CLOSED POPULATION CAPTURE–RECAPTURE MODELS WITH MEASUREMENT ERROR AND MISSING OBSERVATIONS IN COVARIATES

Jakub Stoklosa, Shen-Ming Lee and Wen-Han Hwang

The University of New South Wales, Feng Chia University,

and National Chung Hsing University

Supplementary Material

The following Web Appendix contains: proofs of Theorems 1–5 (Appendix S1), a refined regression calibration method (Appendix S2), an alternative MICS algorithm (Appendix C), and Web Figures 1–6 and Web Tables 1–8 consist of results for simulation studies 1 and 2 (Appendix S4).

S1 Proofs of Theorems 1–5

S1.1 Regularity Conditions for Measurement Error Models

(A1) For any population size N, the covariates $(X_i,\ Z_i)$, $i = 1, \ldots, N$, represent independent observations from a common underlying distribution (Section 4, Huggins, 1989).

(A2) $\mathrm{E}\left\{\Phi_i(\boldsymbol{\theta})\Phi_i^{\mathsf{T}}(\boldsymbol{\theta})\right\}$ is positive definite in a neighborhood for the true $\boldsymbol{\theta}$.

(A3) The first derivative of $U_c(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ exists almost surely in a neighborhood for the true $\boldsymbol{\theta}$. Further, in such a neighborhood, the first derivative is bounded above by a function of $(\mathcal{Y}, \boldsymbol{W}, Z)$, whose expectation exists.

S1.2 Proof of Theorem 1

Denote $\boldsymbol{a}^{\otimes 2} = \boldsymbol{a}\boldsymbol{a}^{\mathsf{T}}$ for a vector $\boldsymbol{a}$. Let $M_c(\boldsymbol{\theta}) = \mathrm{E}\left\{I(\mathcal{C}_i)\Phi_i(\boldsymbol{\theta})^{\otimes 2}\right\}$,

$$G_c(\boldsymbol{\theta}) = \mathrm{E}\left\{-\mathrm{I}(\mathcal{C}_i)\frac{\partial \Phi_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right\}$$

and

$$H_c(\boldsymbol{\theta}) = \mathrm{E}\left\{\frac{\partial}{\partial \boldsymbol{\theta}}\frac{\mathrm{I}(\mathcal{C}_i)}{\bar{P}^*_{i\Delta}(\boldsymbol{\theta})}\right\}$$

which is a row vector of the length of $\boldsymbol{\theta}$.

Theorem 1. *Under the regularity conditions A1–A3, $\hat{\boldsymbol{\theta}}_c$ is a consistent estimator as $N \to \infty$. Moreover, $\sqrt{N}(\hat{\boldsymbol{\theta}}_c - \boldsymbol{\theta})$ converges in distribution to the normal distribution $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_c)$ where $\boldsymbol{\Sigma}_c = G_c^{-1}(\boldsymbol{\theta})M_c(\boldsymbol{\theta})G_c^{-\mathsf{T}}(\boldsymbol{\theta})$.*

Proof: First, recall that

$$\Phi_i(\boldsymbol{\theta}) = \frac{1}{m_i}\sum_{j=1}^{m_i}\begin{pmatrix}\Delta_{ij}\\ Z_i\end{pmatrix}\left\{\mathcal{Y}_i - \mathrm{E}(\mathcal{Y}_i \mid \Delta_{ij}, Z_i, \mathcal{C}_i)\right\}, i = 1, 2, \ldots, D.$$

where $\Delta_{ij} = \mathcal{Y}_i\beta\sigma_u^2 + W_{ij},$ for $j = 1, 2, \ldots, m_i,\ i = 1, 2, \ldots, D.$

Now, $U_c(\boldsymbol{\theta}) = \sum_{i=1}^{N}\mathrm{I}(\mathcal{C}_i)\Phi_i(\boldsymbol{\theta})$ is an unbiased estimating function since $\mathrm{E}\left\{I(\mathcal{C}_i)\Phi_i(\boldsymbol{\theta})\right\} = \mathrm{E}[\mathrm{E}\{\mathrm{I}(\mathcal{C}_i)\Phi_i(\boldsymbol{\theta}) \mid \mathcal{C}_i\}] = 0$ for all i. Using regularity conditions (A1)–(A3) and the inverse function theorem as in Foutz (1977), we can obtain $\hat{\boldsymbol{\theta}}_c$, which is the unique solution for $U_c(\boldsymbol{\theta}) = \mathbf{0}$ in a neighborhood for the true $\boldsymbol{\theta}$, and $\hat{\boldsymbol{\theta}}_c$ is a consistent estimator for $\boldsymbol{\theta}$ as $N \longrightarrow \infty$.

To see the limiting distribution of $\hat{\boldsymbol{\theta}}_c$, we note that $\hat{\boldsymbol{\theta}}_c$ is the solution to $U_c(\boldsymbol{\theta}) = 0$. Consider the first-order Taylor expansion of $U_c(\hat{\boldsymbol{\theta}}_c)$ at $\boldsymbol{\theta}$, then we have

$$0 = \frac{1}{\sqrt{N}} U_c(\hat{\boldsymbol{\theta}}_c) = \frac{1}{\sqrt{N}} U_c(\boldsymbol{\theta}) + \left\{ \frac{\partial U_c(\boldsymbol{\theta})}{N \partial \boldsymbol{\theta}} \right\} \sqrt{N}(\hat{\boldsymbol{\theta}}_c - \boldsymbol{\theta}) + o_p(1).$$

Hence, we obtain that

$$\sqrt{N}\left(\hat{\boldsymbol{\theta}}_c - \boldsymbol{\theta}\right) = -\left\{ \frac{\partial U_c(\boldsymbol{\theta})}{N \partial \boldsymbol{\theta}} \right\}^{-1} \frac{1}{\sqrt{N}} \left\{ \sum_{i=1}^{N} \mathrm{I}(\mathcal{C}_i) \Phi_i(\boldsymbol{\theta}) \right\} + o_p(1).$$

Using the same arguments as in Section 4 of Huggins (1989), we regard $\mathrm{I}(\mathcal{C}_i)\Phi_i(\boldsymbol{\theta})$, for $i = 1, \ldots, N$ as *i.i.d.* random variables, thus the law of large numbers implies that $-\dfrac{\partial U_c(\boldsymbol{\theta})}{N \partial \boldsymbol{\theta}} \rightarrow G_c(\boldsymbol{\theta})$ in probability, and

$$\sqrt{N}\left(\hat{\boldsymbol{\theta}}_c - \boldsymbol{\theta}\right) = G_c^{-1}(\boldsymbol{\theta}) \frac{1}{\sqrt{N}} \left\{ \sum_{i=1}^{N} \mathrm{I}(\mathcal{C}_i) \Phi_i(\boldsymbol{\theta}) \right\} + o_p(1). \tag{S1.1}$$

It then follows (via the central limit theorem) that

$$\sqrt{N}\left(\hat{\boldsymbol{\theta}}_c - \boldsymbol{\theta}\right) \stackrel{d}{\longrightarrow} \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_c),$$

where $\boldsymbol{\Sigma}_c = G_c^{-1}(\boldsymbol{\theta}) M_c(\boldsymbol{\theta}) G_c^{-\mathsf{T}}(\boldsymbol{\theta})$.

S1.3 Proof of Theorem 2

Theorem 2. *Under regularity conditions A1–A3, $\hat{N}_c/N$ converges to one in probability as $N \to \infty$. Moreover, the limiting distribution of $N^{-1/2}(\hat{N}_c - N)$ is $\mathcal{N}(\mathbf{0}, \boldsymbol{\nu}_c)$ where $\boldsymbol{\nu}_c$ is the variance of* $\dfrac{\mathrm{I}(\mathcal{C}_i)}{\bar{P}^*_{i\Delta}(\boldsymbol{\theta})} + H_c(\boldsymbol{\theta})G_c^{-1}(\boldsymbol{\theta})\mathrm{I}(\mathcal{C}_i)\Phi_i(\boldsymbol{\theta})$.

Proof: We consider the first-order Taylor expansion of $\hat{N}_c(\hat{\boldsymbol{\theta}}_c)$ around $\boldsymbol{\theta}$, such that

$$\hat{N}_c(\hat{\boldsymbol{\theta}}_c) = \hat{N}_c(\boldsymbol{\theta}) + \frac{\partial \hat{N}_c(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}_c - \boldsymbol{\theta}) + O_p\left(N(\hat{\boldsymbol{\theta}}_c - \boldsymbol{\theta})^2\right).$$

By the law of large numbers we have $H_c(\boldsymbol{\theta}) = \mathrm{E}\left\{\dfrac{\partial}{\partial \boldsymbol{\theta}}\dfrac{\mathrm{I}(\mathcal{C}_i)}{\bar{P}^*_{i\Delta}(\boldsymbol{\theta})}\right\}$ which is the limit of $\dfrac{1}{N}\dfrac{\partial \hat{N}_c(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ as $N \to \infty$. Since $(\hat{\boldsymbol{\theta}}_c - \boldsymbol{\theta})^2$ is $O_p(1/N)$ and by (S1.1), we have

$$\hat{N}_c(\hat{\boldsymbol{\theta}}_c) = \sum_{i=1}^{N}\left\{\frac{\mathrm{I}(\mathcal{C}_i)}{\bar{P}^*_{i\Delta}(\boldsymbol{\theta})} + H_c(\boldsymbol{\theta})G_c^{-1}(\boldsymbol{\theta})\mathrm{I}(\mathcal{C}_i)\Phi_i(\boldsymbol{\theta})\right\} + O_p(1). \quad \text{(S1.2)}$$

Therefore, we have

$$\frac{\hat{N}_c - N}{N} = \frac{1}{N}\sum_{i=1}^{N}\left\{\frac{\mathrm{I}(\mathcal{C}_i)}{\bar{P}^*_{i\Delta}(\boldsymbol{\theta})} - 1\right\} + H_c(\boldsymbol{\theta})G_c^{-1}(\boldsymbol{\theta})\frac{1}{N}\sum_{i=1}^{N}\mathrm{I}(\mathcal{C}_i)\Phi_i(\boldsymbol{\theta}) + o_p(1).$$

Once again, the law of large numbers implies that both the first two terms on the right hand side of the above equation converges to 0 and thus $(\hat{N}_c - N)/N \longrightarrow 0$ in probability.

By (S1.2), it follows that

$$\frac{\hat{N}_c(\hat{\boldsymbol{\theta}}_c) - N}{\sqrt{N}} = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left\{\frac{\mathrm{I}(\mathcal{C}_i)}{\bar{P}^*_{i\Delta}(\boldsymbol{\theta})} + H_c(\boldsymbol{\theta})G_c^{-1}(\boldsymbol{\theta})\mathrm{I}(\mathcal{C}_i)\Phi_i(\boldsymbol{\theta}) - 1\right\} + o_p(1).$$

Again, the central limit theorem yields

$$\frac{1}{\sqrt{N}}\left\{\hat{N}_c(\hat{\boldsymbol{\theta}}_c) - N\right\} \xrightarrow{d} \mathcal{N}\left(0, \boldsymbol{\nu}_c\right),$$

where $\boldsymbol{\nu}_c$ is the variance of $\dfrac{\mathrm{I}(\mathcal{C}_i)}{\bar{P}^*_{i\Delta}(\boldsymbol{\theta})} + H_c(\boldsymbol{\theta})G_c^{-1}(\boldsymbol{\theta})\mathrm{I}(\mathcal{C}_i)\Phi_i(\boldsymbol{\theta})$. Note that, the covariance of $\mathrm{I}(\mathcal{C}_i)$ and $\mathrm{I}(\mathcal{C}_i)\Phi_i(\boldsymbol{\theta})$ is 0.

S1.4 Additional Regularity Conditions for Measurement Error and Missing Data Models

(B1) Let *supp*(Z) denote the support of Z. For any $y = 1, 2, \ldots, \tau$ and $z \in$ *supp*(Z), the selection probability $\pi(y, z)$ is bounded away from zero.

(B2) $\left\{\dfrac{\Phi_i(\boldsymbol{\theta})\Phi_i^{\mathsf{T}}(\boldsymbol{\theta})}{\pi(\mathcal{Y}, Z)}\right\}$ is positive definite in a neighbourhood for the true $\boldsymbol{\theta}$.

(B3) The first derivative of $U_{wc}(\boldsymbol{\theta}, \boldsymbol{\pi})$ with respect to $\boldsymbol{\theta}$ exists almost surely in a neighbourhood for the true $\boldsymbol{\theta}$. Further, in such a neighbourhood, the first derivative is bounded above by a function of $(\mathcal{Y}, \boldsymbol{W}, Z)$, whose expectation exists.

S1.5 Proof of Theorem 3.

Denote $\Phi_i^*(\boldsymbol{\theta}) = \{\Phi_i(\boldsymbol{\theta}) \mid \mathcal{Y}_i, Z_i\}$, $g_i^*(\boldsymbol{\theta}) = \dfrac{\delta_i}{\pi_i}\Phi_i(\boldsymbol{\theta}) - \dfrac{\delta_i - \pi_i}{\pi_i}\Phi_i^*(\boldsymbol{\theta})$, and $M_{wc}(\boldsymbol{\theta}, \boldsymbol{\pi}) = \mathrm{E}\{\mathrm{I}(\mathcal{C}_i)g_i^*(\boldsymbol{\theta})^{\bigotimes 2}\}$.

