Supplementary materials for: "Analysis of Linear Transformation Models with Covariate Measurement Error and Interval Censoring"

Soutrik Mandal, Suojin Wang, and Samiran Sinha*
Department of Statistics, Texas A&M University, College Station, Texas 77843-3143
sinha@stat.tamu.edu*

S.1 Inconsistency of naive method in measurement error scenario

In the naive approach, unobserved X_i is replaced by $\overline{W}_i = \sum_{j=1}^{m} W_{ij}/m$ in the estimation method where all covariates are assumed to be error free. The reason is explained as follows. First, $f_T(T|\overline{W}_i, \mathbf{Z}_i; \boldsymbol{\beta}, H) = \int f_T(T|\overline{W}_i, X_i, \mathbf{Z}_i; \boldsymbol{\beta}, H) f_X(X_i|\mathbf{Z}_i, \overline{W}_i) dX_i = \int f_T(T|X_i, \mathbf{Z}_i; \boldsymbol{\beta}, H) f_X(X_i|\mathbf{Z}_i, \overline{W}_i) dX_i$. If $f_X(X_i|\mathbf{Z}_i, \overline{W}_i)$ is degenerate at $\overline{W}_i$, the functional form of $f_T(T|\overline{W}_i, \mathbf{Z}_i; \boldsymbol{\beta}, H)$ will be the same as $f_T(T|X_i, \mathbf{Z}_i; \boldsymbol{\beta}, H)$ with X_i being replaced by $\overline{W}_i$ and then the naive approach will consistently estimate $\boldsymbol{\beta}$ and H. However, for a fixed m, $f_X(X_i|\mathbf{Z}_i, \overline{W}_i)$ tends to be degenerate at $\overline{W}_i$ only when $\sigma_u^2 \to 0$. This situation is obviously not much of a concern.

S.2 Details of regression calibration method

Define $\widehat{X}_i(\boldsymbol{\zeta}) = \zeta_0 + \zeta_1 \overline{W}_i + \boldsymbol{\zeta}_2^T \mathbf{Z}_i$. In the regression calibration approach, an unobserved X_i is replaced by $\widehat{X}_i \equiv \widehat{X}_i(\widehat{\boldsymbol{\zeta}})$ in the estimation method where all covariates are assumed to be error free, and $\widehat{\boldsymbol{\zeta}} = (\widehat{\zeta}_0, \widehat{\zeta}_1, \widehat{\boldsymbol{\zeta}}_2)^T$ is given by $\widehat{\zeta}_0 = (s_u^2/\eta m)\overline{W} - (s_u^2/\eta m) s_{wz}^T S_z^{-1} \overline{\mathbf{Z}}$, $\widehat{\zeta}_1 = 1 - (s_u^2/\eta m)$, $\widehat{\boldsymbol{\zeta}}_2 = (s_u^2/\eta m) s_{wz}^T S_z^{-1}$, $\eta = (s_x^2 + s_u^2/m) - s_{wz}^T S_z^{-1} s_{wz}$, $\overline{W} = \sum_{i=1}^n \overline{W}_i/n$, $\overline{\mathbf{Z}} = \sum_{i=1}^n \mathbf{Z}_i/n$, $S_z = (n-1)^{-1} \sum_{i=1}^n (\mathbf{Z}_i - \overline{\mathbf{Z}})^{\otimes 2}$, $s_{wz} = (n-1)^{-1} \sum_{i=1}^n (\mathbf{Z}_i - \overline{\mathbf{Z}})(\overline{W}_i - \overline{W})$, $s_u^2 = \{m(m-1)n\}^{-1} \sum_{i=1}^n \sum_{j<k} (W_{ij} - W_{ik})^2$, $s_x^2 = [m \sum_{i=1}^n (\overline{W}_i - \overline{W})^{\otimes 2} - (n-1) s_u^2]/\{m(n-1)\}$, and $a^{\otimes 2} := a a^T$ for any generic vector a.

S.3 Turnbull's algorithm for estimating survival probability

First we consider a grid of time points $0 = \xi_0 < \xi_1 < \xi_2 < \cdots < \xi_q$ which includes all observed $(L_i, R_i]$ intervals. For the ith subject, we define a weight $\alpha_{ij} = 1$ if $(L_i, R_i]$ contains the interval (ξ_{j-1}, ξ_j) and 0 otherwise. We start with an initial guess at $\widehat{S}(\xi_j)$ and proceed as follows:

Step 1. At time ξ_j, the probability of an event is given by

$$p_j = \widehat{S}(\xi_{j-1}) - \widehat{S}(\xi_j),\ j = 1, \ldots, q;$$

Step 2. The number of events at time ξ_j is estimated by

$$d_j = \sum_{i=1}^{n} \frac{\alpha_{ij} p_j}{\sum_{j'=1}^{q} \alpha_{ij'} p'_j},\ j = 1, \ldots, q;$$

Step 3. The number of at-risk subjects at time ξ_j is estimated by $Y^*(\xi_j) = \sum_{j'=j}^{q} d_{j'}$;
Step 4. Compute the updated product-limit estimator given by

$$\widehat{S}(\xi_j) = \prod_{l=1}^{j}\left\{1 - \frac{d_l}{Y^*(\xi_l)}\right\}.$$

Repeat Steps 1-4 until $\widehat{S}(\xi_j)$ converges with a specified tolerance.

S.4 Conditional distributions used in the Gibbs sampling referenced in Section 3.1

Step 0. Initialize the parameters $\boldsymbol{\gamma}_l, \sigma_l^2, \pi_l$, $l = 1, \ldots, k'$, $\psi_1, \ldots, \psi_n$, and $X_1, \ldots, X_n$.

Then in each iteration of the collapsed Gibbs sampling we do the following steps with the most recently updated values of the parameters, $\psi_1, \ldots, \psi_n$, and $X_1, \ldots, X_n$.
Step 1. Sample σ_u^2 from InvGamma$(mn/2 + a_u, \sum_{i=1}^{n}\sum_{j=1}^{m}(W_{ij} - \overline{W}_i)^2/2 + 1/b_u)$;
Step 2. Sample $(\pi_1, \pi_2, \ldots, \pi_{k'})$ from Dirichlet$(\sum_{i=1}^{n} I(\psi_1 = 1) + \alpha, \ldots, \sum_{i=1}^{n} I(\psi_{k'} = 1) + \alpha)$;
Step 3. For each $i = 1, \ldots, n$, sample ψ_i from Multinomial$(p_{i,1}, \ldots, p_{i,k'})$, where

$$p_{i,l} \propto \frac{\pi_l}{\sqrt{2\pi(\sigma_l^2 + \sigma_u^2/m)}} \exp\{-(X_i - \boldsymbol{\gamma}_l^T \tilde{\mathbf{Z}}_i)^2/2(\sigma_l^2 + \sigma_u^2/m)\}$$

for $l = 1, \cdots, k'$;
Step 4. Sample σ_l^2 from the target density

$$\frac{(\sigma_l^2)^{-a_\sigma - 1}}{(\sigma_u^2/m + \sigma_l^2)^{\sum_{i=1}^{n} I(\psi_i = l)/2}} \exp\left\{-\frac{\sum_{i=1}^{n} I(\psi_i = l)(\overline{W}_i - \gamma_l^T \tilde{\mathbf{Z}}_i)^2}{2(\sigma_u^2/m + \sigma_l^2)} - \frac{1}{b_\sigma \sigma_l^2}\right\};$$

