Supplementary Materials of "Covariate adjustment via propensity scores for recurrent events in the presence of dependent censoring"

Youngjoo Cho *[1] and Debashis Ghosh[2]

[1]Joseph J. Zilber School of Public Health, University of Wisconsin-Milwaukee, Milwaukee, WI 53205

[2]Department of Biostatistics and Informatics, University of Colorado Denver - Anschutz Medical Campus, Aurora, CO 80045

Here we prove the theorems in the main paper. Let $\boldsymbol{H}_i = (1, \boldsymbol{V}_i)^T, i = 1, \ldots, n$. Denote

$$N_i^*(t) = \sum_{k=1}^{\infty} I(T_{ik} \leq t)$$
$$e_i(\boldsymbol{\alpha}) = \frac{\exp(\boldsymbol{\alpha}^T \boldsymbol{H}_i)}{1 + \exp(\boldsymbol{\alpha}^T \boldsymbol{H}_i)}$$
$$w_i(\boldsymbol{\alpha}) = \frac{Z_i}{e_i(\boldsymbol{\alpha})} + \frac{1 - Z_i}{1 - e_i(\boldsymbol{\alpha})}.$$

As defined in the paper, the new estimating equation for the time to the dependent censoring is

$$S_n(\eta^{tr}, \boldsymbol{\alpha}) = n^{-1/2} \sum_{i=1}^{n} \xi_i w_i(\boldsymbol{\alpha}) \left[Z_i - \frac{\sum_{j=1}^{n} I\{\tilde{D}_j^*(\eta^{tr}) \geq \tilde{D}_i^*(\eta^{tr})\} w_j(\boldsymbol{\alpha}) Z_j}{\sum_{j=1}^{n} I\{\tilde{D}_j^*(\eta^{tr}) \geq \tilde{D}_i^*(\eta^{tr})\} w_j(\boldsymbol{\alpha})} \right].$$

*corresponding author, cho23@uwm.edu

The new estimating equation based on Ghosh and Lin (2003) is

$$U_n^{GL}(\beta^{tr}, \boldsymbol{\alpha}) = n^{-1/2} \sum_{i=1}^{n} \int_{-\infty}^{\infty} \left[Z_i - \frac{\sum_{j=1}^{n} I\{\tilde{D}_j^{**}(\beta^{tr}) \geq t\} w_j(\boldsymbol{\alpha}) Z_j}{\sum_{j=1}^{n} I\{\tilde{D}_j^{**}(\beta^{tr}) \geq t\} w_j(\boldsymbol{\alpha})} \right] dN_{2i}(t; \beta^{tr}, \boldsymbol{\alpha}),$$

where

$$\beta^{tr} = (\eta^{tr}, \theta^{tr})$$

$$d(\beta^{tr}) = \max_i \{0, (\theta^{tr} - \eta^{tr}) Z_i\}$$

$$\tilde{D}_i^{**}(\beta^{tr}) = (D_i \wedge C_i) - \eta^{tr} Z_i - d(\beta^{tr})$$

$$N_{2i}^{w}(t; \beta^{tr}, \boldsymbol{\alpha}) = w_i(\boldsymbol{\alpha}) \sum_{k=1}^{\infty} I\{T_{ik} - \theta^{tr} Z_i \leq t \wedge \tilde{D}_i^{**}(\beta^{tr})\}.$$

The new estimating equation based on Hsieh, Ding and Wang (2011) is

$$U_n^{H}(\beta, \boldsymbol{\alpha}) = \frac{2n^{1/2}}{n(n-1)} \sum_{1 \leq i < j \leq n} \sum_{k} (Z_i - Z_j) w_i(\boldsymbol{\alpha}) w_j(\boldsymbol{\alpha}) \phi_{ijk}(\beta^{tr}),$$

where

$$d_{ij}(\beta^{tr}) = \max\{0, (\theta^{tr} - \eta^{tr}) Z_i, (\theta^{tr} - \eta^{tr}) Z_j\}$$

$$\tilde{T}_{i(j)k}^{*}(\beta^{tr}) = (T_{ik} - \theta^{tr} Z_i) \wedge \{(D_i \wedge C_i) - \eta^{tr} Z_i - d_{ij}(\beta^{tr})\}$$

$$\tilde{\delta}_{i(j)k}^{*}(\beta^{tr}) = I[(T_{ik} - \theta^{tr} Z_i) \leq \{(D_i \wedge C_i) - \eta^{tr} Z_i - d_{ij}(\beta)]$$

$$\phi_{ijk}(\beta^{tr}) = \tilde{\delta}_{i(j)k}^{*}(\beta^{tr}) I\{\tilde{T}_{i(j)k}^{*}(\beta^{tr}) \leq \tilde{T}_{j(i)k}^{*}(\beta^{tr})\} - \tilde{\delta}_{j(i)k}^{*}(\beta^{tr}) I\{\tilde{T}_{i(j)k}^{*}(\beta^{tr}) \geq \tilde{T}_{j(i)k}^{*}(\beta^{tr})\}.$$

Let $\boldsymbol{\alpha}_0$ be the true regression coefficient of the logistic regression model. For theoretical results, we make the following assumptions:

(A.1) Parameter space $\boldsymbol{\Gamma}$ is compact and true parameter value $\boldsymbol{\gamma}_0$ is an interior point of $\boldsymbol{\Gamma}$.

(A.2) Denote the filtration as $\mathcal{F}_{t-} = \{N_{1i}(u; \eta^{tr}, \boldsymbol{\alpha}), Y_{1i}(u; \eta^{tr}, \boldsymbol{\alpha}), Z_i;\ i = 1, \ldots, n; 0 \leq u <$

$t\}$, where

$$N_{1i}^{w}(t;\eta^{tr},\boldsymbol{\alpha}) = w_i(\boldsymbol{\alpha})I\{\tilde{D}_i^*(\eta^{tr}) \le t, \xi_i = 1\}$$
$$Y_{1i}^{w}(t;\eta^{tr},\boldsymbol{\alpha}) = w_i(\boldsymbol{\alpha})I\{\tilde{D}_i^*(\eta^{tr}) \ge t\}.$$

Let

$$Q_1^{zw}(t;\eta^{tr},\boldsymbol{\alpha}) = E[w_1(\boldsymbol{\alpha})I\{\tilde{D}_1^*(\eta^{tr}) \ge t\}Z_1] \quad Q_1(t;\eta^{tr},\boldsymbol{\alpha}) = E[w_1(\boldsymbol{\alpha})I\{\tilde{D}_1^*(\eta^{tr}) \ge t\}]$$
$$Q_2^{zw}(t;\beta^{tr},\boldsymbol{\alpha}) = E[w_1(\boldsymbol{\alpha})I\{\tilde{D}_1^{**}(\beta^{tr}) \ge t\}Z_1] \quad Q_2(t;\beta^{tr},\boldsymbol{\alpha}) = E[w_1(\boldsymbol{\alpha})I\{\tilde{D}_1^{**}(\beta^{tr}) \ge t\}].$$

Define

$$m_1^w(t;\eta^{tr},\boldsymbol{\alpha}) = E\left\{\int_{-\infty}^{t}\left[Z_1 - \frac{E\{Q_1^{zw}(u;\eta^{tr},\boldsymbol{\alpha})\}}{E\{Q_1^{w}(u;\eta^{tr},\boldsymbol{\alpha})\}}\right]dN_{11}^{w}(u;\eta^{tr},\boldsymbol{\alpha})\right\}.$$
$$m_2^w(t;\beta^{tr},\boldsymbol{\alpha}) = E\left\{\int_{-\infty}^{t}\left[Z_1 - \frac{E\{Q_2^{zw}(u;\beta^{tr},\boldsymbol{\alpha})\}}{E\{Q_2^{w}(u;\beta^{tr},\boldsymbol{\alpha})\}}\right]dN_{21}^{w}(u;\beta^{tr},\boldsymbol{\alpha})\right\}$$

(A.2) $\boldsymbol{W}$ has finite second moment.