Theorem 3. *Under regularity conditions A1–A2 and B1–B3,* $\hat{\boldsymbol{\theta}}_{wc}$ *is a con-*

sistent estimator as $N \to \infty$*. Moreover,* $\sqrt{N}(\hat{\boldsymbol{\theta}}_{wc} - \boldsymbol{\theta})$ *converges in distribution to the normal distribution* $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{wc})$ *where* $\boldsymbol{\Sigma}_{wc} = G_c^{-1}(\boldsymbol{\theta}) M_{wc}(\boldsymbol{\theta}, \boldsymbol{\pi}) G_c^{-\mathsf{T}}(\boldsymbol{\theta})$*.*

Proof: The IPWCS estimator $\hat{\boldsymbol{\theta}}_{wc}$ solves the following estimating equation:

$$U_{wc}(\boldsymbol{\theta}, \hat{\boldsymbol{\pi}}) = \sum_{i=1}^{D} \frac{\delta_i}{\hat{\pi}_i} \Phi_i(\boldsymbol{\theta}) = \mathbf{0},$$

where $\hat{\pi}_i = \hat{\pi}(\mathcal{Y}_i, Z_i)$ and $\hat{\boldsymbol{\pi}}$ is the vector of $\hat{\pi}_i$ for $i = 1, \ldots, D$. Using a Taylor expansion on $U_{wc}(\boldsymbol{\theta}, \hat{\boldsymbol{\pi}})$ around $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_D)$, we have

$$\begin{aligned} & \frac{1}{\sqrt{N}} U_{wc}(\boldsymbol{\theta}, \hat{\boldsymbol{\pi}}) \\ = \; & \frac{1}{\sqrt{N}} U_{wc}(\boldsymbol{\theta}, \boldsymbol{\pi}) - \frac{1}{\sqrt{N}} \sum_{i=1}^{D} \left[\frac{\hat{\pi}_i - \pi_i}{\pi_i^2} + O_p\left\{ (\hat{\pi}_i - \pi_i)^2 \right\} \right] \delta_i \Phi_i(\boldsymbol{\theta}) \\ = \; & \frac{1}{\sqrt{N}} U_{wc}(\boldsymbol{\theta}, \boldsymbol{\pi}) - \frac{1}{N^{3/2}} \sum_{i=1}^{D} \sum_{s=1}^{D} \left\{ \frac{(\delta_s - \pi_i) \mathrm{I}(\mathcal{Y}_s = \mathcal{Y}_i, Z_s = Z_i)}{P(\mathcal{Y} = \mathcal{Y}_i, Z = Z_i) \pi_i^2} \right\} \delta_i \Phi_i(\boldsymbol{\theta}) + O_p\left(\frac{1}{\sqrt{N}} \right). \end{aligned}$$

Using the equality $(\delta_s - \pi_i)\delta_i = (\delta_s - \pi_i)(\delta_i - \pi_i) + \pi(\delta_s - \pi_i)$, we have

$$\frac{1}{\sqrt{N}} U_{wc}(\boldsymbol{\theta}, \hat{\boldsymbol{\pi}}) = \frac{1}{\sqrt{N}} U_{wc}(\boldsymbol{\theta}, \boldsymbol{\pi}) - A_{1N} - A_{2N} + O_p\left(\frac{1}{\sqrt{N}} \right),$$

where

$$A_{1N} = \frac{1}{N^{3/2}} \sum_{i=1}^{D} \sum_{s=1}^{D} \left\{ \frac{(\delta_s - \pi_i)(\delta_i - \pi_i) \mathrm{I}(\mathcal{Y}_s = \mathcal{Y}_i, Z_s = Z_i)}{P(\mathcal{Y} = \mathcal{Y}_i, Z = Z_i) \pi_i^2} \right\} \Phi_i(\boldsymbol{\theta}),$$

and

$$A_{2N} = \frac{1}{N^{3/2}} \sum_{i=1}^{D} \sum_{s=1}^{D} \left\{ \frac{\pi_i (\delta_s - \pi_i) \mathrm{I}(\mathcal{Y}_s = \mathcal{Y}_i, Z_s = Z_i)}{\pi_i^2 P(\mathcal{Y} = \mathcal{Y}_i, Z = Z_i)} \right\} \Phi_i(\boldsymbol{\theta}).$$

Rearranging the summation and noting that $\pi_i = \pi_s$ whenever $\mathrm{I}(\mathcal{Y}_i = \mathcal{Y}_s, Z_i = Z_s) = 1$, we have

$$\begin{aligned} A_{2N} &= \frac{1}{\sqrt{N}} \sum_{s=1}^{D} \frac{(\delta_s - \pi_s)}{\pi_s} \left\{ \frac{1}{N} \sum_{i=1}^{N} \frac{\mathrm{I}(\mathcal{C}_i)\mathrm{I}(\mathcal{Y}_i = \mathcal{Y}_s, Z_i = Z_s)}{P(\mathcal{Y} = \mathcal{Y}_s, Z = Z_s)} \Phi_i(\boldsymbol{\theta}) \right\} + O_p\left(\frac{1}{\sqrt{N}}\right) \\ &= \frac{1}{\sqrt{N}} \sum_{s=1}^{D} \left\{ \frac{\delta_s - \pi_s}{\pi_s} \right\} \Phi_s^*(\boldsymbol{\theta}) + O_p\left(\frac{1}{\sqrt{N}}\right). \end{aligned}$$

Next, we define $f_{is} = \dfrac{(\delta_i - \pi_i)(\delta_s - \pi_s)\mathrm{I}(\mathcal{Y}_i = \mathcal{Y}_s, Z_i = Z_s)\Phi_i(\boldsymbol{\theta})}{\pi_i^2 P(\mathcal{Y} = \mathcal{Y}_i, Z = Z_i)}$. Therefore, A_{1N} can be expressed as the following:

$$A_{1N} = \frac{1}{N^{3/2}} \sum_{i=1}^{D} \sum_{s=1}^{D} f_{is}.$$

Note that

$$\mathrm{E}\left[\mathrm{E}\left\{f_{is} \mid \boldsymbol{W}_i, \mathcal{Y}_s = \mathcal{Y}_i, Z_s = Z_i\right\}\right] = \begin{cases} 0 & \text{if } i \neq s, \\ \mathrm{E}\left\{ \dfrac{\pi_i(1-\pi_i)\Phi_i(\boldsymbol{\theta}) I[\mathcal{Y}_i = \mathcal{Y}_s, Z_i = Z_s]}{\pi_i^2 P(\mathcal{Y} = \mathcal{Y}_i, Z = Z_i)} \right\} & \text{if } i = s. \end{cases}$$

Furthermore, we can simplify the expression of $\mathrm{E}(f_{ii})$ as

$$\mathrm{E}\left\{ \frac{\pi_i(1-\pi_i)\Phi_i(\boldsymbol{\theta}) I[\mathcal{Y}_i = \mathcal{Y}_s, Z_i = Z_s]}{\pi_i^2 P(\mathcal{Y} = \mathcal{Y}_i, Z = Z_i)} \right\} = \mathrm{E}\left\{ \frac{(1-\pi_i)\Phi_i^*(\boldsymbol{\theta})}{\pi_i} \right\}.$$

Hence we obtain

$$\mathrm{E}(A_{1N}) = \frac{1}{N^{1/2}} \mathrm{E}\left\{ \frac{\mathrm{I}(\mathcal{C}_i)(1-\pi_i)\Phi_i^*(\boldsymbol{\theta})}{\pi_i} \right\} = O\left(\frac{1}{\sqrt{N}}\right).$$

In addition, using the definition of f_{is}, we have

$$\mathrm{Cov}(f_{ik}, f_{sb}) = \mathrm{E}\left\{ \frac{(1-\pi_i)^2 \Phi_i(\boldsymbol{\theta}) \Phi_i^T(\boldsymbol{\theta})}{\pi_i^2 P(\mathcal{Y} = \mathcal{Y}_i, Z = Z_i)} \right\}$$

at $(i,k)=(s,b)$ and $\mathrm{Cov}(f_{ik}, f_{sb}) = 0$ otherwise. Hence, we obtain

$$\mathrm{Var}(A_{1N}) = \frac{1}{N^3}\sum_{i,k}^{N} \mathrm{Var}(f_{ik}) = O\left(\frac{1}{N}\right).$$

Based on $\mathrm{E}(A_{1N}) = O(1/\sqrt{N})$ and $\mathrm{Var}(A_{1N}) = O(1/N)$, we have $A_{1N} = O_p(1/\sqrt{N})$.

As a consequence, we obtain

$$\frac{1}{\sqrt{N}} U_{wc}(\boldsymbol{\theta}, \hat{\boldsymbol{\pi}}) \;\; = \;\; \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \mathrm{I}(\mathcal{C}_i) g_i^*(\boldsymbol{\theta}) + O_p\left(\frac{1}{\sqrt{N}}\right).$$

Since $\hat{\boldsymbol{\theta}}_{wc}$ is the solution to $U_{wc}(\boldsymbol{\theta}, \hat{\boldsymbol{\pi}}) = \mathbf{0}$, the Taylor expansion around $\boldsymbol{\theta}$ yields

$$\mathbf{0} = \frac{1}{\sqrt{N}} U_{wc}(\hat{\boldsymbol{\theta}}_{wc}, \hat{\boldsymbol{\pi}}) = \frac{1}{\sqrt{N}} U_{wc}(\boldsymbol{\theta}, \hat{\boldsymbol{\pi}}) + \left\{\frac{\partial U_{wc}(\boldsymbol{\theta}, \boldsymbol{\pi})}{N \partial \boldsymbol{\theta}}\right\} \sqrt{N}\left(\hat{\boldsymbol{\theta}}_{wc} - \boldsymbol{\theta}\right) + o_p(1).$$

Note that $G_c(\boldsymbol{\theta}) = \mathrm{E}\left\{-\mathrm{I}(\mathcal{C}_i)\frac{\partial \Phi_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right\}$ is the limit of $\frac{\partial U_{wc}(\boldsymbol{\theta}, \boldsymbol{\pi})}{N \partial \boldsymbol{\theta}}$. Therefore, we have

$$\sqrt{N}\left(\hat{\boldsymbol{\theta}}_{wc} - \boldsymbol{\theta}\right) = G_c^{-1}(\boldsymbol{\theta}) \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \mathrm{I}(\mathcal{C}_i) g_i^*(\boldsymbol{\theta}) + o_p(1).$$

As the estimating functions $\mathrm{I}(\mathcal{C}_i) g_i^*(\boldsymbol{\theta})$ are zero unbiased, the central limit theorem implies that

$$\sqrt{N}\left(\hat{\boldsymbol{\theta}}_{wc} - \boldsymbol{\theta}\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}_{wc}\right),$$

where $\boldsymbol{\Sigma}_{wc} = G_c^{-1}(\boldsymbol{\theta}) M_{wc}(\boldsymbol{\theta}, \boldsymbol{\pi}) G_c^{-\mathsf{T}}(\boldsymbol{\theta})$.

S1.6 Proof of Theorem 4.

Theorem 4. *Under regularity conditions A1–A2 and B1–B3, $\hat{N}_{wc}/N$ converges to one in probability as $N \to \infty$. Moreover, let $\kappa_j^*(\boldsymbol{\theta})$ be the expectation of*

$$\frac{\mathrm{I}(\mathcal{C}_j)}{\overline{P}^*_{j\Delta}(\boldsymbol{\theta})}$$

conditional on $(\mathcal{Y}_j, Z_j)$, such that the limiting distribution of $N^{-1/2}(\hat{N}_{wc} - N)$ is $\mathcal{N}(\mathbf{0}, \boldsymbol{\nu}_{wc})$ where $\boldsymbol{\nu}_{wc}$ is the variance of $\mathrm{I}(\mathcal{C}_i)\left\{\frac{\delta_i}{\pi_i}\frac{1}{\overline{P}^*_{i\Delta}(\boldsymbol{\theta})} + H_c(\boldsymbol{\theta})G_c^{-1}(\boldsymbol{\theta})g_i^*(\boldsymbol{\theta}) - \frac{\delta_i - \pi_i}{\pi_i}\kappa_i^*(\boldsymbol{\theta})\right\}$.

Proof: The IPWCS population size estimator is

$$\hat{N}_{wc}(\hat{\boldsymbol{\theta}}_{wc}, \hat{\boldsymbol{\pi}}) = \sum_{i=1}^{N} \frac{\delta_i \mathrm{I}(\mathcal{C}_i)}{\hat{\pi}_i \overline{P}^*_{i\Delta}(\hat{\boldsymbol{\theta}}_{wc})}.$$

Now, the first-order Taylor expansion of $\hat{N}_{wc}$ around $(\boldsymbol{\theta}, \boldsymbol{\pi})$ yields

$$\begin{aligned} \hat{N}_{wc}(\hat{\boldsymbol{\theta}}_{wc}, \hat{\boldsymbol{\pi}}) - \hat{N}_{wc}(\boldsymbol{\theta}, \boldsymbol{\pi}) &= \frac{\partial \hat{N}_{wc}(\boldsymbol{\theta}, \boldsymbol{\pi})}{\partial \boldsymbol{\theta}}\left(\hat{\boldsymbol{\theta}}_{wc} - \boldsymbol{\theta}\right) + O_p\left\{N\left(\hat{\boldsymbol{\theta}}_{wc} - \boldsymbol{\theta}\right)^2\right\} \\ &\quad - \sum_{i=1}^{N}\left\{\frac{\delta_i \mathrm{I}(\mathcal{C}_i)}{\pi_i^2 \overline{P}^*_{i\Delta}(\boldsymbol{\theta})}\right\}\left[(\hat{\pi}_i - \pi_i) + O_p\left\{(\hat{\pi}_i - \pi_i)^2\right\}\right] \end{aligned}$$

Since both $(\hat{\pi}_i - \pi_i)^2$ and $\left(\hat{\boldsymbol{\theta}}_{wc} - \boldsymbol{\theta}\right)^2$ are $O_p(1/N)$, thus

$$\hat{N}_{wc}\left(\hat{\boldsymbol{\theta}}_{wc}, \hat{\boldsymbol{\pi}}\right) - \hat{N}_{wc}(\boldsymbol{\theta}, \boldsymbol{\pi}) = \frac{\partial \hat{N}_{wc}(\boldsymbol{\theta}, \boldsymbol{\pi})}{\partial \boldsymbol{\theta}}\left(\hat{\boldsymbol{\theta}}_{wc} - \boldsymbol{\theta}\right) - \sum_{i=1}^{N}\left\{\frac{\delta_i \mathrm{I}(\mathcal{C}_i)}{\pi_i^2 \overline{P}^*_{i\Delta}(\boldsymbol{\theta})}\right\}(\hat{\pi}_i - \pi_i) + O_p(1).$$

Replacing $\hat{\pi}_i - \pi_i$ with $\frac{\sum_{j=1}^{N} \mathrm{I}(\mathcal{C}_j)(\delta_j - \pi_i)\mathrm{I}(\mathcal{Y}_j = \mathcal{Y}_i, Z_j = Z_i)}{\sum_{s=1}^{N} \mathrm{I}(\mathcal{C}_s)\mathrm{I}(\mathcal{Y}_s = \mathcal{Y}_i, Z_s = Z_i)}$ in the second term, we obtain

$$\sum_{i=1}^{N}\left\{\frac{\delta_i \mathrm{I}(\mathcal{C}_i)}{\pi_i^2 \bar{P}_{i\Delta}^*(\boldsymbol{\theta})}\right\}(\hat{\pi}_i-\pi_i)$$
$$= \sum_{i=1}^{N}\sum_{j=1}^{N}\left\{\frac{\mathrm{I}(\mathcal{C}_i)\delta_i(\delta_j-\pi_i)\mathrm{I}(\mathcal{C}_j)\mathrm{I}(\mathcal{Y}_j=\mathcal{Y}_i, Z_j=Z_i)}{\overline{P}_{i\Delta}^*(\boldsymbol{\theta})\pi_i^2}\right\}\left\{\frac{1}{\sum_{s=1}^{N}\mathrm{I}(\mathcal{C}_s)\mathrm{I}(\mathcal{Y}_s=\mathcal{Y}_i, Z_s=Z_i)}\right\}.$$

Next, we rearrange the summation and note that $\pi_i = \pi_j$ and $\overline{P}_{i\Delta}^*(\boldsymbol{\theta}) = \overline{P}_{j\Delta}^*(\boldsymbol{\theta})$ since $\mathrm{I}(\mathcal{Y}_i=\mathcal{Y}_j, Z_i=Z_j)=1$, then

$$\sum_{i=1}^{N}\left\{\frac{\delta_i \mathrm{I}(\mathcal{C}_i)}{\pi_i^2 \bar{P}_{i\Delta}^*(\boldsymbol{\theta})}\right\}(\hat{\pi}_i-\pi_i) = \sum_{j=1}^{N}\left\{\frac{\mathrm{I}(\mathcal{C}_j)(\delta_j-\pi_j)}{\pi_j^2}\right\}\sum_{i=1}^{N}\left\{\frac{\delta_i \mathrm{I}(\mathcal{C}_i)\mathrm{I}(\mathcal{Y}_i=\mathcal{Y}_j, Z_i=Z_j]}{\overline{P}_{i\Delta}^*(\boldsymbol{\theta})\sum_{s=1}^{N}\mathrm{I}(\mathcal{C}_s)\mathrm{I}(\mathcal{Y}_s=\mathcal{Y}_j, Z_s=Z_j)}\right\}.$$

