'$. Now, we denote the estimating equations $\mathbf{g}_F$ in $\mathbf{G}_F = (\mathbf{g}_F', \mathbf{s}'(\boldsymbol{\theta}))'$ and $\mathbf{g}$ based on WGEE candidate model by

$$\mathbf{g}_F(\mathbf{X}_F, \mathbf{Y}, \widetilde{\boldsymbol{\beta}}, \boldsymbol{\rho}^c, \boldsymbol{\theta}) = \begin{pmatrix} \mathbf{g}_1 \\ \mathbf{g}_2 \\ \mathbf{g}_3 \end{pmatrix} = \begin{pmatrix} \mathbf{g}_{F1} \\ \mathbf{g}_3 \end{pmatrix}, \ \mathbf{g}(\mathbf{X}, \mathbf{Y}, \boldsymbol{\gamma}, \boldsymbol{\omega}) = \begin{pmatrix} \mathbf{g}_1 \\ \mathbf{g}_3^* \end{pmatrix},$$

where $\mathbf{g}_1 = \mathbf{X}'\big\{\partial\boldsymbol{\mu}(\mathbf{X}\boldsymbol{\beta})/\partial(\mathbf{X}\boldsymbol{\beta})'\big\}'\mathbf{V}^{-1}\mathbf{W}\big\{\mathbf{Y}-\boldsymbol{\mu}(\mathbf{X}\boldsymbol{\beta})\big\}$, $\mathbf{g}_2 = \mathbf{Z}'\big\{\partial\boldsymbol{\mu}(\mathbf{X}\boldsymbol{\beta})/\partial(\mathbf{X}\boldsymbol{\beta})'\big\}'\mathbf{V}^{-1}\mathbf{W}\big\{\mathbf{Y}-\boldsymbol{\mu}(\mathbf{X}\boldsymbol{\beta})\big\}$, $\mathbf{g}_3 = \mathbf{U}_i(\widetilde{\boldsymbol{\beta}}) - \boldsymbol{h}(\boldsymbol{\rho}^c)\phi$ with notation $\boldsymbol{\rho}^c$ defined as $(\rho_1^c, \ldots, \rho_{T-1}^c)'$. The notation $\mathbf{g}_3^*$ is the estimating equations for correlation coefficients in WGEE candidate model. Note that both $\mathbf{g}_3$ and $\mathbf{g}_3^*$ involve the sample size n. Next, some conditions are provided following to facilitate the proofs of our main results. It is worth mentioning that condition (2) to (4) are set to simplify the proof for the theorems; however, relaxation of these assumptions could also be possible but with the sacrifice of heavier algebra. Here, we denote $\boldsymbol{\eta} = (\boldsymbol{\beta}', (\boldsymbol{\rho}^c)', \boldsymbol{\theta}')'$ with $\widehat{\boldsymbol{\eta}}_{EL} = (\widehat{\boldsymbol{\beta}}_{EL}', (\widehat{\boldsymbol{\rho}}^c)_{EL}', \widehat{\boldsymbol{\theta}}_{EL}')'$ as the empirical likelihood estimators.

The Conditions for Theorem Proofs

(1) $E(\mathbf{G}_F\mathbf{G}_F')$ is a positive definite matrix. $(\partial^2\mathbf{G}_F)/(\partial\boldsymbol{\eta}'\partial\boldsymbol{\eta})$ is continuous in a neighborhood of the true value $\boldsymbol{\eta}_0$. $\|(\partial\mathbf{G}_F)/(\partial\boldsymbol{\eta}')\|$, $\|(\partial^2\mathbf{G}_F)/(\partial\boldsymbol{\eta}'\partial\boldsymbol{\eta})\|$, and $\|\mathbf{G}_F\|^3$ are bounded by some integrable function around the true value $\boldsymbol{\eta}_0$, and the rank of $E\big\{(\partial\mathbf{G}_F)/(\partial\boldsymbol{\eta}')\big\}$ is $\widetilde{p}$.

(2) $Cov\big\{\mathbf{U}_i(\widetilde{\boldsymbol{\beta}}) - \boldsymbol{h}(\boldsymbol{\rho}^c)\phi\big\} = Diag\Big[\big\{\sigma^2 + o(1)\big\}\phi^2(T - j - p/n)\Big]$ for some finite $\sigma^2 > 0$ and $j = 1, \ldots, T-1$.

(3) $E(\mathbf{g}_{F1}\mathbf{g}_3') = \mathbf{0}$, $E(\mathbf{g}_3\mathbf{s}') = \mathbf{0}$, $E(\partial\mathbf{g}_3/\partial\boldsymbol{\beta}') = \mathbf{0}$.

(4) The covariates $\mathbf{Z}$ is redundant and independent of $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{H}$.

For condition (2), it asks for independent estimating equations $\mathbf{g}_3$ with equal variances for correlation coefficients, which could be easily released by adding a weight matrix into the estimating equation to adjust for its poten- tial heterogeneity. However, the variance-covariance matrix for correlation coefficients is usually unclear in practice. Thus, for simplicity, we assume homogeneity holds for estimating equation of those correlation coefficients. Condition (3) requires more moment restrictions between estimating equations of marginal mean model ($\mathbf{g}_{F1}$), correlation efficients ($\mathbf{g}_3$), and drop-out model for missingness ($\mathbf{s}$). Condition (4) focuses on those redundant covariates which are independent of the outcomes and the other associated covariates. It says the redundant covariates should not carry information about the system. Again, conditions (2)-(4) are exploratory and somewhat strong assumptions to guarantee a stronger result for Theorem 2. They might not hold in practice, but provide some insights on the context under which the plug-in estimators and empirical-likelihood-based estimators are exactly equivalent in asymptotic manner. Here, we regard the plug-in estimators as a reasonable choice for our proposal due to avoid of computational issues and good approximation, which has been justified by extensive numerical analysis via simulation.