(A.3) $N_i(\cdot)$ is bounded for all $i = 1, \ldots, n$.

(A.4) $r_0(\cdot)$ has bounded the first partial derivatives with respect to θ and η.

(A.5) The solutions of $\boldsymbol{G}_n(\boldsymbol{\alpha}) = 0, S_n(\eta^{tr},\boldsymbol{\alpha}_0) = 0, U_n^{GL}(\theta^{tr},\eta_0^{tr},\boldsymbol{\alpha}_0)$ and $U_n^{H}(\theta^{tr},\eta_0^{tr},\boldsymbol{\alpha}_0) = 0$ are unique.

(A.6) For $i = 1, \ldots, n$, there exists $s > 0$ and $r > 0$ such that $0 < s \le w_i(\boldsymbol{\alpha}) \le r < \infty$ for all $\boldsymbol{\alpha}$.

(A.7) Condition 3 in Ying (1993) guarantees that C_is have uniformly bounded densities.

(A.8) Existence of limiting quantities : For every $u > 0$, there exist $\bar{z}^{(1)}(\cdot) > 0$ and $\bar{z}^{(2)}(\cdot) > 0$

such that

$$
\begin{aligned}
\bar{Z}^{(1)}(u;\eta^{tr},\boldsymbol{\alpha}) &= \frac{\sum_{j=1}^{n} I\{\tilde{D}_j^{*}(\eta^{tr}) \geq u\} w_j(\boldsymbol{\alpha}) Z_j}{\sum_{j=1}^{n} I\{\tilde{D}_j^{*}(\eta^{tr}) \geq u\} w_j(\boldsymbol{\alpha})} \\
\bar{Z}^{(2)}(u;\beta^{tr},\boldsymbol{\alpha}) &= \frac{\sum_{j=1}^{n} I\{\tilde{D}_j^{**}(\beta^{tr}) \geq u\} w_j(\boldsymbol{\alpha}) Z_j}{\sum_{j=1}^{n} I\{\tilde{D}_j^{**}(\beta^{tr}) \geq u\} w_j(\boldsymbol{\alpha})} \\
\bar{z}^{(1)}(u) &= \lim_{n\to\infty} \frac{\sum_{j=1}^{n} I\{\tilde{D}_j^{*}(\eta_0^{tr}) \geq u\} w_j(\boldsymbol{\alpha}_0) Z_j}{\sum_{j=1}^{n} I\{\tilde{D}_j^{*}(\eta_0^{tr}) \geq u\} w_j(\boldsymbol{\alpha}_0)} \\
\bar{z}^{(2)}(u) &= \lim_{n\to\infty} \frac{\sum_{j=1}^{n} I\{\tilde{D}_j^{**}(\beta_0^{tr}) \geq u\} w_j(\boldsymbol{\alpha}_0) Z_j}{\sum_{j=1}^{n} I\{\tilde{D}_j^{**}(\beta_0^{tr}) \geq u\} w_j(\boldsymbol{\alpha}_0)}.
\end{aligned}
$$

(A.9) Let $\lambda(\beta^{tr},\boldsymbol{\alpha}) = E\{n^{-1/2}U_n^H(\beta^{tr},\boldsymbol{\alpha})\}$. Assume that $\lambda(\beta^{tr},\boldsymbol{\alpha})$ is differentiable at $\boldsymbol{\alpha}_0$ and β_0^{tr}, and $\left.\frac{\partial\lambda(\beta^{tr},\boldsymbol{\alpha})}{\partial\boldsymbol{\alpha}}\right|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0,\beta^{tr}=\beta_0^{tr}}$, $\left.\frac{\partial\lambda(\beta^{tr},\boldsymbol{\alpha})}{\partial\eta^{tr}}\right|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0,\beta^{tr}=\beta_0^{tr}}$ and $\left.\frac{\partial\lambda(\beta^{tr},\boldsymbol{\alpha})}{\partial\theta^{tr}}\right|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0,\beta^{tr}=\beta_0^{tr}}$ are nonsingular.

Note that conditions (A.1), (A.2), (A.5), (A.6), (A.7), (A.8), (A.9), (A.10) and (A.11) are identical to those in Cho, Hu and Ghosh (2018). There are additional assumptions ((A.3) and (A.4)) because of recurrent events. Moreover, we need assumptions for the logistic regression model for propensity scores. These assumptions are originally from Ferguson (1996, Chapter 17, p.114), Zhu (2013) and Zhu et al. (2014) and also are used in Cho, Hu and Ghosh (2018).

(B.1) The parameter $\boldsymbol{\alpha}$ belongs to a compact subset B of $\mathbb{R}^p$. The likelihood

$$
L^*(z,\boldsymbol{\alpha}) = e(\boldsymbol{\alpha})^z(1-e(\boldsymbol{\alpha}))^{1-z},
$$

is measurable in z for every $\boldsymbol{\alpha}$ in B. Moreover, L^* is continuous in $\boldsymbol{\alpha}$ for every z.

(B.2) For all z and $\boldsymbol{\alpha}$,

$$
\log\ \frac{L^*(z,\boldsymbol{\alpha}|\boldsymbol{H})}{L^*(z,\boldsymbol{\alpha}_0|\boldsymbol{H})} \leq h(z),
$$

where $h(z)$ is a function satisfying $E_{\boldsymbol{\alpha}_0}|h(z)| < \infty$.

1. Proof of Theorem 1

   By assumption of independence of $(\epsilon_{ik}^R, \epsilon_i^D)$ with $(Z_i, \boldsymbol{V}_i)$ along with $Z_i \perp \boldsymbol{V}_i | e(\boldsymbol{V}_i)$, argument of Theorem 1 in Cho, Hu and Ghosh (2018) is applicable. By Cho, Hu and Ghosh (2018), the theorem holds.

   From Theorem 1, we add the following assumptions:

(A.10) Conditions 2 and 4 in Ying (1993). Let $f(\cdot)$ be common density of $\tilde{\epsilon}_i{}^D = D_i - \theta_0^{tr} Z_i$s true given propensity score We assume that by Ghosh (2000), $f(\cdot)$ and its derivative are bounded, and

$$\int_{-\infty}^{\infty} \left(\frac{\dot{f}(t)}{f(t)}\right)^2 f(t)dt < \infty,$$

where $\dot{f}(t) = \dfrac{df}{dt}$. Condition 4 in Ying (1993) implies that $\sup_i E|\min\{\tilde{\epsilon}_i^D, C_i\}|^{\nu_0} < \infty$ for some $\nu_0 > 0$.