Now, by the law of large numbers we have

$$\sum_{i=1}^{N}\frac{\delta_i \mathrm{I}(\mathcal{C}_i)\mathrm{I}(\mathcal{Y}_i=\mathcal{Y}_j, Z_i=Z_j)}{\overline{P}_{i\Delta}^*(\boldsymbol{\theta})\left\{\sum_{s=1}^{N}\mathrm{I}(\mathcal{C}_s)\mathrm{I}(\mathcal{Y}_s=\mathcal{Y}_j, Z_s=Z_j)\right\}} \xrightarrow{p} \pi_j \mathrm{E}\left(\frac{\mathrm{I}(\mathcal{C}_j)}{\overline{P}_{j\Delta}^*(\boldsymbol{\theta})} \mid \mathcal{Y}_j, Z_j\right) = \pi_j\kappa_j^*(\boldsymbol{\theta}).$$

A further calculation yields $\dfrac{\partial \hat{N}_{wc}(\boldsymbol{\theta},\boldsymbol{\pi})}{N\partial\boldsymbol{\theta}}$ which converges to $H_c(\boldsymbol{\theta})$ in probability. It follows that

$$\hat{N}_{wc}(\hat{\boldsymbol{\theta}}_{wc},\hat{\boldsymbol{\pi}}) - \hat{N}_{wc}(\boldsymbol{\theta},\boldsymbol{\pi}) - \sum_{i=1}^{N}\frac{(\delta_i-\pi_i)\mathrm{I}(\mathcal{C}_i)\kappa_i^*(\boldsymbol{\theta})}{\pi_i} + O_p(1).$$

As a consequence, $\hat{N}_{wc}\left(\hat{\boldsymbol{\theta}}_{wc},\hat{\boldsymbol{\pi}}\right)$ equates to

$$\sum_{i=1}^{N}\mathrm{I}(\mathcal{C}_i)\left\{\frac{\delta_i}{\pi_i}\frac{1}{\overline{P}_{i\Delta}^*(\boldsymbol{\theta})} + H_c(\boldsymbol{\theta})G_c^{-1}(\boldsymbol{\theta})g_i^*(\boldsymbol{\theta}) - \frac{\delta_i-\pi_i}{\pi_i}\kappa_i^*(\boldsymbol{\theta})\right\} + O_p(1).$$

By the weak law of large number, it is easy to obtain that $\hat{N}_{wc}(\hat{\boldsymbol{\theta}}_{wc},\hat{\boldsymbol{\pi}})/N \longrightarrow 1$ in probability as $N \longrightarrow \infty$. Once again, when applying the central limit theorem this gives

$$\frac{1}{\sqrt{N}}\left\{\hat{N}_{wc}(\hat{\boldsymbol{\theta}}_{wc},\hat{\boldsymbol{\pi}}) - N\right\} \xrightarrow{d} \mathcal{N}\left(0,\boldsymbol{\nu}_{wc}\right),$$

where $\boldsymbol{\nu}_{wc}$ is the variance of $\mathrm{I}(\mathcal{C}_i)\left\{\dfrac{\delta_i}{\pi_i}\dfrac{1}{\overline{P}^*_{i\Delta}(\boldsymbol{\theta})}+H_c(\boldsymbol{\theta})G_c^{-1}(\boldsymbol{\theta})g_i^*(\boldsymbol{\theta})-\dfrac{\delta_i-\pi_i}{\pi_i}\kappa_i^*(\boldsymbol{\theta})\right\}$.

S1.7 Proof of Theorem 5

Theorem 5. *Under regularity conditions A1–A2 and B1–B3, we have* $\sqrt{N}(\hat{\boldsymbol{\theta}}_{wc}-\hat{\boldsymbol{\theta}}_{mc})$ *which converges to* $\mathbf{0}$ *in probability as both* N *and* M *are increased without bound. Similarly,* $N^{-1/2}(\hat{N}_{wc}-\hat{N}_{mc})$ *converges to 0 in probability as* $N, M \longrightarrow \infty$.

Proof: First, we show that $\sqrt{N}(\hat{\boldsymbol{\theta}}_{wc}-\hat{\boldsymbol{\theta}}_{mc})$ converges to 0 in probability. Let $\widetilde{m}_i = \mathrm{E}(m_i \mid \mathcal{Y}_i, Z_i)$ and define

$$\hat{m}_i = \sum_{s=1}^{D}\frac{\delta_s m_s \mathrm{I}(\mathcal{Y}_s=\mathcal{Y}_i, Z_s=Z_i)}{\sum_{r=1}^{D}\delta_r \mathrm{I}(\mathcal{Y}_r=\mathcal{Y}_i, Z_r=Z_i)}.$$

For completeness, we recall the following notation used in the main text. We denote $\Phi_i(\boldsymbol{\theta}) = (1/m_i)\sum_{j=1}^{m_i}\phi_{ij}(\boldsymbol{\theta})$, where $\phi_{ij}(\boldsymbol{\theta}) = (\Delta_{ij}, Z_i^{\mathsf{T}})^{\mathsf{T}}\left\{\mathcal{Y}_i - \mathrm{E}(\mathcal{Y}_i \mid \Delta_{ij}, Z_i, \mathcal{C}_i)\right\}$. For the generated imputation data, $m_{i,v}$ and $W^{\dagger}_{ij,v}$ with $j=1,2,\ldots,m_{i,v}$, we denote $\phi^{\dagger}_{ij,v}(\boldsymbol{\theta}) = (\Delta^{\dagger}_{ij,v}, Z_i^{\mathsf{T}})^{\mathsf{T}}\{\mathcal{Y}_i - \mathrm{E}(\mathcal{Y}_i \mid \Delta^{\dagger}_{ij,v}, Z_i, \mathcal{C}_i)\}$, where $\Delta^{\dagger}_{ij,v} = \beta\sigma_u^2\mathcal{Y}_i + W^{\dagger}_{ij,v}$.

Finally, we denote $\tilde{\Phi}^{\dagger}_{iv}(\boldsymbol{\theta}) = \dfrac{1}{m_{i,v}}\displaystyle\sum_{j=1}^{m_{i,v}}\phi^{\dagger}_{ij,v}(\boldsymbol{\theta})$ and $\tilde{\Phi}^{\dagger}_{i}(\boldsymbol{\theta}) = \dfrac{1}{M}\displaystyle\sum_{v=1}^{M}\tilde{\Phi}^{\dagger}_{iv}(\boldsymbol{\theta})$.

Next, note that

$$\mathrm{E}_{\widehat{F}}\left\{\phi^{\dagger}_{ij,v}(\boldsymbol{\theta}) \mid \mathcal{Y}_i, Z_i\right\} = \int \phi_{i1}(\boldsymbol{\theta})\mathrm{d}\widehat{F}(w \mid \mathcal{Y}_i, Z_i) = \sum_{s=1}^{D}\frac{\delta_s\sum_{j=1}^{m_s}\phi_{sj}(\boldsymbol{\theta})\mathrm{I}(\mathcal{Y}_s=\mathcal{Y}_i, Z_s=Z_i)}{\sum_{r=1}^{D}m_r\delta_r\mathrm{I}(\mathcal{Y}_r=\mathcal{Y}_i, Z_r=Z_i)}.$$

In addition to the above, we have

$$\begin{aligned}
\mathrm{E}_{\hat{F}}\left\{\tilde{\Phi}_i^{\dagger}(\boldsymbol{\theta}) \mid \mathcal{Y}_i, Z_i\right\} &= \mathrm{E}_{\hat{F}}\left\{\phi_{ij,v}^{\dagger}(\boldsymbol{\theta}) \mid \mathcal{Y}_i, Z_i\right\} \\
&= \sum_{s=1}^{D} \delta_s \left(\frac{m_s}{\hat{m}_i}\right)\left(\frac{1}{\hat{\pi}_i}\right) \Phi_s(\boldsymbol{\theta}) \frac{\mathrm{I}(\mathcal{Y}_s = \mathcal{Y}_i, Z_s = Z_i)}{\sum_{r=1}^{D} \mathrm{I}(\mathcal{Y}_r = \mathcal{Y}_i, Z_r = Z_i)}.
\end{aligned}$$

Thus, we express $U_{mc}(\boldsymbol{\theta})$ as follows:

$$\begin{aligned}
U_{mc}(\boldsymbol{\theta}) &= \sum_{i=1}^{D} \frac{\delta_i}{\hat{\pi}_i} \Phi_i(\boldsymbol{\theta}) + \sum_{i=1}^{D} \left(1 - \frac{1}{\hat{\pi}_i}\right) \delta_i \Phi_i(\boldsymbol{\theta}) + \sum_{i=1}^{D} (1-\delta_i) \tilde{\Phi}_i^{\dagger}(\boldsymbol{\theta}) \\
&= \sum_{i=1}^{D} \frac{\delta_i}{\hat{\pi}_i} \Phi_i(\boldsymbol{\theta}) - \sum_{s=1}^{D} \frac{(1-\delta_s)}{\hat{\pi}_s} \sum_{i=1}^{D} \frac{\delta_i \Phi_i(\boldsymbol{\theta}) \mathrm{I}(\mathcal{Y}_i = \mathcal{Y}_s, Z_i = Z_s)}{\sum_{r=1}^{D} \mathrm{I}(\mathcal{Y}_r = \mathcal{Y}_s, Z_r = Z_s)} + \sum_{i=1}^{D} (1-\delta_i) \tilde{\Phi}_i^{\dagger}(\boldsymbol{\theta}) \\
&= U_{wc}(\boldsymbol{\theta}, \hat{\boldsymbol{\pi}}) + \sum_{i=1}^{D} (1-\delta_i) \left\{\tilde{\Phi}_i^{\dagger}(\boldsymbol{\theta}) - \mathrm{E}_{\hat{F}}\left(\tilde{\Phi}_i^{\dagger}(\boldsymbol{\theta}) \mid \mathcal{Y}_i, Z_i\right)\right\} \\
&\quad + \sum_{i=1}^{D} \frac{(1-\delta_i)}{\hat{\pi}_i} \sum_{s=1}^{D} \delta_s \left(\frac{m_s}{\hat{m}_i} - 1\right) \Phi_s(\boldsymbol{\theta}) \frac{\mathrm{I}(\mathcal{Y}_s = \mathcal{Y}_i, Z_s = Z_i)}{\sum_{r=1}^{D} \mathrm{I}(\mathcal{Y}_r = \mathcal{Y}_i, Z_r = Z_i)}.
\end{aligned}$$

Since $\hat{\pi}_i = \sum_{s=1}^{D} \frac{\delta_s \mathrm{I}(\mathcal{Y}_s = \mathcal{Y}_i, Z_s = Z_i)}{\sum_{r=1}^{D} \mathrm{I}(\mathcal{Y}_r = \mathcal{Y}_i, Z_r = Z_i)}$, this gives

$$\begin{aligned}
\hat{\pi}_i - \pi_i &= \sum_{s=1}^{N} I[\mathcal{C}_s] \frac{(\delta_s - \pi_i) \mathrm{I}(\mathcal{Y}_s = \mathcal{Y}_i, Z_s = Z_i)}{\sum_{r=1}^{N} I[\mathcal{C}_r] \mathrm{I}(\mathcal{Y}_r = \mathcal{Y}_i, Z_r = Z_i)} \\
&= N^{-1} \sum_{s=1}^{D} \frac{(\delta_s - \pi_i) \mathrm{I}(\mathcal{Y}_s = \mathcal{Y}_i, Z_s = Z_i)}{\boldsymbol{P}(\mathcal{Y} = \mathcal{Y}_i, Z = Z_i)} + o_p\left(N^{-1/2}\right).
\end{aligned}$$

For $\hat{m}_i$, we also have

$$\begin{aligned}
\hat{m}_i - \widetilde{m}_i &= \sum_{s=1}^{N} \mathrm{I}(\mathcal{C}_s) \frac{\delta_s (m_s - \widetilde{m}_i) \mathrm{I}(\mathcal{Y}_s = \mathcal{Y}_i, Z_s = Z_i)}{\sum_{r=1}^{N} \delta_r \mathrm{I}(\mathcal{C}_r) \mathrm{I}(\mathcal{Y}_r = \mathcal{Y}_i, Z_r = Z_i)} \\
&= N^{-1} \sum_{s=1}^{D} \frac{\delta_s (m_s - \widetilde{m}_i) \mathrm{I}(\mathcal{Y}_s = \mathcal{Y}_i, Z_s = Z_i)}{\pi_i \boldsymbol{P}(\mathcal{Y} = \mathcal{Y}_i, Z = Z_i)} + o_p\left(N^{-1/2}\right).
\end{aligned}$$

Using the same algebra as above, we have

$$\begin{aligned}
& \sum_{i=1}^{D} \frac{(1-\delta_i)}{\hat{\pi}_i} \sum_{s=1}^{D} \delta_s \left(\frac{m_s}{\hat{m}_i} - 1 \right) \Phi_s(\boldsymbol{\theta}) \frac{\mathrm{I}(\mathcal{Y}_s = \mathcal{Y}_i, Z_s = Z_i)}{\sum_{r=1}^{D} \mathrm{I}(\mathcal{Y}_r = \mathcal{Y}_i, Z_r = Z_i)} \\
= \;& \sum_{i=1}^{D} \frac{(1-\delta_i)}{\hat{\pi}_i} \frac{1}{N} \sum_{s=1}^{D} \delta_s \left(\frac{m_s}{\hat{m}_i} - 1 \right) \Phi_s(\boldsymbol{\theta}) \frac{\mathrm{I}(\mathcal{Y}_s = \mathcal{Y}_i, Z_s = Z_i)}{\boldsymbol{P}(\mathcal{Y} = \mathcal{Y}_i, Z = Z_i) + O_p(N^{-1/2})}.
\end{aligned}$$

As a result, we write

$$\begin{aligned}
& \frac{1}{\sqrt{N}} \left\{ U_{mc}(\boldsymbol{\theta}) - U_{wc}(\boldsymbol{\theta}, \hat{\boldsymbol{\pi}}) \right\} \\
= \;& M^{-1/2} \frac{1}{\sqrt{N}} \sum_{i=1}^{D} (1-\delta_i) \sqrt{M} \left[\tilde{\Phi}_i^{\dagger}(\boldsymbol{\theta}) - \mathrm{E}_{\hat{F}} \left\{ \tilde{\Phi}_i^{\dagger}(\boldsymbol{\theta}) \mid \mathcal{Y}_i, Z_i \right\} \right] \\
& + \frac{1}{\sqrt{N}} \sum_{i=1}^{D} (1-\delta_i) \left\{ \frac{1}{N} \sum_{s=1}^{D} \frac{\delta_s}{\hat{\pi}_i} \left(\frac{m_s}{\hat{m}_i} - 1 \right) \Phi_s(\boldsymbol{\theta}) \frac{\mathrm{I}(\mathcal{Y}_s = \mathcal{Y}_i, Z_s = Z_i)}{\boldsymbol{P}(\mathcal{Y} = \mathcal{Y}_i, Z = Z_i))} \right\} + o_p(1).
\end{aligned}$$