***A.2** Proof of Theorem 2.1*

Proof. Denote

$$Q_{1n}(\boldsymbol{\eta},\boldsymbol{\lambda}) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{1+\boldsymbol{\lambda}'\mathbf{G}_F(\mathbf{X}_{Fi},\mathbf{Y}_i,\widetilde{\boldsymbol{\beta}},\boldsymbol{\rho}^c,\boldsymbol{\theta})}\mathbf{G}_F(\mathbf{X}_{Fi},\mathbf{Y}_i,\widetilde{\boldsymbol{\beta}},\boldsymbol{\rho}^c,\boldsymbol{\theta})$$

$$Q_{2n}(\boldsymbol{\eta},\boldsymbol{\lambda}) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{1+\boldsymbol{\lambda}'\mathbf{G}_F(\mathbf{X}_{Fi},\mathbf{Y}_i,\widetilde{\boldsymbol{\beta}},\boldsymbol{\rho}^c,\boldsymbol{\theta})}\Big\{\frac{\partial\mathbf{G}_F(\mathbf{X}_{Fi},\mathbf{Y}_i,\widetilde{\boldsymbol{\beta}},\boldsymbol{\rho}^c,\boldsymbol{\theta})}{\partial\boldsymbol{\eta}'}\Big\}'\boldsymbol{\lambda}.$$

Along the lines with the proof of Lemma 1 in Qin and Lawless (1994) under the condition (1), we can get $\boldsymbol{\eta}$ and $\boldsymbol{\lambda}$ such that $\boldsymbol{\lambda} = O(\mathbf{n}^{-1/3})$ and $\|\boldsymbol{\eta}-\boldsymbol{\eta}_0\| \leqslant n^{-1/3}$ with probability 1, which satisfies $Q_{1n}(\boldsymbol{\eta},\boldsymbol{\lambda}) = \mathbf{0}$ and $Q_{2n}(\boldsymbol{\eta},\boldsymbol{\lambda}) = \mathbf{0}$. By the first-order Taylor expansion of $Q_{1n}(\boldsymbol{\eta},\boldsymbol{\lambda}) = \mathbf{0}$ and $Q_{2n}(\boldsymbol{\eta},\boldsymbol{\lambda}) = \mathbf{0}$ at $(\boldsymbol{\eta}_0',\mathbf{0}')'$, we can further solve $(\widehat{\boldsymbol{\eta}}_{EL}',\widehat{\boldsymbol{\lambda}}')'$ as

$$\begin{pmatrix}\widehat{\boldsymbol{\lambda}}\\ \widehat{\boldsymbol{\eta}}_{EL}-\boldsymbol{\eta}_0\end{pmatrix} = \boldsymbol{\Lambda}_n^{-1}\begin{pmatrix}-\mathbf{Q}_{1n}(\boldsymbol{\eta}_0,\mathbf{0})+o_p(\boldsymbol{\epsilon}_n)\\ o_p(\boldsymbol{\epsilon}_n)\end{pmatrix}, \tag{A.1}$$

where $\boldsymbol{\epsilon}_n$ is of order $O_p(n^{-1/2})$ and

$$\boldsymbol{\Lambda}_n = \begin{pmatrix} \frac{\partial \mathbf{Q}_{1n}}{\partial \boldsymbol{\lambda}'} & \frac{\partial \mathbf{Q}_{1n}}{\partial \boldsymbol{\eta}'} \\ \frac{\partial \mathbf{Q}_{2n}}{\partial \boldsymbol{\lambda}'} & \mathbf{0} \end{pmatrix}_{(\boldsymbol{\eta}_0, \mathbf{0})} \longrightarrow \boldsymbol{\Lambda} = \begin{pmatrix} \boldsymbol{\Lambda}_{11} & \boldsymbol{\Lambda}_{12} \\ \boldsymbol{\Lambda}_{21} & \boldsymbol{\Lambda}_{22} \end{pmatrix} = \begin{pmatrix} -E(\mathbf{G}_F \mathbf{G}_F') & E(\frac{\partial \mathbf{G}_F}{\partial \boldsymbol{\eta}'}) \\ E(\frac{\partial \mathbf{G}_F}{\partial \boldsymbol{\eta}'})' & \mathbf{0} \end{pmatrix}.$$

Further expansion according to the equation (A.1), we can get

$$\begin{aligned} \widehat{\boldsymbol{\eta}}_{EL} - \boldsymbol{\eta}_0 = & - \Big[E(\frac{\partial \mathbf{G}_F}{\partial \boldsymbol{\eta}'})' \big\{ E(\mathbf{G}_F \mathbf{G}_F') \big\}^{-1} E(\frac{\partial \mathbf{G}_F}{\partial \boldsymbol{\eta}'}) \Big]^{-1} \\ & \times E(\frac{\partial \mathbf{G}_F}{\partial \boldsymbol{\eta}'})' \big\{ E(\mathbf{G}_F \mathbf{G}_F') \big\}^{-1} \mathbf{Q}_{1n}(\boldsymbol{\eta}_0, \mathbf{0}) + o_p(\mathbf{n}^{-\frac{1}{2}}). \end{aligned}$$

Notice that $\mathbf{Q}_{1n}(\boldsymbol{\eta}_0, \mathbf{0}) = \big(1/n \sum_{i=1}^n \mathbf{g}_F'(\mathbf{X}_i, \mathbf{Y}_i, \widetilde{\boldsymbol{\beta}}, \boldsymbol{\rho}^c, \boldsymbol{\theta}), \mathbf{S}_{n\boldsymbol{\theta}}'\big)'$ with $\mathbf{S}_{n\boldsymbol{\theta}}$ defined in (1) from the main body of the paper. After some algebra, we rewrite the formula above as

$$\widehat{\boldsymbol{\eta}}_{EL} - \boldsymbol{\eta}_0 = \begin{pmatrix} \widehat{\boldsymbol{\gamma}}_{EL} - \boldsymbol{\gamma}_0 \\ \widehat{\boldsymbol{\theta}}_{EL} - \boldsymbol{\theta}_0 \end{pmatrix} = \begin{pmatrix} -\mathbf{V}_* \mathbf{A}_* \mathbf{Q}_n^* \\ \boldsymbol{\Omega} \mathbf{S}_{n\boldsymbol{\theta}} \end{pmatrix} + o_p(\mathbf{n}^{-\frac{1}{2}}) \tag{A.2}$$

with notations defined in Theorem 2.1. It completes the first part in Theorem 2.1. For the second part, with Lemma (1), asymptotic normality can be directly derived by (A.2) with the condition (1). Furthermore, the calculations show that $Cov(\mathbf{Q}_n^*, \mathbf{S}_{n\boldsymbol{\theta}})$ converges to zero, which means $\widehat{\boldsymbol{\gamma}}_{EL}$ and $\widehat{\boldsymbol{\theta}}_{EL}$ are asymptotically independent. Asymptotic χ^2 of $R^F(\widehat{\boldsymbol{\beta}}_{EL}, \widehat{\boldsymbol{\rho}}_{EL}^c, \widehat{\boldsymbol{\theta}}_{EL})$ in the third part can be showed by the similar arguments in Theorem 2 in Qin and Lawless (1994). The basic idea is sketched as follows: First we apply the second-order Taylor expansion to $-2 \log R^F(\widehat{\boldsymbol{\beta}}_{EL}, \widehat{\boldsymbol{\rho}}_{EL}^c, \widehat{\boldsymbol{\theta}}_{EL})$ and then, after a little algebra, it reduces to a quadratic form with negligible term. Finally, we show the matrix in the quadratic term is symmetric and idempotent with the rank of $\tilde{L} - \tilde{p}$, which finally justifies that $-2 \log R^F(\widehat{\boldsymbol{\beta}}_{EL}, \widehat{\boldsymbol{\rho}}_{EL}^c, \widehat{\boldsymbol{\theta}}_{EL})$ follows up χ^2 with the degree of freedom $\tilde{L} - \tilde{p}$.

