Supplementary Materials for "Improving Main Analysis by Borrowing Information from Auxiliary Data"

Chixiang Chen[1], Peisong Han[2], Fan He[3]

[1]Department of Biostatistics, Epidemiology and Informatics,
University of Pennsylvania, Philadelphia, PA, U.S.A
[2]Department of Biostatistics, University of Michigan, Ann Arbor, Michigan, U.S.A.
[3]Division of Epidemiology, College of Medicine, Penn State Univerisity, Hershey, PA, USA.
*Contact: Chixiang.Chen@Pennmedicine.upenn.edu

The following context will be organized as follows. Section 1 provides proofs for Theorem 1 and Property 1 from the main manuscript as well as some discussion about the re-weighting estimation scheme. Section 2 includes extra simulation results. Section 3 includes extra note about covariate and efficiency gain and one detailed extension of the proposed method to address missing data problem.

1 Proof

Without lose of generality, among total n samples, we assume that the first $m_1(\leqslant n)$ subjects in the study have auxiliary data, with a constant ρ defined as $\lim_{n\to\infty}(m_1/n)$. In what follows, we use $\tilde{E}(f)$ to denote the limit value of $(1/m_1)\sum_{i=1}^{m_1} f_i$, for any measurable random variable f. The notation $\|\cdot\|$ represents L2 norm. To facilitate the proof, we need the following regularity assumptions that are broadly adopted in empirical likelihood (Qin and Lawless, 1994) and generalized method of moment (Newey and McFadden, 1994).

Assumption 1 $E\{\boldsymbol{g}(\boldsymbol{D}_i^u;\boldsymbol{\beta})\} = \mathbf{0}$ if and only if $\boldsymbol{\beta} = \boldsymbol{\beta}_0$.

Assumption 2 Suppose that $\boldsymbol{D}_i^u$ are independent and identically distributed, for $i = 1,\ldots,n$. The function $\boldsymbol{g}(\boldsymbol{D}_i^u;\boldsymbol{\beta})$ is twice continuously differentiable; $E\|\boldsymbol{g}(\boldsymbol{D}_i^u;\boldsymbol{\beta})\|^2$ is finite; and $\|\partial^2\boldsymbol{g}(\boldsymbol{D}_i^u;\boldsymbol{\beta})/\partial\boldsymbol{\beta}^T\partial\boldsymbol{\beta}\|$ can be bounded by some integrable function in the neighborhood of $\boldsymbol{\beta}_0$ defined in Assumption 1.

Assumption 3 There exist values of parameter $\boldsymbol{\theta}_*$ such that $\tilde{E}\{\boldsymbol{h}(\boldsymbol{D}_i^a;\boldsymbol{\theta}_*)\} = \mathbf{0}$.

Assumption 4 Suppose that $\boldsymbol{D}_i^a$ are independent and identically distributed, for $i = 1,\ldots,n$. For $\boldsymbol{\theta}_*$ defined in Assumption 3, $\tilde{E}\{\boldsymbol{h}(\boldsymbol{D}_i^a;\boldsymbol{\theta}_*)\boldsymbol{h}^T(\boldsymbol{D}_i^a;\boldsymbol{\theta}_*)\}$ is positive definite, and $\{\partial^2\boldsymbol{h}(\boldsymbol{D}_i^a;\boldsymbol{\theta}_*)\}/(\partial\boldsymbol{\theta}^T\partial\boldsymbol{\theta})$ is continuous in the neighborhood of $\boldsymbol{\theta}_*$. Moreover, $\|\{\partial\boldsymbol{h}(\boldsymbol{D}_i^a;\boldsymbol{\theta}_*)\}/(\partial\boldsymbol{\theta}^T)\|$, $\|\{\partial^2\boldsymbol{h}(\boldsymbol{D}_i^a;\boldsymbol{\theta}_*)\}/(\partial\boldsymbol{\theta}^T\partial\boldsymbol{\theta})\|$, and $\|\boldsymbol{h}(\boldsymbol{D}_i^a;\boldsymbol{\theta}_*)\|^3$ are bounded by some integrable function around $\boldsymbol{\theta}_*$.

Assumption 1 and 2 are the moment regularities for the main parameter estimation of interest. Assumption 3 and 4 are regularities for the auxiliary model. Assumption 1 and 3 are important for identifiability and estimation consistency for the main parameter, whereas Assumption 2 and 4 are moment conditions to guarantee valid Taylor expansion of given estimating equations.

1.1 Proof for Theorem 1

In order to solve the problem of constrained maximization in (3), we introduce Lagrange multipliers $\boldsymbol{\lambda}$, and from the theorem of empirical likelihood (Qin and Lawless, 1994; Owen, 2001) under Assumption 3,4, we have

$$\hat{\boldsymbol{\lambda}} = \frac{1}{m_1}\boldsymbol{S}\boldsymbol{Q}_{m_1}(\boldsymbol{\theta}_*) + o_p(\boldsymbol{n}^{-\frac{1}{2}}) \tag{1}$$

with $\boldsymbol{Q}_{m_1}(\boldsymbol{\theta}_*) = \sum_{i=1}^{m_1} \boldsymbol{h}(\boldsymbol{D}_i^a; \boldsymbol{\theta}_*)$. The notation $\boldsymbol{S}$ is defined in the Theorem 1. On the other hand, by empirical likelihood theorem and Assumption 3,4 again, estimated weights can be expressed as $\hat{p}_i = m_1^{-1}[1 - \hat{\boldsymbol{\lambda}}^T \boldsymbol{h}(\boldsymbol{D}_i^a; \boldsymbol{\theta}_*)\{1 + o_p(1)\}]$. Based on (1) and expression of estimated weights and under Assumption 2, 4, we have