It follows that $\sqrt{M} \left[\tilde{\Phi}_i^{\dagger}(\boldsymbol{\theta}) - \mathrm{E}_{\hat{F}} \left\{ \tilde{\Phi}_i^{\dagger}(\boldsymbol{\theta}) \mid \mathcal{Y}_i, Z_i \right\} \right]$ converges in distribution to a normal random vector as $M \longrightarrow \infty$. This implies that

$$M^{-1/2} \frac{1}{\sqrt{N}} \sum_{i=1}^{D} (1-\delta_i) \sqrt{M} \left[\tilde{\Phi}_i^{\dagger}(\boldsymbol{\theta}) - \mathrm{E}_{\hat{F}} \left\{ \tilde{\Phi}_i^{\dagger}(\boldsymbol{\theta}) \mid \mathcal{Y}_i, Z_i \right\} \right] = O_p(M^{-1/2}).$$

Next, we show that

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{D} (1-\delta_i) \left\{ \frac{1}{N} \sum_{s=1}^{D} \frac{\delta_s}{\hat{\pi}_i} \left(\frac{m_s}{\hat{m}_i} - 1 \right) \Phi_s(\boldsymbol{\theta}) \frac{\mathrm{I}(\mathcal{Y}_s = \mathcal{Y}_i, Z_s = Z_i)}{\boldsymbol{P}(\mathcal{Y} = \mathcal{Y}_i, Z = Z_i))} \right\} = o_p(1).$$

Using

$$\begin{aligned}
\left(\frac{m_s}{\hat{m}_i} - \frac{m_s}{\tilde{m}_i} \right) &= -\frac{m_s(\hat{m}_i - \tilde{m}_i)}{\tilde{m}_i^2} + O_p\left((\hat{m}_i - \tilde{m}_i)^2 \right), \\
\left(\frac{\delta_s}{\pi_i} - \frac{\delta_s}{\hat{\pi}_i} \right) &= -\frac{\delta_s(\hat{\pi}_i - \pi_i)}{\pi_i^2} + O_p\left((\hat{\pi}_i - \pi_i)^2 \right),
\end{aligned}$$

and the equality

$$\begin{aligned}&\frac{\delta_s}{\hat{\pi}_i}\left(\frac{m_s}{\hat{m}_i}-1\right)\\ =\;& \frac{\delta_s}{\pi_i}\left(\frac{m_s}{\tilde{m}_i}-1\right)+\frac{\delta_s}{\pi_i}\left(\frac{m_s}{\hat{m}_i}-\frac{m_s}{\tilde{m}_i}\right)+\left(\frac{m_s}{\tilde{m}_i}-1\right)\left(\frac{\delta_s}{\pi_i}-\frac{\delta_s}{\hat{\pi}_i}\right)+\left(\frac{\delta_s}{\pi_i}-\frac{\delta_s}{\hat{\pi}_i}\right)\left(\frac{m_s}{\hat{m}_i}-\frac{m_s}{\tilde{m}_i}\right),\end{aligned}$$

we obtain

$$\begin{aligned}&\frac{1}{N}\sum_{s=1}^{D}\frac{\delta_s}{\hat{\pi}_i}\left(\frac{m_s}{\hat{m}_i}-1\right)\Phi_s(\boldsymbol{\theta})\frac{\mathrm{I}(\mathcal{Y}_s=\mathcal{Y}_i,Z_s=Z_i)}{\boldsymbol{P}(\mathcal{Y}=\mathcal{Y}_i,Z=Z_i))}\\ =\;&\frac{1}{N}\sum_{s=1}^{D}\frac{\delta_s}{\pi_i}\left(\frac{m_s}{\tilde{m}_i}-1\right)\Phi_s(\boldsymbol{\theta})\frac{\mathrm{I}(\mathcal{Y}_s=\mathcal{Y}_i,Z_s=Z_i)}{\boldsymbol{P}(\mathcal{Y}=\mathcal{Y}_i,Z=Z_i))}+O_p(N^{-1/2}).\end{aligned}$$

Define $\xi_{si}(\boldsymbol{\theta}) = \frac{\delta_s}{\pi_i}\left(\frac{m_s}{\tilde{m}_i}-1\right)\Phi_s(\boldsymbol{\theta})\frac{\mathrm{I}(\mathcal{Y}_s=\mathcal{Y}_i,Z_s=Z_i)}{\boldsymbol{P}(\mathcal{Y}=\mathcal{Y}_i,Z=Z_i))}$. Taking the expectation of $\xi_{si}(\boldsymbol{\theta})$ yields

$$\begin{aligned}\mathrm{E}\{\xi_{si}(\boldsymbol{\theta})\} &= \mathrm{E}\left[\mathrm{E}\{\xi_{si}(\boldsymbol{\theta}) \mid \mathcal{Y}_s=\mathcal{Y}_i,Z_s=Z_i,m_s,\boldsymbol{W}_i\}\right]\\ &= \mathrm{E}\left[\mathrm{E}\left\{\frac{\delta_s}{\pi_i}\left(\frac{m_s}{\tilde{m}_i}-1\right)\Phi_s(\boldsymbol{\theta})\frac{\mathrm{I}(\mathcal{Y}_s=\mathcal{Y}_i,Z_s=Z_i)}{\boldsymbol{P}(\mathcal{Y}=\mathcal{Y}_i,Z=Z_i)}\middle|\mathcal{Y}_s=\mathcal{Y}_i,Z_s=Z_i,m_s,\boldsymbol{W}_i\right\}\right]\\ &= \mathrm{E}\left[\phi^*_{i1}(\boldsymbol{\theta})\mathrm{E}\left\{\left(\frac{m_s}{\tilde{m}_i}-1\right)\frac{\mathrm{I}(\mathcal{Y}_s=\mathcal{Y}_i,Z_s=Z_i)}{\boldsymbol{P}(\mathcal{Y}=\mathcal{Y}_i,Z=Z_i)}\middle|\mathcal{Y}_s=\mathcal{Y}_i,Z_s=Z_i\right\}\right]\\ &= 0,\end{aligned}$$

where $\phi^*_{i1}(\boldsymbol{\theta})=\mathrm{E}\left\{\frac{1}{m_s}\sum_{j=1}^{m_s}\phi_{sj}(\boldsymbol{\theta})\middle|\mathcal{Y}_s=\mathcal{Y}_i,Z_s=Z_i,m_s\right\}=\mathrm{E}\{\phi_{s1}(\boldsymbol{\theta})\mid\mathcal{Y}_s=\mathcal{Y}_i,Z_s=Z_i\}$.

Next, we show that

$$\begin{aligned}&\frac{1}{\sqrt{N}}\sum_{i=1}^{D}(1-\delta_i)\left\{\frac{1}{N}\sum_{s=1}^{D}\frac{\delta_s}{\pi_i}\left(\frac{m_s}{\tilde{m}_i}-1\right)\Phi_s(\boldsymbol{\theta})\frac{\mathrm{I}(\mathcal{Y}_s=\mathcal{Y}_i,Z_s=Z_i)}{\boldsymbol{P}(\mathcal{Y}=\mathcal{Y}_i,Z=Z_i)}\right\}\\ =\;&\frac{1}{N}\sum_{i=1}^{D}(1-\delta_i)\left\{\frac{1}{\sqrt{N}}\sum_{s=1}^{D}\xi_{si}(\boldsymbol{\theta})\right\}=o_p(1).\end{aligned}$$

Let $\xi_{sik}(\boldsymbol{\theta})$ be the kth element of $\xi_{si}(\boldsymbol{\theta})$. Then, by Cauchy–Schwarz inequality, we have

$$\mathrm{E}\left|(1-\delta_i)\left\{\frac{1}{\sqrt{N}}\sum_{s=1}^{D}\xi_{sik}(\boldsymbol{\theta})\right\}\right| \leq \left\{\mathrm{E}\{(1-\delta_i)^2\}\mathrm{E}\left[\frac{1}{N}\left\{\sum_{s=1}^{D}\xi_{sik}(\boldsymbol{\theta})\right\}^2\right]\right\}^{1/2},$$

and

$$\begin{aligned}\mathrm{E}\left[\frac{1}{N}\left\{\sum_{s=1}^{D}\xi_{sik}(\boldsymbol{\theta})\right\}^2\right] &= \frac{1}{N}\mathrm{E}\left\{\sum_{s=1}^{D}\xi_{sik}^2(\boldsymbol{\theta})+\sum_{s=1}^{D}\sum_{r\neq s,1}^{D}\xi_{sik}(\boldsymbol{\theta})\xi_{rik}(\boldsymbol{\theta})\right\}\\ &= \frac{1}{N}\mathrm{E}\left\{\sum_{s=1}^{D}\xi_{sik}^2(\boldsymbol{\theta})\right\}<\infty.\end{aligned}$$

Thus, we have $\mathrm{E}\left|(1-\delta_i)\left\{\frac{1}{\sqrt{N}}\sum_{s=1}^{D}\xi_{sik}(\boldsymbol{\theta})\right\}\right|<\infty$. Consequently, by the law of large numbers we have

$$\frac{1}{N}\sum_{i=1}^{D}(1-\delta_i)\left\{\frac{1}{\sqrt{N}}\sum_{s=1}^{D}\xi_{si}(\boldsymbol{\theta})\right\}=o_p(1),$$

and so

$$\frac{1}{\sqrt{N}}\left\{U_{mc}(\boldsymbol{\theta})-U_{wc}(\boldsymbol{\theta},\hat{\boldsymbol{\pi}})\right\}=O_p(M^{-1/2})+o_p(1).$$

Let $\mathcal{O}$ denote the observed data. In the same manner as above, it can be shown that

$$\mathrm{E}\left\{\frac{\partial U_{mc}(\boldsymbol{\theta})}{N\partial\boldsymbol{\theta}}\Big|\mathcal{O}\right\}=\frac{\partial U_{wc}(\boldsymbol{\theta},\hat{\boldsymbol{\pi}})}{N\partial\boldsymbol{\theta}}+o_p(N^{-1/2})+O_p(M^{-1/2}N^{-1/2}),$$

and

$$\mathrm{E}\left\{\frac{\partial U_{mc}(\boldsymbol{\theta})}{N\partial\boldsymbol{\theta}}\right\}=\mathrm{E}\left\{\frac{\partial U_{wc}(\boldsymbol{\theta},\hat{\boldsymbol{\pi}})}{N\partial\boldsymbol{\theta}}\right\}+o(N^{-1/2})+O(M^{-1/2}N^{-1/2}).$$

Since $G_c(\boldsymbol{\theta})$ is the limit of

$$\mathrm{E}\left\{-\frac{\partial U_{wc}(\boldsymbol{\theta},\hat{\boldsymbol{\pi}})}{N\partial\boldsymbol{\theta}}\right\},$$

and let $\hat{\boldsymbol{\theta}}_{mc}^{(M)}$ be the solution to $U_{mc}(\boldsymbol{\theta})=\mathbf{0}$ where the index M is appended for by noting that $\hat{\boldsymbol{\theta}}_{mc}$ is dependent on M. We now write

$$\begin{aligned} N^{-1/2}U_{mc}^{(M)}(\hat{\boldsymbol{\theta}}_{mc}^{(M)}) &= N^{-1/2}U_{mc}^{(M)}(\boldsymbol{\theta})+\left\{\frac{\partial U_{mc}^{(M)}(\boldsymbol{\theta})}{N\partial\boldsymbol{\theta}}\right\}\sqrt{N}\left\{\hat{\boldsymbol{\theta}}_{mc}^{(M)}-\boldsymbol{\theta}\right\} \\ &= N^{-1/2}U_{mc}^{(M)}(\boldsymbol{\theta})-G_c(\boldsymbol{\theta})\sqrt{N}\left\{\hat{\boldsymbol{\theta}}_{mc}^{(M)}-\boldsymbol{\theta}\right\}+o_p(1), \end{aligned}$$

and

$$N^{-1/2}U_{wc}(\hat{\boldsymbol{\theta}}_{wc},\hat{\boldsymbol{\pi}})=N^{-1/2}U_{wc}(\boldsymbol{\theta},\hat{\boldsymbol{\pi}})-G_c(\boldsymbol{\theta})\sqrt{N}(\hat{\boldsymbol{\theta}}_{wc}-\boldsymbol{\theta})+o_p(1).$$

Therefore, we have

$$\begin{aligned} \sqrt{N}\left(\hat{\boldsymbol{\theta}}_{mc}^{(M)}-\hat{\boldsymbol{\theta}}_{wc}\right) &= G_c^{-1}(\boldsymbol{\theta})N^{-1/2}\left\{U_{mc}^{(M)}(\boldsymbol{\theta})-U_{wc}(\boldsymbol{\theta},\hat{\boldsymbol{\pi}})\right\}+o_p(1) \\ &= O_p(M^{-1/2})+o_p(1). \end{aligned}$$

Let $\hat{\boldsymbol{\theta}}_{mc}^{(\infty)}$ be the limit of $\hat{\boldsymbol{\theta}}_{mc}^{(M)}$ as $M\longrightarrow\infty$, hence we have shown that $\sqrt{N}\left(\hat{\boldsymbol{\theta}}_{mc}^{(\infty)}-\hat{\boldsymbol{\theta}}_{wc}\right)$ converges in probability to $\mathbf{0}$ as $N\longrightarrow\infty$.

Next, we show that $N^{-1/2}(\hat{N}_{wc}-\hat{N}_{mc})$ converges in probability to 0. Consider the multiple imputation estimator for the population size, $\hat{N}_{mc}$, where

$$\hat{N}_{mc}(\boldsymbol{\theta}) = \sum_{i=1}^{D}\left\{\delta_i\frac{1}{\overline{P}_{i\Delta}^{*}(\boldsymbol{\theta})}+(1-\delta_i)\frac{1}{M}\sum_{v=1}^{M}\frac{1}{\widetilde{P}_{vi}^{\dagger}(\boldsymbol{\theta})}\right\}$$

Applying the same algebra used in the derivation of $\mathrm{E}_{\hat{F}}(\tilde{\Phi}^{\dagger}_{iv}(\boldsymbol{\theta}) \mid \mathcal{Y}_i, Z_i)$, we obtain

$$
\begin{aligned}
& N^{-1/2}\left\{\hat{N}_{mc}(\boldsymbol{\theta}) - \hat{N}_{wc}(\boldsymbol{\theta}, \hat{\boldsymbol{\pi}})\right\} \\
= \; & M^{-1/2}\frac{1}{\sqrt{N}}\sum_{i=1}^{D}(1-\delta_i)\sqrt{M}\left[\frac{1}{\widetilde{P}^{\dagger}_{i\Delta}(\boldsymbol{\theta})} - \mathrm{E}_{\hat{F}}\left\{\frac{1}{\widetilde{P}^{\dagger}_{1\Delta}(\boldsymbol{\theta})} \mid \mathcal{Y}_i, Z_i\right\}\right] \\
& + \frac{1}{\sqrt{N}}\sum_{i=1}^{D}(1-\delta_i)\left[\frac{1}{N}\sum_{s=1}^{D}\frac{\delta_s}{\hat{\pi}_i}\left(\frac{m_s}{\hat{m}_i} - 1\right)\frac{1}{\overline{P}^{*}_{s\Delta}(\boldsymbol{\theta})}\left\{\frac{\mathrm{I}(\mathcal{Y}_s = \mathcal{Y}_i, Z_s = Z_i)}{\boldsymbol{P}(\mathcal{Y} = \mathcal{Y}_i, Z = Z_i))}\right\}\right] + o_p(1) \\
= \; & O_p(M^{-1/2}) + o_p(1).
\end{aligned}
$$

Hence, we have

$$
\begin{aligned}
& N^{-1/2}\left[\hat{N}_{mc}(\hat{\boldsymbol{\theta}}^{(M)}_{mc}) - \hat{N}_{wc}(\hat{\boldsymbol{\theta}}_{wc}, \hat{\boldsymbol{\pi}})\right] \\
= \; & N^{-1/2}\left[\hat{N}_{wc}(\hat{\boldsymbol{\theta}}^{(M)}_{mc}, \hat{\boldsymbol{\pi}}) - \hat{N}_{wc}(\hat{\boldsymbol{\theta}}_{wc}, \hat{\boldsymbol{\pi}})\right] + O_p(M^{-1/2}) + o_p(1) \\
= \; & \left\{\frac{\partial \hat{N}_{wc}(\boldsymbol{\theta}, \boldsymbol{\pi})}{N\partial\boldsymbol{\theta}}\right\}\sqrt{N}\left(\hat{\boldsymbol{\theta}}^{(M)}_{mc} - \hat{\boldsymbol{\theta}}_{wc}\right) + O_p(M^{-1/2}) + o_p(1) \\
= \; & \{H_c(\boldsymbol{\theta}) + o_p(1)\}\sqrt{N}\left(\hat{\boldsymbol{\theta}}^{(M)}_{mc} - \hat{\boldsymbol{\theta}}_{wc}\right) + O_p(M^{-1/2}) + o_p(1),
\end{aligned}
$$

since $\dfrac{\partial \hat{N}_{wc}(\boldsymbol{\theta}, \boldsymbol{\pi})}{N\partial\boldsymbol{\theta}} \to H_c(\boldsymbol{\theta})$ in probability.