(A.11) For all $i = 1, \ldots, n$, given $e_i(\boldsymbol{\alpha}_0)$. $\tilde{D}_i(\beta_0^{tr})$ has bounded density.

2. Proof of $E\{S_n(\eta_0^{tr}, \boldsymbol{\alpha}_0)\} = 0$, $E\{U_n^{GL}(\beta_0^{tr}, \boldsymbol{\alpha}_0)\} = 0$ and $E\{U_n^H(\beta_0^{tr}, \boldsymbol{\alpha}_0)\} = 0$

   In this part, we will prove $E\{S_n(\eta_0^{tr}, \boldsymbol{\alpha}_0)\} = 0$, $E\{U_n^{GL}(\beta_0^{tr}, \boldsymbol{\alpha}_0)\} = 0$ and $E\{U_n^H(\beta_0^{tr}, \boldsymbol{\alpha}_0)\} = 0$. Let

$$M_{1i}^w(u; \eta_0^{tr}, \boldsymbol{\alpha}_0) = N_{1i}^w(t; \eta_0^{tr}, \boldsymbol{\alpha}_0) - \int_{-\infty}^{u} w_i(\boldsymbol{\alpha}_0) I\{\tilde{D}_i^*(\eta_0^{tr}) \geq t\}\lambda_{10}(t)dt.$$
$$M_{2i}^w(t; \beta_0^{tr}, \boldsymbol{\alpha}_0) = N_{2i}^w(t; \beta_0^{tr}, \boldsymbol{\alpha}_0) - \int_{-\infty}^{t} w_i(\boldsymbol{\alpha}_0) I\{\tilde{D}_i^{**}(\beta_0^{tr}) \geq u\} r_0(u; \theta_0^{tr}, \eta_0^{tr})dt,$$

where $\lambda_{10}(t)$ is the baseline hazard function for the time to the dependent censoring and

$$r_0(u; \theta_0^{tr}, \eta_0^{tr}) = E\{dN_i^*(t + \theta_0^{tr} Z_i) | \tilde{D}_i^{**}(\beta_0^{tr}) \geq t\},$$

and it takes the common value which does not depend on Z_i given the true propensity score by Theorem 1. Note that

$$S_n(\eta^{tr}, \boldsymbol{\alpha}) = n^{-1/2}\sum_{i=1}^{n}\int_{-\infty}^{\infty}\left[Z_i - \frac{\sum_{j=1}^{n} I\{\tilde{D}_j^*(\eta^{tr}) \geq t\} w_j(\boldsymbol{\alpha}) Z_j}{\sum_{j=1}^{n} I\{\tilde{D}_j^*(\eta^{tr}) \geq t\} w_j(\boldsymbol{\alpha})}\right] dN_{1i}^w(t; \eta, \boldsymbol{\alpha}).$$

Then by Cho, Hu and Ghosh (2018), $E\{S_n(\eta_0^{tr}, \boldsymbol{\alpha}_0)\} = 0$. Next, we will show that $E\{U_n^{GL}(\beta_0^{tr}, \boldsymbol{\alpha}_0)\} = 0$. By using a similar argument as the time to dependent censoring, we can show that

$$U_n^{GL}(\beta_0^{tr}, \boldsymbol{\alpha}_0) = n^{-1/2}\sum_{i=1}^{n}\int_{-\infty}^{\infty}\{Z_i - \bar{Z}^{(2)}(u; \beta_0^{tr}, \boldsymbol{\alpha}_0)\} dM_{2i}^w(u; \beta_0^{tr}, \boldsymbol{\alpha}_0),$$

As in Ghosh and Lin (2003), define $r_0(u; \theta_0^{tr}, \eta_0^{tr} | e_i(\boldsymbol{\alpha}_0))dt$ to be the common value of

$$E\{dN_i^*(t + \theta_0^{tr} Z_i) | \tilde{D}_i^{**}(\beta_0^{tr}) \geq t, e_i(\boldsymbol{\alpha}_0)\}$$

which does not depend on Z_i by Theorem 1. By algebra,

$$M_{2i}^w(t; \beta_0^{tr}, \boldsymbol{\alpha}_0) = \int_{-\infty}^{t} w_i(\boldsymbol{\alpha}_0) I\{\tilde{D}_i^{**}(\beta_0^{tr}) \geq u\}[\{dN_i^*(u + \theta_0^{tr} Z_i) | \tilde{D}_i^{**}(\beta_0^{tr}) \geq u\} - r_0(u; \theta_0^{tr}, \eta_0^{tr})dt].$$

Then as in Ghosh and Lin (2003),

$$\begin{aligned}
&E\{M_{2i}(t; \beta_0^{tr}, \boldsymbol{\alpha}_0)\} \\
&= E\bigg(\int_{-\infty}^{t} w_i(\boldsymbol{\alpha}_0) I\{\tilde{D}_i^{**}(\beta_0^{tr}) \geq u\}[\{dN_i^*(t + \theta_0^{tr} Z_i) | \tilde{D}_i^{**}(\beta_0^{tr}) \geq u\} - r_0(t; \theta_0^{tr}, \eta_0^{tr})dt]\bigg) \\
&= E\bigg[\int_{-\infty}^{t} E\bigg\{ w_i(\boldsymbol{\alpha}_0) \times \\
&\quad E\bigg(I\{\tilde{D}_i^{**}(\beta_0^{tr}) \geq u\} \times [E\{dN_i^*(t + \theta_0^{tr} Z_i) | \tilde{D}_i^{**}(\beta_0^{tr}) \geq u\} - r_0(t; \theta_0^{tr}, \eta_0^{tr})dt \bigg| e_i(\boldsymbol{\alpha}_0)]\bigg) \bigg| Z_i \bigg\}\bigg] \\
&= 0
\end{aligned}$$

Thus $U_n^{GL}(\beta_0^{tr}, \boldsymbol{\alpha}_0)$ is a zero-mean process (Ghosh % Lin, 2003; Ghosh (2000). Hence