$$\begin{aligned}
\mathbf{0} =& \frac{m_1}{n}\sum_{i=1}^{m_1} \hat{p}_i \boldsymbol{g}(\boldsymbol{D}_i^u; \hat{\boldsymbol{\beta}}_{EN}) + \frac{1}{n}\sum_{i=m_1+1}^{n} \boldsymbol{g}(\boldsymbol{D}_i^u; \hat{\boldsymbol{\beta}}_{EN}) \\
=& \frac{1}{n}\sum_{i=1}^{m_1} \boldsymbol{g}(\boldsymbol{D}_i^u; \hat{\boldsymbol{\beta}}_{EN}) \left[1 - \hat{\boldsymbol{\lambda}}^T \boldsymbol{h}(\boldsymbol{D}_i^a; \boldsymbol{\theta}_*)\{1 + o_p(1)\}\right] + \frac{1}{n}\sum_{i=m_1+1}^{n} \boldsymbol{g}(\boldsymbol{D}_i^u; \hat{\boldsymbol{\beta}}_{EN}) \\
=& \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{g}(\boldsymbol{D}_i^u; \boldsymbol{\beta}_0) + \frac{1}{n}\sum_{i=1}^{n} \frac{\partial \boldsymbol{g}(\boldsymbol{D}_i^u; \boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}} (\hat{\boldsymbol{\beta}}_{EN} - \boldsymbol{\beta}_0) - \frac{1}{n}\sum_{i=1}^{m_1} \boldsymbol{g}(\boldsymbol{D}_i^u; \boldsymbol{\beta}_0)\boldsymbol{h}(\boldsymbol{D}_i^a; \boldsymbol{\theta}_*)^T \hat{\boldsymbol{\lambda}} + o_p(\boldsymbol{n}^{-\frac{1}{2}}),
\end{aligned}$$

The third equation is based on Taylor expansion with respect to $\boldsymbol{\beta}_0$. Thus, the asymptotic expansion of the estimator $\hat{\boldsymbol{\beta}}_{EN}$ can be derived as

$$\begin{aligned}
& n^{\frac{1}{2}}(\hat{\boldsymbol{\beta}}_{EN} - \boldsymbol{\beta}_0) \\
=& -\left(\frac{1}{n}\sum_{i=1}^{n} \frac{\partial \boldsymbol{g}(\boldsymbol{D}_i^u; \boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}^T}\right)^{-1} \left[n^{-\frac{1}{2}}\sum_{i=1}^{n} \boldsymbol{g}(\boldsymbol{D}_i^u; \boldsymbol{\beta}_0) - n^{-\frac{1}{2}}\sum_{i=1}^{m_1} \left\{\boldsymbol{g}(\boldsymbol{D}_i^u; \boldsymbol{\beta}_0)\boldsymbol{h}(\boldsymbol{D}_i^a; \boldsymbol{\theta}_*)^T\right\}\hat{\boldsymbol{\lambda}}\right] + o_p(\mathbf{1}) \\
=& -\boldsymbol{\Gamma}^{-1}\left\{n^{-\frac{1}{2}}\sum_{i=1}^{n} \boldsymbol{g}(\boldsymbol{D}_i^u; \boldsymbol{\beta}_0) - n^{-\frac{1}{2}}\boldsymbol{\Lambda}\boldsymbol{S}\boldsymbol{Q}_{m_1}(\boldsymbol{\theta}_*)\right\} + o_p(\mathbf{1})
\end{aligned} \tag{2}$$

To complete the proof, it suffices to calculate the asymptotic variance. Based on the

influence function in (2), we have

$$\begin{aligned}\lim_{n\to\infty} var\Big\{n^{\frac{1}{2}}(\hat{\boldsymbol{\beta}}_{EN}-\boldsymbol{\beta}_0)\Big\} = \lim_{n\to\infty} \boldsymbol{\Gamma}^{-1}\Bigg[& var\Big\{n^{-\frac{1}{2}}\sum_{i=1}^{n}\boldsymbol{g}(\boldsymbol{D}_i^u;\boldsymbol{\beta}_0)\Big\} + var\Big\{n^{-\frac{1}{2}}\boldsymbol{\Lambda}\boldsymbol{S}\boldsymbol{Q}_{m_1}(\boldsymbol{\theta}_*)\Big\} \\ & - cov\Big\{n^{-\frac{1}{2}}\sum_{i=1}^{n}\boldsymbol{g}(\boldsymbol{D}_i^u;\boldsymbol{\beta}_0), n^{-\frac{1}{2}}\boldsymbol{\Lambda}\boldsymbol{S}\boldsymbol{Q}_{m_1}(\boldsymbol{\theta}_*)\Big\} \\ & - cov\Big\{n^{-\frac{1}{2}}\boldsymbol{\Lambda}\boldsymbol{S}\boldsymbol{Q}_{m_1}(\boldsymbol{\theta}_*), n^{-\frac{1}{2}}\sum_{i=1}^{n}\boldsymbol{g}(\boldsymbol{D}_i^u;\boldsymbol{\beta}_0)\Big\}\Bigg](\boldsymbol{\Gamma}^T)^{-1}.\end{aligned}$$

Notice that

$$\begin{aligned}\lim_{n\to\infty} var\Big\{n^{-\frac{1}{2}}\sum_{i=1}^{n}\boldsymbol{g}(\boldsymbol{D}_i^u;\boldsymbol{\beta}_0)\Big\} =& \boldsymbol{\Sigma}, \\ \lim_{n\to\infty} var\Big\{n^{-\frac{1}{2}}\boldsymbol{\Lambda}\boldsymbol{S}\boldsymbol{Q}_{m_1}(\boldsymbol{\theta}_*)\Big\} =& \rho\boldsymbol{\Lambda}\boldsymbol{S}\boldsymbol{\Lambda}^T, \\ \lim_{n\to\infty} cov\Big\{n^{-\frac{1}{2}}\sum_{i=1}^{n}\boldsymbol{g}(\boldsymbol{D}_i^u;\boldsymbol{\beta}_0), n^{-\frac{1}{2}}\boldsymbol{\Lambda}\boldsymbol{S}\boldsymbol{Q}_{m_1}(\boldsymbol{\theta}_*)\Big\} =& \rho\boldsymbol{\Lambda}\boldsymbol{S}\boldsymbol{\Lambda}^T.\end{aligned}$$

Thus, based on central limit theory, we have the asymptotic normality of $n^{1/2}(\hat{\boldsymbol{\beta}}_{EN}-\boldsymbol{\beta}_0)$ with the asymptotic variance-covariance matrix equal to $\boldsymbol{\Gamma}^{-1}(\boldsymbol{\Sigma}-\rho\boldsymbol{\Lambda}\boldsymbol{S}\boldsymbol{\Lambda}^T)(\boldsymbol{\Gamma}^T)^{-1}$.