Since $\sqrt{N}\left\{\hat{\boldsymbol{\theta}}^{(M)}_{mc} - \hat{\boldsymbol{\theta}}_{wc}\right\}$ converges in probability to $\mathbf{0}$ as both $N \longrightarrow \infty$ and $M \longrightarrow \infty$, we have established that $N^{-1/2}\left\{\hat{N}_{mc}(\hat{\boldsymbol{\theta}}^{(M)}_{mc}) - \hat{N}_{wc}(\hat{\boldsymbol{\theta}}_{wc}, \hat{\boldsymbol{\pi}})\right\}$ converges in probability to $\mathbf{0}$ as both $N \longrightarrow \infty$ and $M \longrightarrow \infty$.

S2 Regression Calibration Estimation

Regression calibration in capture–recapture studies was originally proposed by Hwang and Huang (2003) under the restriction that $m_i = 1$, that is, surrogate variables consist of homogeneous measurement error variance. Here, we modify this approach that allows for the general case of $m_i > 1$.

Measurement error analysis involving regression calibration is widely used in many applications. For linear regression models, parameters are estimated using least square estimators where the unknown X_i is replaced with $\mathrm{E}(X_i \mid \overline{W}_i)$. For logistic regression type models, a similar but better strategy is used, such that $H(\beta X_i + \gamma^{\mathsf{T}} Z_i)$ is replaced by $\mathrm{E}\{H(\beta X_i + \gamma^{\mathsf{T}} Z_i) \mid Z_i, \overline{W}_i\}$. Let $\mu_x = \mathrm{E}(X_i)$ and $\sigma_x^2 = \mathrm{Var}(X_i)$. Using the approximation $H(x) \approx \Phi(x/1.7)$ and a normality assumption on X_i (Hwang and Huang, 2003), we have

$$\mathrm{E}\{H(\beta X_i + \gamma^{\mathsf{T}} Z_i) \mid Z_i, \overline{W}_i\} \approx H\left(\frac{\beta \tilde{X}_i + \gamma^{\mathsf{T}} Z_i}{\sqrt{1 + 0.346\beta^2 \tilde{\sigma}_i^2}}\right),$$

where $\tilde{X}_i = \mu_x + \dfrac{m_i \sigma_x^2}{m_i \sigma_x^2 + \sigma_u^2}(\overline{W}_i - \mu_x)$ and $\tilde{\sigma}_i^2 = \dfrac{\sigma_x^2 \sigma_u^2}{m_i \sigma_x^2 + \sigma_u^2}$. The above method is also commonly referred to as *refined regression calibration* (RRC, Wang et al., 2000). In practice, μ_x and σ_x^2 are estimated from the observed W_{ij} – e.g., $\hat{\mu}_x = \sum_{i=1}^{D} \overline{W}_i / D$ and $\hat{\sigma}_x^2 = \sum_{i=1}^{D}\{(\overline{W}_i - \hat{\mu}_x)^2 - \hat{\sigma}_u^2/m_i\}/D$. The RRC approach then solves equation (1) by replacing the $\beta \overline{W}_i + \gamma^{\mathsf{T}} Z_i$

with $(\beta\tilde{X}_i + \gamma^\mathsf{T} Z_i)/\sqrt{(1 + 0.346\beta^2\tilde{\sigma}_i^2)}$.

Let $\hat{\boldsymbol{\theta}}_r = (\hat{\beta}_r, \hat{\gamma}_r^\mathsf{T})^\mathsf{T}$ denote the resulting estimate of the regression calibration method, we estimate the population size N by using

$$\hat{N}_r = \frac{1}{\tau}\sum_{i=1}^{D} \frac{\mathcal{Y}_i}{H\{\hat{\beta}_r\overline{W}_i + \hat{\gamma}_r^\mathsf{T} Z_i + \frac{1}{2m_i}\hat{\beta}_r^2\hat{\sigma}_u^2\}}.$$

Note that $\hat{N}_r$ is an unbiased estimator of N if $\hat{\boldsymbol{\theta}}_r$ was replaced by the true $\boldsymbol{\theta}$. Regression calibration usually outperforms the naïve method when measurement error is present. However, in cases where the measurement error variance is large or there is obvious departure from normality in X_i then severe bias can arise in both the regression parameters and population size estimates, see Section 4.1.

S3 An Alternative MICS Algorithm

Algorithm: Multiple imputation with conditional score (MICS) estimation

{**Step 1:**} First, impute the estimating function for each missing of $\delta_i = 0$ and $i \leq D$, and then generate a simple random sample (with replacement) of size M from the set $\{k : \delta_k = 1, \mathcal{Y}_k = \mathcal{Y}_i, Z_k = Z_i\}$ and let $\mathcal{A}_i$ denote the resulting sample. Define $\tilde{\Phi}_i^{\dagger}(\boldsymbol{\theta}) = (1/M) \sum_{v \in \mathcal{A}_i} \Phi_v(\boldsymbol{\theta})$ where the summation has M terms.

{**Step 2:**} Solve the estimating equation:

$$U_{mc}(\boldsymbol{\theta}) = \sum_{i=1}^{D} \{\delta_i \Phi_i(\boldsymbol{\theta}) + (1 - \delta_i) \tilde{\Phi}_i^{\dagger}(\boldsymbol{\theta})\} = 0,$$

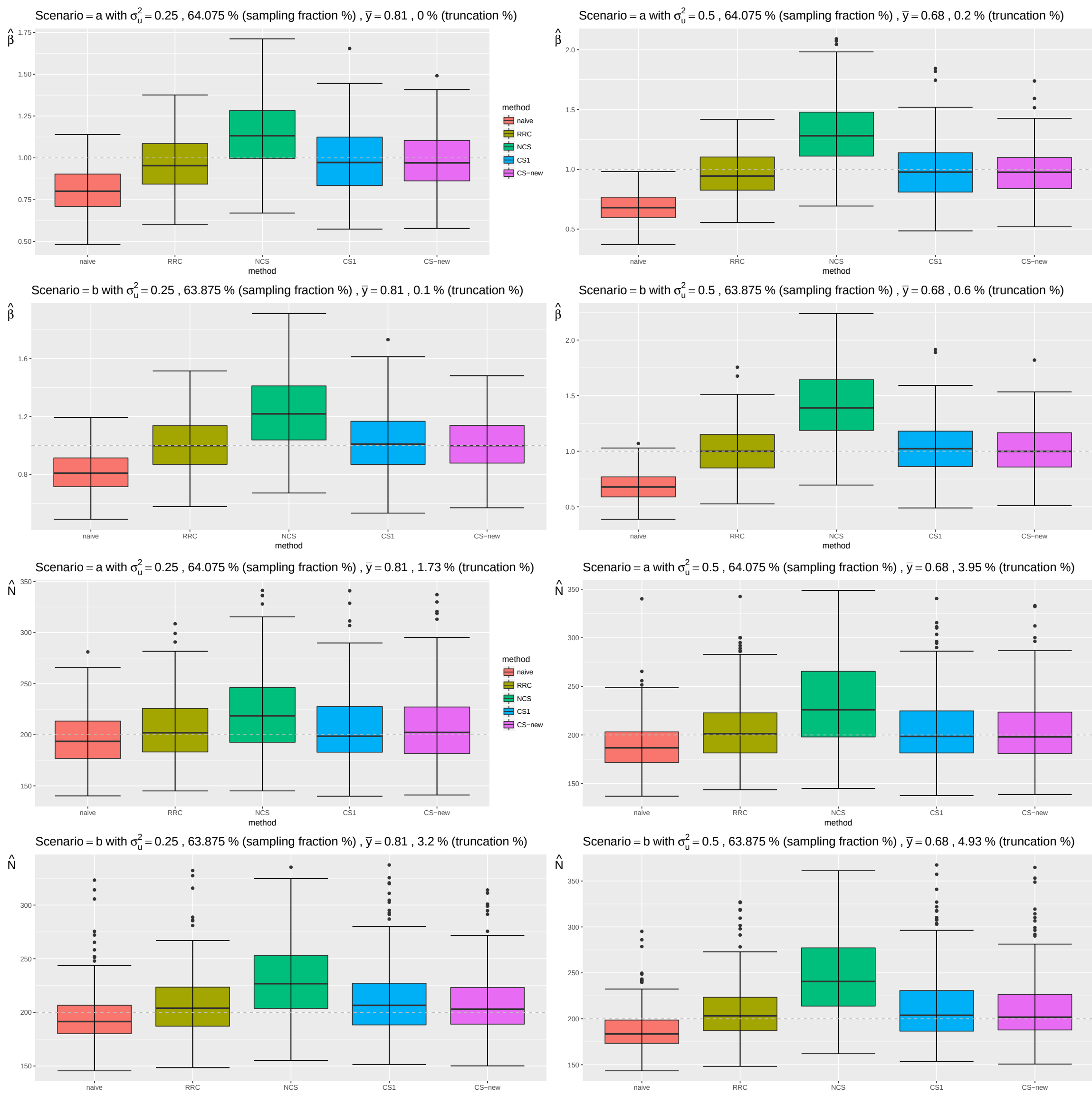
where $\hat{\boldsymbol{\theta}}_{mc}$ is the solution.

{**Step 3:**} For each missing value of $\delta_i = 0$ and $i \leq D$, we define $\tilde{P}_{i\Delta}^{\dagger}(\boldsymbol{\theta})$ to be the harmonic average of $\bar{P}_{v\Delta}^{*}(\boldsymbol{\theta})$, for all $v \in \mathcal{A}_i$. The MICS population size estimator is

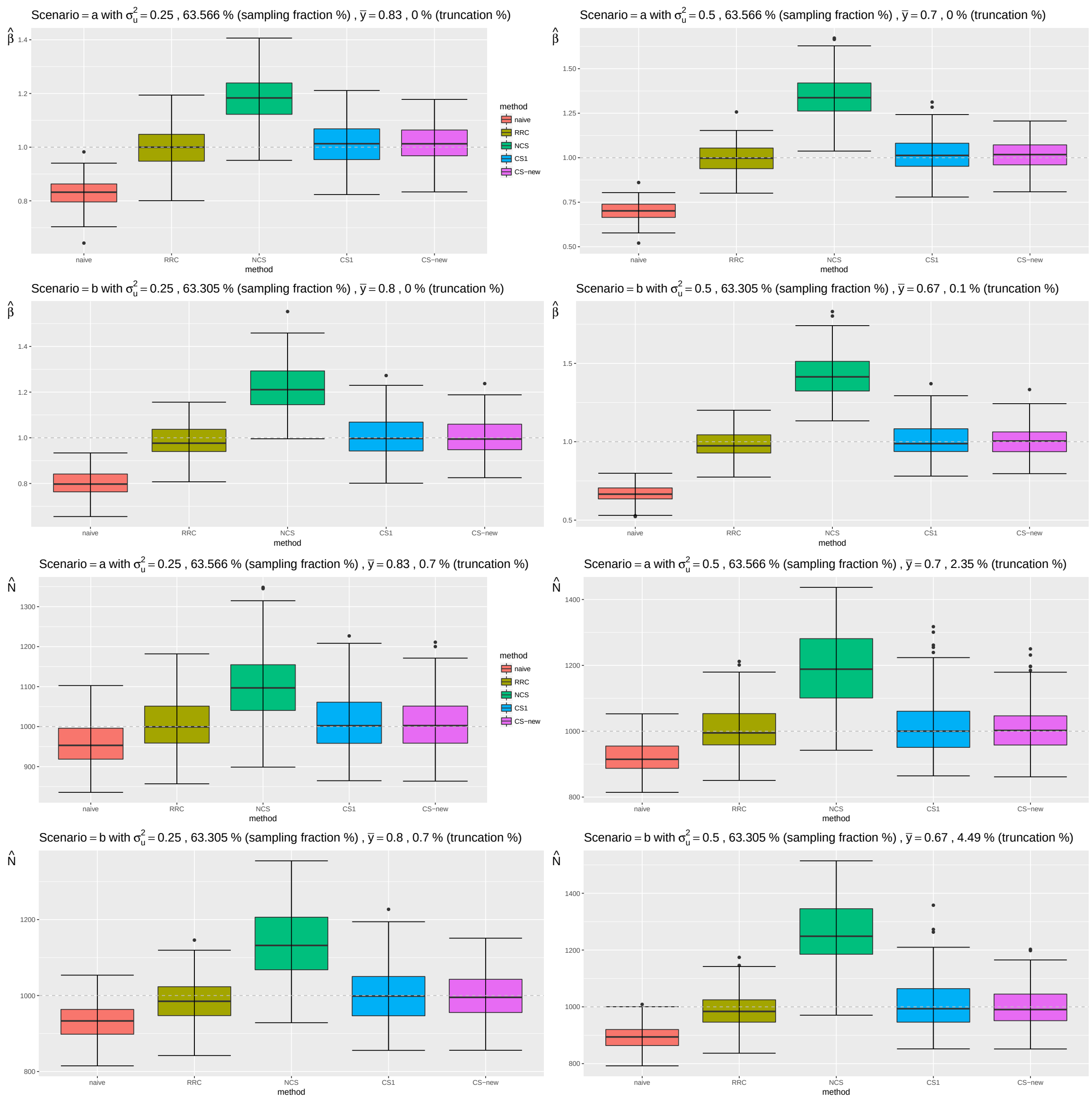
$$\hat{N}_{mc} = \sum_{i=1}^{D} \left\{ \delta_i \frac{1}{\bar{P}_{i\Delta}^{*}(\hat{\boldsymbol{\theta}}_{mc})} + (1 - \delta_i) \frac{1}{\tilde{P}_{i\Delta}^{\dagger}(\hat{\boldsymbol{\theta}}_{mc})} \right\}.$$

