Proportional hazards regression for the analysis of clustered survival data from case-cohort studies: Web Appendix

Hui Zhang*, Douglas E. Schaubel, and John D. Kalbfleisch

Department of Biostatistics, University of Michigan, Ann Arbor, Michigan 48109-2029, U.S.A.

*email: huizh@umich.edu
Supplementary Materials

A.1 Proof of Theorem 1

Evaluated at the true values, the estimating function is given by

\[ U(\beta_0, p_0) = \sum_{i=1}^{n} \sum_{j=1}^{m_i} \int_0^r \left\{ Z_{ij}(u) - \frac{S^{(1)}(\beta_0, p_0, u)}{S^{(0)}(\beta_0, p_0, u)} \right\} dN_{ij}(u). \]

By some simple algebra, we have

\[ n^{-1/2} U(\beta, p_0) = n^{-1/2} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \int_0^r \left\{ Z_{ij}(u) - e(\beta, u) \right\} dN_{ij}(u) \]

\[ -n^{-1/2} \int_0^r \left\{ \frac{S^{(1)}(\beta_0, p_0, u)}{S^{(0)}(\beta_0, p_0, u)} - e(\beta, u) \right\} dF(u) \]

\[ -n^{-1/2} \int_0^r \left\{ \frac{S^{(1)}(\beta_0, p_0, u)}{S^{(0)}(\beta_0, p_0, u)} - e(\beta, u) \right\} \{ dN(u) - dF(u) \}. \]

By a functional Taylor expansion of \( S^{(1)}(\beta_0, p_0, u)/S^{(0)}(\beta_0, p_0, u) \) with respect to \( S^{(1)}(\beta, p_0, u) \) and \( S^{(0)}(\beta, p_0, u) \) around \( \mu^{-1}s^{(1)}(\beta, u) \) and \( \mu^{-1}s^{(0)}(\beta, u) \), respectively, combined with Conditions (d), (e) and the fact that \( n^{1/2} \{ N(u) - F(u) \} \) converges in distribution to a zero-mean Gaussian process, \( n^{-1/2} U(\beta, p_0) \) can be written as

\[ n^{-1/2} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \int_0^r \left\{ Z_{ij}(u) - e(\beta, u) \right\} dN_{ij}(u) \]

\[ -n^{-1/2} \int_0^r \sum_{i=1}^{n} \sum_{j=1}^{m_i} \left\{ Z_{ij}(u) - e(\beta, u) \right\} Y_{ij}(u) e^{\beta^T Z_{ij}(u)} \left\{ \frac{p_0}{N_1} \delta_{ij} + \frac{1 - p_0}{n_0} (1 - \delta_{ij}) H_i H_{ij} \right\} \]

\[ \times \left\{ \frac{1}{\mu} s^{(0)}(\beta, u) \right\}^{-1} dF(u) + o_p(1) \]

\[ = n^{-1/2} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \int_0^r \left\{ Z_{ij}(u) - e(\beta, u) \right\} dN_{ij}(u) \]

\[ -n^{-1/2} \frac{p_0}{N_1/n} \int_0^r \sum_{i=1}^{n} \sum_{j=1}^{m_i} \left\{ Z_{ij}(u) - e(\beta, u) \right\} \delta_{ij} Y_{ij} e^{\beta^T Z_{ij}(u)} \left\{ \frac{1}{\mu} s^{(0)}(\beta, u) \right\}^{-1} dF(u) \]

\[ -n^{-1/2} \frac{1 - p_0}{n_0/n} \int_0^r \sum_{i=1}^{n} \sum_{j=1}^{m_i} \left\{ Z_{ij}(u) - e(\beta, u) \right\} (1 - \delta_{ij}) H_i H_{ij} Y_{ij} e^{\beta^T Z_{ij}(u)} \]

\[ \times \left\{ \frac{1}{\mu} s^{(0)}(\beta, u) \right\}^{-1} dF(u) + o_p(1) \]

following a parallel setting described by van der Vaart & Wellner (1996) (example 2.11.16)
on p.215). By another functional Taylor expansion, we get

\[
\frac{p_0}{N_1/n} = \frac{1}{\mu} - \frac{1}{\mu^2 p_0} \left( \frac{N_1}{n} - \mu p_0 \right) + o_p(1)
\]

\[
\frac{1 - p_0}{n_0/n} = \frac{1}{\mu^2 \theta} - \frac{1}{(\mu^2 \theta)^2} \left( \frac{n_0}{n} - \mu \gamma \theta (1 - p_0) \right) + o_p(1),
\]

such that (1) can be written as

\[
-n^{-1/2} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \int_0^\tau \left\{ \mathbf{Z}_{ij}(u) - \mathbf{e}(\mathbf{\beta}, u) \right\} \frac{\delta_{ij} Y_{ij}(u)}{\mu} e^{\mathbf{\beta}^T \mathbf{Z}_{ij}(u)} \left\{ \frac{1}{\mu} s^{(0)}(\mathbf{\beta}, u) \right\}^{-1} dF(u)
\]

\[
+ n^{-1/2} \left( \frac{N_1}{n} - \mu p_0 \right) \sum_{i=1}^{n} \sum_{j=1}^{m_i} \int_0^\tau \left\{ \mathbf{Z}_{ij}(u) - \mathbf{e}(\mathbf{\beta}, u) \right\} \frac{\delta_{ij} Y_{ij}(u)}{\mu^2 p_0} e^{\mathbf{\beta}^T \mathbf{Z}_{ij}(u)} \left\{ \frac{1}{\mu} s^{(0)}(\mathbf{\beta}, u) \right\}^{-1} dF(u) + o_p(1).
\]

It is easy to show that

\[
n^{-1/2} \left( \frac{N_1}{n} - \mu p_0 \right) = n^{-1/2} \sum_{i=1}^{n} \frac{\sum_{j=1}^{m_i} \delta_{ij} - \mu p_0}{n}
\]

\[
= n^{-1/2} \sum_{i=1}^{n} G_1(p_0),
\]

(3)

where \( G_1(p) \) is as defined in Theorem 1, and that

\[
\sum_{i=1}^{n} \sum_{j=1}^{m_i} \int_0^\tau \left\{ \mathbf{Z}_{ij}(u) - \mathbf{e}(\mathbf{\beta}, u) \right\} \frac{1}{\mu^2 p_0} \delta_{ij} Y_{ij}(u) e^{\mathbf{\beta}^T \mathbf{Z}_{ij}(u)} \left\{ \frac{1}{\mu} s^{(0)}(\mathbf{\beta}, u) \right\}^{-1} dF(u)
\]

\[
= \sum_{i=1}^{n} \sum_{j=1}^{m_i} \int_0^\tau \left\{ \mathbf{Z}_{ij}(u) - \mathbf{e}(\mathbf{\beta}, u) \right\} \frac{1}{\mu^2 p_0} \delta_{ij} Y_{ij}(u) e^{\mathbf{\beta}^T \mathbf{Z}_{ij}(u)} \left\{ \frac{1}{\mu} s^{(0)}(\mathbf{\beta}, u) \right\}^{-1} dN(u)
\]

\[
- \sum_{i=1}^{n} \sum_{j=1}^{m_i} \int_0^\tau \left\{ \mathbf{Z}_{ij}(u) - \mathbf{e}(\mathbf{\beta}, u) \right\} \frac{1}{\mu^2 p_0} \delta_{ij} Y_{ij}(u) e^{\mathbf{\beta}^T \mathbf{Z}_{ij}(u)} \left\{ \frac{1}{\mu} s^{(0)}(\mathbf{\beta}, u) \right\}^{-1} d(\overline{N}(u) - F(u))
\]

\[
= n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \int_0^\tau \left\{ \mathbf{Z}_{ij}(u) - \mathbf{e}(\mathbf{\beta}, u) \right\} \frac{1}{\mu^2 p_0} \delta_{ij} Y_{ij}(u) e^{\mathbf{\beta}^T \mathbf{Z}_{ij}(u)} \left\{ \frac{1}{\mu} s^{(0)}(\mathbf{\beta}, u) \right\}^{-1} \sum_{k=1}^{n} \sum_{l=1}^{m_i} dN_{kl}(u)
\]

\[
+ o_p(1).
\]

(4)

Combining (3), (4) and using the fact that (4) converges in probability to \( \mathbf{D}_1(\mathbf{\beta}) \), (1) can be written as

\[
-n^{-1/2} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \int_0^\tau \left\{ \mathbf{Z}_{ij}(u) - \mathbf{e}(\mathbf{\beta}, u) \right\} \frac{1}{\mu} \delta_{ij} Y_{ij}(u) e^{\mathbf{\beta}^T \mathbf{Z}_{ij}(u)} \left\{ \frac{1}{\mu} s^{(0)}(\mathbf{\beta}, u) \right\}^{-1} dF(u)
\]

\[
+ \mathbf{D}_1(\mathbf{\beta}) \times n^{-1/2} \sum_{i=1}^{n} G_1(p_0) + o_p(1).
\]
Similarly, we can show that (2) can be written as
\[ -n^{-1/2} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \int_0^\tau \{ Z_{ij}(u) - \mathbf{e}(\beta, u) \} \frac{1}{\mu \gamma \theta} (1 - \delta_{ij}) Y_{ij}(u) e^{\beta^T Z_{ij}(u)} \left\{ \frac{1}{\mu} s^{(0)}(\beta, u) \right\}^{-1} dF(u) \]

\[ + D_2(\beta) \times n^{-1/2} \sum_{i=1}^{n} G_{2i}(p_0) + o_p(1). \]