$E\{U_n^{GL}(\beta_0^{tr}, \boldsymbol{\alpha}_0)\} = 0$. For a proof that $E\{U_n^{H}(\beta_0^{tr}, \boldsymbol{\alpha}_0)\} = 0$,

$$E\{U_n^{H}(\beta_0^{tr}, \boldsymbol{\alpha}_0)\} = E[w_1(\boldsymbol{\alpha}_0)w_2(\boldsymbol{\alpha}_0)(Z_1 - Z_2)E\{U_{12}(\beta_0^{tr}|Z_1, Z_2, e_1(\boldsymbol{\alpha}_0), e_2(\boldsymbol{\alpha}_0))\}],$$

where $U_{ij} = \sum_k \phi_{ijk}(\beta^{tr})$. Then

$$\begin{aligned}
&E\{U_{12}(\beta_0^{tr}|Z_1, Z_2, e_1(\boldsymbol{\alpha}_0), e_2(\boldsymbol{\alpha}_0))\}\\
=&\sum_k P\{(T_{1k} - \theta_0^{tr} Z_1) \leq \tilde{D}_1^{**}(\beta_0^{tr}) \wedge (T_{2k} - \theta_0^{tr} Z_2) \wedge \tilde{D}_2^{**}(\beta_0^{tr})|Z_1, Z_2, e_1(\boldsymbol{\alpha}_0), e_2(\boldsymbol{\alpha}_0))\}\\
&- P\{(T_{2k} - \theta_0^{tr} Z_2) \leq \tilde{D}_1^{**}(\beta_0^{tr}) \wedge (T_{1k} - \theta_0^{tr} Z_1) \wedge \tilde{D}_2^{**}(\beta_0^{tr})|Z_1, Z_2, e_1(\boldsymbol{\alpha}_0), e_2(\boldsymbol{\alpha}_0))\} = 0.
\end{aligned}$$

Hence $E\{U_n^{H}(\beta_0^{tr}, \boldsymbol{\alpha}_0)\} = 0$.

2. Proof of Theorem 2 :

By assumptions from (B.1) and (B.2), $\hat{\boldsymbol{\alpha}}$ is strongly consistent. For strong consistency of $(\hat{\eta}^{catr}, \hat{\beta}^{GLcatr})^T$, we can mimic the proofs in Ghosh (2000). Note that we consider

$$U_n^{GL}(\beta_0^{tr}, \boldsymbol{\alpha}_0, t) = n^{-1/2} \sum_{i=1}^{n} \int_{-\infty}^{\infty} \{Z_i - \bar{Z}^{(2)}(u; \beta_0^{tr}, \boldsymbol{\alpha}_0)\} dM_{2i}^{w}(u; \beta_0^{tr}, \boldsymbol{\alpha}_0),$$

on interval $[0, b]$, where b is chosen as $lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} I\{\tilde{D}_i^{**}(\beta_0^{tr}) \geq b\} > 0$. Let $\bar{G}_Z = \sum_{i=1}^{n} \int_0^t Z_i dM_{2i}^{w}(u)$ and $\bar{G}(t) = \sum_{i=1}^{n} M_{2i}^{w}(t)$. Then by multivariate central limit theorem, $n^{-1/2}\bar{G}_Z(t)$ and $n^{-1/2}\bar{G}(t)$ converge to zero-mean Gaussian processes. Denote such processes as $\mathcal{G}_Z$ and $\mathcal{G}$. By the Skorokhod-Dudley-Wichura theorem (Shorack and Wellner, pp.47-48 ; Ghosh, 2000, Chapter 6), there is another probability space such tat $n^{-1/2}\bar{G}_Z(t)$ and $n^{-1/2}\bar{G}(t)$ converges almost surely to $\mathcal{G}_Z$ and $\mathcal{G}$. By the uniform strong law of large numbers (Pollard, 1990, p41), $\bar{Z}^{(2)}(t; \beta_0^{tr}, \boldsymbol{\alpha}_0)$ converges to $\bar{z}^{(2)}(t)$ in probability uniformly in t and hence $U_n^{GL}(t; \beta_0^{tr}, \boldsymbol{\alpha}_0)$ converges to $\mathcal{G}_Z(t) - \int_0^t \bar{z}^{(2)}(u) d\mathcal{G}(u)$ almost surely. Thus in the original probability space, $U_n^{GL}(u; \beta_0^{tr}, \boldsymbol{\alpha}_0)$ converges weakly $\mathcal{G}_Z(t) - \int_0^t \bar{z}^{(2)}(u) d\mathcal{G}(u)$. By extending the bound a to ∞ (Ying, 1993), we have

$$S_n(\eta_0^{tr}, \boldsymbol{\alpha}_0) = n^{-1/2}\sum_{i=1}^{n}\int_{-\infty}^{\infty}\{Z_i - \bar{z}^{(1)}(u)\}dM_{1i}^{w}(u; \eta_0^{tr}, \boldsymbol{\alpha}_0) + o_p(1) \quad (1)$$

$$U_n^{GL}(\beta_0^{tr}, \boldsymbol{\alpha}_0) = n^{-1/2}\sum_{i=1}^{n}\int_{-\infty}^{\infty}\{Z_i - \bar{z}^{(2)}(u)\}dM_{2i}^{w}(u; \beta_0^{tr}, \boldsymbol{\alpha}_0) + o_p(1). \quad (2)$$

For all positive M_1, M_2 and M_3, by assumption (A.2) and arguments in Chapter 6 in Ghosh (2000),

$$\sup_{|\eta^{tr}|\le M_1, ||\boldsymbol{\alpha}||\le M_2} |n^{-1/2}S_n(\eta^{tr}, \hat{\boldsymbol{\alpha}}) - m_1^w(\eta^{tr}, \boldsymbol{\alpha})| = o(n^{-1/2+\epsilon_1^*})$$

$$\sup_{||\beta^{tr}||\le M_3, ||\boldsymbol{\alpha}||\le M_2} |n^{-1/2}U_n^{GL}(\beta^{tr}, \hat{\boldsymbol{\alpha}}) - m_2^w(\beta^{tr}, \boldsymbol{\alpha})| = o(n^{-1/2+\epsilon_2^*}),$$

for any $\epsilon_1^* > 0$ and $\epsilon_2^* > 0$. Let $\mathcal{N}_0$ be a neighborhood of η_0^{tr} and $\mathcal{B}^*$ be any neighborhood of $\boldsymbol{\alpha}_0$. By Ying (1993), for any $\eta^{tr} \in \mathcal{N}_0$ and $\boldsymbol{\alpha} \in \mathcal{B}^*$,

$$\sup_{\boldsymbol{\alpha}\in\mathcal{B}^*, \eta^{tr}\in\mathcal{N}_0} |n^{-1/2}S_n(\eta^{tr}, \boldsymbol{\alpha}) - m_1(\eta^{tr}, \boldsymbol{\alpha})| \xrightarrow{p} 0$$

Let $\mathcal{N}_1$ be any neighborhood of β_0^{tr}. Similarly, for any $\beta^{tr} \in \mathcal{N}_1$ and $\boldsymbol{\alpha} \in \mathcal{B}^*$,

$$\sup_{\boldsymbol{\alpha}\in\mathcal{B}, \beta^{tr}\in\mathcal{N}_1} |n^{-1/2}U_n^{GL}(\beta^{tr}, \boldsymbol{\alpha}) - m_2(\beta^{tr}, \boldsymbol{\alpha})| \xrightarrow{p} 0.$$

Hence $\hat{\eta}^{catr}$ and $\hat{\theta}^{GLcatr}$ are strongly consistent.