S4 Web Figures and Tables for Simulation Studies

Web Figures for Simulation Study 2

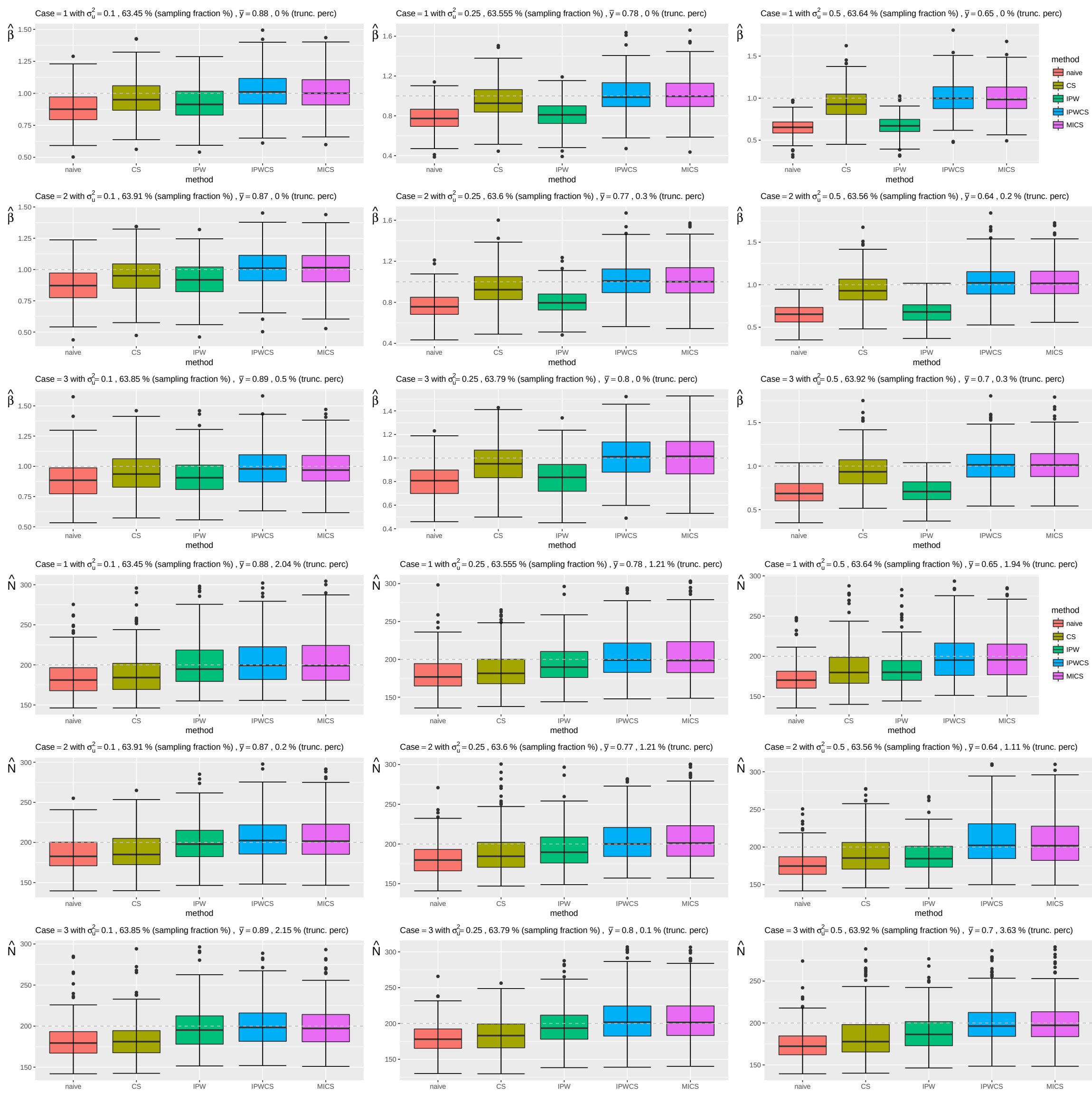


Web Figure 1: Simulation study 1. Comparative boxplots for $\hat{\beta}$ (first top two panels) and $\hat{N}$ (last bottom two panels) with two measurement error variances for two scenarios of different distribution for the covariate. The left-hand side column gives the results for $\sigma_u^2 = 0.25$ and the right-hand side column for $\sigma_u^2 = 0.5$. In each panel, the top row is scenario (a) and bottom row is scenario (b). In this simulation study we used $\beta = 1$, $N = 200$ and $\tau = 5$.

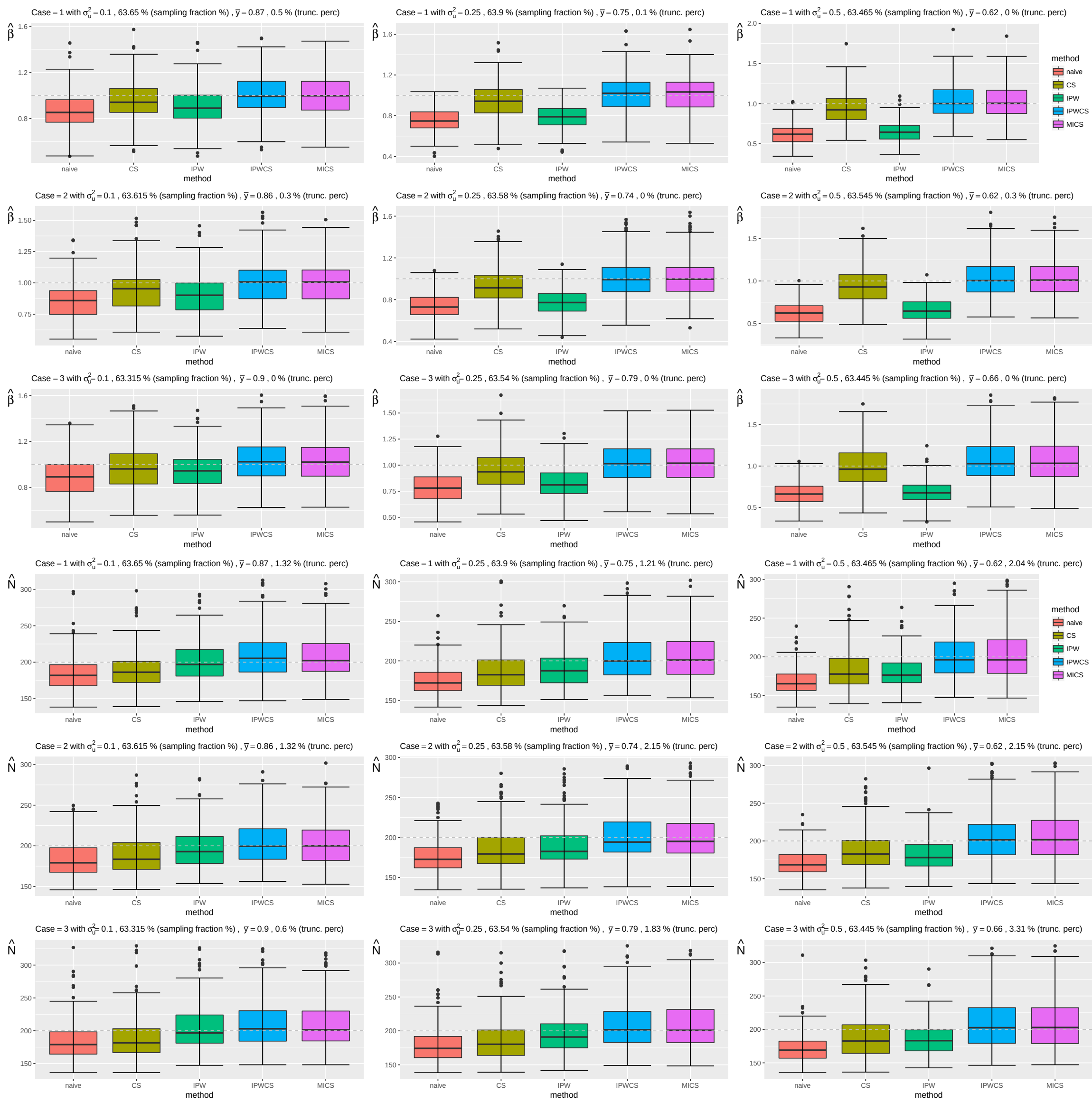


Web Figure 2: Comparative boxplots for $\hat{\beta}$ (first top two panels) and $\hat{N}$ (bottom last two panels) with two measurement error variances for two scenarios of different distribution for the covariate. The left-hand side column gives the results for $\sigma_u^2 = 0.25$ and the right-hand side column for $\sigma_u^2 = 0.5$. In each panel, the top row is scenario (a) and bottom row is scenario (b). In this simulation study we used $\beta = 1$, $N = 1000$ and $\tau = 5$.

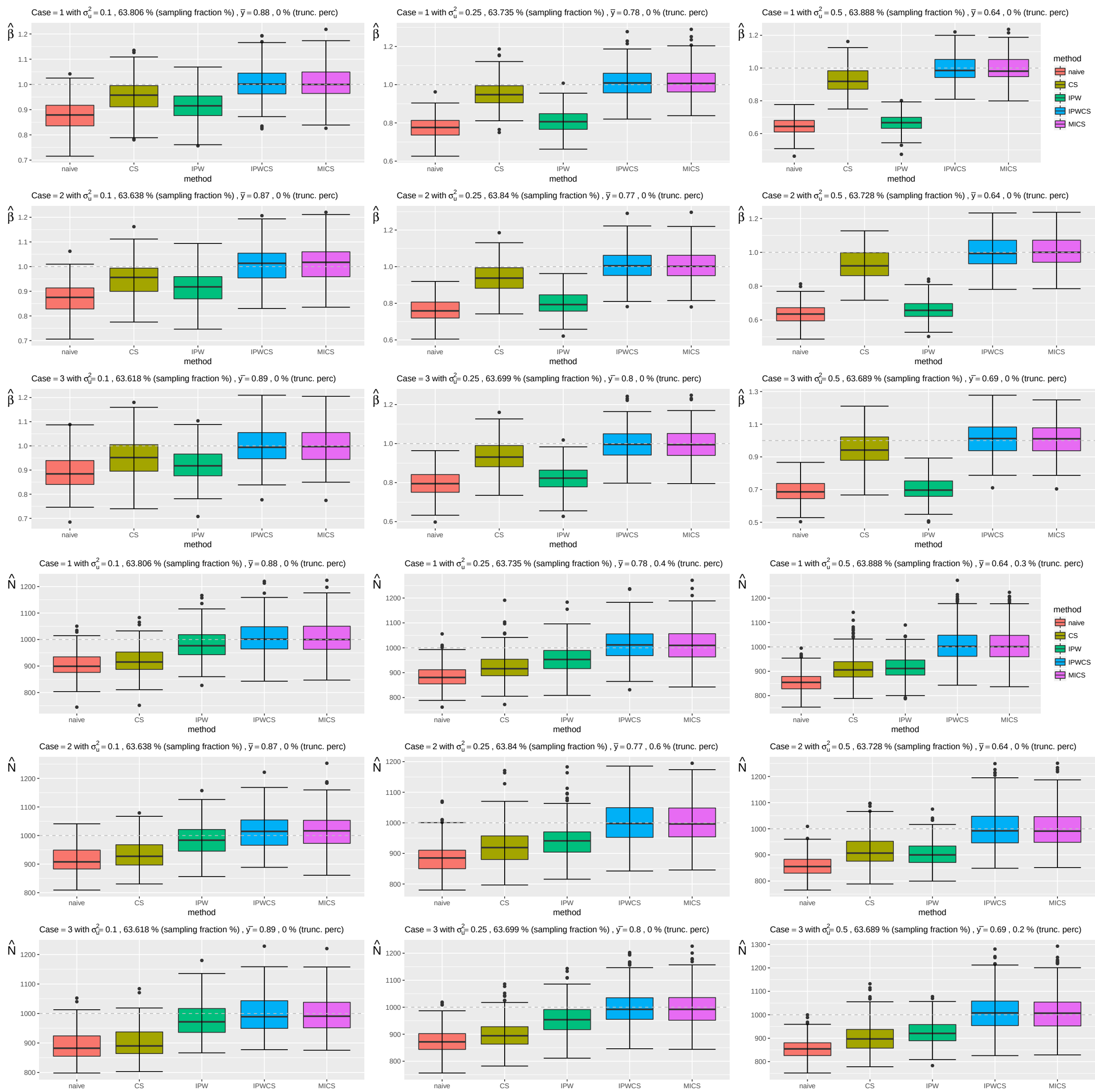
Web Figures for Simulation Study 2



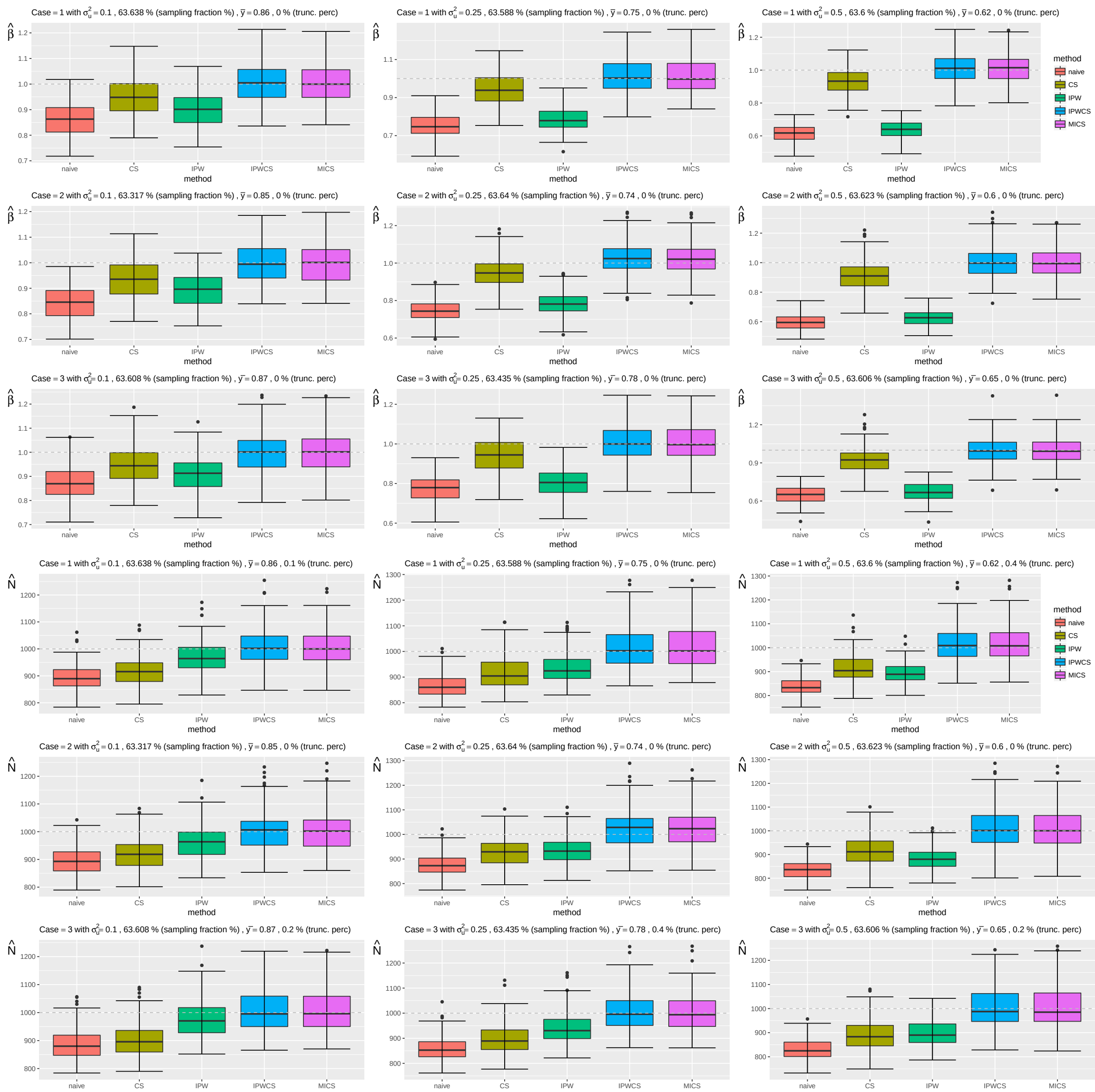
Web Figure 3: Simulation study 2. Comparative boxplots for $\hat{\beta}$ (first top three panels) and $\hat{N}$ (bottom last three panels) with three measurement error variances (by columns) for three missing data cases (by rows). In this simulation study we used $\beta = 1$, $N = 200$ and $\tau = 5$ for scenario (a).