Step 5. Sample $\boldsymbol{\gamma}_l$ from the multivariate normal distribution with variance and mean

$$\Omega_l = \left\{\frac{\sum_{i=1}^{n} I(\psi_i = l)\tilde{\mathbf{Z}}_i^{\otimes 2}}{\sigma_l^2 + \sigma_u^2/m} + \Sigma_{\gamma_j}^{-1}\right\}^{-1},$$

$$\tilde{\boldsymbol{\mu}}_l = \Omega_l\left\{\frac{\sum_{i=1}^{n} I(\psi_i = l)(\overline{W}_i - \boldsymbol{\gamma}_l^T \tilde{\mathbf{Z}}_i)\tilde{\mathbf{Z}}_i}{\sigma_l^2 + \sigma_u^2/m} + \Sigma_{\gamma_j}^{-1}\boldsymbol{\mu}_{\gamma_j}\right\},$$

respectively;
Step 6. Finally, sample X_i's from the Normal distribution with variance $v_x = \{\sum_{l=1}^{k'} I(\psi_i = l)/\sigma_l^2 + m/\sigma_u^2\}^{-1}$ and mean $m_x = v_x\{\sum_{l=1}^{k'} I(\psi_i = l)\boldsymbol{\gamma}_l^T \tilde{\mathbf{Z}}_i/\sigma_l^2 + m\overline{W}_i/\sigma_u^2\}$.

S.5 Standard error for the regression calibration method

Let $\widehat{\boldsymbol{\beta}}_{rc,k_t} = (\widehat{\beta}_{rc,k_t,1}, \widehat{\boldsymbol{\beta}}_{rc,k_t,2}^T)^T$ and $\widehat{H}_{rc,k_t}$ be the estimators of $\boldsymbol{\beta}$ and H for the k_tth imputed data with T being imputed according to the method in Section 3.2, and X_i being replaced by $\widehat{X}_i$ whose formula is given in Section

2. After a few steps of algebra we can obtain a consistent estimator of the asymptotic variance of $\widehat{\boldsymbol{\beta}}_{rc,k_t}$ given by

$$\Sigma_{rc}(\widehat{\boldsymbol{\beta}}_{rc,k_t}, \widehat{H}_{rc,k_t}) = A_{rc}^{-1}\{A_{rc,M} + D\widehat{\text{var}}(\sqrt{n}\widehat{\boldsymbol{\zeta}})D^T\}A_{rc}^{-T},$$

where A_{rc} and $A_{rc,M}$ are the same as A and A_M, respectively, given in Section 3.3 with $\widehat{\boldsymbol{\beta}}$, $\widehat{H}$, and X_i being replaced by $\widehat{\boldsymbol{\beta}}_{rc,k_t}$ and $\widehat{H}_{rc,k_t}$, and $\widehat{X}_i$, respectively, and $D = \sum_{i=1}^n \psi_i/n$,

$$\psi_i = \begin{pmatrix} \widehat{X}_i \\ \mathbf{Z}_i \end{pmatrix} \lambda\{\widehat{\beta}_{rc,k_t,1}\widehat{X}_i + \widehat{\boldsymbol{\beta}}_{rc,k_t,2}^T\mathbf{Z}_i + \widehat{H}_{rc,k_t}(V_{i,k_t})\}\left\{\frac{\partial}{\partial\boldsymbol{\zeta}}\widehat{H}_{rc,k_t}(V_{i,k_t}, \widehat{\boldsymbol{\beta}}_{rc,k_t}, \widehat{\boldsymbol{\zeta}}) + \widehat{\beta}_{rc,k_t,1}\frac{\partial}{\partial\boldsymbol{\zeta}}\widehat{X}_i(\widehat{\boldsymbol{\zeta}})\right\},$$

$$\frac{\partial}{\partial\boldsymbol{\zeta}}\widehat{H}_{rc,k_t}(t, \widehat{\boldsymbol{\beta}}_{rc,k_t}, \widehat{\boldsymbol{\zeta}}) = -\widehat{\beta}_{rc,k_t,1}\frac{\int_0^t \{C_3(s)/C_2(s)\}\exp[\int_0^s\{C_1(u)/C_2(u)\}d\widehat{H}_{rc,k_t}(u)]d\widehat{H}_{rc,k_t}(s)}{\exp[\int_0^t\{C_1(u)/C_2(u)\}d\widehat{H}_{rc,k_t}(u)]},$$

$C_1(u) = (1/n)\sum_{i=1}^n Y_i(u)\dot{\lambda}\{\widehat{\beta}_{rc,k_t,1}\widehat{X}_i + \widehat{\boldsymbol{\beta}}_{rc,k_t,2}^T\mathbf{Z}_i + \widehat{H}_{rc,k_t}(u)\}$, $C_2(u) = (1/n)\sum_{i=1}^n Y_i(u)\lambda\{\widehat{\beta}_{rc,k_t,1}\widehat{X}_i + \widehat{\boldsymbol{\beta}}_{rc,k_t,2}^T\mathbf{Z}_i + \widehat{H}_{rc,k_t}(u)\}$, $C_3(u) = (1/n)\sum_{i=1}^n Y_i(u)\{\partial\widehat{X}_i(\widehat{\boldsymbol{\zeta}})/\partial\boldsymbol{\zeta}\}\dot{\lambda}\{\widehat{\beta}_{rc,k_t,1}\widehat{X}_i + \widehat{\boldsymbol{\beta}}_{rc,k_t,2}^T\mathbf{Z}_i + \widehat{H}_{rc,k_t}(u)\}$, and $\widehat{\text{var}}(\widehat{\boldsymbol{\zeta}})$ is obtained by the bootstrap resampling method. Finally, the RC estimators for $\boldsymbol{\beta}$ and H are defined as $\widehat{\boldsymbol{\beta}}_{rc} = \sum_{k_t=1}^{m_t^*}\widehat{\boldsymbol{\beta}}_{rc,k_t}/m_t^*$ and $\widehat{H}_{rc} = \sum_{k_t=1}^{m_t^*}\widehat{H}_{rc,k_t}/m_t^*$, respectively, and the asymptotic variance of $\widehat{\boldsymbol{\beta}}_{rc}$ can be estimated by

$$\Omega_{rc} = \frac{1}{m_t^*}\sum_{k_t=1}^{m_t^*}\Sigma_{rc}(\widehat{\boldsymbol{\beta}}_{rc,k_t}, \widehat{H}_{rc,k_t}) + (1 + \frac{1}{m_t^*})\sum_{k_t=1}^{m_t^*}\frac{(\widehat{\boldsymbol{\beta}}_{rc,k_t} - \widehat{\boldsymbol{\beta}}_{rc})^{\otimes 2}}{m_t^* - 1}.$$