Therefore, it follows that
\[ n^{-1/2} \mathbf{U}(\beta, p_0) = n^{-1/2} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \int_0^\tau \{ Z_{ij}(u) - \mathbf{e}(\beta, u) \} dN_{ij}(u) \]

\[ -n^{-1/2} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \int_0^\tau \{ Z_{ij}(u) - \mathbf{e}(\beta, u) \} \left\{ \frac{1}{\mu} \delta_{ij} + \frac{1}{\mu \gamma \theta} (1 - \delta_{ij}) H_i H_{ij} \right\} Y_{ij}(u) e^{\beta^T Z_{ij}(u)} \]

\[ \times \left\{ \frac{1}{\mu} s^{(0)}(\beta, u) \right\}^{-1} dF(u) \]

\[ + D_1(\beta) \times n^{-1/2} \sum_{i=1}^{n} G_{1i}(p_0) + D_2(\beta) \times n^{-1/2} \sum_{i=1}^{n} G_{2i}(p_0) + o_p(1) \]

\[ = n^{-1/2} \sum_{i=1}^{n} W_i(\beta, p_0) + o_p(1), \]

with \( W_i(\beta, p) \) as defined in Theorem 1.

The quantity \( W_i(\beta_0, p_0) \) can be written as
\[ W_i(\beta_0, p_0) = \sum_{j=1}^{m_i} \int_0^\tau \{ Z_{ij}(u) - \mathbf{e}(\beta_0, u) \} dM_{ij}(u) \]

\[ + \sum_{j=1}^{m_i} \int_0^\tau \{ Z_{ij}(u) - \mathbf{e}(\beta_0, u) \} (1 - \delta_{ij}) (1 - \frac{1}{\gamma \theta} H_i H_{ij}) Y_{ij}(u) e^{\beta^T Z_{ij}(u)} \lambda_0(u) du \]

\[ + D_1(\beta_0) G_{1i}(p_0) + D_2(\beta_0) G_{2i}(p_0), \]

where \( M_{ij}(t) = N_{ij}(t) - \int_0^t Y_{ij}(u) e^{\beta^T Z_{ij}(u)} \lambda_0(u) du \) is a mean-zero process.

Note that \( \mathcal{E} \{ 1 - (\gamma \theta)^{-1} H_i H_{ij} \} = 0, \mathcal{E} \{ dM_{ij}(u) \} = 0, \mathcal{E} \{ G_{1i}(p_0) \} = 0 \) and \( \mathcal{E} \{ G_{2i}(p_0) \} = 0 \), such that \( \mathcal{E} \{ W_i(\beta_0, p_0) \} = 0, \) for \( i = 1, \ldots, n \). Hence under the assumed conditions, asymptotically, \( \{ W_i(\beta_0, p_0) \}_{i=1}^n \) are independent and identically distributed random quantities with mean zero and finite variance, \( \mathcal{E} \{ W_i(\beta_0, p_0) \} \equiv 2 \). By the Multivariate Central Limit Theorem (MCLT), \( n^{-1/2} \mathbf{U}(\beta_0, p_0) \xrightarrow{D} \mathcal{N} \left( 0, \Sigma(\beta_0, p_0) \right) \), where \( \Sigma(\beta_0, p_0) \) is defined in Theorem 1.

A.2 Proof of Theorem 2
To prove the consistency of $\hat{\beta}_t$, we use the Inverse Function Theorem (Foutz, 1977) by verifying the following conditions:

(i) $\partial U(\beta, p_0) / \partial \beta^T$ exists and is continuous in an open neighborhood $\mathcal{B}$ of $\beta_0$.

(ii) $-n^{-1} \partial U(\beta, p_0) / \partial \beta^T |_{\beta = \beta_0}$ is positive definite with probability 1 as $n \to \infty$.

(iii) $-n^{-1} \partial U(\beta, p_0) / \partial \beta^T$ converges in probability to a fixed function, $A(\beta)$, uniformly in an open neighborhood $\mathcal{B}$ of $\beta_0$.

(iv) Asymptotic unbiasedness of the estimating function: $-n^{-1} U(\beta_0, p_0) \xrightarrow{P} 0$.

Conditions (i), (ii) and (iii) follow from Conditions (d), (e), (f) and (g). Using the result in the proof of Theorem 1, $n^{-1} U(\beta_0, p_0) \xrightarrow{P} 0$ by Chebyshev’s inequality. Then, Condition (iv) holds under the assumed model. Having now verified conditions (i) to (iv), we conclude that $\hat{\beta}_t$ converges in probability to $\beta_0$.

A.3 Proof of Theorem 3

Here we prove results for $\hat{\beta}_s$ only, since results for $\hat{\beta}_w$ can be proved similarly. By a Taylor expansion of the score function $U(\hat{\beta}_s, \hat{p}_s)$ with respect to $\beta$ and around $\beta_0$, and by a Taylor expansion of $U(\beta_0, \hat{p}_s)$ with respect to $p$ around $p_0$,

$$n^{-1/2} U(\hat{\beta}_s, \hat{p}_s) = n^{-1/2} U(\beta_0, \hat{p}_s) - \hat{A}(\beta_s, \hat{p}_s) n^{1/2} (\hat{\beta}_s - \beta_0)$$

$$n^{-1/2} U(\beta_0, \hat{p}_s) = n^{-1/2} U(\beta_0, p_0) + \hat{B}(\beta_0, p_s) n^{1/2} (\hat{p}_s - p_0),$$

where $\beta_s$ is on the line segment between $\hat{\beta}_s$ and $\beta_0$, $p_s$ is on the line segment between $\hat{p}_s$ and $p_0$, and

$$\hat{A}(\beta, p) = -n^{-1} \frac{\partial}{\partial \beta} U(\beta, p)$$

$$= n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \int_0^\tau \left[ \frac{S(2)(\beta, p, u)}{S(0)(\beta, p, u)} - \left( \frac{S(1)(\beta, p, u)}{S(0)(\beta, p, u)} \right)^2 \right] dN_{ij}(u)$$
\[
\hat{B}(\beta, p) = n^{-1} \frac{\partial}{\partial p} U(\beta, p) = n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \left\{ \frac{S^{(1)}(\beta, p, u)}{S^{(0)}(\beta, p, u)^2} \frac{\partial}{\partial p} S^{(0)}(\beta, p, u) \right\} dN_{ij}(u).
\]

Since \( \hat{\beta}_s \xrightarrow{P} \beta_0 \) and \( \| \beta_s - \beta_0 \| \leq \| \hat{\beta}_s - \beta_0 \|, \beta_s \xrightarrow{P} \beta_0 \). Using the fact that \( \hat{p}_s \xrightarrow{P} p_0 \), Condition (c) and continuity,

\[
\hat{A}(\beta_s, \hat{p}_s) \xrightarrow{P} \int_0^\tau \left\{ \frac{s^{(2)}(\beta_0, u)}{s^{(0)}(\beta_0, u)} - e(\beta_0, u)^{\otimes 2} \right\} dF(u) = A(\beta_0).
\]

Since \( \hat{p}_s \xrightarrow{P} p_0 \) and \( \| p_s - p_0 \| \leq \| \hat{p}_s - p_0 \| \), we obtain that \( p_s \xrightarrow{P} p_0 \). We can express \( R^{(d)}(\beta, p, u) \) as follows,

\[
R^{(d)}(\beta, p, u) = \frac{\partial}{\partial p} S^{(d)}(\beta, p, u) = \sum_{i=1}^{n} \sum_{j=1}^{m_i} \left\{ \frac{1}{N_1} \delta_{ij} - \frac{1}{n_0} (1 - \delta_{ij}) H_i H_j \right\} Y_{ij}(u) e^{\beta^T Z_{ij}(u)} Z_{ij}(u)^{\otimes d}
\]

\[
= \frac{1}{N_1 n} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \delta_{ij} Y_{ij}(u) e^{\beta^T Z_{ij}(u)} Z_{ij}(u)^{\otimes d} - \frac{1}{n_0 / n} \sum_{i=1}^{n} \sum_{j=1}^{m_i} (1 - \delta_{ij}) H_i H_j Y_{ij}(u) e^{\beta^T Z_{ij}(u)} Z_{ij}(u)^{\otimes d},
\]

such that

\[
R^{(d)}(\beta_0, p_0, u) \xrightarrow{P} \frac{1}{\mu p_0} \times \mu \times \mathcal{E} \left\{ \delta_{11} Y_{11}(u) e^{\beta^T z_{11}(u)} Z_{11}(u)^{\otimes d} \right\}
\]

\[
- \frac{1}{\mu \gamma \theta (1 - p_0)} \times \mu \gamma \theta \times \mathcal{E} \left\{ (1 - \delta_{11}) Y_{11}(u) e^{\beta^T z_{11}(u)} Z_{11}(u)^{\otimes d} \right\}
\]

\[
= \frac{1}{p_0} \mathcal{E} \left\{ \delta_{11} Y_{11}(u) e^{\beta^T z_{11}(u)} Z_{11}(u)^{\otimes d} \right\}
\]

\[
- \frac{1}{1 - p_0} \mathcal{E} \left\{ (1 - \delta_{11}) Y_{11}(u) e^{\beta^T z_{11}(u)} Z_{11}(u)^{\otimes d} \right\}
\]

\[
= r^{(d)}(\beta_0, u).
\]
Then, by continuous mapping,

\[
\hat{B}^\ast(B_0, p) \xrightarrow{P} \int_0^\tau \left\{ \frac{s^{(1)}(B_0, u)}{s^{(0)}(B_0, u)} r^{(0)}(B_0, u) - \frac{1}{s^{(0)}(B_0, u)} r^{(1)}(B_0, u) \right\} dF(u)
\]

\[\equiv B(B_0).\]