1.2 Proof of Property 1

In the following proof, we will utilize $\boldsymbol{S}_{11}$ and $\boldsymbol{S}_{12}$ to represent $\boldsymbol{S}_{11}(\boldsymbol{\theta}_*)$ and $\boldsymbol{S}_{12}(\boldsymbol{\theta}_*)$, respectively. The same strategy from the proof of Theorem 1 will be applied to the proof of Property 1. For convenience, let us denote the solution to the estimating equations $\boldsymbol{G}_n^*(\boldsymbol{\beta})=\boldsymbol{0}$ as $\hat{\boldsymbol{\beta}}$. Now applying Taylor expansion to the estimating equations at $\boldsymbol{\beta}_0$, we have

$$n^{\frac{1}{2}}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0) = -\boldsymbol{\Gamma}^{-1}n^{\frac{1}{2}}\Big\{\boldsymbol{G}_n(\boldsymbol{\beta}_0)-\boldsymbol{C}\boldsymbol{Q}_{m_1}(\hat{\boldsymbol{\theta}})\Big\}+o_p(\boldsymbol{1}), \tag{3}$$

with $\boldsymbol{G}_n(\boldsymbol{\beta}_0)$ defined as $1/n\sum_{i=1}^{n}\boldsymbol{g}(\boldsymbol{D}_i^u;\boldsymbol{\beta}_0)$ and $\boldsymbol{Q}_{m_1}(\hat{\boldsymbol{\theta}})$ re-defined as $1/n\sum_{i=1}^{m_1}\boldsymbol{h}(\boldsymbol{D}_i^a;\hat{\boldsymbol{\theta}})$.

Notice that, by empirical likelihood theorem (Qin and Lawless, 1994) and Taylor expansion at $\boldsymbol{\theta}_*$, $\boldsymbol{Q}_{m_1}(\hat{\boldsymbol{\theta}})$ can be expressed as

$$\begin{aligned}\boldsymbol{Q}_{m_1}(\hat{\boldsymbol{\theta}}) =& \boldsymbol{Q}_{m_1}(\boldsymbol{\theta}_*)+\frac{\partial\boldsymbol{Q}_{m_1}(\boldsymbol{\theta}_*)}{\partial\boldsymbol{\theta}^T}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_*)+o_p(\boldsymbol{n}^{-\frac{1}{2}}), \quad \text{with} \\ n^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_*) =& -n^{\frac{1}{2}}\boldsymbol{\Omega}\boldsymbol{S}_{21}\boldsymbol{S}_{11}^{-1}\frac{n}{m_1}\boldsymbol{Q}_{m_1}(\boldsymbol{\theta}_*)+o_p(\boldsymbol{1}).\end{aligned}$$

By applying the above results, the expression in (3) becomes

$$n^{\frac{1}{2}}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0) = -\boldsymbol{\Gamma}^{-1}\Big\{n^{\frac{1}{2}}\boldsymbol{G}_{m_1}(\boldsymbol{\beta}_0)+n^{\frac{1}{2}}\boldsymbol{G}_{m_2}(\boldsymbol{\beta}_0)-n^{\frac{1}{2}}\boldsymbol{C}\boldsymbol{A}(\boldsymbol{\theta}_*)\boldsymbol{Q}_{m_1}(\boldsymbol{\theta}_*)\Big\}+o_p(\boldsymbol{1}) \tag{4}$$

with $\boldsymbol{A}(\boldsymbol{\theta}_*) = \boldsymbol{I} - \boldsymbol{S}_{12}\boldsymbol{\Omega}\boldsymbol{S}_{21}\boldsymbol{S}_{11}^{-1}$, $\boldsymbol{G}_{m_1}(\boldsymbol{\beta}_0) = 1/n\sum_{i=1}^{m_1}\boldsymbol{g}(\boldsymbol{D}_i^u;\boldsymbol{\beta}_0)$, and $\boldsymbol{G}_{m_2}(\boldsymbol{\beta}_0) = 1/n\sum_{i=m_1+1}^{n}\boldsymbol{g}(\boldsymbol{D}_i^u;\boldsymbol{\beta}_0)$.

To obtain the estimator $\hat{\boldsymbol{\beta}}$ in (4) with smallest variance, it suffices to project $\boldsymbol{G}_{m_1}(\boldsymbol{\beta}_0) + \boldsymbol{G}_{m_2}(\boldsymbol{\beta}_0)$ onto the linear subspace spanned by $\boldsymbol{A}(\boldsymbol{\theta}_*)\boldsymbol{Q}_{m_1}(\boldsymbol{\theta}_*)$ based on projection theory (Qin, 2017), which leads the matrix $\boldsymbol{C}$ equal to $E\{(\boldsymbol{G}_{m_1} + \boldsymbol{G}_{m_2})\tilde{\boldsymbol{Q}}_{m_1}^T\}E^-(\tilde{\boldsymbol{Q}}_{m_1}\tilde{\boldsymbol{Q}}_{m_1}^T) = E(\boldsymbol{G}_{m_1}\tilde{\boldsymbol{Q}}_{m_1}^T)E^-(\tilde{\boldsymbol{Q}}_{m_1}\tilde{\boldsymbol{Q}}_{m_1}^T)$ with $\tilde{\boldsymbol{Q}}_{m_1}(\boldsymbol{\theta}_*) = \boldsymbol{A}(\boldsymbol{\theta}_*)\boldsymbol{Q}_{m_1}(\boldsymbol{\theta}_*)$. Here, the notation E^- represents generalized inverse of the expectation. Therefore, under such a $\boldsymbol{C}$ matrix, we have