S.6 Large sample properties of the proposed estimator

Proofs of Theorem 1 and Corollary 1. Following Lemma 1 of Wang and Robins (1998), the proposed estimator $\widehat{\boldsymbol{\beta}}_c$ is asymptotically equivalent to $\widetilde{\boldsymbol{\beta}}$, the solution of

$$\begin{aligned} \mathbf{0} &= \frac{1}{m_x^* m_t^* n}\sum_{k_t=1}^{m_t^*}\sum_{k_x=1}^{m_x^*}\boldsymbol{S}_\beta(\boldsymbol{\beta}, \widehat{H}(\cdot, \boldsymbol{\beta}), k_x, k_t) \\ &= \frac{1}{m_x^* m_t^* n}\sum_{i=1}^n\sum_{k_t=1}^{m_t^*}\sum_{k_x=1}^{m_x^*}\mathbf{Z}_{i,k_x}^*\left[\Delta_{i,k_t,k_x} - \Lambda\{\widehat{H}_{k_t,k_x}(V_{i,k_t,k_x}; \boldsymbol{\beta}) + X_{i,k_x}^*\beta_1 + \mathbf{Z}_i^T\boldsymbol{\beta}_2\}\right], \end{aligned}$$

where $\mathbf{Z}_{i,k_x}^* = (X_{i,k_x}^*, \mathbf{Z}_i^T)^T$, $X_{i,k_x}^* = X_{i,k_x}^*(\widetilde{\boldsymbol{\theta}}_{k_x})$ denotes the k_xth posterior sample of X_i when $\boldsymbol{\theta} = \widetilde{\boldsymbol{\theta}}_{k_x}$. Also, note that when $\Delta_i = 1$, $V_{i,k_t,k_x} = T_{i,k_t,k_x}$ the k_tth sampled value of the ith time-to-event drawn from the conditional density $f(T_{i,k_t,k_x}|X_{i,k_x}^*, \mathbf{Z}_i, L_i < T_{i,k_t,k_x} \le R_i; \boldsymbol{\beta})$, and when $\Delta_i = 0$, $V_{i,k_t,k_x} = L_i$. Also, note that $\Delta_i = \Delta_{i,k_t,k_x}$ for every k_t and k_x.

Additionally, define

$$S_i(\boldsymbol{\beta}, V_{i,k_t,k_x}(\boldsymbol{\beta}), X^*_{i,k_x}(\boldsymbol{\theta})) = \int_0^\tau \{\mathbf{Z}^*_{i,k_x} - \boldsymbol{\mu}_{z,k_t,k_x}(u;\boldsymbol{\beta})\} d\mathcal{M}_{i,k_t,k_x}(u;\boldsymbol{\beta},H),$$

where $\mathcal{M}_{i,k_t,k_x}(u;\boldsymbol{\beta},H) = N_{i,k_t,k_x}(u) - \int_0^u Y_{i,k_t,k_x}(\vartheta) d\Lambda\{H_{k_t,k_x}(\vartheta) + X^*_{i,k_x}\beta_1 + \mathbf{Z}_i^T\boldsymbol{\beta}_2\}$, $N_{i,k_t,k_x}(u) = I(V_{i,k_t,k_x} \leq u, \Delta_{i,k_t,k_x} = 1)$ and $Y_{i,k_t,k_x}(u) = I(V_{i,k_t,k_x} \geq u)$,

$$\begin{aligned}
\boldsymbol{\mu}_{z,k_t,k_x}(t;\boldsymbol{\beta}) &= \frac{\sum_{i=1}^n \mathbf{Z}^*_i \lambda\{\beta_1 X^*_{i,k_x} + \mathbf{Z}_i^T\boldsymbol{\beta}_2 + \widehat{H}_{k_t,k_x}(V_{i,k_t,k_x};\boldsymbol{\beta})\} Y_{i,k_t,k_x}(t) B_{k_t,k_x}(t, V_{i,k_t,k_x};\boldsymbol{\beta})}{\sum_{i=1}^n \lambda\{\beta_1 X^*_{i,k_x} + \mathbf{Z}_i^T\boldsymbol{\beta}_2 + \widehat{H}_{k_t,k_x}(t;\boldsymbol{\beta})\} Y_{i,k_t,k_x}(t)}, \\
B_{k_t,k_x}(t,s;\boldsymbol{\beta}) &= \exp\Big[- \sum_{k: t \leq \vartheta_{k-1} < \vartheta_k \leq s} \frac{\sum_{i=1}^n \dot{\lambda}\{\beta_1 X^*_{i,k_x} + \mathbf{Z}_i^T\boldsymbol{\beta}_2 + \widehat{H}_{k_t,k_x}(\vartheta_k;\boldsymbol{\beta})\} Y_{i,k_t,k_x}(\vartheta_k)}{\sum_{i=1}^n \lambda\{\widehat{\beta}_{c,1} X^*_{i,k_x} + \mathbf{Z}_i^T\widehat{\boldsymbol{\beta}}_{c,2} + \widehat{H}_{k_t,k_x}(\vartheta_k;\boldsymbol{\beta})\} Y_{i,k_t,k_x}(\vartheta_k)} \\
&\quad \times \{\widehat{H}_{k_t,k_x}(\vartheta_k;\boldsymbol{\beta}) - \widehat{H}_{k_t,k_x}(\vartheta_{k-1};\boldsymbol{\beta})\}\Big],
\end{aligned}$$

and $\widehat{H}_{k_t,k_x}(t;\boldsymbol{\beta})$ is the solution of $H(t)$ for the equation

$$\sum_{i=1}^n [dN_{i,k_t,k_x}(t) - Y_{i,k_t,k_x}(t) d\Lambda\{H(t) + X^*_{i,k_x}\beta_1 + \mathbf{Z}_i^T\boldsymbol{\beta}_2\}] = 0. \tag{S.1}$$

Note that due to the martingale property, $E\{S_i(\boldsymbol{\beta}^*, V_{i,k_t,k_x}(\boldsymbol{\beta}^*), X^*_{i,k_x}(\boldsymbol{\theta}))\} = 0$. We assume that standard regularity conditions hold (Hartigan, 1983; p. 108) and that our flexible prior allows that the posterior distribution of $\boldsymbol{\theta}$ is asymptotically normal with mean $\widehat{\boldsymbol{\theta}}_{ML}$ and variance $\{n\widehat{\mathcal{I}}_1(\widehat{\boldsymbol{\theta}}_{ML})\}^{-1}$, where $\widehat{\boldsymbol{\theta}}_{ML}$ is the maximum likelihood estimate of $\boldsymbol{\theta}$, $\widehat{\mathcal{I}}_1(\boldsymbol{\theta}) = \{-\partial^2 \log(\mathcal{L}_1)/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T\}/n$, and $\mathcal{L}_1 = \prod_{i=1}^n \mathcal{L}_{1,i}$, where

$$\begin{aligned}
\mathcal{L}_{1,i} &= \int f(W_{i1}|X_i) \times \cdots \times f(W_{im}|X_i) f(X_i|\tilde{\mathbf{Z}}_i) dX_i \\
&= \frac{1}{(\sqrt{2\pi}\sigma_u)^m} \exp\Big\{ - \frac{\sum_{j=1}^m (W_{ij} - \overline{W}_i)^2}{2\sigma_u^2} \Big\} \times \sum_{l=1}^{k'} \frac{\pi_l}{\sqrt{2\pi(\sigma_u^2/m + \sigma_l^2)}} \exp\Big\{ - \frac{(\overline{W}_i - \gamma_l^T \tilde{\mathbf{Z}}_i)^2}{2(\sigma_u^2/m + \sigma_l^2)} \Big\}.
\end{aligned}$$