Using the fact that

\[
\hat{p} - p_0 = \frac{\sum_{i=1}^n \sum_{j=1}^{m_i} H_i H_{ij} \delta_{ij}}{\sum_{i=1}^n \sum_{j=1}^{m_i} H_i H_{ij}} - p_0
\]

it follows that

\[
n^{1/2}(\hat{p} - p_0) = n^{-1/2} \sum_{i=1}^n \sum_{j=1}^{m_i} \frac{H_i H_{ij} (\delta_{ij} - p_0)}{\mu \gamma \theta} + o_p(1)
\]

\[
= n^{-1/2} \sum_{i=1}^n \left\{ \frac{1}{\mu \gamma \theta} \sum_{j=1}^{m_i} H_i H_{ij} (\delta_{ij} - p_0) \right\} + o_p(1)
\]

\[
= n^{-1/2} \sum_{i=1}^n Q_i(p_0) + o_p(1).
\]

Note that \( E \{ H_i H_{ij} (\delta_{ij} - p_0) \} = 0 \), such that \( E \{ Q_i(p_0) \} = 0 \). Therefore,

\[
n^{-1/2} U(B_0, \hat{p}) = n^{-1/2} \sum_{i=1}^n \{ W_i(B_0, p_0) + B(B_0) Q_i(p_0) \} + o_p(1)
\]

\[
= n^{-1/2} \sum_{i=1}^n \psi_i(B_0, p_0) + o_p(1),
\]

where \( \psi_i(B, p) \) is as defined in Theorem 3.

Since \( E \{ \psi_i(B_0, p_0) \} = 0 \), by the MCLT,

\[
n^{-1/2} U(B_0, \hat{p}) \xrightarrow{D} N(0, \Omega(B_0)),
\]

where \( \Omega(B_0) = E \{ \psi_i(B_0, p_0)^\otimes 2 \} \). We then have

\[
n^{1/2}(\hat{\beta} - \beta_0) = \hat{A}(\hat{\beta}, \hat{p})^{-1} \times n^{-1/2} U(B_0, \hat{p}),
\]
since $U(\hat{\beta}_s, \hat{\rho}_s) = 0$. Note that $\hat{\Sigma}(\beta, \hat{\rho}) \xrightarrow{D} \Sigma(\beta)$. Therefore by Slutsky’s Theorem, $n^{1/2}(\hat{\beta}_s - \beta_0) \xrightarrow{D} N(0, \Sigma(\beta_0)^{-1} \Omega(\beta_0) \Sigma(\beta_0)^{-1})$, completing the proof.

A.4 Covariance Matrix Estimators

We now describe the consistent estimates of the covariance matrices in Theorems 2 and 3. Let $\hat{\gamma} = n^{-1} \sum_{i=1}^n H_i$, and $\hat{\theta} = \sum_{i=1}^n \sum_{j=1}^{m_i} H_{ij} / \sum_{i=1}^n H_{ii}$. The covariance matrices $A(\beta_0)^{-1} \Sigma(\beta_0, p_0) A(\beta_0)^{-1}$ and $A(\beta_0)^{-1} \Omega(\beta_0, p_0) A(\beta_0)^{-1}$ can be consistently estimated by $\hat{A}(\hat{\beta}_t, \hat{\rho}_0)^{-1} \hat{\Sigma}(\hat{\beta}_t, \hat{\rho}_0) \hat{A}(\hat{\beta}_t, \hat{\rho}_0)^{-1}$ and $\hat{A}(\hat{\beta}_a, \hat{\rho}_a)^{-1} \hat{\Omega}(\hat{\beta}_a, \hat{\rho}_a) \hat{A}(\hat{\beta}_a, \hat{\rho}_a)^{-1}$, respectively, where $\hat{\Sigma}(\hat{\beta}_t, \hat{\rho}_0) = n^{-1} \sum_{i=1}^n \hat{W}_i(\hat{\beta}_i, \hat{\rho}_0)$, $\hat{\Omega}(\hat{\beta}_a, \hat{\rho}_a) = n^{-1} \sum_{i=1}^n \hat{\psi}_i^p(\hat{\beta}_i, \hat{\rho}_a)$, $\hat{\psi}_i^a(\hat{\beta}_i, \hat{\rho}_a) = \hat{W}_i(\hat{\beta}_a, \hat{\rho}_a)$, and

\[
\hat{W}_{ij}(\beta, p) = \left\{ Z_{ij}(X_{ij}) - \hat{E}(\beta, p, X_{ij}) \right\} \delta_{ij} - n^{-1} \sum_{k=1}^n \sum_{l=1}^n \left\{ \frac{1}{\hat{\mu}} \delta_{ij} + \frac{1}{\hat{\mu} \hat{\gamma} \hat{\theta}} (1 - \delta_{ij}) H_{ij} \right\} Y_{ij}(X_{kl}) e^{\beta^T Z_{ij}(X_{kl})} \times S^{(0)}(\beta, p, X_{ij})^{-1} \left\{ Z_{ij}(X_{kl}) - \hat{E}(\beta, p, X_{ij}) \right\} \delta_{kl} + \hat{D}_1(\beta) G_{1i}(p) + \hat{D}_2(\beta) G_{2i}(p)
\]

\[
\hat{D}_1(\beta) = n^{-1} \sum_{i=1}^n \sum_{j=1}^{m_i} \sum_{k=1}^{m_k} \sum_{l=1}^{m_l} \left\{ Z_{ij}(X_{kl}) - \hat{E}(\beta, p, X_{ij}) \right\} \frac{\delta_{ij}}{\hat{\mu}^2 p} Y_{ij}(X_{kl}) e^{\beta^T Z_{ij}(X_{kl})} \times S^{(0)}(\beta, p, X_{ij})^{-1} \delta_{kl} \]

\[
\hat{D}_2(\beta) = n^{-1} \sum_{i=1}^n \sum_{j=1}^{m_i} \sum_{k=1}^{m_k} \sum_{l=1}^{m_l} \left\{ Z_{ij}(X_{kl}) - \hat{E}(\beta, p, X_{ij}) \right\} \frac{1}{\hat{\mu} \hat{\gamma} \hat{\theta}} (1 - \delta_{ij}) H_{ij} Y_{ij}(X_{kl}) \times e^{\beta^T Z_{ij}(X_{kl})} \left\{ \frac{1}{\hat{\mu}} S^{(0)}(\beta, p, X_{ij}) \right\}^{-1} \delta_{kl},
\]
and $\mathbf{R}^{(d)}(\boldsymbol{\beta}, p, u)$ is as defined in (5).

**A.5 Proof of Theorem 4**

We can decompose $\alpha_n(t) = \hat{\Lambda}_0(\hat{\boldsymbol{\beta}}, \hat{p}, t) - \Lambda_0(t)$ into three parts, $\alpha_n(t) = \alpha_{1,n}(t) + \alpha_{2,n}(t) + \alpha_{3,n}(t)$, where

\[
\begin{align*}
\alpha_{1,n}(t) &= \hat{\Lambda}_0(\hat{\boldsymbol{\beta}}, \hat{p}, t) - \hat{\Lambda}_0(\boldsymbol{\beta}_0, p_0, t) \\
\alpha_{2,n}(t) &= \hat{\Lambda}_0(\boldsymbol{\beta}_0, p_0, t) - \hat{\Lambda}_0(\boldsymbol{\beta}_0, p_0, t) \\
\alpha_{3,n}(t) &= \hat{\Lambda}_0(\boldsymbol{\beta}_0, p_0, t) - \Lambda_0(t).
\end{align*}
\]

(6)

Taking a Taylor expansion of $\alpha_{1,n}(t)$,

\[
\alpha_{1,n}(t) = \left. \frac{\partial \hat{\Lambda}_0(\hat{\boldsymbol{\beta}}, p, t)}{\partial p} \right|_{p=p_0} \times (\hat{p} - p_0) = -\int_0^t \frac{1}{\mu\overline{S}^{(0)}(\hat{\beta}, p, u)^2} \left. \frac{\partial \overline{S}^{(0)}(\beta, p, u)d\overline{N}(u)}{\partial p} \right|_{p=p_0} \times (\hat{p} - p_0),
\]

where $p_0$ lies between $\hat{p}$ and $p_0$, and $\mathbf{R}^{(0)}(\boldsymbol{\beta}, p, u)$ is as defined in (5). Under assumptions (a)-(g), $\overline{S}^{(0)}(\boldsymbol{\beta}, p, u)$, $\mathbf{R}^{(0)}(\boldsymbol{\beta}, p, u)$ and $\overline{N}(u)$ are all bounded and $\overline{S}^{(0)}(\beta, p, u)$ is bounded away from 0. Using the fact that $\hat{p}$ converges in probability to $p_0$ implies that $\alpha_{1,n}(t) \xrightarrow{P} 0$.