Let $\lambda(\beta^{tr}, \boldsymbol{\alpha}_0) = E\{n^{-1/2}U_n^H(\beta^{tr}, \boldsymbol{\alpha}_0)\}$ and $h(Z_i, Z_j, \boldsymbol{V}_i, \boldsymbol{V}_j, \beta^{tr}, \boldsymbol{\alpha}_0) = w_i(\boldsymbol{\alpha}_0)w_j(\boldsymbol{\alpha}_0)(Z_i - Z_j)\phi_{ij}(\beta^{tr})$. Let $\mathcal{W}$ be compact set. From U-statistics law of large numbers,

$$|n^{-1/2}U_n^H(\beta^{tr}, \boldsymbol{\alpha}_0) - \lambda(\beta^{tr}, \boldsymbol{\alpha}_0)| \xrightarrow{p} 0,$$

for all $\beta^{tr} \in \mathcal{W}$. By decomposing compact set $\mathcal{W}$ into several finite subsets $\mathcal{W}_1, \ldots, \mathcal{W}_m$

such that $\mathcal{W} \in \cup_{i=1}^m \mathcal{W}_i$, for $(\beta^{tr})^i \in \mathcal{W}_i$,

$$\max_{1 \le i \le m} |n^{-1/2} U_n^H((\beta^{tr})^i, \boldsymbol{\alpha}_0) - \lambda((\beta^{tr})^i, \boldsymbol{\alpha}_0)| \xrightarrow{p} 0.$$

Since $w_i(\boldsymbol{\alpha}_0), i = 1 \dots n$ are bounded, by the Appendix of Hsieh, Ding and Wang (2011),

$$\sup_{||\beta^{tr} - \tilde{\beta}^{tr}|| \le \nu} n^{-1/2} |U_n^H(\beta^{tr}, \boldsymbol{\alpha}_0) - U_n^H(\tilde{\beta}^{tr}, \boldsymbol{\alpha}_0)| \le \frac{2}{n(n-1)} [\sum_{1 \le i < j \le n} |Z_i - Z_j| K_{ij}(\tilde{\beta}^{tr}, \boldsymbol{\alpha}_0, \nu)],$$

where

$$L_{ijk}^{(1)}(\tilde{\beta}^{tr}, \boldsymbol{\alpha}_0, \nu) = w_i(\boldsymbol{\alpha}_0) w_j(\boldsymbol{\alpha}_0) I[\{\beta^{tr} : ||\beta^{tr} - \tilde{\beta}^{tr}|| \le \nu, \tilde{X}_{i(j)}^*(\beta^{tr}) = \tilde{X}_{j(i)}^*(\beta^{tr})\} \neq \emptyset]$$
$$L_{ijk}^{(2)}(\tilde{\beta}^{tr}, \boldsymbol{\alpha}_0, \nu) = w_i(\boldsymbol{\alpha}_0) w_j(\boldsymbol{\alpha}_0) I[\{\beta^{tr} : ||\beta^{tr} - \tilde{\beta}^{tr}|| \le \nu, \tilde{\delta}_{i(j)}^*(\beta) \neq \tilde{\delta}_{i(j)}^*(\tilde{\beta}^{tr})\} \neq \emptyset]$$
$$K_{ij}(\tilde{\beta}^{tr}, \boldsymbol{\alpha}_0, \nu) = \sum_{k=1}^{K_i \vee K_j} \{L_{ijk}^{(1)}(\tilde{\beta}^{tr}, \boldsymbol{\alpha}_0, \nu) + L_{ijk}^{(2)}(\tilde{\beta}^{tr}, \boldsymbol{\alpha}_0, \nu) + L_{jik}^{(1)}(\tilde{\beta}^{tr}, \boldsymbol{\alpha}_0, \nu) + L_{jik}^{(2)}(\tilde{\beta}^{tr}, \boldsymbol{\alpha}_0, \nu)\}.$$

Let $H_{ij}(\tilde{\beta}^{tr}, \boldsymbol{\alpha}_0, \nu) = |Z_i - Z_j| K_{ij}(\tilde{\beta}^{tr}, \boldsymbol{\alpha}_0, \nu)$ and $H(\tilde{\beta}^{tr}, \boldsymbol{\alpha}_0, \nu) = \sum_{1 \le i < j \le n} |Z_i - Z_j| K_{ij}(\tilde{\beta}^{tr}, \boldsymbol{\alpha}_0, \nu)$. The Hoeffding decomposition yields

$$H(\tilde{\beta}^{tr}, \boldsymbol{\alpha}_0, \nu) - E\{H(\tilde{\beta}^{tr}, \boldsymbol{\alpha}_0, \nu)\} = \sum_{i=1}^n B_i(\tilde{\beta}^{tr}, \boldsymbol{\alpha}_0, \nu) + \sum_{i<j} B_{ij}(\tilde{\beta}^{tr}, \boldsymbol{\alpha}_0, \nu),$$

where

$$B_i(\tilde{\beta}^{tr}, \boldsymbol{\alpha}_0, \nu) = \sum_{j \neq i} [E\{H_{ij}(\tilde{\beta}^{tr}, \boldsymbol{\alpha}_0, \nu) | Z_i, e_i(\boldsymbol{\alpha}_0)\} - E\{H_{ij}(\tilde{\beta}^{tr}, \boldsymbol{\alpha}_0, \nu)\}]$$
$$B_{ij}(\tilde{\beta}^{tr}, \boldsymbol{\alpha}_0, \nu) = H_{ij}(\tilde{\beta}^{tr}, \boldsymbol{\alpha}_0, \nu) - E\{H_{ij}(\tilde{\beta}^{tr}, \boldsymbol{\alpha}_0, \nu) | Z_i, e_i(\boldsymbol{\alpha}_0)\} - E\{H_{ij}(\tilde{\beta}^{tr}, \boldsymbol{\alpha}_0, \nu) | Z_j, e_j(\boldsymbol{\alpha}_0)\}$$
$$+ E\{H_{ij}(\tilde{\beta}^{tr}, \boldsymbol{\alpha}_0, \nu)\}.$$

Similar to Peng and Fine (2006), Hsieh, Ding and Wang (2011) and Cho, Hu and Ghosh (2018), we have the following result.

Lemma 1. *There exists a constant* b_0 *such that* $E\{H_{ij}(\tilde{\beta}^{tr}, \boldsymbol{\alpha}_0, \nu)\} \le b_0 \nu$ *and* $E\{H(\tilde{\beta}^{tr}, \boldsymbol{\alpha}_0, \nu)\} \le$

$c_0 \nu n^2$

We will prove this lemma later. By Lemma 1 and a similar argument as Hsieh, Ding and Wang (2011) and Peng and Fine (2006), there exist $v_{10} > 0$ and $v_{20} > 0$ such that

$$Var\{H(\tilde{\beta}^{tr}, \boldsymbol{\alpha}_0, \nu)\} = \sum_{i=1}^{n} Var\{B_i(\tilde{\beta}^{tr}, \boldsymbol{\alpha}_0, \nu)\} + \sum_{i<j} Var\{B_{ij}(\tilde{\beta}^{tr}, \boldsymbol{\alpha}_0, \nu)\} = O(n^3).$$