Suppose that standard regularity conditions hold and $\sqrt{n}(\widehat{\boldsymbol{\theta}}_{ML} - \boldsymbol{\theta}) = n^{-1/2} \sum_{i=1}^n \psi_i(\boldsymbol{\theta}) + o_p(1)$, where $\psi_i(\boldsymbol{\theta}) = \mathcal{I}_1^{-1}(\boldsymbol{\theta}) \partial \log(\mathcal{L}_{1,i})/\partial\boldsymbol{\theta}$ and all elements of $\mathrm{cov}\{\psi_i(\boldsymbol{\theta})\}$ are assumed to be finite, and $\mathcal{I}_1(\boldsymbol{\theta}) = E\{-\partial^2 \log(\mathcal{L}_{1,i})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T\}$.

Define $S_i(\boldsymbol{\beta}, \boldsymbol{\beta}, \boldsymbol{\theta}) = \sum_{k_t=1}^{m_t^*} \sum_{k_x=1}^{m_x^*} S_i(\boldsymbol{\beta}, V_{i,k_t,k_x}(\boldsymbol{\beta}), X^*_{i,k_x}(\boldsymbol{\theta}))/m_t^* m_x^*$. Assume that $\tilde{\boldsymbol{\beta}}$ converges to $\boldsymbol{\beta}^*$. We can now write

$$\begin{aligned}
\mathbf{0} &= \frac{1}{m_x^* m_t^* n} \sum_{i=1}^n \sum_{k_t=1}^{m_t^*} \sum_{k_x=1}^{m_x^*} \mathbf{Z}^*_{i,k_x} \Big[\Delta_{i,k_t,k_x} - \Lambda\{\widehat{H}_{k_t,k_x}(V_{i,k_t,k_x};\widetilde{\boldsymbol{\beta}}) + X^*_{i,k_x}(\widetilde{\boldsymbol{\theta}}_{k_x})\widetilde{\beta}_1 + \mathbf{Z}_i^T \widetilde{\boldsymbol{\beta}}_2\} \Big] \\
&= \frac{1}{m_t^* m_x^* n} \sum_{i=1}^n \sum_{k_t=1}^{m_t^*} \sum_{k_x=1}^{m_x^*} S_i(\widetilde{\boldsymbol{\beta}}, V_{i,k_t,k_x}(\widetilde{\boldsymbol{\beta}}), X^*_{i,k_x}(\widehat{\boldsymbol{\theta}}_{ML})) + \frac{1}{m_x^*} \sum_{k_x=1}^{m_x^*} D_2(\widetilde{\boldsymbol{\theta}}_{k_x} - \widehat{\boldsymbol{\theta}}_{ML}) + o_p(n^{-1/2})
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{n}\sum_{i=1}^{n} S_i(\widetilde{\boldsymbol\beta},\widetilde{\boldsymbol\beta},\boldsymbol\theta) + D_2(\widehat{\boldsymbol\theta}_{ML}-\boldsymbol\theta) + \frac{1}{m_x^*}\sum_{k_x=1}^{m_x^*} D_2(\widetilde{\boldsymbol\theta}_{k_x}-\widehat{\boldsymbol\theta}_{ML}) + o_p(n^{-1/2}) \\
&= \frac{1}{n}\sum_{i=1}^{n} S_i(\boldsymbol\beta^*,\boldsymbol\beta^*,\boldsymbol\theta) + D_1(\widetilde{\boldsymbol\beta}-\boldsymbol\beta^*) + D_2(\widehat{\boldsymbol\theta}_{ML}-\boldsymbol\theta) + \frac{1}{m_x^*}\sum_{k_x=1}^{m_x^*} D_2(\widetilde{\boldsymbol\theta}_{k_x}-\widehat{\boldsymbol\theta}_{ML}) + o_p(n^{-1/2}), \quad \text{(S.2)}
\end{aligned}
$$

where

$$
\begin{aligned}
D_1 &= -\Bigg(E\Bigg[\frac{1}{m_t^* m_x^*}\sum_{k_t=1}^{m_t^*}\sum_{k_x=1}^{m_x^*}\int_0^\tau \{\mathbf{Z}^*_{k_x}-\boldsymbol\mu_{z,k_t,k_x}(u;\boldsymbol\beta)\}Y_{k_t,k_x}(u)\Big[\mathbf{Z}^{*T}_{k_x}\dot\lambda\{\widehat{H}_{k_t,k_x}(u;\boldsymbol\beta)+X^*_{k_x}\beta_1+\mathbf{Z}^T\boldsymbol\beta_2\}d\widehat{H}_{k_t,k_x}(u;\boldsymbol\beta) \\
&\quad +\dot\lambda\{\widehat{H}_{k_t,k_x}(u;\boldsymbol\beta)+X^*_{k_x}\beta_1+\mathbf{Z}^T\boldsymbol\beta_2\}\frac{\partial H_{k_t,k_x}(u;\boldsymbol\beta)}{\partial\boldsymbol\beta}d\widehat{H}_{k_t,k_x}(u;\boldsymbol\beta)+\lambda\{\widehat{H}_{k_t,k_x}(u;\boldsymbol\beta)+X^*_{k_x}\beta_1+\mathbf{Z}_i^T\boldsymbol\beta_2\} \\
&\quad \times\frac{\partial d\widehat{H}_{k_t,k_x}(u;\boldsymbol\beta)}{\partial\boldsymbol\beta}\Big]\Bigg]+E\Bigg[\int\frac{1}{m_x^*}\sum_{k_x=1}^{m_x^*}\int_0^\tau\{\mathbf{Z}^*_{k_x}-\boldsymbol\mu_{z,k_t,k_x}(u;\boldsymbol\beta)\}d\mathcal{M}_{k_t,k_x}(u;\boldsymbol\beta,H) \\
&\quad \times\Delta\frac{\partial}{\partial\boldsymbol\beta}\log\{f(T_{k_t,k_x}|X^*_{k_x},\mathbf{Z},L<T_{k_t,k_x}\le R;\boldsymbol\beta)\}f(T_{k_t,k_x}|X^*_{k_x},\mathbf{Z},L<T_{k_t,k_x}\le R;\boldsymbol\beta)dT_{k_t,k_x}\Bigg]\Bigg)_{\boldsymbol\beta=\boldsymbol\beta^*},
\end{aligned}
$$

$$
\begin{aligned}
D_2 &= \Bigg(E\Bigg[\frac{1}{m_t^*}\sum_{k_t=1}^{m_t^*}\int S(\boldsymbol\beta,V_{k_t,k_x}(\boldsymbol\beta),X^*_{k_x})\frac{\partial}{\partial\boldsymbol\theta^T}\log\{f(X^*_{k_x}|L<T<R,W_1,\ldots,W_m,\mathbf{Z};\boldsymbol\theta)\} \\
&\quad \times f(X^*_{k_x}|L<T<R,W_1,\ldots,W_m,\mathbf{Z};\boldsymbol\theta)dX^*_{k_x}\Bigg]\Bigg)_{\boldsymbol\beta=\boldsymbol\beta^*}.
\end{aligned}
$$