With respect to the second term of (6), applying a Taylor expansion,

\[
\alpha_{2,n}(t) = (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^T \left. \frac{\partial \hat{\Lambda}_0(\beta, p_0, t)}{\partial \beta} \right|_{\beta=\hat{\beta}} = -(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^T \int_0^t \left. \frac{\overline{S}^{(1)}(\beta, p_0, u)}{\mu\overline{S}^{(0)}(\beta, p_0, u)^2} \right|_{\beta=\hat{\beta}} \times \overline{N}(u),
\]

where $\hat{\beta}$ lies between $\hat{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}_0$. Since $\overline{E}(\beta, p_0, u)$ and $\overline{N}(u)$ are bounded, $\overline{S}^{(0)}(\beta, p_0, u)$ is bounded away from 0, and $\hat{\beta} \xrightarrow{P} \beta_0$, it follows that $\alpha_{2,n}(t) \xrightarrow{P} 0$. 
Now, considering the last term in (6),

\[
\alpha_{3,n}(t) = \int_0^t \frac{dN(u)}{\mu S^{(0)}(\beta_0, p_0, u)} - \int_0^t \lambda_0(u) du
\]

\[
= \int_0^t \frac{dN(u)}{\mu S^{(0)}(\beta_0, p_0, u)} - \int_0^t \frac{n^{-1} \sum_{i=1}^n \sum_{j=1}^{m_i} Y_{ij}(u) e^{\beta_0^T Z_{ij}(u)} \lambda_0(u)}{S^{(0)}(\beta_0, u)} du
\]

\[
= \int_0^t \frac{dN(u)}{\mu S^{(0)}(\beta_0, p_0, u)} - \int_0^t \frac{1}{S^{(0)}(\beta_0, u)} n^{-1} \sum_{i=1}^n \sum_{j=1}^{m_i} \{dN_{ij}(u) - dM_{ij}(u)\}
\]

\[
= \int_0^t \left\{ \frac{1}{\mu S^{(0)}(\beta_0, p_0, u)} - \frac{1}{S^{(0)}(\beta_0, u)} \right\} dN(u) + \int_0^t \frac{1}{S^{(0)}(\beta_0, u)} d\tilde{M}(u)
\]

\[
= \int_0^t \left\{ \frac{1}{\mu S^{(0)}(\beta_0, p_0, u)} - \frac{1}{S^{(0)}(\beta_0, u)} \right\} dF(u) + \int_0^t \frac{1}{S^{(0)}(\beta_0, u)} d\tilde{M}(u)
\]

Since \( \left\{ \frac{1}{\mu S^{(0)}(\beta_0, p_0, u)} \right\} - S^{(0)}(\beta_0, u)^{-1} \xrightarrow{p} 0 \), with \( F(u) \) bounded, \( S^{(0)}(\beta_0, u) \xrightarrow{P} s^{(0)}(\beta_0, u) \), which is bounded away from 0, and since \( \int_0^t d\tilde{M}(u) = n^{-1} \sum_{i=1}^n \sum_{j=1}^{m_i} \int_0^t dM_{ij}(u) \xrightarrow{P} 0 \) for \( t \in [0, \tau] \), it follows that \( \alpha_{3,n}(t) \xrightarrow{P} 0 \). Combining results for \( \alpha_{1,n}(t) \), \( \alpha_{2,n}(t) \) and \( \alpha_{3,n}(t) \), it follows that \( \tilde{\Lambda}_0(\hat{\beta}, \hat{p}, t) \xrightarrow{P} \Lambda_0(t) \).

With respect to convergence to a Gaussian process, note that, by the consistency of \( \hat{\beta}, p_\ast \) and Lemma 1 in the Appendix in Lin et al. (2000), \(- \int_0^t R^{(0)}(\hat{\beta}, p_\ast, u) \left\{ \mu S^{(0)}(\hat{\beta}, p_\ast, u)^2 \right\}^{-1} d\bar{N}(u) \xrightarrow{P} k(\beta_0, p_0, t)\), where \( k(\beta_0, p_0, t) = - \int_0^t \mu r^{(0)}(\beta_0, p_0, u) s^{(0)}(\beta_0, u) d\Lambda_0(u) \). It then follows that

\[
n^{1/2} \alpha_{1,n} = k(\beta_0, p_0, t) n^{1/2}(\hat{p} - p_0) + o_p(1)
\]

\[
= n^{-1/2} \sum_{i=1}^n k(\beta_0, p_0, t) Q_i(p_0) + o_p(1).
\]

Similarly, \(- \int_0^t \bar{E}(\beta_0, p_0, u) \left\{ \mu S^{(0)}(\beta_0, p_0, u) \right\}^{-1} d\bar{N}(u) \xrightarrow{P} h(\beta_0, p_0, t)\), where \( h(\beta_0, p_0, t) = \)
\[-\int_0^t e(\mathbf{\beta}_0, u)/s^{(0)}(\mathbf{\beta}_0, u) dF(u).\] We then have that

\[n^{1/2} \alpha_{3:n} = h^T(\mathbf{\beta}_0, p_0, t) n^{1/2}(\hat{\mathbf{\beta}} - \mathbf{\beta}_0) + o_p(1)\]

\[= h^T(\mathbf{\beta}_0, p_0, t) A(\mathbf{\beta}_0, \hat{\rho})^{-1} n^{-1/2} U(\mathbf{\beta}_0, \hat{\rho}) + o_p(1)\]

\[= n^{-1/2} \sum_{i=1}^n h^T(\mathbf{\beta}_0, p_0, t) A(\mathbf{\beta}_0)^{-1} \psi_i(\mathbf{\beta}_0, p_0) + o_p(1).\]

Considering \(n^{1/2} \alpha_{3:n},\)

\[n^{1/2} \alpha_{3:n} = n^{1/2} \int_0^t \left\{ \frac{1}{\mu S^{(0)}(\mathbf{\beta}_0, p_0, u)} - \frac{1}{S^{(0)}(\mathbf{\beta}_0, u)} \right\} dF(u)\]

\[+ n^{1/2} \int_0^t \frac{1}{S^{(0)}(\mathbf{\beta}_0, u)} d\overline{M}(u) + o_p(1)\]

\[= n^{1/2} \int_0^t \left\{ \frac{1}{\mu S^{(0)}(\mathbf{\beta}_0, p_0, u)} - \frac{1}{S^{(0)}(\mathbf{\beta}_0, u)} \right\} dF(u)\]

\[+ n^{-1/2} \sum_{i=1}^n \sum_{j=1}^{m_i} \int_0^t \frac{1}{S^{(0)}(\mathbf{\beta}_0, u)} dM_{ij}(u) + o_p(1).\]

Applying Taylor expansions of \(\left\{ \mu S^{(0)}(\mathbf{\beta}_0, p_0, u) \right\}^{-1}\) and \(S^{(0)}(\mathbf{\beta}_0, u)^{-1},\)

\[\frac{1}{\mu S^{(0)}(\mathbf{\beta}_0, p_0, u)} - \frac{1}{S^{(0)}(\mathbf{\beta}_0, u)}\]

\[= \left\{ \frac{1}{\mu S^{(0)}(\mathbf{\beta}_0, p_0, u)} - \frac{S^{(0)}(\mathbf{\beta}_0, p_0, u) - \mu^{-1}s^{(0)}(\mathbf{\beta}_0, u)}{\mu(\mu^{-1}s^{(0)}(\mathbf{\beta}_0, u))^2} \right\}\]

\[+ o_p(1)\]

\[= \frac{\mu^{-1}s^{(0)}(\mathbf{\beta}_0, u) - S^{(0)}(\mathbf{\beta}_0, p_0, u)}{\mu^{-1}s^{(0)}(\mathbf{\beta}_0, u)^2} + o_p(1)\]

\[= \frac{1}{\mu^{-1}s^{(0)}(\mathbf{\beta}_0, u)^2} \times n^{-1} \sum_{i=1}^n \sum_{j=1}^{m_i} \left\{ \frac{1}{\mu} - \frac{p_0}{N_1/n} \delta_{ij} - \frac{1 - p_0}{n_0/n} (1 - \delta_{ij}) H_i H_j \right\} Y_{ij} e^{\beta_0^T Z_{ij}(u)}\]

\[+ o_p(1)\]

\[= \frac{1}{s^{(0)}(\mathbf{\beta}_0, u)^2} \times n^{-1} \sum_{i=1}^n \sum_{j=1}^{m_i} \left\{ 1 - \delta_{ij} - \frac{1}{\gamma \theta} (1 - \delta_{ij}) H_i H_j \right\} Y_{ij} e^{\beta_0^T Z_{ij}(u)} + o_p(1).\]

It then follows that

\[n^{1/2} \alpha_{3:n} = n^{-1/2} \sum_{i=1}^n \partial_i(\mathbf{\beta}_0, p_0, t) + o_p(1),\]
where

\[ \vartheta_i(\beta_0, p_0, t) = \sum_{j=1}^{m_i} \int_0^t \frac{1}{s_i^{(0)}(\beta_0, u)} dM_{ij}(u) \]

\[ + \sum_{j=1}^{m_i} \int_0^t \frac{1}{s_i^{(0)}(\beta_0, u)^2} \left\{ 1 - \delta_{ij} - \frac{1}{\gamma \theta} (1 - \delta_{ij}) H_i H_{ij} \right\} Y_{ij} e^{\beta_0 Z_i(u)} dF(u). \]