Take $\epsilon^* > 0$ and let $0 < \nu < \epsilon^*/(3b_0)$ for any $b_0 > 0$. Then by Markov inequality,

$$\begin{aligned} P\{[n(n-1)]^{-1} H(\tilde{\beta}^{tr}, \boldsymbol{\alpha}_0, \nu) \geq \epsilon\} &\leq P\{[n(n-1)]^{-1}[H(\tilde{\beta}^{tr}, \boldsymbol{\alpha}_0, \nu) - E\{H(\tilde{\beta}^{tr}, \boldsymbol{\alpha}_0, \nu)]\} \geq \epsilon/3\} \\ &\leq \frac{9Var\{H(\tilde{\beta}^{tr}, \boldsymbol{\alpha}_0, \nu)\}}{[n(n-1)]^2\epsilon^2} \to 0. \end{aligned}$$

Hence

$$\sup_{||\beta^{tr} - \tilde{\beta}^{tr}|| \leq \nu} n^{-1/2} ||U_n^H(\beta^{tr}, \boldsymbol{\alpha}_0) - U_n^H(\tilde{\beta}^{tr}, \boldsymbol{\alpha}_0)|| \xrightarrow{p} 0. \tag{3}$$

Note that for any $\nu^* > 0$,

$$\begin{aligned} &\sup_{||\boldsymbol{\alpha} - \boldsymbol{\alpha}_0|| \leq \nu^*, ||\beta^{tr} - \tilde{\beta}^{tr}|| \leq \nu} n^{-1/2} |U_n^H(\beta^{tr}, \boldsymbol{\alpha}) - U_n^H(\tilde{\beta}^{tr}, \boldsymbol{\alpha}_0)| \\ &\leq \sup_{||\beta^{tr} - \tilde{\beta}^{tr}|| \leq \nu, ||\boldsymbol{\alpha} - \boldsymbol{\alpha}_0|| \leq \nu^*} n^{-1/2} |U_n^H(\beta^{tr}, \boldsymbol{\alpha}) - U_n^H(\beta^{tr}, \hat{\boldsymbol{\alpha}})| \\ &+ \sup_{||\beta^{tr} - \tilde{\beta}^{tr}|| \leq \nu, ||\boldsymbol{\alpha} - \boldsymbol{\alpha}_0|| \leq \nu^*} n^{-1/2} |U_n^H(\beta^{tr}, \hat{\boldsymbol{\alpha}}) - U_n^H(\beta^{tr}, \boldsymbol{\alpha}_0)| \\ &+ \sup_{||\beta^{tr} - \tilde{\beta}^{tr}|| \leq \nu, ||\boldsymbol{\alpha} - \boldsymbol{\alpha}_0|| \leq \nu^*} n^{-1/2} ||U_n^H(\beta^{tr}, \boldsymbol{\alpha}_0) - U_n^H(\tilde{\beta}^{tr}, \boldsymbol{\alpha}_0)||. \end{aligned} \tag{4}$$

By strong consistency of $\hat{\boldsymbol{\alpha}}$, first and second term in the righthand side of (4) converges to 0 in probability. Combined with (3), we can conclude

$$\sup_{||\boldsymbol{\alpha} - \tilde{\boldsymbol{\alpha}}|| \leq \nu^*, ||\beta^{tr} - \tilde{\beta}^{tr}|| \leq \nu} n^{-1/2} ||U_n^H(\beta^{tr}, \boldsymbol{\alpha}) - U_n^H(\tilde{\beta}^{tr}, \tilde{\boldsymbol{\alpha}})|| \xrightarrow{p} 0.$$

Thus $\hat{\theta}^{Hcatr}$ is strongly consistent.

Proof of Lemma 1. We can use arguments in Peng and Fine (2006) and Appendix of

Hsieh, Ding and Wang (2011) and Cho, Hu and Ghosh (2018). Note that the set

$$\{||\beta^{tr} - \tilde{\beta}^{tr}|| \le \nu, \tilde{\delta}^*_{i(j)}(\beta^{tr}) \ne \tilde{\delta}^*_{i(j)}(\tilde{\beta}^{tr})\} \in D_1(\tilde{\beta}^{tr}, \nu) \cup D_2(\tilde{\beta}^{tr}, \nu)$$

where

$$D_1(\tilde{\beta}^{tr}, \nu) = \{||\beta^{tr} - \tilde{\beta}^{tr}|| < \nu, T_{ik} - \theta^{tr} Z_i = [D_i + \eta^{tr}(Z_j - Z_i)] - \theta^{tr} Z_j\}$$
$$D_2(\tilde{\beta}^{tr}, \nu) = \{||\beta^{tr} - \tilde{\beta}^{tr}|| < \nu, T_{ik} - \theta^{tr} Z_i = [C_i + \eta^{tr}(Z_j - Z_i)] - \theta^{tr} Z_j\}$$

Then

$$\begin{aligned}
& D_1(\tilde{\beta}^{tr}, \nu) \\
& = \{||\beta^{tr} - \tilde{\beta}^{tr}|| < \nu, T_{ik} + \theta^{tr}(Z_j - Z_i) = D_i + \eta^{tr}(Z_j - Z_i)\} \\
& = \{||\beta^{tr} - \tilde{\beta}^{tr}|| < \nu, \epsilon^R_{ik} + \boldsymbol{\theta}_0^T \boldsymbol{W}_i + \theta^{tr}(Z_j - Z_i) = \epsilon^D_i + \boldsymbol{\eta}_0^T \boldsymbol{W}_i + \eta^{tr}(Z_j - Z_i)\} \\
& = \{||\beta^{tr} - \tilde{\beta}^{tr}|| < \nu, \epsilon^R_{ik} + (\boldsymbol{\theta}_0^{cfd})^T \boldsymbol{W}_i + [\tilde{\theta}^{tr} - (\tilde{\theta}^{tr} - \theta^{tr})](Z_j - Z_i) \\
& = \epsilon^D_i + \boldsymbol{\eta}_0^T \boldsymbol{W}_i + [\tilde{\eta}^{tr} - (\tilde{\eta}^{tr} - \eta^{tr})](Z_j - Z_i)\} \\
& \subseteq \{||\epsilon^R_{ik} - \epsilon^D_i + (\boldsymbol{\theta}_0 - \boldsymbol{\eta}_0)^T \boldsymbol{W}_i + (\tilde{\theta}^{tr} - \tilde{\eta}^{tr})(Z_j - Z_i)|| < 2\nu |Z_j - Z_i|\}.
\end{aligned}$$

Thus there exists $d_0 > 0$ such that $w_i(\boldsymbol{\alpha}_0) w_j(\boldsymbol{\alpha}_0) P\{D_1(\tilde{\beta}^{tr}, \nu) | e_i(\boldsymbol{\alpha}_0), e_j(\boldsymbol{\alpha}_0), Z_i, Z_j\} \le 2 w_i(\boldsymbol{\alpha}_0) w_j(\boldsymbol{\alpha}_0) d_0 |Z_j - Z_i| \nu$ by the assumption. Similarly,

$$w_i(\boldsymbol{\alpha}_0) w_j(\boldsymbol{\alpha}_0) P\{D_2(\tilde{\beta}^{tr}, \nu) | e_i(\boldsymbol{\alpha}_0), e_j(\boldsymbol{\alpha}_0), Z_i, Z_j\} \le 2 w_i(\boldsymbol{\alpha}_0) w_j(\boldsymbol{\alpha}_0) |Z_j - Z_i| \nu.$$