Now, after rearranging the terms of (S.2), and using $\sqrt{n}(\widehat{\boldsymbol\theta}_{ML}-\boldsymbol\theta)=n^{-1/2}\sum_{i=1}^n\psi_i(\boldsymbol\theta)+o_p(1)$, where $E\{\psi(\boldsymbol\theta)\}=0$ and all elements of $\text{cov}\{\psi_i(\boldsymbol\theta)\}$ are finite, we obtain the following influence function representation of $\widetilde{\boldsymbol\beta}$,

$$
\sqrt{n}(\widetilde{\boldsymbol\beta}-\boldsymbol\beta^*) = -\frac{1}{\sqrt{n}}\sum_{i=1}^n\Big\{D_1^{-1}S_i(\boldsymbol\beta^*,\boldsymbol\beta^*,\boldsymbol\theta)+D_1^{-1}D_2\psi_i(\boldsymbol\theta)\Big\}-\frac{D_1^{-1}D_2\sqrt{n}}{m_x^*}\sum_{k_x=1}^{m_x^*}(\widetilde{\boldsymbol\theta}_{k_x}-\widehat{\boldsymbol\theta}_{ML})+o_p(1). \quad \text{(S.3)}
$$

The second component of (S.3) arises due to posterior samples $(\widetilde{\boldsymbol\theta}_1,\ldots,\widetilde{\boldsymbol\theta}_{m_x^*})$. Conditional on the observed data, $D_1^{-1}D_2\sqrt{n}\sum_{k_x=1}^{m_x^*}(\widetilde{\boldsymbol\theta}_{k_x}-\widehat{\boldsymbol\theta}_{ML})/m_x^*$ converges to a normal distribution with mean 0 and variance $D_1^{-1}D_2\{\mathcal{I}_1(\boldsymbol\theta)\}^{-1}D_2^TD_1^{-T}/m_x^*$ (which is actually independent of the observed data), where $\mathcal{I}_1(\boldsymbol\theta)=E\{\widehat{\mathcal{I}}_1(\boldsymbol\theta)\}$, and the first part of (S.3) converges to a mean zero normal distribution due to the Central Limit Theorem. Therefore, $\sqrt{n}(\widetilde{\boldsymbol\beta}-\boldsymbol\beta^*)$ converges to a convolution of two independent normal distributions which is normal. In other words,

$$
\sqrt{n}\Sigma_{\widetilde{\boldsymbol\beta}}^{-1/2}(\widetilde{\boldsymbol\beta}-\boldsymbol\beta^*)\xrightarrow{d}\text{Normal}(\mathbf{0},\mathbf{I}),
$$

where $\mathbf{I}$ is the identity matrix and $\Sigma_{\widetilde{\boldsymbol{\beta}}} = E\left\{D_1^{-1}S(\boldsymbol{\beta}^*, \boldsymbol{\beta}^*, \boldsymbol{\theta}) + D_1^{-1}D_2\psi(\boldsymbol{\theta})\right\}^{\otimes 2} + (m_x^*)^{-1}D_1^{-1}D_2\{\mathcal{I}_1(\boldsymbol{\theta})\}^{-1}D_2^TD_1^{-T}$. The asymptotic variance of $\sqrt{n}\widetilde{\boldsymbol{\beta}}$ (or $\sqrt{n}\widehat{\boldsymbol{\beta}}_c$) can be consistently estimated by

$$\frac{1}{n}\sum_{i=1}^{n}\left\{\widehat{D}_1^{-1}S_i(\widehat{\boldsymbol{\beta}}_c, \widehat{\boldsymbol{\beta}}_c, \widehat{\boldsymbol{\theta}}_{ML}) + \widehat{D}_1^{-1}\widehat{D}_2\psi_i(\widehat{\boldsymbol{\theta}}_{ML})\right\}^{\otimes 2} + (m_x^*)^{-1}\widehat{D}_1^{-1}\widehat{D}_2\{\widehat{\mathcal{I}}_1(\widehat{\boldsymbol{\theta}}_{ML})\}^{-1}\widehat{D}_2^T\widehat{D}_1^{-T},$$

where

$$\begin{aligned}
\widehat{D}_1 \;=\; & -\frac{1}{n}\sum_{i=1}^{n}\Bigg(\frac{1}{m_t^* m_x^*}\sum_{k_t=1}^{m_t^*}\sum_{k_x=1}^{m_x^*}\int_0^{\tau}\{\mathbf{Z}_{i,k_x}^* - \boldsymbol{\mu}_{z,k_t,k_x}(u;\boldsymbol{\beta})\}Y_{i,k_t,k_x}(u) \\
& \times\Bigg[\mathbf{Z}^{*T}_{i,k_x}\dot{\lambda}\{\widehat{H}_{k_t,k_x}(u;\boldsymbol{\beta}) + X_{i,k_x}^*\beta_1 + \mathbf{Z}_i^T\boldsymbol{\beta}_2\}d\widehat{H}_{k_t,k_x}(u;\boldsymbol{\beta}) \\
& + \dot{\lambda}\{\widehat{H}_{k_t,k_x}(u;\boldsymbol{\beta}) + X_{i,k_x}^*\beta_1 + \mathbf{Z}_i^T\boldsymbol{\beta}_2\}\frac{\partial \widehat{H}_{k_t,k_x}(u;\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}d\widehat{H}_{k_t,k_x}(u;\boldsymbol{\beta}) + \lambda\{\widehat{H}_{k_t,k_x}(u;\boldsymbol{\beta}) + X_{i,k_x}^*\beta_1 + \mathbf{Z}_i^T\boldsymbol{\beta}_2\} \\
& \times\frac{\partial d\widehat{H}_{k_t,k_x}(u;\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\Bigg]\Bigg)_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}_c, \boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}_{ML}} \\
& + \frac{1}{n}\sum_{i=1}^{n}\Bigg[\int\frac{1}{m_x^*}\sum_{k_x=1}^{m_x^*}\int_0^{\tau}\{\mathbf{Z}_{i,k_x}^* - \boldsymbol{\mu}_{z,k_t,k_x}(u;\boldsymbol{\beta})\}d\mathcal{M}_{i,k_t,k_x}(u;\boldsymbol{\beta}, H) \\
& \times\Delta_i\frac{\partial}{\partial\boldsymbol{\beta}}\log\{f(T_{i,k_t,k_x}|X_{i,k_x}^*, \mathbf{Z}_i, L_i < T_{i,k_t,k_x} \le R_i;\boldsymbol{\beta})\} \\
& \times f(T_{i,k_t,k_x}|X_{i,k_x}^*, \mathbf{Z}_i, L_i < T_{i,k_t,k_x} \le R_i;\boldsymbol{\beta})dT_{i,k_t,k_x}\Bigg]_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}_c, \boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}_{ML}},
\end{aligned}$$