$$\begin{aligned}\lim_{n\to\infty} var\{n^{\frac{1}{2}}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)\} = \lim_{n\to\infty} \boldsymbol{\Gamma}^{-1}\Big[&var\{n^{\frac{1}{2}}\boldsymbol{G}_{m_1}(\boldsymbol{\beta}_0) + n^{\frac{1}{2}}\boldsymbol{G}_{m_2}(\boldsymbol{\beta}_0)\} + var\{n^{\frac{1}{2}}\boldsymbol{C}\tilde{Q}_{m_1}(\boldsymbol{\theta}_*)\}\\ &- cov\{n^{\frac{1}{2}}\boldsymbol{G}_{m_1}(\boldsymbol{\beta}_0) + n^{\frac{1}{2}}\boldsymbol{G}_{m_2}(\boldsymbol{\beta}_0), n^{\frac{1}{2}}\boldsymbol{C}\tilde{Q}_{m_1}(\boldsymbol{\theta}_*)\}\\ &- cov\{n^{\frac{1}{2}}\boldsymbol{C}\tilde{Q}_{m_1}(\boldsymbol{\theta}_*), n^{\frac{1}{2}}\boldsymbol{G}_{m_1}(\boldsymbol{\beta}_0) + n^{\frac{1}{2}}\boldsymbol{G}_{m_2}(\boldsymbol{\beta}_0)\}\Big](\boldsymbol{\Gamma}^T)^{-1}\\ =&\boldsymbol{\Gamma}^{-1}\Big\{\boldsymbol{\Sigma} - \lim_{n\to\infty} nE(\boldsymbol{G}_{m_1}\tilde{\boldsymbol{Q}}_{m_1}^T)E^-(\tilde{\boldsymbol{Q}}_{m_1}\tilde{\boldsymbol{Q}}_{m_1}^T)E(\tilde{\boldsymbol{Q}}_{m_1}\boldsymbol{G}_{m_1}^T)\Big\}(\boldsymbol{\Gamma}^T)^{-1}.\end{aligned}$$

The second equality holds by the fact that

$$\begin{aligned}var\big(n^{\frac{1}{2}}\boldsymbol{C}\tilde{Q}_{m_1}(\boldsymbol{\theta}_*)\big) =&nE(\boldsymbol{G}_{m_1}\tilde{\boldsymbol{Q}}_{m_1}^T)E^-(\tilde{\boldsymbol{Q}}_{m_1}\tilde{\boldsymbol{Q}}_{m_1}^T)E(\tilde{\boldsymbol{Q}}_{m_1}\boldsymbol{G}_{m_1}^T),\\ cov\big(n^{\frac{1}{2}}\boldsymbol{G}_{m_1}(\boldsymbol{\beta}_0) + n^{\frac{1}{2}}\boldsymbol{G}_{m_2}(\boldsymbol{\beta}_0), n^{\frac{1}{2}}\boldsymbol{C}\tilde{Q}_{m_1}(\boldsymbol{\theta}_*)\big) =&nE(\boldsymbol{G}_{m_1}\tilde{\boldsymbol{Q}}_{m_1}^T)E^-(\tilde{\boldsymbol{Q}}_{m_1}\tilde{\boldsymbol{Q}}_{m_1}^T)E(\tilde{\boldsymbol{Q}}_{m_1}\boldsymbol{G}_{m_1}^T),\\ \lim_{n\to\infty} var\big(n^{\frac{1}{2}}\boldsymbol{G}_{m_1}(\boldsymbol{\beta}_0) + n^{\frac{1}{2}}\boldsymbol{G}_{m_2}(\boldsymbol{\beta}_0)\big) =&\boldsymbol{\Sigma}.\end{aligned}$$

Finally, by realizing that $E(\boldsymbol{G}_{m_1}\tilde{\boldsymbol{Q}}_{m_1}^T) = (m_1/n^2)\boldsymbol{\Lambda}(I - \boldsymbol{S}_{11}^{-1}\boldsymbol{S}_{12}\boldsymbol{\Omega}\boldsymbol{S}_{21})$ and $E(\tilde{\boldsymbol{Q}}_{m_1}\tilde{\boldsymbol{Q}}_{m_1}^T) = (m_1/n^2)(\boldsymbol{S}_{11} - \boldsymbol{S}_{12}\boldsymbol{\Omega}\boldsymbol{S}_{21})$, we have

$$\lim_{n\to\infty} var\big(n^{\frac{1}{2}}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)\big) = \boldsymbol{\Gamma}^{-1}\left(\boldsymbol{\Sigma} - \rho\boldsymbol{\Lambda}\boldsymbol{S}\boldsymbol{\Lambda}^T\right)(\boldsymbol{\Gamma}^{-1})^T = \boldsymbol{V}_{EN},$$

which completes the proof.

1.3 An alternative equivalent formulation of the proposed estimation procedure

let us consider another formulation of the proposed method. Let η_i be the indicator which equals one if the i^{th} subject has the auxiliary records, and is equal to zero otherwise. Then, the reweighting scheme in (2) from the main manuscript can be summarized as

$$\sum_{i=1}^{n}\hat{p}_i^*\boldsymbol{g}(\boldsymbol{D}_i^u;\boldsymbol{\beta}) = \boldsymbol{0}, \tag{5}$$

where the non-negative weights $\hat{p}_i^*$ on subjects $i = 1,\ldots,n$ are obtained by maximizing $\prod_{i=1}^{n} p_i^*$ under the constraints

$$\sum_{i=1}^{n} p_i^* = 1, \sum_{i=1}^{n} p_i^*\eta_i\boldsymbol{h}(\boldsymbol{D}_i^a;\boldsymbol{\theta}) = \boldsymbol{0}. \tag{6}$$

It can be shown that the above scheme is equal to the one in (2) from the main manuscript, given that the first m_1 subjects in the study have auxiliary data, and the rest subjects have no auxiliary records. To this end, notice that the estimated weights from (6) are $\hat{p}_i^* = n^{-1}/\{1 + \hat{\boldsymbol{\lambda}}^T \eta_i \boldsymbol{h}(\boldsymbol{D}_i^a; \hat{\boldsymbol{\theta}})\}$ by applying Lagrange multiplier technique from Section 1.1, which equals $1/n$ for $i = m_1 + 1, \ldots, n$. The constrains in (6) then become $\sum_{i=1}^{m_1} p_i^* = m_1/n$ and $\sum_{i=1}^{m_1} p_i^* \boldsymbol{h}(\boldsymbol{D}_i^a; \boldsymbol{\theta}) = \mathbf{0}$, and the estimating equation in (4) becomes $\sum_{i=1}^{m_1} \hat{p}_i^* \boldsymbol{g}(\boldsymbol{D}_i^u; \boldsymbol{\beta}) + \sum_{i=m_1+1}^{n} (1/n) \boldsymbol{g}(\boldsymbol{D}_i^u; \boldsymbol{\beta}) = \mathbf{0}$. Thus, (5) and (6) are reduced to (2) and (3) from the main manuscript by realizing the fact that $\hat{p}_i^* = (m_1/n)\hat{p}_i$, where $\hat{p}_i$ are the estimated weights solving (3) from the main manuscript.