Combining the above results, one obtains \( n^{1/2} \left\{ \hat{\Lambda}_0(\hat{\beta}, \hat{\nu}, t) - \Lambda_0(t) \right\} = n^{-1/2} \sum_{i=1}^{n} \phi_i(\beta_0, p_0, t) + o_p(1) \), where \( \phi_i(\beta_0, p_0, t) = k(\beta_0, p_0, t) Q_i(p_0) + h^T(\beta_0, p_0, t) A(\beta_0) \psi_i(\beta_0, p_0) + \vartheta_i(\beta_0, p_0, t). \)

It then follows from the MCLT that \( n^{1/2} \left\{ \hat{\Lambda}_0(\hat{\beta}, \hat{\nu}, t) - \Lambda_0(t) \right\} \) converges to a multivariate normal with mean zero and covariance function at \((s, t)\) given by \( \mathcal{E} \left\{ \phi_1(\beta_0, p_0, s) \phi_1(\beta_0, p_0, t) \right\}. \)

Using similar arguments to Spiekerman et al. (1998), tightness can be verified. Therefore, by the Functional Central Limit Theorem (Pollard, 1990), \( n^{1/2} \left\{ \hat{\Lambda}_0(\hat{\beta}, \hat{\nu}, t) - \Lambda_0(t) \right\} \) converges to a Gaussian process with mean zero and covariance function at \((s, t)\) given by \( \mathcal{E} \left\{ \phi_1(\beta_0, p_0, s) \phi_1(\beta_0, p_0, t) \right\}. \)

**A.6 Derivation of Equation (3)**

Let \( \tilde{Z}(u) = \{Z(s) : 0 < s \leq u\} \). Equation (3) of the article can be derived as follows:

\[ \mathcal{E} \left\{ Z(u) | X = u, \delta = 1 \right\} = \int z(u) \frac{\int_{X = u, \delta = 1} f_{\tilde{Z}(u)|\tilde{Z}(u)} dF_{\tilde{Z}(u)}}{\int_{X = u, \delta = 1} f_{\tilde{Z}(u)|\tilde{Z}(u)} dF_{\tilde{Z}(u)}} dF_{\tilde{Z}(u)} \]

\[ = \frac{\int z(u) \lambda_0(u) e^{\beta T \tilde{Z}(u)} P(T \geq u | \tilde{Z}(u)) P(C \geq u | \tilde{Z}(u)) dF_{\tilde{Z}(u)}}{\int \lambda_0(u) e^{\beta T \tilde{Z}(u)} P(T \geq u | \tilde{Z}(u)) P(C \geq u | \tilde{Z}(u)) dF_{\tilde{Z}(u)}} \]

\[ = \frac{\int z(u) e^{\beta T \tilde{Z}(u)} \mathcal{E} \left\{ Y(u) | \tilde{Z}(u) \right\} dF_{\tilde{Z}(u)}}{\int e^{\beta T \tilde{Z}(u)} \mathcal{E} \left\{ Y(u) | \tilde{Z}(u) \right\} dF_{\tilde{Z}(u)}} \]

\[ = \mathcal{E} \left\{ Y(u) Z(u) e^{\beta T \tilde{Z}(u)} \right\}. \]

**A.7 Extension of proposed methods to a stratified model**

Let \( V_{ij} \) denote the stratum for subject \((i, j)\) and set \( V_{ijk} = I \{ V_{ij} = k \}, k = 1, \ldots, K \), where there are \( K \) mutually exclusive strata. If subject \((i, j)\) is in the \( k \)th stratum, the marginal
hazard of failure is specified as

\[
\lambda_{ij}(t | V_{ijk} = 1) = \lambda_{0k}(t)e^{\beta_{ik}^T Z_{ij}(t)},
\]

where \(\lambda_{0k}(\cdot)\) is an unspecified stratum-specific baseline hazard function. Under model (7), let \(p_{0k} = Pr(\delta_{ij} = 1 | V_{ij} = k)\) for \(k = 1, \ldots, K\) and set \(p_0 = (p_{01}, \ldots, p_{0K})^T\). Let \(N_{ik} = \sum_{i=1}^{n} \sum_{j=1}^{m_i} V_{ijk}\delta_{ij}\) be the total number of failures in stratum \(k\) in the full cohort and let \(n_{0k} = \sum_{i=1}^{n} \sum_{j=1}^{m_i} V_{ijk}(1 - \delta_{ij})H_i H_{ij}\) be the total number of non-failures in stratum \(k\) in the subcohort. We assume that \(\{N_{ij}(\cdot), Y_{ij}(\cdot), Z_{ij}(\cdot), V_{ij}, m_i : j = 1, \ldots, m_i\}, i = 1, \ldots, n\) are independent and identically distributed, and for each \(k\), let \(m_{ik} = \sum_{j=1}^{m_i} V_{ijk}\), and \(E[m_{ik}] = \mu_k\).

The parameter \(\beta_0\) can be estimated by \(\tilde{\beta}\), the solution to \(\tilde{U}(\beta, p) = 0\), where

\[
\tilde{U}(\beta, p) = \sum_{k=1}^{K} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \int_0^\tau V_{ijk} \left\{ Z_{ij}(u) - \tilde{E}_k(\beta, p_k, u) \right\} dN_{ij}(u)
\]

with \(\tilde{E}_k(\beta, p_k, u) = S_k^{(1)}(\beta, p_k, u)/\tilde{\lambda}_k^{(0)}(\beta, p_k, u), \tilde{S}_k^{(d)}(\beta, p_k, u) = \sum_{i=1}^{n} \sum_{j=1}^{m_i} V_{ijk}[N_{1k}^{-1} p_k \delta_{ij} + n_{0k}^{-1}(1 - p_k)(1 - \delta_{ij})H_i H_{ij}]Y_{ij}(u)e^{\beta_{ik}^T Z_{ij}(u)} Z_{ij}(u)^{\otimes d}\) for \(d = 0, 1, 2\). We can estimate \(p_{0k}\) by the subcohort case percentage in stratum \(k\), \(\tilde{p}_{ks}\), or by the full cohort case percentage in stratum \(k\), \(\tilde{p}_{kw}\), or \(p_{0k}\) itself if it is known. Let \(\tilde{\beta}_s, \tilde{\beta}_w\) and \(\tilde{\beta}_t\) be the solutions of corresponding estimating equations, respectively. The cumulative baseline hazard function, \(\Lambda_{0k}(t) = \int_0^t \lambda_{0k}(u) du\), can be estimated by \(\tilde{\Lambda}_{0k}(t; \tilde{\beta}, \tilde{p}_k)\), where

\[
\tilde{\Lambda}_{0k}(t; \beta, p_k) = \int_0^t \frac{dN_k(u)}{\mu_k S_k^{(0)}(\beta, p_k, u)}.
\]

with \(N_k(u) = \sum_{i=1}^{n} \sum_{j=1}^{m_i} V_{ijk}N_{ij}(u)\).

To establish the asymptotic properties of \(\tilde{\beta}_t, \tilde{\beta}_s\) and \(\tilde{\beta}_w\), we need to modify Conditions (a), (e), (f) and (g) as follows:

(a') \(\{N_{ij}(\cdot), Y_{ij}(\cdot), Z_{ij}(\cdot), V_{ij}, m_i : j = 1, \ldots, m_i\}, i = 1, \ldots, n\) are independently and identically distributed.

(e') For each \(k\), \(\sup_{u \in [0, \tau], \beta \in B} \| \tilde{S}_k^{(d)}(\beta, p_k, u) - \mu_k^{-1} S_k^{(d)}(\beta, u) \| \overset{P}{\longrightarrow} 0\) for \(d = 0, 1, 2\), where \(S_k^{(d)}(\beta, u)\)
is an absolutely continuous function of $\beta \in \mathcal{B}$ and uniformly in $u \in (0, \tau]$, and $s_k^{(0)}(\beta, u)$ is bounded away from zero, $k = 1, \ldots, K$.

\[(f') \quad I(\beta_0) = \sum_{k=1}^{K} \int_0^\tau \left[ s_{k}^{(2)}(\beta_0, u)/s_{k}^{(0)}(\beta_0, u) - \left\{ s_{k}^{(1)}(\beta_0, u)/s_{k}^{(0)}(\beta_0, u) \right\}^2 \right] dF_k(u) \text{ is positive definite, where } F_k(u) = \mathcal{E} \{ \mathcal{N}_k(u) \}. \]

\[(g') \quad \Lambda_0(\tau) < \infty \text{ for each } k, \text{ and } \lambda_0(t) \text{ is absolutely continuous for } t \in (0, \tau]. \]

Conditions $(b')$, $(c')$, and $(d')$ are the same as $(b)$, $(c)$, and $(d)$ respectively.