$$\begin{aligned}
\widehat{D}_2 \;=\; & \frac{1}{n}\sum_{i=1}^{n}\Bigg[\frac{1}{m_t^*}\sum_{k_t=1}^{m_t^*}\int S_i(\boldsymbol{\beta}, V_{i,k_t,k_x}(\boldsymbol{\beta}), X_{i,k_x}^*)\frac{\partial}{\partial\boldsymbol{\theta}^T}\log\{f(X_{i,k_x}^*|L_i < T_i < R_i, W_{i,1}, \ldots, W_{i,m}, \mathbf{Z}_i;\boldsymbol{\theta})\} \\
& \times f(X_{i,k_x}^*|L_i < T_i < R_i, W_{i,1}, \ldots, W_{i,m}, \mathbf{Z}_i;\boldsymbol{\theta})dX_{i,k_x}^*\Bigg]_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}_c, \boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}_{ML}}.
\end{aligned}$$

For a sufficiently large n and flexible prior distributions on the parameters, the posterior mode and the maximum likelihood estimator of $\boldsymbol{\theta}$ will be very close, so we can replace $\widehat{\boldsymbol{\theta}}_{ML}$ by $\widehat{\boldsymbol{\theta}}_{MAP}$, the posterior mode of $\boldsymbol{\theta}$. Additionally, we shall use

$$\frac{\partial \widehat{H}_{k_t,k_x}(t,\boldsymbol{\beta})}{\partial\boldsymbol{\beta}} \;=\; -\frac{1}{\lambda^*\{\widehat{H}_{k_t,k_x}(t;\boldsymbol{\beta})\}}\int_0^t \lambda^*\{\widehat{H}_{k_t,k_x}(s;\boldsymbol{\beta})\}\frac{\sum_{i=1}^n \mathbf{Z}_i^*\dot{\lambda}\{\mathbf{Z}_i^{*T}\boldsymbol{\beta} + \widehat{H}_{k_t,k_x}(s;\boldsymbol{\beta})\}Y_{i,k_t,k_x}(s)}{\sum_{i=1}^n \lambda\{\mathbf{Z}_i^{*T}\boldsymbol{\beta} + \widehat{H}_{k_t,k_x}(s;\boldsymbol{\beta})\}Y_{i,k_t,k_x}(s)}d\widehat{H}_{k_t,k_x}(s;\boldsymbol{\beta}),$$

where $\lambda^*\{\widehat{H}_{k_t,k_x}(t;\boldsymbol{\beta})\} = B_{k_t,k_x}(t,a;\boldsymbol{\beta})$.

S.7 Additional simulation study

To study sensitivity of the proposed method we considered some additional scenarios. We considered an additional case by simulating X from a mixture of normal distributions and U from the modified gamma distribution with all the other settings of the simulation remaining the same as in the last case in the manuscript where we mimicked the AIDS clinical trial data set. The results are presented in Table S.1 for the $r = 0$ case. As in the manuscript, here also IM shows superior performance over the other methods.

Next we simulated data similar to the previous scenario except that X followed $\{\text{Gamma}(1,1) - 1\}$ and the measurement error U was allowed to depend on the time-to-event (differential measurement error) and the true covariate X. Specifically, U was simulated from the following mixture distribution: Normal($\mu = 0$, $\sigma = 0.5^{(0.5+\Delta)}$)$I(X < 0)$ + Uniform(-1, 1)$I(X > 0)$. The results given in Table S.2 indicate quite satisfactory performance of IM compared to NV and RC in terms of bias and coverage probability. Thus, all of these results indicate that IM is not sensitive towards a moderate violation of the model assumptions.

To address a comment from a reviewer, we included an additional simulation study to assess the sensitivity of IM towards the independence assumption of the measurement error U and the true covariate X. First, X was generated from Normal$(0, 1)$ and depending on whether X is less than 0 or greater than 0, U was generated from Normal$(0, 0.5^2)$ or Normal$(0, 0.71^2)$, respectively. Second, X was generated from a centered and scaled gamma distribution and depending on whether X is less than or greater than 0, U was generated from a modified gamma distribution with mean zero and variance 0.25 or 0.5 respectively. In this second scenario, three assumptions, independence of X and U, mixture normal distribution for X, and normal distribution for U, are violated simultaneously. The results are given in Table S.3. In this situation where model assumptions are grossly violated, the bias for the estimator of β_1 under the IM method is somewhat larger than that the RC method. However, IM is still much better than NV in terms of bias. This result confirms that even for somewhat severe degree of model violations, IM still performs much better than NV.

S.8 Plots from the MCMC chains for the real data analysis

This supplementary material contains the trace plots and Gelman-Rubin diagnostic plots of the parameter estimates involved in the ACTG data analysis (see pages 10, 11, 12). Here $\gamma_{i,j}$ represents the jth element of the γ parameter involved in the ith component of the mixture normal model. For the ACTG data, since the BIC chose a two-component mixture normal model, we have $i = 1, 2$ and $j = 1, \ldots, 7$. We also report the Gelman-Rubin factor (F) for each of the parameters, and usually values close to 1 suggest convergence of the chain.

Although $\gamma_{1,7}$ has an F value of 1.37, the corresponding trace plot shows satisfactory convergence of the chain. Similarly, despite a slow mixing of the chain of $\sigma^2_{x,1}$ and $\sigma^2_{x,2}$ in their trace plots, the corresponding F values indicate satisfactory convergence.

References

Hartigan, J. A. (1983). *Bayes Theory*. Springer, New York.

Wang, N. and Robins, J. (1998). Large-sample theory for parametric multiple imputation procedures. *Biometrika*, **85**, 935–948.

Table S.1: Simulation results based on 1000 replications for $r = 0$ with $n = 500$, unequal-length intervals and 90% right censoring on average. Here X follows a mixture of normals and measurement error $U = \sigma_u U^*$, and U^* follows the modified gamma distribution. All entries are multiplied by 100. B ≡ bias, S ≡ standard deviation, E ≡ estimated standard error, C ≡ 95% coverage probability, MN ≡ Mixture Normal, NM ≡ No measurement error, NV ≡ Naive, RC ≡ Regression calibration, IM ≡ Imputation method.