A.3 *Proof of Theorem 2.2*

In order to prove Theorem 2.2, we need the lemma below:

LEMMA (2): *Under the conditions in Section A.1, we have:* $\mathbf{V}_*^{-1} = \mathbf{V}_{**}^{-1}$ *and* $\mathbf{A}_* \mathbf{Q}_n^* =$

$\mathbf{A}_{**}\mathbf{Q}_n^{**}$ *as* $n \to \infty$, *where*

$$\mathbf{V}_{**}^{-1} = E\Big(\frac{\partial \mathbf{g}}{\partial \boldsymbol{\gamma}'}\Big)'\Big\{E\mathbf{g}\mathbf{g}' - E\Big(\frac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}'}\Big)\big(E\mathbf{s}\mathbf{s}'\big)^{-1}E\Big(\frac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}'}\Big)'\Big\}^{-1}E\Big(\frac{\partial \mathbf{g}}{\partial \boldsymbol{\gamma}'}\Big),$$

$$\mathbf{A}_{**} = E\Big(\frac{\partial \mathbf{g}}{\partial \boldsymbol{\gamma}'}\Big)'\Big\{E\mathbf{g}\mathbf{g}' - E\Big(\frac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}'}\Big)\big(E\mathbf{s}\mathbf{s}'\big)^{-1}E\Big(\frac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}'}\Big)'\Big\}^{-1},$$

where $\mathbf{Q}_n^{**} = 1/n\sum_{i=1}^n \mathbf{g}(\mathbf{X}_i, \mathbf{Y}_i, \boldsymbol{\gamma}_0) + E\big(\partial \mathbf{g}/\partial \boldsymbol{\theta}'\big)E(\mathbf{s}\mathbf{s}')^{-1}\mathbf{S}_{n\boldsymbol{\theta}}$ *with* $\mathbf{V}_*^{-1}$, $\mathbf{A}_*$, $\mathbf{Q}_n^*$, *and* $\mathbf{S}_{n\boldsymbol{\theta}}$ *defined in the main body of the paper. The notation* $*$ *and* $**$ *correspond to the estimators from the empirical likelihood and WGEE candidate model, respectively.*

Proof. Here, we only present the proof of $\mathbf{V}_*^{-1} = \mathbf{V}_{**}^{-1}$ in Lemma (2), and $\mathbf{A}_* = \mathbf{A}_{**}$ can be obtained based on upon similar argument. In addition, without loss of generality, we assume all covariates (excluding intercept) are centered by their means.

First, by generalized information equality, we have $E\big(\partial \mathbf{g}_F/\partial \boldsymbol{\theta}'\big) = -E(\mathbf{g}_F\mathbf{s}')$ (Pierce, 1982). Thus, by $E(\mathbf{g}_{F1}\mathbf{g}_3') = \mathbf{0}$, $E(\mathbf{g}_3\mathbf{s}') = \mathbf{0}$, $E(\partial \mathbf{g}_3/\partial \boldsymbol{\beta}') = \mathbf{0}$ in the condition (3) and some algebra, we have

$$\mathbf{V}_*^{-1} = \begin{pmatrix} \mathbf{V}_{*11} & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_{*22} \end{pmatrix}$$

with $\mathbf{V}_{*11} = E\big(\partial \mathbf{g}_{F1}/\partial \boldsymbol{\beta}'\big)'\Big\{E\mathbf{g}_{F1}\mathbf{g}_{F1}' - E\big(\partial \mathbf{g}_{F1}/\partial \boldsymbol{\theta}'\big)\big(E\mathbf{s}\mathbf{s}'\big)^{-1}E\big(\partial \mathbf{g}_{F1}/\partial \boldsymbol{\theta}'\big)'\Big\}^{-1}E\big(\partial \mathbf{g}_{F1}/\partial \boldsymbol{\beta}'\big)$ and $\mathbf{V}_{*22} = E\big\{\partial \mathbf{g}_3/\partial (\boldsymbol{\rho}^c)'\big\}'\big(E\mathbf{g}_3\mathbf{g}_3'\big)^{-1}E\big\{\partial \mathbf{g}_3/\partial (\boldsymbol{\rho}^c)'\big\}$. Also, by the result $E\big(\partial \mathbf{g}_2/\partial \boldsymbol{\beta}'\big) = E(\mathbf{g}_1\mathbf{g}_2') = E(\mathbf{g}_2\mathbf{s}') = \mathbf{0}$ induced by the condition (4), $\mathbf{V}_{*11}$ can be further simplified as

$$\mathbf{V}_{*11} = E\Big(\frac{\partial \mathbf{g}_1}{\partial \boldsymbol{\beta}'}\Big)'\Big\{E\mathbf{g}_1\mathbf{g}_1' - E\Big(\frac{\partial \mathbf{g}_1}{\partial \boldsymbol{\theta}'}\Big)\big(E\mathbf{s}\mathbf{s}'\big)^{-1}E\Big(\frac{\partial \mathbf{g}_1}{\partial \boldsymbol{\theta}'}\Big)'\Big\}^{-1}E\Big(\frac{\partial \mathbf{g}_1}{\partial \boldsymbol{\beta}'}\Big).$$

Similarly, applying the condition (3), we can derive

$$\mathbf{V}_{**}^{-1} = \begin{pmatrix} \mathbf{V}_{**11} & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_{**22} \end{pmatrix},$$

where $\mathbf{V}_{**11} = \mathbf{V}_{*11}$ and $\mathbf{V}_{**22} = E\big\{\partial \mathbf{g}_3^*/\partial (\boldsymbol{\rho}^c)'\big\}'\big(E\mathbf{g}_3^*\mathbf{g}_3^{*T}\big)^{-1}E\big\{\partial \mathbf{g}_3^*/\partial (\boldsymbol{\rho}^c)'\big\}$.

Now it remains to show $\mathbf{V}_{*22} = \mathbf{V}_{**22}$ as $n \to \infty$. In this following proof, we will consider two specific cases with c denoted as exchangeable (EXC) and AR1 correlation structures since these are commonly used in practice and of most interest for researchers. Then $g_3^* =$

$\sum_{j=1}^{T-1}\Big\{U_{ij}(\boldsymbol{\beta})/\phi(T-j-p/n)-\rho^{EXC}\Big\}$ can be applied to estimate ρ^{EXC}. By calculation and the condition (2), we have

$$\lim_{n\to\infty} E\Big(\frac{\partial \mathbf{g}_3}{\partial \rho^{EXC}}\Big)'\big(E\mathbf{g}_3\mathbf{g}_3'\big)^{-1}E\Big(\frac{\partial \mathbf{g}_3}{\partial \rho^{EXC}}\Big)=\lim_{n\to\infty} E\Big(\frac{\partial g_3^*}{\partial \rho^{EXC}}\Big)'\Big\{E(g_3^*)^2\Big\}^{-1}E\Big(\frac{\partial g_3^*}{\partial \rho^{EXC}}\Big),$$

which justifies that $\mathbf{V}_*^{-1}=\mathbf{V}_{**}^{-1}$ under EXC case as $n\to\infty$.