**Theorem A.7.1**: Under conditions $(a') - (g')$, $\tilde{\beta}_t$ converges in probability to $\beta_0$, and $n^{1/2}(\tilde{\beta}_t - \beta_0)$ converges in distribution to a mean zero Normal with covariance matrix $I(\beta_0)^{-1}\Omega_t(\beta_0, p_0)I(\beta_0)^{-1}$, where $\Omega_t(\beta_0, p_0) = \mathcal{E} \{ W_t(\beta_0, p_0)^{\otimes 2} \}$, $W_t(\beta, p) = \sum_{k=1}^{K} W_{ik}(\beta, p_k)$, and $W_{ik}(\beta, p_k)$ is the same as $W_i(\beta, p)$ except that $W_{ik}(\beta, p_k)$ is calculated within stratum $k$.

**Theorem A.7.2**: Under conditions $(a') - (g')$, both $\tilde{\beta}_s$ and $\tilde{\beta}_w$ converge in probability to $\beta_0$, and each of $n^{1/2}(\tilde{\beta}_s - \beta_0)$ and $n^{1/2}(\tilde{\beta}_w - \beta_0)$ is asymptotically a zero-mean Normal with covariance matrix $I(\beta_0)^{-1}\Omega_s(\beta_0, p_0)I(\beta_0)^{-1}$ and $I(\beta_0)^{-1}\Omega_w(\beta_0, p_0)I(\beta_0)^{-1}$, respectively, where for $a = s$ and $w$, $\Omega_a(\beta_0, p_0) = \mathcal{E} \{ \varphi_a^2(\beta_0, p_0) \}$, $\varphi_a(\beta, p) = W_a(\beta, p) + \sum_{k=1}^{K} B_k(\beta)Q_{ik}(p)$, $Q_{ik}(p) = (\mu_k\gamma_0)^{-1}\sum_{j=1}^{m_k} V_{ij}H_iH_{ij}(\delta_{ij} - p_k$, $Q_{ik}(p) = \mu_k^{-1}\sum_{j=1}^{m_k} V_{ijk} \times(\delta_{ij} - p_k$, $B_k(\beta) = \int_0^\tau \left\{ s_{k}^{(1)}(\beta, u)r_{k}^{(0)}(\beta, u)/s_{k}^{(0)}(\beta, u)^2 - r_{k}^{(1)}(\beta, u)/s_{k}^{(0)}(\beta, u) \right\} dF_k(u)$, with

\[r_k^{(d)}(\beta, u) = p_k^{-1} \mathcal{E} \left\{ \delta_{ij} Y_{ij}(u)e^{\beta T}Z_{ij}(u) \otimes d \right| V_{ijk} = 1 \right\} - (1 - p_k)^{-1} \mathcal{E} \left\{ (1 - \delta_{ij}) Y_{ij}(u)e^{\beta T}Z_{ij}(u) \otimes d \right| V_{ijk} = 1 \right\}. \]

The proofs of Theorems A.7.1 and A.7.2 are very similar to those of Theorems 2 and 3, respectively. The asymptotic properties of $\tilde{\Lambda}_0(\beta, \tilde{p}_k, t)$ and the derivations thereof are analogous to those of $\tilde{\Lambda}_0(\beta, \tilde{p}, t)$.

**A.8 Additional simulation results**
Table 1 gives some results with a continuous covariate. The proposed methods appear to perform well.

[Table 1 about here.]

In the article, we considered $\alpha = 0.8$, which corresponds to Kendall’s $\tau$ of 0.2 for weak intracluster association. Here we conducted some simulation studies with $\alpha = 0.5$, which leads to Kendall’s $\tau$ of 0.5 for fairly strong intracluster association. (Table 2). The proposed methods still perform well, at least in the examples we considered.

[Table 2 about here.]

We conducted some simulation studies with smaller number of clusters, smaller number of subjects within clusters, and smaller subcohort size. Table 3 summarized these results. This illustrates that the proposed method generally works well, though there is some slight under-coverage for Designs B and C, which is reduced as the number of clusters increases.

[Table 3 about here.]

We also did some simulation studies with smaller marginal event rate of $p_0 = 0.03$ (Table 4). The results display that even when the event rate is small, $\hat{\beta}_s$ still performs well.

[Table 4 about here.]

Next, we examined the stratified method proposed in Section (A.7). As shown in Table 5, the proposed stratified method appears to perform well with a reasonable small number of strata.

[Table 5 about here.]

The results in Table 6 show that the efficiency gain of the proposed method over that of Lu and Shih (2006) is more obvious when the covariate is cluster-specific.

[Table 6 about here.]
Next, the performance of an inverse sampling probability weighted (ISPW) estimator and the proposed estimator were compared through simulation (Table 7). The ISPW method used the true sampling probability, while the methods proposed here used estimates of the sampling probability. The results show that the ESD of the ISPW method is generally comparable to that of our proposed method.

In addition, the point estimates based on simple random samples (SRS) for some non-rare event settings are provided (Table 7). This investigation showed that the ESDs of the point estimates based on SRS are very close to those based on Bernoulli sampling. Therefore, one does not gain much efficiency by using SRS, at least for the examples we considered.

References


Table 1
Simulation results to evaluate the estimate of $\beta_0$ with a continuous covariate based on 1000 replications.

<table>
<thead>
<tr>
<th>Design &amp; Method</th>
<th>$\beta_0 = \log(0.5)$</th>
<th>Bias</th>
<th>ESD</th>
<th>ASE</th>
<th>ARE</th>
<th>CP</th>
</tr>
</thead>
<tbody>
<tr>
<td>FC</td>
<td>-0.002</td>
<td>0.033</td>
<td>0.034</td>
<td>1.000</td>
<td>0.947</td>
<td></td>
</tr>
<tr>
<td>A SC</td>
<td>-0.002</td>
<td>0.042</td>
<td>0.044</td>
<td>0.597</td>
<td>0.955</td>
<td></td>
</tr>
<tr>
<td>WC</td>
<td>-0.003</td>
<td>0.041</td>
<td>0.042</td>
<td>0.655</td>
<td>0.943</td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>-0.002</td>
<td>0.037</td>
<td>0.038</td>
<td>0.801</td>
<td>0.935</td>
<td></td>
</tr>
<tr>
<td>LS</td>
<td>-0.005</td>
<td>0.057</td>
<td>0.057</td>
<td>0.356</td>
<td>0.957</td>
<td></td>
</tr>
<tr>
<td>B SC</td>
<td>-0.003</td>
<td>0.049</td>
<td>0.048</td>
<td>0.502</td>
<td>0.940</td>
<td></td>
</tr>
<tr>
<td>WC</td>
<td>-0.004</td>
<td>0.046</td>
<td>0.045</td>
<td>0.571</td>
<td>0.923</td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>-0.004</td>
<td>0.044</td>
<td>0.041</td>
<td>0.688</td>
<td>0.916</td>
<td></td>
</tr>
<tr>
<td>LS</td>
<td>-0.006</td>
<td>0.066</td>
<td>0.063</td>
<td>0.291</td>
<td>0.939</td>
<td></td>
</tr>
<tr>
<td>C SC</td>
<td>-0.003</td>
<td>0.046</td>
<td>0.046</td>
<td>0.546</td>
<td>0.944</td>
<td></td>
</tr>
<tr>
<td>WC</td>
<td>-0.004</td>
<td>0.043</td>
<td>0.043</td>
<td>0.625</td>
<td>0.939</td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>-0.003</td>
<td>0.040</td>
<td>0.040</td>
<td>0.723</td>
<td>0.937</td>
<td></td>
</tr>
<tr>
<td>LS</td>
<td>-0.008</td>
<td>0.061</td>
<td>0.060</td>
<td>0.321</td>
<td>0.944</td>
<td></td>
</tr>
</tbody>
</table>

Estimate of $\beta_0$ from 5 methods with a continuous covariate: Method FC = full cohort analysis; SC = estimating $p_0$ using the subcohort, $\hat{p}_s$; WC = estimating $p_0$ using whole cohort, $\hat{p}_w$; T = using true value, $p_0$; LS = Lu and Shih (2006) estimator. 100 clusters, $m_i$ follows a Bin(50,0.8) distribution, $\alpha=0.8$, $\lambda_0=1$, censoring time $C=1$, $\beta=\log(0.5)$, $Z$ follows a N(0,1) distribution. The number of individuals in the subcohort is $n_s = 800$. 
Table 2
Simulation results with $\alpha = 0.5$ based on 1000 replications.