Web Figure 4: Simulation study 2. Comparative boxplots for $\hat{\beta}$ (first top three panels) and $\hat{N}$ (bottom last three panels) with three measurement error variances (by columns) for three missing data cases (by rows). In this simulation study we used $\beta = 1$, $N = 200$ and $\tau = 5$ for scenario (b).



Web Figure 5: Simulation study 2. Comparative boxplots for $\hat{\beta}$ (first top three panels) and $\hat{N}$ (bottom last three panels) with three measurement error variances (by columns) for three missing data cases (by rows). In this simulation study we used $\beta = 1$, $N = 1000$ and $\tau = 5$ for scenario (a).



Web Figure 6: Simulation study 2. Comparative boxplots for $\hat{\beta}$ (first top three panels) and $\hat{N}$ (bottom last three panels) with three measurement error variances (by columns) for three missing data cases (by rows). In this simulation study we used $\beta = 1,\ N = 1000$ and $\tau = 5$ for scenario (b).

Web Tables for Simulation Study 1

Web Table 1: Simulation study 1: Summary of results for α and γ (top rows for each panel) and β and N (bottom rows for each panel). We used $N = 200$, $\tau = 5$ for scenario (a) with two different measurement error variances. The true parameter values $(\alpha, \gamma, \beta) = (-1, -1, 1)$. We fitted the naïve conditional likelihood model of Section 2.1 (labelled here as naïve), a naïve conditional score (NCS) model, the conditional score 1 (CS1) model, a refined regression calibration (RRC) approach (see Web Appendix S3), and the proposed conditional score approach (CS-new) of Section 2.2.

Scenario (a)			$\sigma^2 = 0.25$							
			α					γ		
Model	Mean	SD	M.SE	RMSE	CP	Mean	SD	M.SE	RMSE	CP
naïve	-0.91	0.15	0.15	0.18	0.88	-0.96	0.29	0.27	0.29	0.92
RRC	-1.02	0.17	0.16	0.17	0.94	-0.96	0.29	0.27	0.29	0.92
NCS	-1.08	0.18	0.18	0.20	0.94	-1.07	0.32	0.30	0.32	0.94
CS1	-0.99	0.18	0.17	0.18	0.92	-1.01	0.31	0.29	0.31	0.92
CS-new	-0.99	0.17	0.16	0.17	0.92	-1.01	0.30	0.28	0.30	0.94
			β					N		
Model	Mean	SD	M.SE	RMSE	CP	Mean	SD	M.SE	RMSE	CP
naïve	0.81	0.14	0.13	0.24	0.62	196.83	28.53	26.40	28.63	0.85
RRC	0.97	0.17	0.16	0.17	0.93	207.96	36.57	23.52	37.33	0.84
NCS	1.15	0.22	0.21	0.26	0.89	234.52	70.08	55.22	77.97	0.94
CS1	0.98	0.19	0.18	0.19	0.94	210.16	44.53	37.42	45.57	0.90
CS-new	0.98	0.18	0.17	0.18	0.92	208.38	39.43	34.19	40.21	0.90
D	128.15	6.48	-	-	-	-	-	-	-	-
$\bar{Y}$	1.91	0.10	-	-	-	-	-	-	-	-
			$\sigma^2 = 0.5$							
			α					γ		
Model	Mean	SD	M.SE	RMSE	CP	Mean	SD	M.SE	RMSE	CP
naïve	-0.85	0.15	0.14	0.21	0.78	-0.93	0.30	0.28	0.30	0.92
RRC	-1.03	0.17	0.17	0.18	0.95	-0.93	0.30	0.27	0.30	0.92
NCS	-1.16	0.22	0.22	0.27	0.94	-1.13	0.36	0.35	0.39	0.93
CS1	-1.00	0.20	0.19	0.20	0.93	-1.01	0.33	0.32	0.33	0.94
CS-new	-0.99	0.18	0.17	0.18	0.94	-1.02	0.32	0.30	0.31	0.94
			β					N		
naïve	0.68	0.13	0.12	0.35	0.28	189.37	25.33	23.30	27.41	0.76
RRC	0.97	0.19	0.18	0.19	0.94	208.53	40.50	23.95	41.29	0.82
NCS	1.32	0.29	0.28	0.43	0.85	284.97	223.00	124.50	238.12	0.95
CS1	0.99	0.23	0.22	0.23	0.92	213.81	57.94	46.65	59.43	0.88
CS-new	0.98	0.20	0.19	0.20	0.94	210.50	47.74	39.45	48.76	0.87
D	128.15	6.48	-	-	-	-	-	-	-	-
$\bar{Y}$	1.91	0.10	-	-	-	-	-	-	-	-

Web Table 2: Simulation study 1: Summary of results for α and γ (top rows for each panel) and β and N (bottom rows for each panel). We used $N = 200,\ \tau = 5$ for scenario (b) with two different measurement error variances. The true parameter values $(\alpha, \gamma, \beta) = (-1, -1, 1)$. We fitted the naïve conditional likelihood model of Section 2.1 (labelled here as naïve), a naïve conditional score (NCS) model, the conditional score 1 (CS1) model, a refined regression calibration (RRC) approach (see Web Appendix S3), and the proposed conditional score approach (CS-new) of Section 2.2.

Scenario (b)			$\sigma^2 = 0.25$							
			α					γ		
Model	Mean	SD	M.SE	RMSE	CP	Mean	SD	M.SE	RMSE	CP
naïve	-0.91	0.16	0.15	0.18	0.86	-0.99	0.28	0.27	0.28	0.94
RRC	-1.03	0.18	0.17	0.18	0.95	-0.99	0.28	0.27	0.28	0.94
NCS	-1.16	0.24	0.22	0.29	0.94	-1.07	0.30	0.29	0.31	0.98
CS1	-1.03	0.21	0.19	0.21	0.93	-1.03	0.29	0.28	0.29	0.96
CS-new	-1.02	0.18	0.18	0.18	0.94	-1.02	0.28	0.27	0.28	0.96
			β					N		
Model	Mean	SD	M.SE	RMSE	CP	Mean	SD	M.SE	RMSE	CP
naïve	0.81	0.14	0.13	0.23	0.63	197.21	27.79	26.87	27.86	0.88
RRC	1.00	0.18	0.17	0.18	0.93	210.77	36.49	24.73	37.95	0.90
NCS	1.23	0.25	0.24	0.34	0.90	252.81	75.07	72.55	91.63	0.99
CS1	1.03	0.21	0.20	0.21	0.94	215.04	40.90	41.65	43.48	0.94
CS-new	1.01	0.18	0.18	0.18	0.96	211.30	35.75	36.31	37.41	0.94
D	127.75	5.77	-	-	-	-	-	-	-	-
$\bar{Y}$	1.93	0.09	-	-	-	-	-	-	-	-
			$\sigma^2 = 0.5$							
			α					γ		
Model	Mean	SD	M.SE	RMSE	CP	Mean	SD	M.SE	RMSE	CP
naïve	-0.84	0.15	0.14	0.22	0.74	-0.96	0.28	0.27	0.28	0.94
RRC	-1.05	0.20	0.18	0.20	0.93	-0.96	0.28	0.27	0.28	0.94
NCS	-1.29	0.32	0.30	0.44	0.96	-1.12	0.35	0.34	0.37	0.98
CS1	-1.04	0.25	0.22	0.25	0.94	-1.03	0.32	0.31	0.32	0.96
CS-new	-1.02	0.20	0.19	0.20	0.96	-1.02	0.29	0.28	0.29	0.95
			β					N		
naïve	0.68	0.13	0.12	0.34	0.32	188.31	23.89	23.08	26.54	0.82
RRC	1.01	0.21	0.19	0.21	0.92	211.82	40.84	26.04	42.42	0.88
NCS	1.45	0.37	0.35	0.59	0.82	330.47	283.21	229.69	311.17	1.00
CS1	1.04	0.27	0.25	0.27	0.92	219.31	55.92	54.50	59.03	0.94
CS-new	1.01	0.21	0.21	0.21	0.94	211.70	39.76	40.97	41.35	0.91
D	127.75	5.77	-	-	-	-	-	-	-	-
$\bar{Y}$	1.93	0.09	-	-	-	-	-	-	-	-

Web Table 3: Simulation study 1: Summary of results for α and γ (top rows for each panel) and β and N (bottom rows for each panel). We used $N = 1000$, $\tau = 5$ for scenario (a) with two different measurement error variances. The true parameter values $(\alpha, \gamma, \beta) = (-1, -1, 1)$. We fitted the naïve conditional likelihood model of Section 2.1 (labelled here as naïve), a naïve conditional score (NCS) model, the conditional score 1 (CS1) model, a refined regression calibration (RRC) approach (see Web Appendix S3), and the proposed conditional score approach (CS-new) of Section 2.2.

Scenario (a)	$\sigma^2 = 0.25$									
	α					γ				
Model	Mean	SD	M.SE	RMSE	CP	Mean	SD	M.SE	RMSE	CP
naïve	-0.93	0.07	0.07	0.10	0.78	-0.94	0.12	0.12	0.13	0.94
RRC	-1.04	0.07	0.07	0.08	0.94	-0.94	0.12	0.12	0.13	0.94
NCS	-1.10	0.09	0.08	0.13	0.80	-1.05	0.13	0.13	0.14	0.94
CS1	-1.01	0.08	0.08	0.08	0.96	-0.99	0.13	0.13	0.13	0.97
CS-new	-1.01	0.08	0.07	0.08	0.96	-0.99	0.12	0.13	0.12	0.98
	β					N				
Model	Mean	SD	M.SE	RMSE	CP	Mean	SD	M.SE	RMSE	CP
naïve	0.83	0.06	0.06	0.18	0.20	957.27	54.10	53.80	68.83	0.80
RRC	1.00	0.07	0.07	0.07	0.96	1007.08	66.14	54.33	66.35	0.90
NCS	1.18	0.09	0.09	0.20	0.48	1110.39	102.01	100.46	150.13	0.94
CS1	1.01	0.08	0.08	0.08	0.96	1008.50	72.58	70.44	72.90	0.94
CS-new	1.01	0.07	0.08	0.07	0.97	1007.93	67.71	67.43	68.00	0.94
D	635.66	14.46	-	-	-	-	-	-	-	-
$\bar{Y}$	1.91	0.04	-	-	-	-	-	-	-	-

	$\sigma^2 = 0.5$									
	α					γ				
Model	Mean	SD	M.SE	RMSE	CP	Mean	SD	M.SE	RMSE	CP
naïve	-0.87	0.06	0.06	0.15	0.48	-0.90	0.12	0.12	0.16	0.88
RRC	-1.06	0.08	0.08	0.10	0.90	-0.90	0.12	0.12	0.16	0.88
NCS	-1.19	0.11	0.10	0.21	0.56	-1.10	0.15	0.15	0.18	0.90
CS1	-1.02	0.09	0.09	0.09	0.94	-0.99	0.14	0.14	0.14	0.98
CS-new	-1.02	0.08	0.08	0.09	0.93	-0.99	0.13	0.13	0.13	0.98
	β					N				
naïve	0.70	0.05	0.06	0.30	0.00	921.14	48.29	47.33	92.41	0.52
RRC	1.00	0.08	0.08	0.08	0.95	1007.62	69.20	55.32	69.44	0.88
NCS	1.34	0.12	0.12	0.36	0.14	1243.91	183.83	170.11	305.15	0.94
CS1	1.02	0.10	0.10	0.10	0.96	1011.69	83.90	79.77	84.51	0.94
CS-new	1.02	0.08	0.09	0.08	0.96	1009.77	74.90	73.75	75.35	0.94
D	635.66	14.46	-	-	-	-	-	-	-	-
$\bar{Y}$	1.91	0.04	-	-	-	-	-	-	-	-

Web Table 4: Simulation study 1: Summary of results for α and γ (top rows for each panel) and β and N (bottom rows for each panel). We used $N = 1000$, $\tau = 5$ for scenario (b) with two different measurement error variances. The true parameter values $(\alpha, \gamma, \beta) = (-1, -1, 1)$. We fitted the naïve conditional likelihood model of Section 2.1 (labelled here as naïve), a naïve conditional score (NCS) model, the conditional score 1 (CS1) model, a refined regression calibration (RRC) approach (see Web Appendix S3), and the proposed conditional score approach (CS-new) of Section 2.2.