Now consider the second scenario when the true correlation is AR1. Theoretically, $g_3^*=\sum_{j=1}^{T-1}j(\rho^{AR1})^{j-1}\big\{\widehat{U}_{ij}(\boldsymbol{\beta})/\phi(T-j-p/n)-(\rho^{AR1})^j\big\}$ can be applied to estimate ρ^{AR1}. Then, after some algebra and the condition (2), we have

$$\lim_{n\to\infty} E\Big(\frac{\partial \mathbf{g}_3}{\partial \rho^{AR1}}\Big)'\big(E\mathbf{g}_3\mathbf{g}_3{}'\big)^{-1}E\Big(\frac{\partial \mathbf{g}_3}{\partial \rho^{AR1}}\Big)=\lim_{n\to\infty} E\Big(\frac{\partial g_3^*}{\partial \rho^{AR1}}\Big)\big\{E(g_3^*)^2\big\}^{-1}E\Big(\frac{\partial g_3^*}{\partial \rho^{AR1}}\Big).$$

Hence, $\mathbf{V}_*^{-1}=\mathbf{V}_{**}^{-1}$ as $n\to\infty$.

Along with the result from Lemma (2), it is ready to prove Theorem 2.2. Please see the details of the proof below:

PROOF. Based upon Theorem 2.1, we have derived the empirical likelihood estimators as

$$\widehat{\boldsymbol{\gamma}}_{EL}-\boldsymbol{\gamma}_0=-\mathbf{V}_*\mathbf{A}_*\mathbf{Q}_n^*+o_p(\mathbf{n}^{-1/2}),$$

On the other hand, by the Taylor expansion, the estimators from WGEE candidate model (Robins et al., 1995) can be written as

$$\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0=-\Big\{E\big(\frac{\partial \mathbf{g}}{\partial \boldsymbol{\gamma}'}\big)'\Big\}^{-1}\mathbf{Q}_n^{**}+o_p(\mathbf{n}^{-\frac{1}{2}}).$$

Based upon some algebra, we have $\Big\{E\big(\partial \mathbf{g}/\partial\boldsymbol{\gamma}'\big)'\Big\}^{-1}\mathbf{Q}_n^{**}=\mathbf{V}_{**}\mathbf{A}_{**}\mathbf{Q}_n^{**}$. Thus, together with Lemma (2) of $\mathbf{V}_*=\mathbf{V}_{**}$ and $\mathbf{A}_*\mathbf{Q}_n^*=\mathbf{A}_{**}\mathbf{Q}_n^{**}$, the proof is completed.

Appendix B. Additional Simulation Studies

B.1 *The case with EXC and Gaussian outcomes*

In this case, we consider Gaussian outcomes. The true marginal mean model is

$$\mu_{ij}=\beta_0+x_{i1}\beta_1+x_{ij2}\beta_2,\ \text{for } i=1,\dots,n; j=1,\dots,T,$$

where all the setups for marginal mean model including the parameters $\boldsymbol{\beta}$, covariates and the true correlation structure are the same as the binary scenario in the main material. The dropout model is also similar as above but considers different parameter set-ups with $\boldsymbol{\theta} = (0.4, 0.5, -1)'$ for $m = 0.3$ and $\boldsymbol{\theta} = (1.15, 0.5, -1)'$ for $m = 0.2$. The results are summarized in table 1

[Table 1 about here.]

B.2 *The case with AR1 with Binary outcomes*

The setups arethe same to Table 1 in the main manuscript except that the true correlation structure is AR1. The results are summarized in table 2

[Table 2 about here.]

B.3 *The case with AR1 with Gaussian outcomes*

The setups are the same to subsection B.1 in the Supporting Information except that the true correlation structure is AR1. The results are summarized in table 3

[Table 3 about here.]

B.4 *The case under missing not at random*

The marginal model is for binary outcomes with the true marginal mean structure as

$$\log\big(\frac{\mu_{ij}}{1-\mu_{ij}}\big) = \beta_0 + x_{i1}\beta_1 + x_{ij2}\beta_2, \text{ for } i = 1, ..., n, j = 1, ..., T,$$

where x_{i1} is the subject (cluster) level covariate generated from $U[0, 1]$ and $x_{ij2} = j - 1$ is a time-dependent covariate. The number of observations (i.e., cluster size) is $T = 3$. The true parameter vector $\boldsymbol{\beta} = (-1, 1, 0.4)'$. The true correlation structure is EXC with a correlation coefficient $\rho_0 = 0.5$. The sample size (n) is 100 and 200, respectively. The true dropout model is given by

$$\log\big(\frac{\lambda_{ij}}{1-\lambda_{ij}}\big) = \theta_0 + y_{i(j-1)}\theta_1 + h_{ij}\theta_2 + y_{ij}\theta_3, \text{ for } i = 1, ..., n, j = 2, ..., T.$$

where h_{ij} follows up $U[-0.5, 0.5]$. We adjust $\boldsymbol{\theta} = (1.29, 0.5, -0.8, -0.5)'$ for missing proportion (m) around 0.3 and $\boldsymbol{\theta} = (1.97, 0.5, -0.8, -0.5)'$ for missing proportion (m) around 0.2. After data generation, we still use the original dropout model shown below in the main manuscript to fit the data.

$$\log\left(\frac{\lambda_{ij}}{1-\lambda_{ij}}\right) = \theta_0 + y_{i(j-1)}\theta_1 + h_{ij}\theta_2, \text{ for } i = 1, ..., n, j = 2, ..., T.$$

The results are summarized in table 4

[Table 4 about here.]

B.5 *The case with more variables*

The true marginal mean is given below:

$$\log\left(\frac{\mu_{ij}}{1-\mu_{ij}}\right) = \beta_0 + x_{ij1}\beta_1 + x_{ij2}\beta_2, \text{ for } i = 1, ..., n, j = 1, ..., T,$$

where $x_{ij1} = j - 1$ and $(x_{ij2}, x_{ij3}, ..., x_{ij8})'$ follows up $MVN(\mathbf{0}, \boldsymbol{\Sigma})$ where Σ is a var-covariance matrix of AR(1) with both variance and correlation coefficient as 0.5. Obviously, the variables except x_{ij1} and x_{ij2} are redundant. The number of observations (i.e., cluster size) is $T = 3$, and the sample size is 200. The parameters in the true marginal mean model is $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)' = (-0.25, 0.15, 0.25)'$. The true correlation structure for longitudinal data is exchangeable (EXC) with a correlation coefficient $\rho_0 = 0.5$. The true dropout model is given by

$$\log\left(\frac{\lambda_{ij}}{1-\lambda_{ij}}\right) = \theta_0 + y_{i(j-1)}\theta_1 + h_{ij}\theta_2, \text{ for } i = 1, ..., n, j = 2, ..., T,$$

where h_{ij} follows up $U[-0.5, 0.5]$. $\boldsymbol{\theta} = (1.7, 0.5, -0.8)'$ is set for missing proportion around 0.2 and $\boldsymbol{\theta} = (1.05, 0.5, -0.8)'$ is set for missing proportion around 0.3. The results are summarized in table 5

[Table 5 about here.]