<table>
<thead>
<tr>
<th>Method</th>
<th>Design &amp; $Z \sim \text{Bernoulli}(0.5)$</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>Design &amp; $Z \sim \mathcal{N}(0,1)$</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bias</td>
<td>ESD</td>
<td>ASE</td>
<td>ARE</td>
<td>CP</td>
<td>Bias</td>
<td>ESD</td>
<td>ASE</td>
<td>ARE</td>
</tr>
<tr>
<td>FC</td>
<td>-0.004</td>
<td>0.068</td>
<td>0.070</td>
<td>1.000</td>
<td>0.949</td>
<td>-0.003</td>
<td>0.049</td>
<td>0.049</td>
<td>1.000</td>
</tr>
<tr>
<td>A</td>
<td>SC</td>
<td>-0.003</td>
<td>0.094</td>
<td>0.093</td>
<td>0.567</td>
<td>0.946</td>
<td>-0.002</td>
<td>0.057</td>
<td>0.057</td>
</tr>
<tr>
<td></td>
<td>WC</td>
<td>-0.004</td>
<td>0.093</td>
<td>0.092</td>
<td>0.579</td>
<td>0.950</td>
<td>-0.003</td>
<td>0.056</td>
<td>0.055</td>
</tr>
<tr>
<td></td>
<td>T</td>
<td>-0.006</td>
<td>0.090</td>
<td>0.089</td>
<td>0.619</td>
<td>0.947</td>
<td>-0.003</td>
<td>0.053</td>
<td>0.051</td>
</tr>
<tr>
<td></td>
<td>LS</td>
<td>-0.005</td>
<td>0.099</td>
<td>0.099</td>
<td>0.500</td>
<td>0.948</td>
<td>-0.004</td>
<td>0.067</td>
<td>0.067</td>
</tr>
<tr>
<td>B</td>
<td>SC</td>
<td>-0.001</td>
<td>0.102</td>
<td>0.102</td>
<td>0.471</td>
<td>0.937</td>
<td>-0.007</td>
<td>0.066</td>
<td>0.066</td>
</tr>
<tr>
<td></td>
<td>WC</td>
<td>-0.006</td>
<td>0.109</td>
<td>0.108</td>
<td>0.420</td>
<td>0.928</td>
<td>-0.009</td>
<td>0.063</td>
<td>0.064</td>
</tr>
<tr>
<td></td>
<td>T</td>
<td>-0.008</td>
<td>0.107</td>
<td>0.105</td>
<td>0.444</td>
<td>0.923</td>
<td>-0.010</td>
<td>0.061</td>
<td>0.059</td>
</tr>
<tr>
<td></td>
<td>LS</td>
<td>-0.007</td>
<td>0.106</td>
<td>0.107</td>
<td>0.428</td>
<td>0.954</td>
<td>-0.013</td>
<td>0.085</td>
<td>0.083</td>
</tr>
<tr>
<td>C</td>
<td>SC</td>
<td>-0.005</td>
<td>0.098</td>
<td>0.097</td>
<td>0.521</td>
<td>0.929</td>
<td>-0.002</td>
<td>0.061</td>
<td>0.060</td>
</tr>
<tr>
<td></td>
<td>WC</td>
<td>-0.008</td>
<td>0.100</td>
<td>0.099</td>
<td>0.500</td>
<td>0.942</td>
<td>-0.003</td>
<td>0.060</td>
<td>0.059</td>
</tr>
<tr>
<td></td>
<td>T</td>
<td>-0.011</td>
<td>0.098</td>
<td>0.095</td>
<td>0.543</td>
<td>0.925</td>
<td>-0.003</td>
<td>0.056</td>
<td>0.055</td>
</tr>
<tr>
<td></td>
<td>LS</td>
<td>-0.007</td>
<td>0.104</td>
<td>0.103</td>
<td>0.462</td>
<td>0.946</td>
<td>-0.010</td>
<td>0.076</td>
<td>0.074</td>
</tr>
</tbody>
</table>

Estimate of $\beta_0$ from 5 methods: Method FC = full cohort analysis; SC = estimating $p_0$ using the subcohort, $\hat{p}_s$; WC = estimating $p_0$ using whole cohort, $\hat{p}_w$; T = using true value, $p_0$; LS = Lu and Shih (2006) estimator.

100 clusters, $m_i$ follows a Bin$(50,0.8)$ distribution, $\alpha=0.5$, $\lambda_0=1$, censoring time $C=1$, $\beta=\log(0.5)$, $Z$ follows either a Bernoulli$(0.5)$ distribution or a $\mathcal{N}(0,1)$ distribution. The number of individuals in the subcohort is $n_s = 800$. 

Table 3
Simulation results to evaluate the performance of the proposed method with a smaller number of clusters and a smaller cluster size based on 1000 replications.

<table>
<thead>
<tr>
<th>Design &amp; Method</th>
<th>$Z \sim Bernoulli(0.5)$</th>
<th>$Z \sim N(0,1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bias ESD ASE ARE CP</td>
<td>Bias ESD ASE ARE CP</td>
</tr>
<tr>
<td>FC</td>
<td>-0.005 0.099 0.098 1.000 0.938</td>
<td>-0.006 0.055 0.056 1.000 0.950</td>
</tr>
<tr>
<td>A SC</td>
<td>-0.009 0.161 0.159 0.380 0.943</td>
<td>-0.006 0.077 0.079 0.502 0.951</td>
</tr>
<tr>
<td>WC</td>
<td>-0.009 0.158 0.158 0.385 0.940</td>
<td>-0.007 0.074 0.075 0.558 0.947</td>
</tr>
<tr>
<td>T</td>
<td>-0.010 0.156 0.155 0.400 0.938</td>
<td>-0.006 0.070 0.070 0.640 0.933</td>
</tr>
<tr>
<td>LS</td>
<td>-0.013 0.175 0.173 0.321 0.940</td>
<td>-0.015 0.107 0.108 0.269 0.947</td>
</tr>
<tr>
<td>B SC</td>
<td>-0.003 0.166 0.155 0.400 0.923</td>
<td>-0.009 0.084 0.081 0.478 0.921</td>
</tr>
<tr>
<td>WC</td>
<td>-0.006 0.172 0.156 0.395 0.907</td>
<td>-0.012 0.082 0.076 0.543 0.912</td>
</tr>
<tr>
<td>T</td>
<td>-0.007 0.169 0.153 0.410 0.900</td>
<td>-0.011 0.077 0.071 0.622 0.899</td>
</tr>
<tr>
<td>LS</td>
<td>-0.008 0.185 0.172 0.325 0.929</td>
<td>-0.028 0.127 0.111 0.255 0.910</td>
</tr>
<tr>
<td>C SC</td>
<td>-0.014 0.166 0.156 0.395 0.920</td>
<td>-0.005 0.086 0.079 0.502 0.924</td>
</tr>
<tr>
<td>WC</td>
<td>-0.014 0.168 0.156 0.395 0.913</td>
<td>-0.008 0.079 0.075 0.558 0.929</td>
</tr>
<tr>
<td>T</td>
<td>-0.015 0.165 0.153 0.410 0.911</td>
<td>-0.007 0.075 0.070 0.640 0.918</td>
</tr>
<tr>
<td>LS</td>
<td>-0.013 0.185 0.170 0.332 0.909</td>
<td>-0.021 0.120 0.109 0.264 0.924</td>
</tr>
</tbody>
</table>

Estimate of $\beta_0$ from 5 methods: Method FC = full cohort analysis; SC = estimating $p_0$ using the subcohort, $\hat{p}_s$; WC = estimating $p_0$ using whole cohort, $\hat{p}_w$; T = using true value, $p_0$; LS = Lu and Shih (2006) estimator.
50 clusters, $m_i$ follows a Bin(25,0.8) distribution, $\alpha=0.8$, $\lambda_0=1$, censoring time $C=1$, $\beta=\log(0.5)$, $Z$ follows either a Bernoulli(0.5) distribution or a $N(0,1)$ distribution. The number of individuals in the subcohort is $n_s=200$. 
Table 4
Simulation results with $p_0 = 0.03$ based on 1000 replications.

<table>
<thead>
<tr>
<th>Design &amp; Method</th>
<th>$Z \sim \text{Bernoulli}(0.5)$</th>
<th>$Z \sim \mathcal{N}(0, 1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\text{Bias}$</td>
<td>$\text{ESD}$</td>
</tr>
<tr>
<td>FC</td>
<td>-0.041</td>
<td>0.242</td>
</tr>
<tr>
<td>A</td>
<td>-0.043</td>
<td>0.248</td>
</tr>
<tr>
<td>SC</td>
<td>-0.042</td>
<td>0.248</td>
</tr>
<tr>
<td>WC</td>
<td>-0.042</td>
<td>0.247</td>
</tr>
<tr>
<td>T</td>
<td>-0.042</td>
<td>0.249</td>
</tr>
<tr>
<td>LS</td>
<td>-0.042</td>
<td>0.249</td>
</tr>
<tr>
<td>B</td>
<td>-0.041</td>
<td>0.248</td>
</tr>
<tr>
<td>SC</td>
<td>-0.041</td>
<td>0.249</td>
</tr>
<tr>
<td>WC</td>
<td>-0.041</td>
<td>0.249</td>
</tr>
<tr>
<td>T</td>
<td>-0.042</td>
<td>0.249</td>
</tr>
<tr>
<td>LS</td>
<td>-0.042</td>
<td>0.250</td>
</tr>
</tbody>
</table>

Estimate of $\beta_0$ from 5 methods: Method FC = full cohort analysis; SC = estimating $p_0$ using the subcohort, $\tilde{p}_s$; WC = estimating $p_0$ using whole cohort, $\tilde{p}_w$; T = using true value, $p_0$; LS = Lu and Shih (2006) estimator.

100 clusters, $m_i$ follows a Bin(50, 0.8) distribution, $\alpha = 0.8$, censoring time $C = 1$, $\beta = \log(0.5)$, $\lambda_0 = 0.04$ when $Z$ follows a Bernoulli(0.5) distribution, $\lambda_0 = 0.0025$ when $Z$ follows a $\mathcal{N}(0, 1)$ distribution, the marginal event rate is $p_0 = 0.03$. The number of individuals in the subcohort is $n_s = 800$. 
Table 5
Simulation results to evaluate the performance of the proposed stratified methods based on 1000 replications.