Hence, $E\{L^{(2)}_{ijk}(\tilde{\beta}^{tr}, \boldsymbol{\alpha}_0, \nu)\} \le 2 d_0 E\{w_i(\boldsymbol{\alpha}_0) w_j(\boldsymbol{\alpha}_0) |Z_j - Z_i|\} \nu$. Similarly, there exists $f_0 > 0$ and $E\{L^{(1)}_{ijk}(\tilde{\beta}^{tr}, \boldsymbol{\alpha}_0, \nu)\} \le 2 f_0 E\{w_i(\boldsymbol{\alpha}_0) w_j(\boldsymbol{\alpha}_0) |Z_j - Z_i|\} \nu$. We can apply similar arguments to $L^{(1)}_{jik}(\tilde{\beta}^{tr}, \boldsymbol{\alpha}_0, \nu)$ and $L^{(2)}_{jik}(\tilde{\beta}^{tr}, \boldsymbol{\alpha}_0, \nu)$. Since $w_i(\boldsymbol{\alpha}_0)$ are bounded, there exists $K_0 > 0$ such that

$$E\{K_{ij}(\tilde{\beta}^{tr}, \boldsymbol{\alpha}_0, \nu)\} = E[\sum_{k=1}^{K_i \vee K_j} \{L_{ijk}^{(1)}(\tilde{\beta}^{tr}, \boldsymbol{\alpha}_0, \nu) + L_{ijk}^{(2)}(\tilde{\beta}^{tr}, \boldsymbol{\alpha}_0, \nu) + L_{jik}^{(1)}(\tilde{\beta}^{tr}, \boldsymbol{\alpha}_0, \nu) + L_{jik}^{(2)}(\tilde{\beta}^{tr}, \boldsymbol{\alpha}_0, \nu)\}]$$
$$\leq (K_i \vee K_j) r^2 E|Z_i - Z_j| \nu \leq K_0 \nu.$$

Similarly, there exists $K_1 > 0$ such that

$$E\{K_{ij}^2(\tilde{\beta}^{tr}, \boldsymbol{\alpha}_0, \nu)\} \leq K_1 \nu.$$

By the Cauchy-Schwarz inequality,

$$E\{H_{ij}(\tilde{\beta}^{tr}, \boldsymbol{\alpha}_0, \nu)\} \leq \sqrt{E\{K_{ij}^2(\tilde{\beta}^{tr}, \boldsymbol{\alpha}_0, \nu)\} E|Z_i - Z_j|^2} \leq K_1 \nu \sqrt{E|Z_i - Z_j|^2}.$$

Hence there exists $b_0 > 0$ such that $E\{H_{ij}(\tilde{\beta}^{tr}, \nu)\} \leq b_0 \nu$. Finally, $E\{H(\tilde{\beta}^{tr}, \nu)\} \leq K_1 b_0 n^2$. Thus there exists $c_0 > 0$ such that $E\{H(\tilde{\beta}^{tr}, \nu)\} \leq c_0 \nu n^2$. $\square$

3. Proof of Theorem 3 :

Let $\boldsymbol{G}_n(\boldsymbol{\alpha})$ be the score function of parameters in logistic regression, where

$$\boldsymbol{G}_n(\boldsymbol{\alpha}) = n^{-1/2} \sum_{i=1}^{n} \boldsymbol{H}_i [Z_i - \frac{\exp(\boldsymbol{\alpha}^T \boldsymbol{H}_i)}{1 + \exp(\boldsymbol{\alpha}^T \boldsymbol{H}_i)}].$$

Let $\Psi_i(\boldsymbol{\alpha}) = \boldsymbol{H}_i [Z_i - \frac{\exp(\boldsymbol{\alpha}^T \boldsymbol{H}_i)}{1 + \exp(\boldsymbol{\alpha}^T \boldsymbol{H}_i)}]$. Then $\boldsymbol{G}_n(\boldsymbol{\alpha}) = n^{-1/2} \sum_{i=1}^{n} \Psi_i(\boldsymbol{\alpha})$. By (1) and

(2), and U-statistic theory,

$$\begin{aligned}
\boldsymbol{G}_n(\boldsymbol{\alpha}_0) &= n^{-1/2}\sum_{i=1}^{n}\Psi_i(\boldsymbol{\alpha}_0)\\
S_n(\eta_0^{tr},\boldsymbol{\alpha}_0) &= n^{-1/2}\sum_{i=1}^{n}\int_{-\infty}^{\infty}(Z_i-\bar{z}^{(1)})dM_{1i}^{w}(u;\eta_0^{tr},\boldsymbol{\alpha}_0)+o_p(1)\\
U_n^{GL}(\beta_0^{tr},\boldsymbol{\alpha}_0) &= n^{-1/2}\sum_{i=1}^{n}\int_{-\infty}^{\infty}(Z_i-\bar{z}^{(2)})dM_{2i}^{w}(u;\beta_0^{tr},\boldsymbol{\alpha}_0)+o_p(1)\\
U_n^{H}(\beta_0^{tr},\boldsymbol{\alpha}_0) &= n^{-1/2}\sum_{i=1}^{n}2h_1(Z_i,\boldsymbol{\alpha}_0,\beta_0^{tr})+o_p(1).
\end{aligned}$$

Let

$$\begin{aligned}
h(Z_i,Z_j,\beta_0^{tr},\boldsymbol{\alpha}_0) &= w_i(\boldsymbol{\alpha}_0)w_j(\boldsymbol{\alpha}_0)(Z_i-Z_j)\phi_{ij}(\beta_0^{tr})\\
h_1(z,\beta_0^{tr},\boldsymbol{\alpha}_0) &= E\{h(z,Z_2,\beta_0^{tr},\boldsymbol{\alpha}_0)\}
\end{aligned}$$

Let $\boldsymbol{\psi}_1=\Psi_1(\boldsymbol{\alpha}_0)$, $\boldsymbol{\psi}_2^{GL}=(\psi_{21},\psi_{22})^T$ and $\boldsymbol{\psi}_2^{H}=(\psi_{21},\psi_{23})^T$, where

$$\begin{aligned}
\psi_{21} &= \int_{-\infty}^{\infty}\{Z_1-\bar{z}^{(1)}(u)\}dM_{1i}^{w}(u;\eta_0^{tr},\boldsymbol{\alpha}_0)\\
\psi_{22} &= \int_{-\infty}^{\infty}\{Z_1-\bar{z}^{(2)}(u)\}dM_{2i}^{w}(u;\beta_0^{tr},\boldsymbol{\alpha}_0)\\
\psi_{23} &= 2h_1(Z_1,\beta_0^{tr},\boldsymbol{\alpha}_0).
\end{aligned}$$