B.6 *Impact on Marginal Mean Selection*

The model set-ups are exactly the same as Table 1 in the main manuscript. Here, Here, JEAIC and JEBIC are used for sole marginal mean selection under the full estimating equations (3) in the main manuscript, given a pre-specified correlation structure (e.g., AR1, EXC, and IND) with EXC as the true correlation structure. The selected rates are summarized in Table 6. We can find that both JEAIC and JEBIC under EXC correlation structure perform much better and more stable compared to the ones under AR1 and IND. It implies that correctly specifying the correlation structure is essential in terms of improving the marginal mean structure selection rates. More interestingly, in Table 1 in the main manuscript, the marginal selection rates for mean structures (column total) regardless of correlation structure selection is as high as, sometimes even a little better than the oracle one under which the true correlation structure is specified in Table 6 here. This findings strongly favor our joint selection that, even if the marginal mean structure is our sole interest, implementing joint selection would promise a high selection rate.

[Table 6 about here.]

Appendix C. Code and Data Resources

The IMPS data example analyzed in Section 4.2 and R codes implementing our method are available with this article at the Biometrics website on Wiley Online Library.

References

Pierce, D. A. (1982). The asymptotic effect of substituting estimators for parameters in certain types of statistics. The Annals of Statistics **10(2),** 475–478.

Qin, J. and Lawless, J. (1994). Empirical likelihood and general estimating equations. The Annals of Statistics **22(1),** 300–325.

Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1995). Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. Journal of the American Statistical Association **90(429),** 106–121.

Table 1

Performance of JEAIC and JEBIC compared with MLIC and QICW$_r$: Percentage of selecting six candidate models with Gaussian outcomes over 500 runs; $T = 3$, $\rho = 0.3$. The model with $\{x_1, x_2\}$ and an EXC correlation structure is the true model. Notation n and m denote the sample size and the missing probability, respectively.

Setups	Method	$\mathbf{C}(\boldsymbol{\rho})$	x_1	x_3	$\mathbf{x}_1, \mathbf{x}_2$	x_1, x_3	x_2, x_3	x_1, x_2, x_3	Total
n=100	JEAIC	AR1	0	0	0.144	0	0.006	0.038	0.188
m=0.2		**EXC**	0	0	**0.578**	0	0.016	0.218	**0.812**
		IND	0	0	0	0	0	0	0
		Total	0	0	**0.722**	0	0.022	0.256	1
	JEBIC	AR1	0	0	0.172	0	0.008	0.008	0.188
		EXC	0	0	**0.71**	0	0.024	0.078	**0.812**
		IND	0	0	0	0	0	0	0
		Total	0	0	**0.882**	0	0.032	0.086	1
	MLIC	AR1	0	0	0.334	0	0.084	0.018	0.436
		EXC	0	0	**0.452**	0	0.088	0.022	**0.562**
		IND	0	0	0.002	0	0	0	0
		Total	0	0	**0.788**	0	0.172	0.040	1
	QICW$_r$	AR1	0	0	0.314	0	0.044	0.024	0.382
		EXC	0	0	**0.500**	0	0.082	0.034	**0.616**
		IND	0	0	0.002	0	0	0	0
		Total	0	0	**0.816**	0	0.126	0.058	1
n=100	JEAIC	AR1	0	0	0.19	0.002	0.006	0.072	0.271
m=0.3		**EXC**	0.002	0	**0.507**	0	0.022	0.196	**0.727**
		IND	0	0	0.002	0	0	0	0.002
		Total	0.002	0	**0.699**	0.002	0.028	0.269	1
	JEBIC	AR1	0.006	0.002	0.234	0	0.008	0.02	0.271
		EXC	0.004	0	**0.623**	0	0.028	0.068	**0.723**
		IND	0.002	0	0.002	0.002	0	0	0.006
		Total	0.012	0.002	**0.86**	0.002	0.036	0.088	1
	MLIC	AR1	0.014	0.008	0.354	0	0.096	0.01	0.482
		EXC	0.006	0.002	**0.39**	0	0.106	0.008	**0.512**
		IND	0	0	0.002	0	0.002	0.002	0.006
		Total	0.02	0.01	**0.746**	0	0.204	0.02	1
	QICW$_r$	AR1	0.006	0	0.318	0	0.054	0.03	0.408
		EXC	0.004	0	**0.476**	0	0.088	0.022	**0.59**
		IND	0	0	0.002	0	0	0	0.002
		Total	0.01	0	**0.796**	0	0.142	0.052	1
n=200	JEAIC	AR1	0	0	0.08	0	0	0.03	0.11
m=0.2		**EXC**	0	0	**0.712**	0	0	0.178	**0.89**
		IND	0	0	0	0	0	0	0
		Total	0	0	**0.792**	0	0	0.208	1
	JEBIC	AR1	0	0	0.106	0	0	0.006	0.112
		EXC	0	0	**0.85**	0	0	0.038	**0.888**
		IND	0	0	0	0	0	0	0
		Total	0	0	**0.956**	0	0	0.044	1
	MLIC	AR1	0	0	0.432	0	0.008	0.018	0.458
		EXC	0	0	**0.53**	0	0.004	0.004	**0.538**
		IND	0	0	0	0	0	0.004	0.004
		Total	0	0	**0.962**	0	0.012	0.026	1
	QICW$_r$	AR1	0	0	0.36	0	0.004	0.024	0.388
		EXC	0	0	**0.584**	0	0.006	0.018	**0.608**
		IND	0	0	0.002	0	0	0.002	0.004
		Total	0	0	**0.946**	0	0.01	0.044	1
n=200	JEAIC	AR1	0	0	0.154	0	0	0.04	0.194
m=0.3		**EXC**	0	0	**0.629**	0	0	0.176	**0.806**
		IND	0	0	0	0	0	0	0
		Total	0	0	**0.784**	0	0	0.216	1
	JEBIC	AR1	0	0	0.188	0	0	0.014	0.202
		EXC	0	0	**0.745**	0	0.002	0.05	**0.798**
		IND	0	0	0	0	0	0	0
		Total	0	0	**0.934**	0	0.002	0.064	1
	MLIC	AR1	0.002	0	0.422	0	0.006	0.02	0.45
		EXC	0	0	**0.516**	0	0.02	0.01	**0.546**
		IND	0	0	0.004	0	0	0	0.004
		Total	0.002	0	**0.942**	0	0.026	0.03	1
	QICW$_r$	AR1	0	0	0.384	0	0.004	0.03	0.418
		EXC	0	0	**0.538**	0	0.006	0.034	**0.578**
		IND	0	0	0	0	0.002	0.002	0.004
		Total	0	0	**0.922**	0	0.012	0.066	1

Table 2

Performance of JEAIC and JEBIC compared with MLIC and QICW$_r$: Percentage of selecting six candidate models with Binary outcomes across 500 Monte Carlo datasets; $T = 3$, $\rho = 0.5$. The model with $\{x_1, x_2\}$ and an AR1 correlation structure is the true model. Notation n and m denote the sample size and the missing probability, respectively.