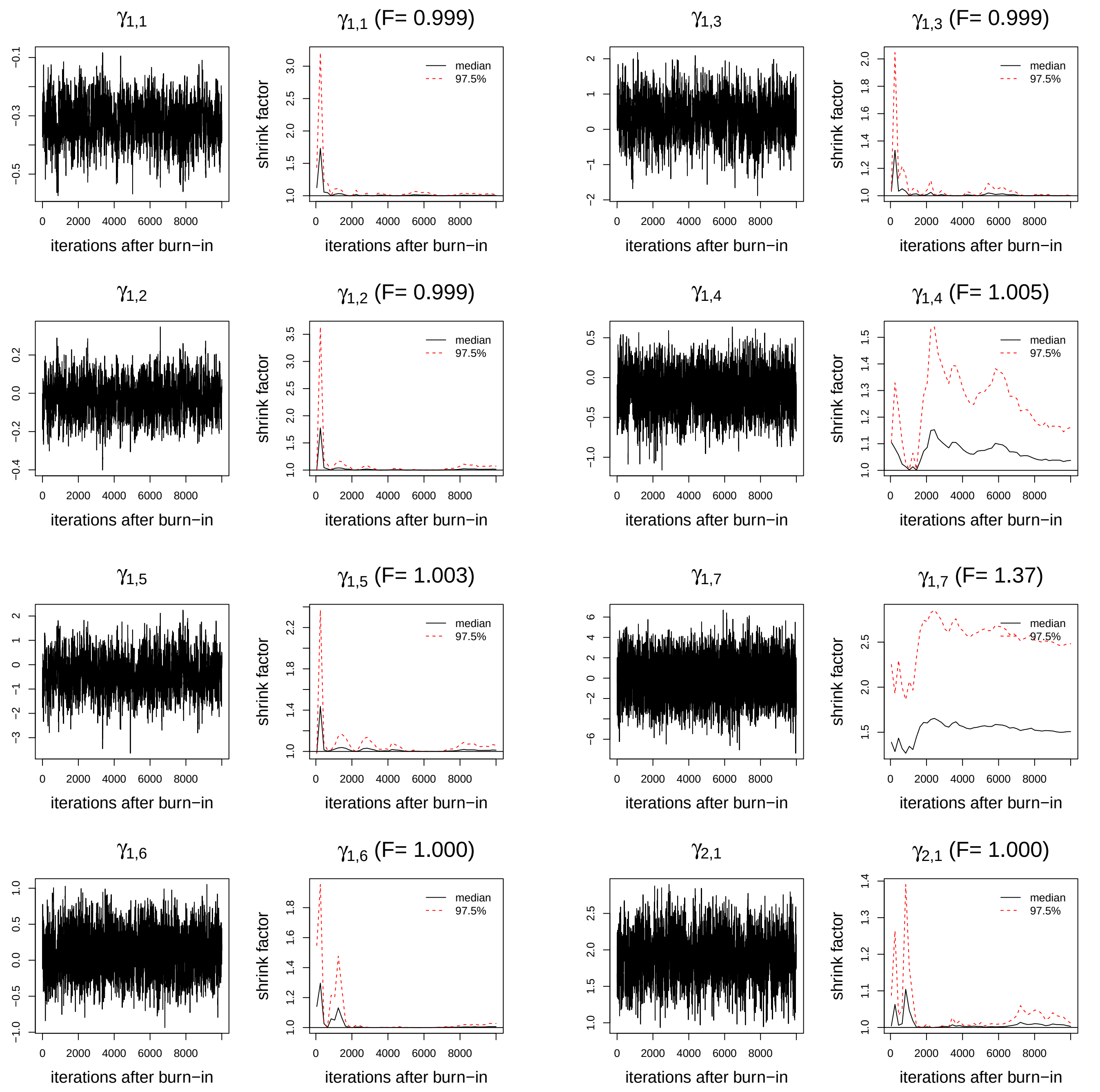
		β_1				β_2			
σ_u^2		NM	NV	RC	IM	NM	NV	RC	IM
0.25	B	−2.3	21.0	11.1	-4.8	3.7	3.2	3.1	3.7
	S	24.3	19.4	21.9	29.1	33.4	33.6	33.4	33.4
	E	24.1	20.0	22.7	29.3	32.1	32.1	34.3	32.5
	C	94.7	78.2	91.6	94.4	95.6	95.3	95.3	95.4
0.5	B	−2.3	33.9	17.3	−7.9	3.7	2.9	2.8	3.5
	S	24.3	17.1	21.5	33.9	33.4	33.6	33.4	33.5
	E	24.1	17.7	22.7	34.7	32.1	32.1	35.0	32.8
	C	94.7	49.4	86.2	95.1	95.6	95.1	95.3	95.5

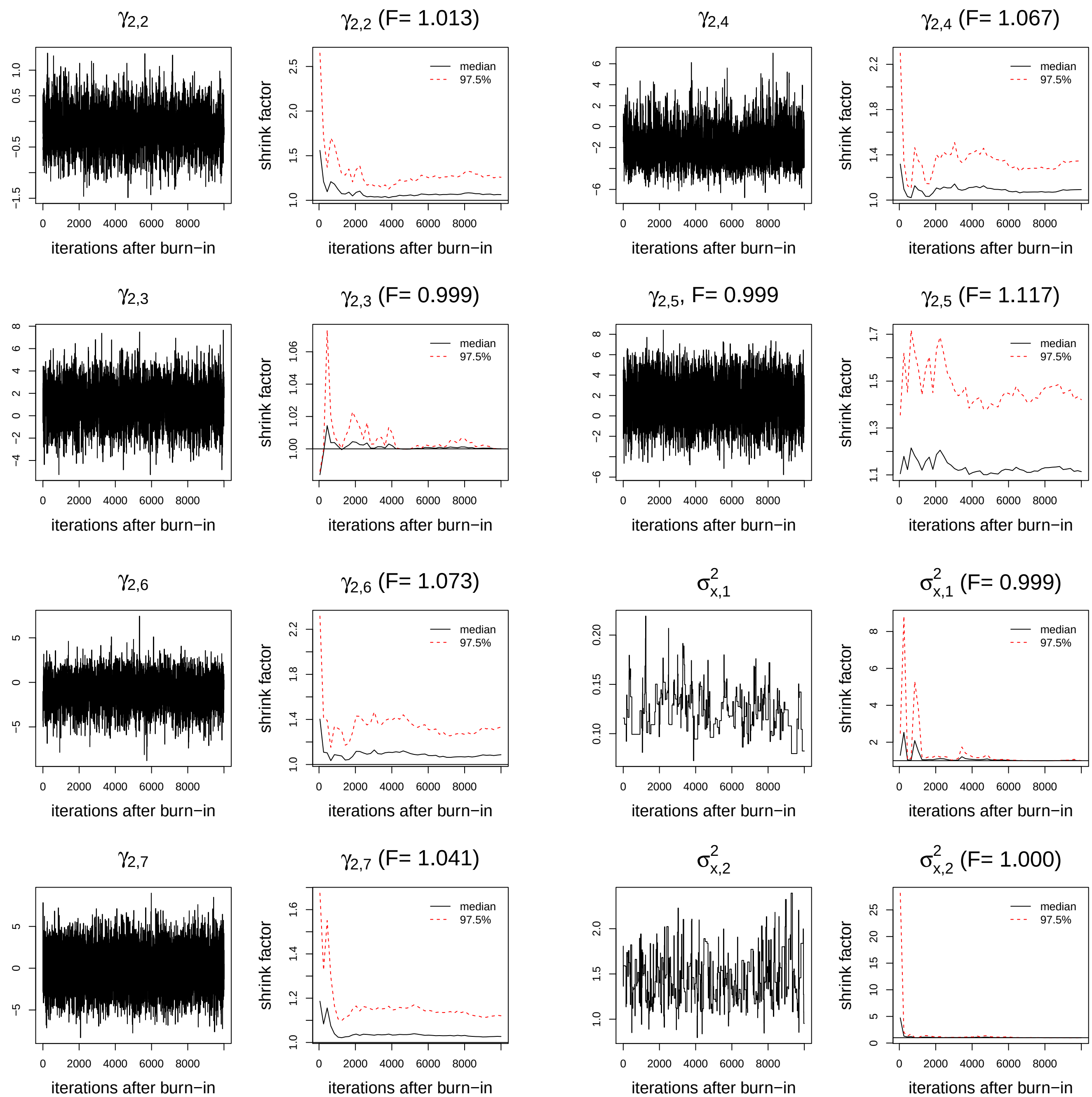
Table S.2: Simulation results based on 1000 replications for $r = 0$ and 1 with $n = 500$, unequal-length intervals and 90% right censoring on average. Here $X \sim \text{Gamma}(1, 1) - 1$, and U depends on X and the time-to-event outcome. All entries are multiplied by 100. B ≡ bias, S ≡ standard deviation, E ≡ estimated standard error, C ≡ 95% coverage probability, NM ≡ No measurement error, NV ≡ Naive, RC ≡ Regression calibration, IM ≡ Imputation method.