2 Extra simulation results

In this section, we will examine the performance of our proposed estimator by considering a linear model for the main analysis, in the presence of longitudinal measurements as auxiliary data. To be specific, suppose, for $i = 1, \ldots, m_1$, the auxiliary data $\boldsymbol{D}_i^a$ contains the repeated measurements $\tilde{\boldsymbol{Y}}_i$ with length $T = 5$ generated by the model $\tilde{\boldsymbol{Y}}_i = \tilde{\boldsymbol{X}}_i \boldsymbol{\theta} + \tilde{\boldsymbol{\epsilon}}_i$ with $\boldsymbol{\theta} = (-1, 1, 2)^T$ and $\tilde{\boldsymbol{X}}_i = (\tilde{\boldsymbol{X}}_{i1}, \ldots, \tilde{\boldsymbol{X}}_{iT})^T$, where $\tilde{\boldsymbol{X}}_{it} = (1, \tilde{X}_{it1}, \tilde{X}_{it2})^T$ with $\tilde{X}_{it1}$ following independent and identical uniform distribution within support $[0, 1]$ and $\tilde{X}_{it2}$ following independent and identical standard normal distribution, for $t = 1, \ldots, T$. On the other hand, for $i = 1, \ldots, n$, the outcomes Y_i in the main data $\boldsymbol{D}_i^u$ is generated by the linear model $Y_i = \tilde{\boldsymbol{X}}_{iT}^T \boldsymbol{\beta} + \epsilon_i$ with $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)^T = (1, -1, -1)^T$. In order to make feasible of borrowing information from the auxiliary data, we require residuals $\bar{\boldsymbol{\epsilon}}_i = (\tilde{\boldsymbol{\epsilon}}_i^T, \epsilon_i)^T$ in both data sets follow 6-dimensional multivariate normal distribution with mean zeros and covariance matrix $\sigma^2 \boldsymbol{C}$, given the variance $\sigma^2 = 1$ and correlation coefficient $\rho = 0.4, 0.6, 0.8$, respectively. Thereafter, we investigate three different situations by specifying the functions $\boldsymbol{h}(\boldsymbol{D}_i^a; \boldsymbol{\theta})$ in (4) for the auxiliary data. In situation 1 (S1), we select base matrices $\boldsymbol{V}_j$ such that $\sum_{j=1}^{\tau} a_j \boldsymbol{V}_j = \boldsymbol{C}^{-1}$. In particular, when the true correlation structure $\boldsymbol{C}$ is exchangeable, we have two base matrices, i.e., identity matrix $\boldsymbol{V}_1$ and a matrix $\boldsymbol{V}_2$ with 0 on the diagonal and 1 off the diagonal; when AR1 structure is applied, we have three base matrices, i.e., $\boldsymbol{V}_1$ defined above, $\boldsymbol{V}_3$ with 1 on the two main off-diagonal and 0 otherwise, and $\boldsymbol{V}_4$ with 1 on the left-up and right-bottom corners and 0 elsewhere. Situation 2 (S2) aims to incorporate all $\boldsymbol{V}_1$, $\boldsymbol{V}_2$, $\boldsymbol{V}_3$, and $\boldsymbol{V}_4$ base matrices into (4). In situation 3 (S3), we evaluate the behavior of mis-specification for the mean structure $\boldsymbol{\mu}_i(\boldsymbol{\theta})$ in (4), i.e., instead of utilizing $\tilde{\boldsymbol{X}}_i$, we use covariate matrix $\tilde{\boldsymbol{Z}}_i = (\tilde{\boldsymbol{Z}}_{i1}, \ldots, \tilde{\boldsymbol{Z}}_{iT})$ where $\tilde{\boldsymbol{Z}}_{it} = (1, \tilde{X}_{it1} + 0.5\tilde{X}_{it2})$. For all situations above, we will only present the results where the underlying true correlation structure $\boldsymbol{C}$ is exchangeable. To evaluate the performance of our proposed estimator, we implemented 1000 Monte Carlo runs, where 75% and 100% subjects have auxiliary data for sample size

$n = 200, 600$, respectively. To further compare with the estimator from classic least square method applied to the main data, we record certain measurements, including empirical relative efficiency of the proposed estimator versus ordinary least square estimator, absolute value of bias, empirical standard error, estimated asymptotic standard error, and empirical coverage proportion of the proposed estimator with IIB recorded for each setup.

The results are summarized in Table S.1. Overall, the proposed estimator has a satisfactory performance as evidenced by the low bias and high relative efficiency. The empirical coverage proportion becomes closer to 95% nominal level as sample size increases. We can also observe that, in general, more auxiliary data would enhance the efficiency in the main analysis, which has been notified in Theorem 1 from the main manuscript. For all situations, the relative efficiency increases as the correlation coefficient of residuals between two outcomes increases. This phenomenon is expected as well, since stronger association implies more information sharing between the auxiliary and main data. In addition, for any given correlation coefficient, the relative efficiency in situation 2 is always higher than that in situation 1. This result supports the theoretical conclusion that adding extra set of estimating equations would further enhance the estimation efficiency. In situation 3, in which the mean structure is mis-specified, the proposed estimator has little bias and better efficiency compared to the classic least square approach. This further indicates the flexibility and robustness of our proposed estimator. Besides, all situations have showed that IIB has the potential to provide a fair evaluation on how well and how much amount of information is borrowed from the auxiliary data.

Moreover, Table S.2 summarizes the results in Example 1 from the main manuscript, where the main outcome is binary and only partial auxiliary data are observed. Note that in Table 1 from the main manuscript, there is little efficiency gain for the estimated parameter ($\hat{\beta}_3$) corresponding to a time-independent covariate. By selecting proper basis matrices, it is possible to substantially increase the estimation efficiency for parameters associated with time-independent covariates. To see this, we consider four basis matrices: $\tilde{V}_1$, $\tilde{V}_2$, $\tilde{V}_3$, and $\tilde{V}_4$, where i^{th} diagonal of $\tilde{V}_i$ is equal to one, and zero otherwise, for $i = 1, 2, 3, 4$. All other setups remain the same to Section 3.1 in the main manuscript, and the results with $r_0 = 0.9$ and $\rho = 0.4$ (defined in Section 3.1 in the main manuscript) are summarized in Table S.3 It can be seen that there is considerable efficiency gain for parameters associated with both time-independent and time-dependent covariates, due to larger than one ERE.