Scenario (b)			$\sigma^2 = 0.25$							
			α					γ		
Model	Mean	SD	M.SE	RMSE	CP	Mean	SD	M.SE	RMSE	CP
naïve	-0.89	0.06	0.07	0.12	0.60	-0.96	0.11	0.12	0.11	0.94
RRC	-1.02	0.07	0.08	0.07	0.96	-0.96	0.11	0.12	0.11	0.94
NCS	-1.15	0.09	0.10	0.17	0.73	-1.05	0.11	0.13	0.13	0.97
CS1	-1.01	0.08	0.09	0.08	0.97	-1.00	0.11	0.13	0.11	0.97
CS-new	-1.01	0.07	0.08	0.07	0.96	-1.00	0.11	0.12	0.11	0.96
			β					N		
Model	Mean	SD	M.SE	RMSE	CP	Mean	SD	M.SE	RMSE	CP
naïve	0.80	0.06	0.06	0.21	0.13	933.04	48.09	50.63	82.37	0.68
RRC	0.99	0.08	0.08	0.08	0.96	988.14	59.94	54.18	60.95	0.90
NCS	1.22	0.10	0.11	0.24	0.52	1148.58	110.24	121.41	184.85	0.96
CS1	1.01	0.09	0.09	0.09	0.95	1001.60	69.17	73.23	69.02	0.94
CS-new	1.00	0.08	0.08	0.08	0.96	998.74	63.59	68.89	63.44	0.94
D	633.05	12.41	-	-	-	-	-	-	-	-
$\bar{Y}$	1.94	0.04	-	-	-	-	-	-	-	-
			$\sigma^2 = 0.5$							
			α					γ		
Model	Mean	SD	M.SE	RMSE	CP	Mean	SD	M.SE	RMSE	CP
naïve	-0.82	0.06	0.06	0.19	0.18	-0.94	0.11	0.12	0.13	0.92
RRC	-1.03	0.08	0.08	0.08	0.96	-0.92	0.11	0.12	0.13	0.90
NCS	-1.27	0.12	0.13	0.29	0.42	-1.11	0.13	0.15	0.17	0.94
CS1	-1.01	0.09	0.10	0.09	0.96	-1.01	0.12	0.14	0.12	0.96
CS-new	-1.01	0.08	0.09	0.08	0.96	-1.00	0.11	0.13	0.11	0.97
			β					N		
naïve	0.67	0.06	0.05	0.33	0.00	893.92	42.62	43.31	114.28	0.33
RRC	0.99	0.09	0.08	0.09	0.94	987.57	63.77	55.30	64.81	0.88
NCS	1.43	0.14	0.15	0.45	0.10	1359.08	246.29	239.58	435.08	0.96
CS1	1.01	0.11	0.11	0.11	0.96	1005.88	84.24	84.40	84.23	0.94
CS-new	1.01	0.09	0.09	0.09	0.97	1000.58	72.61	76.19	72.44	0.93
D	633.05	12.41	-	-	-	-	-	-	-	-
$\bar{Y}$	1.94	0.04	-	-	-	-	-	-	-	-

Web Tables for Comparisons with Other Simulation Studies

Web Table 5: Comparison simulation study 1: Summary of results for α, β and N (top to bottom). We used $N = 500$, $\tau = 5$ with $\sigma_u = 0.6$. The true parameter values $(\alpha, \beta) = (0.2, 1)$. We fitted the naïve conditional likelihood model of Section 2.1 (labelled here as naïve), a naïve conditional score (NCS) model, the conditional score 1 (CS1) model, a refined regression calibration (RRC) approach (see Web Appendix S3), and the proposed conditional score approach (CS-new) of Section 2.2. The results for the generalized method of moments (GMM1) and semiparametric efficient score (Semi-Nor) methods are taken directly from Table 1 of Xu and Ma (2014). The relative efficiency (RE) is defined as the ratio between the mean square errors of Semi-Nor with each model (RE=(MSE of Semi-Nor)/(MSE of model) for each setting.

$\alpha = 0.2$						
Model	Mean	SD	M.SE	MSE	CP	RE
naïve	0.2067	0.0626	0.0636	0.0040	0.958	1.0335
NCS	0.1428	0.0696	0.0711	0.0081	0.875	0.5047
CS1	0.1961	0.0687	0.0714	0.0047	0.964	0.8651
RRC	0.1420	0.0664	0.0674	0.0078	0.858	0.5270
CS-new	0.1960	0.0649	0.0663	0.0042	0.957	0.9689
GMM1	0.1970	0.0688	0.0690	0.0047	0.950	0.8638
Semi-Nor	0.2006	0.0640	0.0658	0.0041	0.954	1.0000
$\beta = 1$						
Model	Mean	SD	M.SE	MSE	CP	RE
naïve	0.9377	0.0487	0.0497	0.0063	0.740	0.4629
NCS	1.0906	0.0601	0.0610	0.0118	0.701	0.2449
CS1	1.0056	0.0577	0.0607	0.0034	0.965	0.8613
RRC	1.0340	0.0576	0.0584	0.0045	0.933	0.6470
CS-new	1.0047	0.0545	0.0558	0.0030	0.951	0.9673
GMM1	1.0091	0.0591	0.0585	0.0036	0.958	0.8095
Semi-Nor	1.0033	0.0537	0.0546	0.0029	0.951	1.0000
$N = 500$						
Model	Mean	SD	M.SE	MSE	CP	RE
naïve	503.5307	22.3323	22.7456	511.1975	0.956	1.0366
NCS	525.7275	28.8415	29.2725	1493.7364	0.972	0.3548
CS1	502.7117	22.8782	22.9453	530.7654	0.939	0.9984
RRC	524.1309	26.5265	2.6095	1285.9555	0.116	0.4121
CS-new	502.3458	22.1295	22.2775	495.2175	0.944	1.0700
GMM1	504.4400	23.9600	23.4000	593.7952	0.953	0.8924
Semi-Nor	502.1400	22.9200	22.2800	529.9060	0.933	1.0000
D	416.9180	8.4684	-	-	-	-
$\bar{Y}$	3.1780	0.0739	-	-	-	-

Web Table 6: Comparison simulation study 1: Summary of results for α, β and N (top to bottom). We used $N = 500$, $\tau = 3$ with $\sigma_u = 0.6$. The true parameter values $(\alpha, \beta) = (-1, 1)$. We fitted the naïve conditional likelihood model of Section 2.1 (labelled here as naïve), a naïve conditional score (NCS) model, the conditional score 1 (CS1) model, a refined regression calibration (RRC) approach (see Web Appendix S3), and the proposed conditional score approach (CS-new) of Section 2.2. The results for the generalized method of moments (GMM1) and semiparametric efficient score (Semi-Nor) methods are taken directly from Table 2 of Xu and Ma (2014). The relative efficiency (RE) is defined as the ratio between the mean square errors of Semi-Nor with each model (RE=(MSE of Semi-Nor)/(MSE of model) for each setting).

			$\alpha = 0.2$			
Model	Mean	SD	M.SE	MSE	CP	RE
naïve	-0.8570	0.1205	0.1213	0.0350	0.748	0.6438
NCS	-1.1440	0.1788	0.1796	0.0527	0.932	0.4271
CS1	-1.0235	0.1608	0.1614	0.0264	0.958	0.8524
RRC	-1.0149	0.1434	0.1444	0.0208	0.957	1.0831
CS-new	-1.0198	0.1485	0.1500	0.0224	0.957	1.0030
GMM1	-1.0686	0.1782	0.1676	0.0365	0.956	0.6174
Semi-Nor	-1.0101	0.1497	0.1465	0.0225	0.957	1.0000
			$\beta = 1$			
Model	Mean	SD	M.SE	MSE	CP	RE
naïve	0.7313	0.1086	0.1057	0.0840	0.285	0.2951
NCS	1.2027	0.2147	0.2074	0.0872	0.898	0.2843
CS1	1.0200	0.1774	0.1738	0.0319	0.943	0.7777
RRC	0.9986	0.1553	0.1479	0.0241	0.938	1.0275
CS-new	1.0154	0.1641	0.1597	0.0272	0.948	0.9123
GMM1	1.0692	0.2082	0.1800	0.0481	0.934	0.5149
Semi-Nor	1.0102	0.1571	0.1556	0.0248	0.950	1.0000
			$N = 500$			
Model	Mean	SD	M.SE	MSE	CP	RE
naïve	469.2274	42.2850	41.3299	2734.9741	0.780	1.7537
NCS	584.6809	135.3284	116.4296	25484.6307	0.981	0.1882
CS1	518.3482	77.0364	71.8022	6271.2634	0.929	0.7648
RRC	510.1921	62.1588	15.0108	3967.5953	0.414	1.2089
CS-new	514.7939	68.6821	65.2863	4936.0903	0.939	0.9717
GMM1	544.2200	142.2100	94.3400	22179.0925	0.962	0.2163
Semi-Nor	512.2700	68.1600	63.8500	4796.3385	0.933	1.0000
D	298.2930	10.9330	-	-	-	-
$\bar{Y}$	1.5241	0.0401	-	-	-	-

Web Table 7: Comparison simulation study 2: Summary of results for N for $\sigma_u^2 = 0.5$ and $\sigma_u^2 = 1$ with no missing data cases $P_{\text{meas}} = 1$ and some missing data with $P_{\text{meas}} = 0.9$. Here, we used $N = 200,\ \tau = 5$ and the true parameter values were $(\alpha, \beta) = (-1, 0.5)$. We fitted the naïve conditional likelihood (labelled here as naïve CL), naïve inverse probability weighting (IPW), the conditional score (CS) approach with complete case only inverse probability weighting conditional score (IPWCS), and multiple imputation conditional score (MICS). The results for the parametric maximum likelihood (ML) are taken directly from Table 1 of Xi *et al.* (2009). The relative efficiency (RE) is defined as the ratio between the mean square errors of Xi *et al.* (2009) with each model (RE=(MSE of Xi *et al.* (2009))/(MSE of model) for each setting).

	$\sigma_u^2 = 0.5$ and $P_{\text{meas}} = 1$						$\sigma_u^2 = 0.5$ and $P_{\text{meas}} = 0.9$					
Model	Mean	SD	M.SE	MSE	CP	RE	Mean	SD	M.SE	MSE	CP	RE
naïve CL	199.50	13.15	12.51	173.173	0.93	1.155	194.90	12.16	11.57	173.876	0.86	1.150
IPW	199.50	13.15	12.51	173.173	0.93	1.155	199.36	13.13	14.10	172.806	0.95	1.157
CS	202.14	14.35	13.68	210.502	0.95	0.950	197.13	13.15	12.93	181.159	0.90	1.104
IPWCS	202.14	14.35	13.53	210.502	0.95	0.950	202.23	14.45	13.68	213.775	0.95	0.936
MICS	202.15	14.36	13.69	210.832	0.95	0.949	202.20	14.43	13.80	213.065	0.95	0.939
Xi *et al.*	202.00	14.00	14.00	200.000	0.95	1.000	202.00	14.00	14.00	200.000	0.95	1.000
D/N	0.77	0.03	-	-	-	-	0.77	0.03	-	-	-	-
$\bar{D}_\delta/N$	0.00	0.00	-	-	-	-	7.58	2.81	-	-	-	-
$\bar{Y}$	1.81	0.07	-	-	-	-	1.81	0.07	-	-	-	-
	$\sigma_u^2 = 1$ and $P_{\text{meas}} = 1$						$\sigma_u^2 = 1$ and $P_{\text{meas}} = 0.9$					
naïve CL	195.05	11.75	11.02	162.565	0.87	1.439	190.91	10.99	10.25	203.408	0.75	1.303
IPW	195.05	11.75	11.02	162.565	0.87	1.439	194.77	11.73	12.22	164.946	0.88	1.607
CS	202.89	16.32	15.45	274.694	0.95	0.852	197.18	14.39	14.18	215.024	0.89	1.232
IPWCS	202.89	16.32	15.11	274.694	0.94	0.852	203.06	16.60	15.57	284.924	0.95	0.930
MICS	202.90	16.34	15.47	275.406	0.95	0.850	202.96	16.44	15.93	279.035	0.95	0.950
Xi *et al.*	203.00	15.00	15.00	234.000	0.95	1.000	203.00	16.00	15.00	265.000	0.95	1.000
D/N	0.77	0.03	-	-	-	-	0.77	0.03	-	-	-	-
$\bar{D}_\delta/N$	0.00	0.00	-	-	-	-	7.58	2.81	-	-	-	-
$\bar{Y}$	1.81	0.07	-	-	-	-	1.81	0.07	-	-	-	-

Web Table 8: Comparison simulation study 2: Summary of results for N for $\sigma_u^2 = 0.5$ and $\sigma_u^2 = 1$ with no missing data cases $P_{\text{meas}} = 1$ and some missing data with $P_{\text{meas}} = 0.9$. Here, we used $N = 200,\ \tau = 5$ and the true parameter values were $(\alpha, \beta) = (-0.5, 1)$. We fitted the naïve conditional likelihood (labelled here as naïve CL), naïve inverse probability weighting (IPW), the conditional score (CS) approach with complete case only inverse probability weighting conditional score (IPWCS) and multiple imputation conditional score (MICS) and the Xi *et al.* (2009) approach. The results for the parametric maximum likelihood (ML) are taken directly from Table 1 of Xi *et al.* (2009). The relative efficiency (RE) is defined as the ratio between the mean square errors of Xi *et al.* (2009) with each model (RE=(MSE of Xi *et al.* (2009))/(MSE of model) for each setting).

	$\sigma_u^2 = 0.5$ and $P_{\text{meas}} = 1$						$\sigma_u^2 = 0.5$ and $P_{\text{meas}} = 0.9$					
Model	Mean	SD	M.SE	MSE	CP	RE	Mean	SD	M.SE	MSE	CP	RE
naïve CL	197.38	11.12	10.95	130.519	0.89	1.793	193.42	10.03	9.81	143.897	0.82	1.842
IPW	197.38	11.12	10.95	130.519	0.89	1.793	196.84	10.97	11.77	130.327	0.90	2.033
CS	201.51	13.18	12.93	175.992	0.93	1.330	197.04	11.81	11.73	148.238	0.89	1.788
IPWCS	201.51	13.18	12.74	175.992	0.93	1.330	201.58	13.34	13.07	180.452	0.93	1.469
MICS	201.51	13.18	12.93	175.992	0.93	1.330	201.60	13.43	13.25	182.925	0.93	1.449
Xi *et al.*	203.00	15.00	15.00	234.000	0.92	1.000	203.00	16.00	14.00	265.000	0.93	1.000
D/N	0.83	0.03	-	-	-	-	0.83	0.03	-	-	-	-
$\bar{D}_\delta/N$	0.00	0.00	-	-	-	-	5.28	2.29	-	-	-	-
$\bar{Y}$	2.39	0.09	-	-	-	-	2.39	0.09	-	-	-	-

	$\sigma_u^2 = 1$ and $P_{\text{meas}} = 1$						$\sigma_u^2 = 1$ and $P_{\text{meas}} = 0.9$					
Model	Mean	SD	M.SE	MSE	CP	RE	Mean	SD	M.SE	MSE	CP	RE
naïve CL	189.45	9.13	8.41	194.659	0.65	1.361	186.24	8.27	7.46	257.730	0.48	1.028
IPW	189.45	9.13	8.41	194.659	0.65	1.361	188.59	8.89	8.71	209.220	0.65	1.267
CS	201.81	19.88	17.00	398.490	0.90	0.665	196.54	18.06	16.72	338.135	0.82	0.784
IPWCS	201.81	19.88	16.90	398.490	0.90	0.665	202.11	21.47	18.17	465.413	0.90	0.569
MICS	201.81	19.88	17.25	398.490	0.90	0.665	202.09	21.27	19.04	456.781	0.90	0.580
Xi *et al.*	203.00	16.00	15.00	265.000	0.93	1.000	203.00	16.00	15.00	265.000	0.93	1.000
D/N	0.83	0.03	-	-	-	-	0.83	0.03	-	-	-	-
$\bar{D}_\delta/N$	0.00	0.00	-	-	-	-	5.28	2.29	-	-	-	-
$\bar{Y}$	2.39	0.09	-	-	-	-	2.39	0.09	-	-	-	-

Bibliography

Foutz, R. V. (1977). On the unique consistent solution to the likelihood equations. *Journal of the American Statistical Association* **72**, 147-148.

Huggins, R. M. (1989). On the statistical analysis of capture experiments. *Biometrika* **76**, 113-140.

Hwang, W. H. and Huang, S. Y. H. (2003). Estimation in capture–recapture models when covariates are subject to measurement errors. *Biometrics* **59**, 1113-1122.

Wang, C. Y., Wang, N. and Wang, S. (2000). Regression analysis when covariates are regression parameters of a random effects model for observed longitudinal measurements. *Biometrics* **56**, 487-495.

Xi, L., Watson, R., Wang, J. P. and Yip, P. S. F. (2009). Estimation in capture–recapture models when covariates are subject to measurement errors and missing data. *The Canadian Journal of Statistics* **37**, 645-658.

Xu, K. and Ma, Y. (2014). Effective use of multiple error-prone covariate measurements in capture–recapture models. *Statistica Sinica* **24**, 1529-1546.