	$r = 0$								$r = 1$							
	β_1				β_2				β_1				β_2			
	NM	NV	RC	IM	NM	NV	RC	IM	NM	NV	RC	IM	NM	NV	RC	IM
B	−4.9	45.9	26.5	10.7	6.1	5.5	5.8	6.6	−4.7	37.1	24.1	7.8	5.2	4.3	4.4	5.8
S	30.7	14.6	17.7	25.1	34.3	35.2	34.9	35.0	29.3	15.0	18.2	26.3	32.6	33.1	33.0	33.3
E	28.8	17.3	22.1	29.8	32.7	32.8	33.1	33.2	28.1	18.4	22.2	30.6	32.4	32.3	32.6	33.0
C	94.0	22.6	78.9	93.7	95.0	94.5	94.9	95.1	94.8	46.6	81.7	95.2	96.4	95.7	95.6	96.0

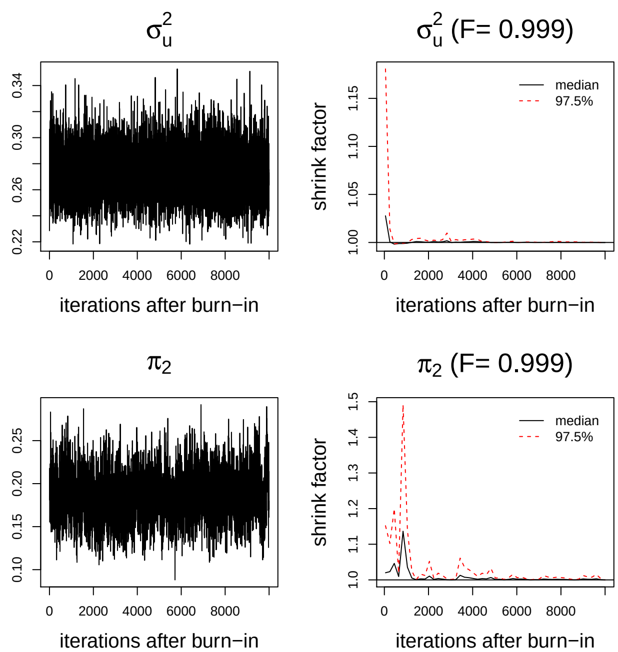
Table S.3: Simulation results based on 1000 replications for $r = 0$ with $n = 500$, unequal-length intervals and 90% right censoring on average. Here U depends on X. All entries are multiplied by 100. B ≡ bias, S ≡ standard deviation, E ≡ estimated standard error, C ≡ 95% coverage probability, NM ≡ No measurement error, NV ≡ Naive, RC ≡ Regression calibration, IM ≡ Imputation method.

	$X \sim N$, $U^*\vert X \sim N$								$X \sim MG$, $U^*\vert X \sim MG$							
	β_1				β_2				β_1				β_2			
	NM	NV	RC	IM	NM	NV	RC	IM	NM	NV	RC	IM	NM	NV	RC	IM
B	−1.3	12.5	−4.1	−6.9	3.4	2.0	2.0	5.4	−4.2	21.5	7.5	−11.3	5.4	5.0	4.8	5.5
S	14.8	14.3	17.2	30.3	30.9	30.8	30.8	34.2	25.5	20.1	23.7	32.5	34.2	34.7	34.3	34.3
E	14.7	13.8	16.6	30.5	30.5	30.4	30.5	32.7	24.2	19.7	23.8	32.2	32.2	32.2	33.6	32.8
C	94.6	82.2	93.9	95.3	95.7	95.7	95.9	94.3	93.4	76.4	92.5	95.0	94.3	93.4	94.4	94.7


γ1,1
γ1,1 (F= 0.999)
γ1,3
γ1,3 (F= 0.999)
γ1,2
γ1,2 (F= 0.999)
γ1,4
γ1,4 (F= 1.005)
γ1,5
γ1,5 (F= 1.003)
γ1,7
γ1,7 (F= 1.37)
γ1,6
γ1,6 (F= 1.000)
γ2,1
γ2,1 (F= 1.000)
shrink factor
median
97.5%
iterations after burn−in



$\gamma_{2,2}$
$\gamma_{2,2}$ (F= 1.013)
$\gamma_{2,4}$
$\gamma_{2,4}$ (F= 1.067)
$\gamma_{2,3}$
$\gamma_{2,3}$ (F= 0.999)
$\gamma_{2,5}$, F= 0.999
$\gamma_{2,5}$ (F= 1.117)
$\gamma_{2,6}$
$\gamma_{2,6}$ (F= 1.073)
$\sigma^2_{x,1}$
$\sigma^2_{x,1}$ (F= 0.999)
$\gamma_{2,7}$
$\gamma_{2,7}$ (F= 1.041)
$\sigma^2_{x,2}$
$\sigma^2_{x,2}$ (F= 1.000)
shrink factor
iterations after burn−in
median
97.5%


σ_u^2
σ_u^2 (F= 0.999)
median
97.5%
shrink factor
iterations after burn−in
iterations after burn−in
π_2
π_2 (F= 0.999)
median
97.5%
shrink factor
iterations after burn−in
iterations after burn−in