3 Extra discussions and one extension to missing data

3.1 One note about covariates and efficiency gain

In the simulation study from the main manuscript, we find that the covariates in the pool of $\tilde{\boldsymbol{X}}_i$ in Example 2 have little efficiency gain. Here, we will provide some mathematical insight onto this phenomenon. Let us consider a special case in Example 2 with $\boldsymbol{d}(\tilde{\boldsymbol{X}}_i, \tilde{\boldsymbol{Z}}_i; \boldsymbol{\theta}) = (\tilde{\boldsymbol{X}}_i^T, \tilde{\boldsymbol{Z}}_i^T)^T$ for illustration. In the auxiliary data, we have a contiunous-scale auxiliary variable $\tilde{Y}_i$ and covariates $\tilde{\boldsymbol{X}}_i$ involved in the mean structure $\mu(\tilde{\boldsymbol{X}}_i; \boldsymbol{\theta})$ as well as some redundant covariates $\tilde{\boldsymbol{Z}}_i$; in the main analysis, the covariates we utilize are $\boldsymbol{X}_i = \tilde{\boldsymbol{X}}_i$, i.e., the same covariates used in the mean structure $\mu(\tilde{\boldsymbol{X}}_i; \boldsymbol{\theta})$ in the working model. Furthermore, we assume that the covariance between two outcomes (Y_i and $\tilde{Y}_i$) in the main data and auxiliary data is constant, denoted by a, and we adopt the score functions for $\boldsymbol{g}(\boldsymbol{D}_i^u; \boldsymbol{\beta})$ to solve the parameters of interest $\boldsymbol{\beta}$. Then, by some algebra, we can check that the matrix $\boldsymbol{\Lambda}\boldsymbol{S}\boldsymbol{\Lambda}^T$ is equal to zero matrix. The underlying reason leading to this result is that the association matrix $\boldsymbol{\Lambda}$ in this situation becomes $a\boldsymbol{S}_{21}$, thus resulting in $\boldsymbol{\Lambda}\boldsymbol{S}_{11}^{-1}(\boldsymbol{\theta}_*)\boldsymbol{S}_{12}(\boldsymbol{\theta}_*)\boldsymbol{\Omega}(\boldsymbol{\theta}_*)\boldsymbol{S}_{21}(\boldsymbol{\theta}_*)\boldsymbol{S}_{11}^{-1}(\boldsymbol{\theta}_*)\boldsymbol{\Lambda}^T = \boldsymbol{\Lambda}\boldsymbol{S}_{11}^{-1}(\boldsymbol{\theta}_*)\boldsymbol{\Lambda}^T$ and then $\boldsymbol{\Lambda}\boldsymbol{S}\boldsymbol{\Lambda}^T = \boldsymbol{0}$. But it would not be the case if the covariates in the main analysis become $\boldsymbol{X}_i = \tilde{\boldsymbol{Z}}_i$. These findings may provide some guidance for researchers to select proper covariates in the working model, in order to achieve a satisfactory efficiency gain in the main estimation. More simulation studies can be conducted to check this, which is omitted in this paper.

3.2 Extension to missing data problem

In simulation studies and the real data application from the main manuscript, we assume all the main data are observed or missing completely at random, which may not be the case in some applications. However, our method can be easily accommodated to the data missing at random. Herein, we present an extension by incorporating inverse probability weight for illustration. To be specific, let us denote observing indicator $\tilde{R}_i$, which is equal to one if subject i is observed and 0 otherwise. Then, we define the probability of observing subject i as $\pi_i(\boldsymbol{\alpha}) = E(\tilde{R}_i|\boldsymbol{X}_i)$, where $\boldsymbol{\alpha}$ are parameters involved in the missing data model. Furthermore, let us define the score function for parameters $\boldsymbol{\alpha}$ in the missing data model as $\sum_{i=1}^n \boldsymbol{L}_i(\boldsymbol{\alpha})$, which can be obtained from logistic regression by modeling $\tilde{R}_i$ with covariates $\boldsymbol{X}_i$. We keep other setups the same to Section 2 in the main manuscript. Then, we simultaneously solve $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ by the following modified weighted estimating equations:

$$\sum_{i=1}^{m_1} \hat{p}_i \tilde{\boldsymbol{g}}(\boldsymbol{D}_i^u, \boldsymbol{\alpha}; \boldsymbol{\beta}) + \sum_{i=m_1+1}^{n} \tilde{\boldsymbol{g}}(\boldsymbol{D}_i^u, \boldsymbol{\alpha}; \boldsymbol{\beta}) = \boldsymbol{0}, \tag{7}$$

where $\tilde{\boldsymbol{g}}(\boldsymbol{D}_i^u, \boldsymbol{\alpha}; \boldsymbol{\beta}) = (R_i/\pi_i(\boldsymbol{\alpha})\boldsymbol{g}^T(\boldsymbol{D}_i^u; \boldsymbol{\beta}), \boldsymbol{L}_i^T(\boldsymbol{\alpha}))^T$. Here $\boldsymbol{g}(\boldsymbol{D}_i^u; \boldsymbol{\beta})$ is the same in (2) from the main manuscript. The final asymptotic variance-covariance matrix for joint $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ parameters will be the same in Theorem 1 by replacing $\boldsymbol{g}(\boldsymbol{D}_i^u; \boldsymbol{\beta}_0)$ with $\tilde{\boldsymbol{g}}(\boldsymbol{D}_i^u, \boldsymbol{\alpha}_0; \boldsymbol{\beta}_0)$, where $\boldsymbol{\alpha}_0$ and $\boldsymbol{\beta}_0$ are true parameter values. Weights $\hat{p}_i$ are the same from solving constrained optimization problem in (3) from the main manuscript.