Setups	Method	$\mathbf{C}(\boldsymbol{\rho})$	x_1	x_3	$\mathbf{x}_1, \mathbf{x}_2$	x_1, x_3	x_2, x_3	x_1, x_2, x_3	Total
n=100	JEAIC	**AR1**	0.068	0.038	**0.487**	0.004	0.174	0.07	**0.842**
m=0.3		EXC	0.006	0.008	0.078	0.004	0.05	0.012	0.158
		IND	0	0	0	0	0	0	0
		Total	0.074	0.046	**0.565**	0.008	0.224	0.082	1
	JEBIC	**AR1**	0.182	0.078	**0.423**	0.004	0.154	0.008	**0.85**
		EXC	0.018	0.022	0.064	0	0.04	0.006	0.15
		IND	0	0	0	0	0	0	0
		Total	0.2	0.1	**0.487**	0.004	0.194	0.014	1
	MLIC	**AR1**	0.054	0.036	**0.288**	0.006	0.174	0.088	**0.646**
		EXC	0.026	0.026	0.118	0.002	0.088	0.034	0.294
		IND	0.002	0.006	0.03	0	0.018	0.004	0.06
		Total	0.082	0.068	**0.436**	0.008	0.28	0.126	1
	QICWr	**AR1**	0.02	0.016	**0.23**	0.002	0.116	0.094	**0.478**
		EXC	0.008	0.012	0.228	0.004	0.126	0.09	0.468
		IND	0.002	0.002	0.022	0	0.016	0.012	0.054
		Total	0.03	0.03	**0.48**	0.006	0.258	0.196	1
n=200	JEAIC	**AR1**	0.014	0	**0.652**	0	0.102	0.156	**0.924**
m=0.3		EXC	0	0	0.054	0	0.008	0.014	0.076
		IND	0	0	0	0	0	0	0
		Total	0.014	0	**0.706**	0	0.11	0.17	1
	JEBIC	**AR1**	0.074	0.004	**0.692**	0.002	0.138	0.014	**0.924**
		EXC	0.004	0	0.058	0	0.014	0	0.076
		IND	0	0	0	0	0	0	0
		Total	0.078	0.004	**0.75**	0.002	0.152	0.014	1
	MLIC	**AR1**	0.006	0.004	**0.422**	0.002	0.126	0.12	**0.68**
		EXC	0.002	0	0.154	0.002	0.05	0.052	0.26
		IND	0	0	0.032	0.002	0.008	0.018	0.06
		Total	0.008	0.004	**0.608**	0.006	0.184	0.19	1
	QICWr	**AR1**	0	0	**0.28**	0.002	0.068	0.104	**0.454**
		EXC	0.002	0	0.308	0	0.07	0.124	0.504
		IND	0	0	0.022	0	0.006	0.014	0.042
		Total	0.002	0	**0.61**	0.002	0.144	0.242	1
n=100	JEAIC	**AR1**	0.046	0.014	**0.552**	0.01	0.166	0.094	**0.882**
m=0.2		EXC	0.004	0.002	0.08	0.002	0.014	0.016	0.118
		IND	0	0	0	0	0	0	0
		Total	0.05	0.016	**0.632**	0.012	0.18	0.11	1
	JEBIC	**AR1**	0.174	0.046	**0.496**	0	0.166	0.006	**0.888**
		EXC	0.02	0.006	0.07	0	0.016	0	0.112
		IND	0	0	0	0	0	0	0
		Total	0.194	0.052	**0.566**	0	0.182	0.006	1
	MLIC	**AR1**	0.04	0.03	**0.382**	0.002	0.192	0.096	**0.742**
		EXC	0.006	0.004	0.088	0.004	0.072	0.04	0.214
		IND	0.002	0.002	0.026	0	0.006	0.008	0.044
		Total	0.048	0.036	**0.496**	0.006	0.27	0.144	1
	QICWr	**AR1**	0.022	0.01	**0.33**	0	0.152	0.088	**0.602**
		EXC	0.002	0.008	0.172	0.004	0.08	0.076	0.342
		IND	0	0.002	0.028	0	0.012	0.014	0.056
		Total	0.024	0.02	**0.53**	0.004	0.244	0.178	1
n=200	JEAIC	**AR1**	0.006	0	**0.706**	0	0.09	0.156	**0.958**
m=0.2		EXC	0	0	0.036	0	0	0.006	0.042
		IND	0	0	0	0	0	0	0
		Total	0.006	0	**0.742**	0	0.09	0.162	1
	JEBIC	**AR1**	0.04	0.004	**0.792**	0.002	0.11	0.012	**0.96**
		EXC	0	0	0.038	0	0	0.002	0.04
		IND	0	0	0	0	0	0	0
		Total	0.04	0.004	**0.83**	0.002	0.11	0.014	1
	MLIC	**AR1**	0.004	0	**0.496**	0	0.114	0.14	**0.754**
		EXC	0.002	0	0.12	0	0.018	0.044	0.184
		IND	0	0	0.034	0.002	0.008	0.018	0.062
		Total	0.006	0	**0.65**	0.002	0.14	0.202	1
	QICWr	**AR1**	0	0	**0.37**	0.002	0.086	0.134	**0.592**
		EXC	0.002	0	0.24	0	0.032	0.074	0.348
		IND	0	0	0.036	0.002	0.006	0.016	0.06
		Total	0.002	0	**0.646**	0.004	0.124	0.224	1

Table 3

Performance of JEAIC and JEBIC compared with MLIC and QICW$_r$: Percentage of selecting six candidate linear models across 500 Monte Carlo datasets; $T = 3$, $\rho = 0.5$. The model with $\{x_1, x_2\}$ and an AR1 correlation structure is the true model. Notation n and m denote the sample size and the missing probability, respectively.