<table>
<thead>
<tr>
<th>Design &amp; Method</th>
<th>$Z \sim \text{Bernoulli}(0.5)$</th>
<th>$Z \sim \mathcal{N}(0, 1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bias</td>
<td>ESD</td>
</tr>
<tr>
<td>FC</td>
<td>-0.014</td>
<td>0.110</td>
</tr>
<tr>
<td>A SC</td>
<td>-0.011</td>
<td>0.133</td>
</tr>
<tr>
<td>WC</td>
<td>-0.012</td>
<td>0.134</td>
</tr>
<tr>
<td>T</td>
<td>-0.011</td>
<td>0.132</td>
</tr>
<tr>
<td>LS</td>
<td>-0.011</td>
<td>0.135</td>
</tr>
<tr>
<td>B SC</td>
<td>-0.016</td>
<td>0.133</td>
</tr>
<tr>
<td>WC</td>
<td>-0.016</td>
<td>0.134</td>
</tr>
<tr>
<td>T</td>
<td>-0.016</td>
<td>0.132</td>
</tr>
<tr>
<td>LS</td>
<td>-0.016</td>
<td>0.134</td>
</tr>
<tr>
<td>C SC</td>
<td>-0.013</td>
<td>0.131</td>
</tr>
<tr>
<td>WC</td>
<td>-0.013</td>
<td>0.131</td>
</tr>
<tr>
<td>T</td>
<td>-0.013</td>
<td>0.129</td>
</tr>
<tr>
<td>LS</td>
<td>-0.014</td>
<td>0.131</td>
</tr>
</tbody>
</table>

Estimate of $\beta_0$ from 5 methods: Method FC = full cohort analysis; SC = estimating $p_0$ using the subcohort, $\hat{p}_s$; WC = estimating $p_0$ using whole cohort, $\hat{p}_w$; T = using true value, $p_0$; LS = Lu and Shih (2006) estimator. 100 clusters, $n_i$ follows a Bin$(50, 0.8)$ distribution, $\alpha = 0.8$, censoring time $C = 1$, $\beta = \log(0.5)$, $Z$ follows either a Bernoulli$(0.5)$ distribution or a $\mathcal{N}(0, 1)$ distribution, $Z_2 \sim U(0, 1)$, stratum $k = 1, 2$ or $3$ if $Z_2 \leq 0.33$, $0.33 < Z_2 \leq 0.67$ or $Z_2 > 0.67$, respectively. $\lambda_{0k} = 0.1 \times k$, for $k = 1, 2, 3$. The number of individuals in the subcohort is $n_s = 800$. 

$Z \sim \text{Bernoulli}(0.5)$, $Z \sim \mathcal{N}(0, 1)$.
Table 6
Simulation results with a cluster-level covariate based on 1000 replications.

<table>
<thead>
<tr>
<th>Design &amp; Method</th>
<th>$\beta_0 = \log(0.5)$</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>$\beta_0 = 0$</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bias</td>
<td>ESD</td>
<td>ASE</td>
<td>ARE</td>
<td>CP</td>
<td>Bias</td>
<td>ESD</td>
<td>ASE</td>
<td>ARE</td>
<td>CP</td>
</tr>
<tr>
<td>FC</td>
<td>0.000</td>
<td>0.145</td>
<td>0.147</td>
<td>1.000</td>
<td>0.939</td>
<td>0.002</td>
<td>0.126</td>
<td>0.127</td>
<td>1.000</td>
<td>0.940</td>
</tr>
<tr>
<td>A</td>
<td>SC</td>
<td>0.002</td>
<td>0.160</td>
<td>0.159</td>
<td>0.855</td>
<td>0.944</td>
<td>0.003</td>
<td>0.141</td>
<td>0.141</td>
<td>0.811</td>
</tr>
<tr>
<td></td>
<td>WC</td>
<td>0.002</td>
<td>0.160</td>
<td>0.159</td>
<td>0.855</td>
<td>0.946</td>
<td>0.003</td>
<td>0.141</td>
<td>0.141</td>
<td>0.811</td>
</tr>
<tr>
<td></td>
<td>T</td>
<td>0.001</td>
<td>0.159</td>
<td>0.157</td>
<td>0.877</td>
<td>0.948</td>
<td>0.003</td>
<td>0.141</td>
<td>0.141</td>
<td>0.811</td>
</tr>
<tr>
<td></td>
<td>LS</td>
<td>0.000</td>
<td>0.163</td>
<td>0.163</td>
<td>0.813</td>
<td>0.949</td>
<td>0.002</td>
<td>0.147</td>
<td>0.146</td>
<td>0.757</td>
</tr>
<tr>
<td>B</td>
<td>SC</td>
<td>0.008</td>
<td>0.365</td>
<td>0.350</td>
<td>0.176</td>
<td>0.947</td>
<td>0.005</td>
<td>0.315</td>
<td>0.302</td>
<td>0.177</td>
</tr>
<tr>
<td></td>
<td>WC</td>
<td>0.013</td>
<td>0.353</td>
<td>0.334</td>
<td>0.194</td>
<td>0.934</td>
<td>0.005</td>
<td>0.315</td>
<td>0.300</td>
<td>0.179</td>
</tr>
<tr>
<td></td>
<td>T</td>
<td>0.013</td>
<td>0.353</td>
<td>0.333</td>
<td>0.195</td>
<td>0.934</td>
<td>0.005</td>
<td>0.315</td>
<td>0.300</td>
<td>0.179</td>
</tr>
<tr>
<td></td>
<td>LS</td>
<td>0.015</td>
<td>0.483</td>
<td>0.465</td>
<td>0.100</td>
<td>0.962</td>
<td>0.019</td>
<td>0.483</td>
<td>0.463</td>
<td>0.075</td>
</tr>
<tr>
<td>C</td>
<td>SC</td>
<td>-0.003</td>
<td>0.253</td>
<td>0.247</td>
<td>0.354</td>
<td>0.936</td>
<td>-0.001</td>
<td>0.222</td>
<td>0.215</td>
<td>0.349</td>
</tr>
<tr>
<td></td>
<td>WC</td>
<td>0.001</td>
<td>0.246</td>
<td>0.240</td>
<td>0.375</td>
<td>0.934</td>
<td>-0.001</td>
<td>0.222</td>
<td>0.215</td>
<td>0.349</td>
</tr>
<tr>
<td></td>
<td>T</td>
<td>0.000</td>
<td>0.246</td>
<td>0.238</td>
<td>0.381</td>
<td>0.935</td>
<td>-0.001</td>
<td>0.222</td>
<td>0.215</td>
<td>0.349</td>
</tr>
<tr>
<td></td>
<td>LS</td>
<td>-0.004</td>
<td>0.313</td>
<td>0.306</td>
<td>0.231</td>
<td>0.944</td>
<td>-0.002</td>
<td>0.306</td>
<td>0.299</td>
<td>0.180</td>
</tr>
</tbody>
</table>

Estimate of $\beta_0$ from 5 methods: Method FC = full cohort analysis; SC = estimating $p_0$ using the subcohort, $\hat{p}_s$; WC = estimating $p_0$ using whole cohort, $\hat{p}_w$; T = using true value, $p_0$; LS = Lu and Shih (2006) estimator. 

100 clusters, $m_i$ follows a Bin(50,0.8) distribution, $\alpha=0.8$, $\lambda=1$, censoring time $C=1$, $\beta=\log(0.5)$, $Z$ follows a Bernoulli(0.5) distribution. The number of individuals in the subcohort is $n_s = 800$. 


Table 7
Simulation results to compare the proposed methods with the ISPW and SRS methods based on 1000 replications.

<table>
<thead>
<tr>
<th>Design</th>
<th>SC</th>
<th>WC</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BER</td>
<td>SRS</td>
<td>BER</td>
</tr>
<tr>
<td>$Z \sim Ber(0.5)$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>0.053</td>
<td>0.083</td>
<td>0.082</td>
</tr>
<tr>
<td>B</td>
<td>0.053</td>
<td>0.083</td>
<td>0.086</td>
</tr>
<tr>
<td>C</td>
<td>0.053</td>
<td>0.084</td>
<td>0.083</td>
</tr>
<tr>
<td>$Z \sim N(0,1)$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>0.033</td>
<td>0.042</td>
<td>0.041</td>
</tr>
<tr>
<td>B</td>
<td>0.033</td>
<td>0.049</td>
<td>0.045</td>
</tr>
<tr>
<td>C</td>
<td>0.033</td>
<td>0.046</td>
<td>0.043</td>
</tr>
</tbody>
</table>

Empirical standard deviation of the estimate of $\beta_0$ from following methods: Method FC = full cohort analysis; SC = estimating $p_0$ using the subcohort, $\hat{p}_s$; WC = estimating $p_0$ using whole cohort, $\hat{p}_w$; T = using true value, $p_0$; ISPW = inverse sampling probability method; BER = Bernoulli sampling; SRS = simple random sampling.

100 clusters, $m_i$ follows a Bin(50,0.8) distribution, $\alpha=0.8$, $\lambda_0=1$, censoring time $C=1$, $\beta=\log(0.5)$, $Z$ follows either a Bernoulli(0.5) distribution or a N(0,1) distribution. The number of individuals in the subcohort is $n_s = 800$. 