In the presence of high missingness, we can see that the only modification is to construct an extended estimating functions $\tilde{\boldsymbol{g}}(\boldsymbol{D}_i^u, \boldsymbol{\alpha}; \boldsymbol{\beta})$ for the main analysis. The first component in $\tilde{\boldsymbol{g}}(\boldsymbol{D}_i^u, \boldsymbol{\alpha}; \boldsymbol{\beta})$ is a typical estimating function under inverse probability weight framework. The second term is to make the constructed estimating functions $\tilde{\boldsymbol{g}}(\boldsymbol{D}_i^u, \boldsymbol{\alpha}; \boldsymbol{\beta})$ independently and identically distributed, in order to guarantee the efficiency gain for $\boldsymbol{\beta}$ estimates in theory. Numerical evaluations can be done by generating the main data with missing at random. Since this paper does not specifically focus on the missing data problem, we omit the details and regard it as a future work.

References

Newey, W. K. and McFadden, D. (1994). Large sample estimation and hypothesis testing. *Handbook of econometrics* **4,** 2111–2245.

Owen, A. B. (2001). *Empirical likelihood*. CRC press.

Qin, J. (2017). *Biased sampling, over-identified parameter problems and beyond*. Springer.

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

Table S.1: Simulation results for linear regression in the main analysis by borrowing information from continuous scaled auxiliary records.

			β_1					β_2					
Set-up			Bias	ESE	ASE	ERE	95%CP	Bias	ESE	ASE	ERE	95%CP	IIB
P=100%	S2	ρ=0.8	0.3	21	19	1.48	93	0.2	6.2	5.6	1.35	91	0.49
n=200		ρ=0.6	0.2	22	21	1.26	94	0.3	6.7	6.1	1.15	91	0.35
		ρ=0.4	0.2	23	22	1.12	94	0.3	7.2	6.5	1.03	91	0.23
	S1	ρ=0.8	0.6	23	22	1.17	94	0.3	6.6	6.5	1.17	94	0.20
		ρ=0.6	0.5	24	23	1.12	94	0.3	6.9	6.7	1.12	94	0.14
		ρ=0.4	0.5	24	23	1.07	95	0.2	7.1	6.8	1.07	93	0.09
	S3	ρ=0.8	0.2	25	23	1.01	94	0.4	6.5	6.2	1.22	94	0.28
		ρ=0.6	0.3	25	24	1.00	94	0.4	6.9	6.5	1.12	93	0.20
		ρ=0.4	0.3	25	24	0.99	94	0.3	7.1	6.7	1.04	92	0.13
P=100%	S2	ρ=0.8	0.1	11	11	1.56	94	0.2	3.4	3.3	1.51	95	0.42
n=600		ρ=0.6	0.0	13	12	1.28	94	0.2	3.7	3.6	1.28	95	0.28
		ρ=0.4	0.0	14	14	1.12	94	0.2	3.9	4.1	1.12	94	0.16
	S1	ρ=0.8	0.1	13	13	1.22	94	0.2	3.9	3.7	1.11	94	0.18
		ρ=0.6	0.1	13	13	1.14	94	0.2	4.0	3.9	1.07	94	0.12
		ρ=0.4	0.0	14	14	1.07	95	0.2	4.0	3.9	1.03	94	0.07
	S3	ρ=0.8	0.2	14	14	1.06	94	0.1	3.7	3.6	1.23	95	0.24
		ρ=0.6	0.0	14	14	1.03	94	0.1	3.9	3.8	1.14	94	0.16
		ρ=0.4	0.1	14	14	1.00	94	0.1	4.0	3.9	1.07	94	0.09
P=75%	S2	ρ=0.8	0.4	22	20	1.31	93	0.2	6.5	5.9	1.21	93	0.52
n=200		ρ=0.6	0.2	23	22	1.19	94	0.2	6.9	6.3	1.09	92	0.38
		ρ=0.4	0.1	24	23	1.09	94	0.2	7.2	6.6	1.02	92	0.26
	S1	ρ=0.8	0.4	24	23	1.12	93	0.2	6.8	6.6	1.13	94	0.22
		ρ=0.6	0.4	24	23	1.09	94	0.2	7.0	6.7	1.09	94	0.16
		ρ=0.4	0.4	24	24	1.06	95	0.2	7.1	6.8	1.05	93	0.10
	S3	ρ=0.8	0.4	25	24	1.01	95	0.2	6.9	6.5	1.09	94	0.23
		ρ=0.6	0.4	24	24	1.00	95	0.2	7.2	6.7	1.03	93	0.17
		ρ=0.4	0.4	24	24	1.00	95	0.2	7.3	6.8	0.99	92	0.13
P=75%	S2	ρ=0.8	0.0	12	12	1.36	94	0.2	3.6	3.5	1.31	94	0.44
n=600		ρ=0.6	0.0	13	13	1.19	95	0.2	3.8	3.7	1.17	94	0.29
		ρ=0.4	0.1	14	13	1.08	94	0.2	4.0	3.9	1.07	93	0.18
	S1	ρ=0.8	0.2	13	13	1.16	95	0.2	4.0	3.8	1.06	93	0.18
		ρ=0.6	0.1	14	14	1.11	94	0.2	4.1	3.9	1.03	93	0.13
		ρ=0.4	0.0	14	14	1.06	95	0.2	4.1	4.0	1.01	94	0.08
	S3	ρ=0.8	0.3	14	14	1.01	94	0.2	4.0	3.8	1.12	94	0.18
		ρ=0.6	0.4	14	14	0.99	94	0.2	4.0	4.0	1.07	94	0.12
		ρ=0.4	0.4	15	14	0.98	95	0.1	4.1	4.0	1.03	94	0.07

1000 Monte Carlo simulation runs are implemented under three situations (S1,S2,S3) with correlation coefficient ρ equal to 0.8, 0.6, and 0.4, respectively. The sample size $n = 200, 600$, where 75%,100% subjects have auxiliary data ($P = 75\%, 100\%$), respectively. The measurements such as absolute value of bias, empirical standard error (ESE), estimated asymptotic standard error (ASE), empirical relative efficiency (ERE), coverage probability (CP), and IIB are recorded. All values except ERE and IIB are multiplied by 100.