Setups	Method	Cor.Str.	x_1	x_3	$\mathbf{x_1, x_2}$	x_1, x_3	x_2, x_3	x_1, x_2, x_3	Total
n=100	JEAIC	**AR1**	0.004	0	**0.606**	0	0.014	0.187	**0.811**
m=0.3		EXC	0	0	0.143	0	0.008	0.032	0.183
		IND	0.004	0	0.002	0	0	0	0.006
		Total	0.008	0	**0.751**	0	0.022	0.219	1
	JEBIC	**AR1**	0.014	0	**0.717**	0	0.018	0.058	**0.807**
		EXC	0	0	0.161	0	0.008	0.014	0.183
		IND	0.004	0	0.006	0	0	0	0.01
		Total	0.018	0	**0.884**	0	0.026	0.072	1
	MLIC	**AR1**	0.018	0.006	**0.48**	0.002	0.098	0.012	**0.616**
		EXC	0.002	0	0.3	0	0.064	0.01	0.376
		IND	0.002	0	0.004	0	0.002	0	0.008
		Total	0.022	0.006	**0.784**	0.002	0.164	0.022	1
	QICWr	**AR1**	0.006	0	**0.476**	0.002	0.064	0.036	**0.584**
		EXC	0	0.002	0.342	0	0.048	0.016	0.408
		IND	0	0	0.002	0	0.002	0.004	0.008
		Total	0.006	0.002	**0.82**	0.002	0.114	0.056	1
n=200	JEAIC	**AR1**	0	0	**0.665**	0	0	0.182	**0.848**
m=0.3		EXC	0	0	0.122	0	0	0.03	0.152
		IND	0	0	0	0	0	0	0
		Total	0	0	**0.788**	0	0	0.212	1
	JEBIC	**AR1**	0	0	**0.816**	0	0.002	0.036	**0.854**
		EXC	0	0	0.138	0	0	0.008	0.146
		IND	0	0	0	0	0	0	0
		Total	0	0	**0.954**	0	0.002	0.044	1
	MLIC	**AR1**	0.002	0	**0.506**	0	0.01	0.006	**0.524**
		EXC	0.002	0	0.442	0	0.014	0.014	0.472
		IND	0	0	0.004	0	0	0	0.004
		Total	0.004	0	**0.952**	0	0.024	0.02	1
	QICWr	**AR1**	0	0	**0.482**	0	0.004	0.026	**0.512**
		EXC	0.002	0	0.442	0	0.01	0.028	0.482
		IND	0	0	0.006	0	0	0	0.006
		Total	0.002	0	**0.93**	0	0.014	0.054	1
n=100	JEAIC	**AR1**	0	0	**0.644**	0	0.014	0.202	**0.86**
m=0.2		EXC	0	0	0.098	0	0.006	0.036	0.14
		IND	0	0	0	0	0	0	0
		Total	0	0	**0.742**	0	0.02	0.238	1
	JEBIC	**AR1**	0	0	**0.764**	0	0.026	0.08	**0.87**
		EXC	0	0	0.114	0	0.008	0.008	0.13
		IND	0	0	0	0	0	0	0
		Total	0	0	**0.878**	0	0.034	0.088	1
	MLIC	**AR1**	0.002	0.002	**0.486**	0	0.078	0.016	**0.584**
		EXC	0	0	0.354	0	0.058	0.002	0.414
		IND	0	0	0.002	0	0	0	0.002
		Total	0.002	0.002	**0.842**	0	0.136	0.018	1
	QICWr	**AR1**	0.002	0.002	**0.47**	0	0.06	0.024	**0.558**
		EXC	0	0	0.39	0	0.042	0.008	0.44
		IND	0	0	0.002	0	0	0	0.002
		Total	0.002	0.002	**0.862**	0	0.102	0.032	1
n=200	JEAIC	**AR1**	0	0	**0.73**	0	0	0.19	**0.92**
m=0.2		EXC	0	0	0.058	0	0	0.022	0.08
		IND	0	0	0	0	0	0	0
		Total	0	0	**0.788**	0	0	0.212	1
	JEBIC	**AR1**	0	0	**0.902**	0	0	0.024	**0.926**
		EXC	0	0	0.066	0	0	0.008	0.074
		IND	0	0	0	0	0	0	0
		Total	0	0	**0.968**	0	0	0.032	1
	MLIC	**AR1**	0	0	**0.544**	0	0.008	0.012	**0.564**
		EXC	0	0	0.424	0	0.004	0.008	0.436
		IND	0	0	0	0	0	0	0
		Total	0	0	**0.968**	0	0.012	0.02	1
	QICWr	**AR1**	0	0	**0.518**	0	0.006	0.02	**0.544**
		EXC	0	0	0.438	0	0.002	0.016	0.456
		IND	0	0	0	0	0	0	0
		Total	0	0	**0.956**	0	0.008	0.036	1

Table 4

Performance of JEAIC and JEBIC when the missing mechanism is missing not at random (MNAR): Percentage of selecting six candidate logistic models across 500 Monte Carlo datasets; $T = 3$, $\rho = 0.5$. The model with $\{x_1, x_2\}$ and an EXC correlation structure is the true model. Notation n and m denote the sample size and the missing probability, respectively.

Setups	Method	$\mathbf{C}(\boldsymbol{\rho})$	x_1	x_3	$\mathbf{x}_1, \mathbf{x}_2$	x_1, x_3	x_2, x_3	x_1, x_2, x_3	Total
n=100,m=0.2	JEAIC	AR1	0.002	0.006	0.072	0.002	0.014	0.006	0.102
		EXC	0.058	0.024	**0.51**	0.008	0.198	0.1	**0.898**
		IND	0	0	0	0	0	0	0
		Total	0.06	0.03	**0.582**	0.01	0.212	0.106	1
	JEBIC	AR1	0.028	0.014	0.058	0	0.014	0.002	0.116
		EXC	0.188	0.062	**0.456**	0.002	0.17	0.006	**0.884**
		IND	0	0	0	0	0	0	0
		Total	0.216	0.076	**0.514**	0.002	0.184	0.008	1
n=200,m=0.2	JEAIC	AR1	0	0	0.028	0	0.006	0.008	0.042
		EXC	0.01	0	**0.712**	0	0.112	0.124	**0.958**
		IND	0	0	0	0	0	0	0
		Total	0.01	0	**0.74**	0	0.118	0.132	1
	JEBIC	AR1	0.002	0	0.03	0.002	0.008	0	0.042
		EXC	0.036	0.004	**0.764**	0.002	0.14	0.012	**0.958**
		IND	0	0	0	0	0	0	0
		Total	0.038	0.004	**0.794**	0.004	0.148	0.012	1
n=100, m=0.3	JEAIC	AR1	0.04	0.008	0.062	0.01	0.018	0.006	0.144
		EXC	0.104	0.042	**0.418**	0.036	0.174	0.082	**0.856**
		IND	0	0	0	0	0	0	0
		Total	0.144	0.05	**0.48**	0.046	0.192	0.088	1
	JEBIC	AR1	0.064	0.028	0.05	0.002	0.016	0	0.16
		EXC	0.242	0.108	**0.334**	0.012	0.134	0.008	**0.838**
		IND	0	0	0	0.002	0	0	0.002
		Total	0.306	0.136	**0.384**	0.016	0.15	0.008	1
n=200, m=0.3	JEAIC	AR1	0.012	0.006	0.07	0.002	0.014	0.006	0.11
		EXC	0.042	0.004	**0.622**	0.006	0.112	0.104	**0.89**
		IND	0	0	0	0	0	0	0
		Total	0.054	0.01	**0.692**	0.008	0.126	0.11	1
	JEBIC	AR1	0.04	0.018	0.054	0	0.008	0.002	0.122
		EXC	0.176	0.022	**0.556**	0.006	0.11	0.008	**0.878**
		IND	0	0	0	0	0	0	0
		Total	0.216	0.04	**0.61**	0.006	0.118	0.01	1

Table 5

Performance of JEAIC and JEBIC compared with MLIC and $QICW_r$ when there are more variables: Percentage of selecting six candidate logistic models across 500 Monte Carlo datasets; $T = 3$, $\rho = 0.5$. The model with $\{x_1, x_2\}$ and an EXC correlation structure is the true model. Notation n and m denote the sample size and the missing probability, respectively.

Setups	Method	$\mathbf{C}(\boldsymbol{\rho})$	x_1	x_3	$\mathbf{x}_1, \mathbf{x}_2$	x_1, x_3	x_2, x_3	x_1, x_2, x_3	Total
n=200, m=0.2	JEAIC	AR1	0.004	0.004	0.036	0	0.06	0.002	0.052
		EXC	0.114	0.11	**0.546**	0.076	0.042	0.06	**0.948**
		IND	0	0	0	0	0	0	0
		Total	0.118	0.114	**0.582**	0.076	0.048	0.062	1
	JEBIC	AR1	0.004	0.006	0.042	0	0	0	0.052
		EXC	0.138	0.13	**0.67**	0.01	0	0	**0.948**
		IND	0	0	0	0	0	0	0
		Total	0.142	0.136	**0.712**	0.01	0	0	1
	MLIC	AR1	0.046	0.02	0.134	0.022	0.022	0.024	0.268
		EXC	0.122	0.094	**0.348**	0.056	0.034	0.032	**0.686**
		IND	0.01	0.006	0.022	0.004	0	0.004	0.046
		Total	0.178	0.12	**0.504**	0.082	0.056	0.06	1
	$QICW_r$	AR1	0.018	0.024	0.088	0.024	0.01	0.036	0.2
		EXC	0.092	0.088	**0.33**	0.056	0.036	0.07	**0.672**
		IND	0.02	0.002	0.058	0.014	0.014	0.02	0.128
		Total	0.13	0.114	**0.476**	0.094	0.06	0.126	1
n=200, m=0.3	JEAIC	AR1	0.018	0.018	0.068	0.002	0.002	0.004	0.112
		EXC	0.116	0.12	**0.478**	0.072	0.048	0.054	**0.888**
		IND	0	0	0	0	0	0	0
		Total	0.134	0.138	**0.546**	0.074	0.05	0.058	1
	JEBIC	AR1	0.02	0.02	0.074	0	0	0	0.114
		EXC	0.142	0.144	**0.586**	0.01	0.002	0.002	**0.886**
		IND	0	0	0	0	0	0	0
		Total	0.162	0.164	**0.66**	0.01	0.002	0.002	1
	MLIC	AR1	0.034	0.05	0.126	0.016	0.01	0.014	0.25
		EXC	0.106	0.094	**0.31**	0.054	0.036	0.028	**0.628**
		IND	0.028	0.014	0.048	0.014	0.008	0.01	0.122
		Total	0.168	0.158	**0.484**	0.084	0.054	0.052	1
	$QICW_r$	AR1	0.042	0.044	0.104	0.026	0.02	0.026	0.262
		EXC	0.062	0.056	**0.218**	0.038	0.032	0.052	**0.458**
		IND	0.032	0.026	0.102	0.032	0.02	0.068	0.28
		Total	0.136	0.126	**0.424**	0.096	0.072	0.146	1

Table 6

Performance of JEAIC and JEBIC in sole marginal mean selection given AR1, EXC, and IND correlation structures, respectively: Percentage of selecting six candidate models with binary outcomes over 500 runs; $T = 3$, $\rho = 0.5$. The model with $\{x_1, x_2\}$ and an EXC correlation structure is the true model. Notation n and m denote the sample size and the missing probability, respectively.

Setups	Method	$\mathbf{C}(\boldsymbol{\rho})$	x_1	x_3	$\mathbf{x_1, x_2}$	x_1, x_3	x_2, x_3	x_1, x_2, x_3
n=100,m=0.2	JEAIC	AR1	0.134	0.008	0.602	0.004	0.19	0.062
		EXC	0.096	0.008	**0.658**	0.002	0.208	0.028
		IND	0.156	0.056	0.426	0.034	0.188	0.14
	JEBIC	AR1	0.044	0.046	0.526	0.004	0.182	0.198
		EXC	0.004	0.036	**0.638**	0.002	0.218	0.102
		IND	0.066	0.122	0.366	0.014	0.146	0.286
n=100,m=0.3	JEAIC	AR1	0.136	0.036	0.536	0.008	0.204	0.08
		EXC	0.112	0.03	**0.57**	0.004	0.236	0.048
		IND	0.132	0.062	0.412	0.038	0.186	0.17
	JEBIC	AR1	0.026	0.092	0.466	0.008	0.19	0.218
		EXC	0.014	0.08	**0.526**	0.002	0.226	0.152
		IND	0.052	0.124	0.344	0.01	0.162	0.308
n=200,m=0.2	JEAIC	AR1	0.154	0	0.742	0	0.1	0.004
		EXC	0.134	0	**0.76**	0	0.106	0
		IND	0.134	0.014	0.65	0.012	0.118	0.072
	JEBIC	AR1	0.016	0.006	0.82	0	0.13	0.028
		EXC	0.018	0	**0.846**	0	0.126	0.01
		IND	0.016	0.066	0.55	0.006	0.1	0.262
n=200,m=0.3	JEAIC	AR1	0.162	0	0.68	0	0.142	0.016
		EXC	0.138	0	**0.708**	0	0.148	0.006
		IND	0.12	0.046	0.506	0.024	0.15	0.154
	JEBIC	AR1	0.018	0.022	0.714	0	0.16	0.086
		EXC	0.01	0.006	**0.776**	0.002	0.168	0.038
		IND	0.024	0.108	0.418	0.014	0.122	0.314