Table S.2: Simulation results for Example 1 in the main manuscript. Half of the subjects have auxiliary data ($\rho = 50\%$)

			n=300						n=600					
			Bias	ESE	ASE	ERE	95%CP	*IIB*	Bias	ESE	ASE	ERE	95%CP	*IIB*
S2	$\tilde{\rho}$=0.4	β_0	2	25	24	0.98	95.0	0.166	1	18	17	0.99	94.6	0.111
	r_0=0.5	β_1	-2	29	29	0.98	95.3		-2	20	20	1.01	96.0	
		β_2	-3	17	16	0.97	94.5		-1	12	12	0.99	94.8	
		β_3	1	30	29	0.98	93.7		0	21	20	1.00	94.1	
	$\tilde{\rho}$=0.4	β_0	2	25	24	0.98	94.4	0.293	2	17	17	1.01	95.5	0.237
	r_0=0.9	β_1	-2	28	28	1.01	93.9		-1	20	20	1.04	95.3	
		β_2	-3	17	16	1.01	93.7		-1	12	11	1.04	94.4	
		β_3	1	29	29	1.01	95.1		0	20	20	1.02	95.2	
	$\tilde{\rho}$=0.8	β_0	2	25	24	0.97	95.4	0.181	1	17	17	1.00	95.7	0.126
	r_0=0.5	β_1	-2	28	28	0.99	95.7		-2	20	20	1.03	95.9	
		β_2	-3	17	16	0.97	94.7		-1	12	12	1.01	94.7	
		β_3	1	30	29	0.97	93.8		0	21	20	0.99	93.6	
	$\tilde{\rho}$=0.8	β_0	2	25	24	0.98	94.4	0.344	2	18	17	1.00	94.4	0.290
	r_0=0.9	β_1	-2	28	28	1.01	93.9		-1	19	19	1.08	95.4	
		β_2	-3	17	16	1.01	93.7		-1	12	11	1.06	94.0	
		β_3	1	29	29	1.01	95.1		0	20	20	0.99	94.6	
S1	$\tilde{\rho}$=0.4	β_0	2	25	25	0.99	95.3	0.042	1	17	17	1.00	95.3	0.028
	r_0=0.5	β_1	-2	29	29	0.99	95.6		-2	20	20	1.01	96.6	
		β_2	-3	17	17	0.99	95.0		-1	12	12	1.00	94.9	
		β_3	1	30	29	0.98	94.6		0	21	20	1.00	94.1	
	$\tilde{\rho}$=0.4	β_0	3	25	25	0.99	95.4	0.075	2	17	17	1.00	95.0	0.060
	r_0=0.9	β_1	-2	28	29	1.00	95.1		-1	20	20	1.01	95.8	
		β_2	-3	17	17	1.00	94.4		-1	12	12	1.02	94.1	
		β_3	1	29	29	0.99	95.7		0	20	20	1.00	95.6	
	$\tilde{\rho}$=0.8	β_0	2	24	25	0.99	95.6	0.072	1	17	17	1.01	95.9	0.056
	r_0=0.5	β_1	-2	28	29	1.01	95.8		-2	20	20	1.02	96.1	
		β_2	-3	17	17	0.98	94.8		-1	12	12	1.02	94.6	
		β_3	1	30	29	0.98	94.6		0	21	20	1.00	93.7	
	$\tilde{\rho}$=0.8	β_0	2	25	25	1.01	96.2	0.170	2	17	17	1.01	94.7	0.155
	r_0=0.9	β_1	-2	28	28	1.07	95.7		-1	19	20	1.06	95.2	
		β_2	-3	17	16	1.03	93.4		-1	12	11	1.04	94.7	
		β_3	2	29	29	0.98	94.4		0	20	20	0.99	94.8	
S3	$\tilde{\rho}$=0.4	β_0	2	25	25	0.98	94.9	0.131	1	18	17	0.99	94.5	0.083
	r_0=0.5	β_1	-2	29	29	0.99	95.4		-2	20	20	1.00	96.1	
		β_2	-3	17	17	0.98	94.9		-1	12	12	1.00	94.9	
		β_3	1	30	29	0.98	94.2		0	21	20	0.99	94.0	
	$\tilde{\rho}$=0.4	β_0	3	25	24	0.97	95.2	0.216	2	17	17	1.01	95.7	0.167
	r_0=0.9	β_1	-2	28	28	1.03	94.7		-1	20	20	1.04	96.0	
		β_2	-3	17	16	1.03	94.5		-1	12	11	1.04	94.2	
		β_3	1	30	29	0.98	94.8		0	20	20	0.98	95.2	
	$\tilde{\rho}$=0.8	β_0	2	25	25	0.97	95.6	0.148	1	17	17	1.00	95.4	0.100
	r_0=0.5	β_1	-2	28	29	0.99	96.1		-2	20	20	1.02	96.2	
		β_2	-3	17	16	0.98	95.0		-1	12	12	1.01	94.9	
		β_3	1	30	29	0.97	94.4		0	21	20	0.99	93.8	
	$\tilde{\rho}$=0.8	β_0	3	25	24	0.98	95.6	0.270	2	17	17	1.00	94.6	0.226
	r_0=0.9	β_1	-2	28	28	1.06	95.4		-1	19	20	1.07	95.0	
		β_2	-3	17	16	1.05	94.1		-1	12	11	1.05	94.2	
		β_3	1	29	29	0.97	94.2		0	20	20	0.98	94.4	

S1, S2, S3: scenarios 1, 2, 3. ESE: empirical standard error. ASE: estimated asymptotic standard error. ERE: empirical relative efficiency, the empirical variance of the maximum likelihood estimator using the main study data alone divided by the empirical variance of the proposed estimator. CP: coverage probability. All values except ERE and *IIB* are multiplied by 100.

Table S.3: Simulation results for Example 1 in the main manuscript with different basis matrices

Sample Size	Parameter	Bias	ESE	ASE	ERE	95%CP	IIB
n=300	β_0	2.1	23	21.8	1.14	95	0.966
	β_1	-1.5	25.4	24.8	1.25	95	
	β_2	-2.7	15.9	14.7	1.19	94	
	β_3	1.5	27.7	26.1	1.11	93	
n=600	β_0	1.3	15.9	15.4	1.19	94	0.9
	β_1	-0.8	17.7	17.6	1.26	95	
	β_2	-1.3	10.6	10.4	1.28	94	
	β_3	0.3	18.8	18.5	1.15	95	

These results are based on the setup where $r_0 = 0.9$ and $\rho = 0.4$. ESE: empirical standard error. ASE: estimated asymptotic standard error. ERE: empirical relative efficiency, the empirical variance of the maximum likelihood estimator using the main study data alone divided by the empirical variance of the proposed estimator. CP: coverage probability. All values except ERE and *IIB* are multiplied by 100.