Let $\boldsymbol{\gamma}^{GL}=(\boldsymbol{\alpha}^T,\eta^{tr},\theta^{tr})^T$, $\boldsymbol{\gamma}^{H}=(\boldsymbol{\alpha}^T,\eta^{tr},\theta^{tr})^T$ and $\boldsymbol{\gamma}_0=(\boldsymbol{\alpha}_0^T,\eta_0^{tr},\theta_0^{tr})^T$ and

$$\begin{aligned}
\boldsymbol{J}_n^{GL}(\boldsymbol{\gamma}_0) &= [S_n^T(\eta_0^{tr},\boldsymbol{\alpha}_0),\{U_n^{GL}(\beta_0^{tr},\boldsymbol{\alpha}_0)\}^T]^T\\
\boldsymbol{J}_n^{H}(\boldsymbol{\gamma}_0) &= [S_n^T(\eta_0^{tr},\boldsymbol{\alpha}_0),\{U_n^{H}(\beta_0^{tr},\boldsymbol{\alpha}_0)\}^T]^T
\end{aligned}$$

By the multivariate central limit theorem, $\boldsymbol{Q}_n^{GL}(\boldsymbol{\gamma}_0)=\{\boldsymbol{G}_n^T(\boldsymbol{\alpha}_0),(\boldsymbol{J}_n^{GL}(\boldsymbol{\gamma}_0))^T\}^T$ and $\boldsymbol{Q}_n^{H}(\boldsymbol{\gamma}_0)=\{\boldsymbol{G}_n^T(\boldsymbol{\alpha}_0),(\boldsymbol{J}_n^{H}(\boldsymbol{\gamma}_0))^T\}^T$ converge to normal distributions with covariance matrices $\mathbf{\Omega}_0^{GL}$ and $\mathbf{\Omega}_0^{H}$, respectively, where

$$\mathbf{\Omega}_0^{GL} = E\begin{pmatrix} \boldsymbol{\psi}_1\boldsymbol{\psi}_1^T & \boldsymbol{\psi}_1\boldsymbol{\psi}_2^T \\ \boldsymbol{\psi}_2\boldsymbol{\psi}_1^T & \boldsymbol{\psi}_2\boldsymbol{\psi}_2^T \end{pmatrix}.$$

Let

$$\mathbf{\Lambda}_0^{GL} = \begin{pmatrix} \boldsymbol{L}_1 & 0 & 0 \\ \boldsymbol{L}_2 & E_1 & 0 \\ \boldsymbol{L}_3 & E_2 & E_3 \end{pmatrix} \quad \mathbf{\Lambda}_0^{H} = \begin{pmatrix} \boldsymbol{L}_1 & 0 & 0 \\ \boldsymbol{L}_2 & E_1 & 0 \\ \boldsymbol{L}_4 & E_4 & E_5 \end{pmatrix},$$

where

$$\boldsymbol{L}_1 = E\left[\frac{\partial \Psi_1(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}}\bigg|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0}\right] \quad \boldsymbol{L}_2 = \int_{-\infty}^{\infty} E\left[\frac{\partial}{\partial \boldsymbol{\alpha}}\{Z_1 - \bar{z}^{(1)}(t;\eta^{tr},\boldsymbol{\alpha})\}dN_{11}^{w}(t;\eta^{tr},\boldsymbol{\alpha})\bigg|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0,\eta^{tr}=\eta_0^{tr}}\right]$$

$$\boldsymbol{L}_3 = \int_{-\infty}^{\infty} E\left[\frac{\partial}{\partial \boldsymbol{\alpha}}\{Z_1 - \bar{z}^{(2)}(t;\beta^{tr},\boldsymbol{\alpha})\}I\{\tilde{D}_1^{**}(\beta^{tr}) \geq t\}w_1(\boldsymbol{\alpha})r_0(t;\theta^{tr},\eta^{tr})dt\bigg|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0,\beta^{tr}=\beta_0^{tr}}\right]$$

$$\boldsymbol{L}_4 = \frac{\partial \lambda(\beta^{tr},\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}}\bigg|_{\beta^{tr}=\beta_0^{tr},\boldsymbol{\alpha}=\boldsymbol{\alpha}_0}$$

$$E_1 = \int_{-\infty}^{\infty} E\left[\{Z_1 - \bar{z}^{(1)}(t;\eta^{tr},\boldsymbol{\alpha})\}I\{\tilde{D}_1^{*}(\eta^{tr}) \geq t\}w_1(\boldsymbol{\alpha})\bigg|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0,\eta^{tr}=\eta_0^{tr}}\frac{\lambda_{10}'(t)}{\lambda_{10}(t)}f(t)\right]dt$$

$$E_2 = \int_{-\infty}^{\infty} E\left[\frac{\partial}{\partial \eta^{tr}}\{Z_1 - \bar{z}^{(2)}(t;\beta^{tr},\boldsymbol{\alpha})\}I\{\tilde{D}_1^{**}(\beta^{tr}) \geq t\}w_1(\boldsymbol{\alpha})r_0(t;\theta^{tr},\eta^{tr})dt\bigg|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0,\beta^{tr}=\beta_0^{tr}}\right.$$

$$E_3 = \int_{-\infty}^{\infty} E\left[\frac{\partial}{\partial \theta^{tr}}\{Z_1 - \bar{z}^{(2)}(t;\beta^{tr},\boldsymbol{\alpha})\}I\{\tilde{D}_1^{**}(\beta^{tr}) \geq t\}w_1(\boldsymbol{\alpha})r_0(t;\theta^{tr},\eta^{tr})dt\bigg|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0,\beta^{tr}=\beta_0^{tr}}\right.$$

$$E_4 = \frac{\partial \lambda(\beta^{tr},\boldsymbol{\alpha})}{\partial \eta^{tr}}\bigg|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0,\beta^{tr}=\beta_0^{tr}} \quad E_5 = \frac{\partial \lambda(\beta^{tr},\boldsymbol{\alpha})}{\partial \theta^{tr}}\bigg|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0,\beta^{tr}=\beta_0^{tr}}.$$

where $\lambda_{10}(t)$ is a true common hazard function of error term for $\{\tilde{D}_i^{*}(\eta_0^{tr})\}_{i=1}^{n}$ given the true propensity score. Similar to Cho, Hu and Ghosh (2018), we have the following lemma.

Lemma 2.

$$\sup_{||\boldsymbol{\gamma}-\boldsymbol{\gamma}_0||\le c_n} \frac{||\boldsymbol{Q}_n^{GL}(\boldsymbol{\gamma}) - \boldsymbol{Q}_n^{GL}(\boldsymbol{\gamma}_0) - n^{1/2}\boldsymbol{\Lambda}_0^{GL}(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0)||}{1+n^{1/2}||\boldsymbol{\gamma}-\boldsymbol{\gamma}_0||} = o_p(1),$$
$$\sup_{||\boldsymbol{\gamma}-\boldsymbol{\gamma}_0||\le c_n} \frac{||\boldsymbol{Q}_n^{H}(\boldsymbol{\gamma}) - \boldsymbol{Q}_n^{H}(\boldsymbol{\gamma}_0) - n^{1/2}\boldsymbol{\Lambda}_0^{H}(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0)||}{1+n^{1/2}||\boldsymbol{\gamma}-\boldsymbol{\gamma}_0||} = o_p(1),$$

whenever c_n *converges to 0 in probability and*

$$\sup_{||\boldsymbol{\gamma}-\boldsymbol{\gamma}_0||\le c_n} \frac{||\boldsymbol{Q}_n^{GL}(\boldsymbol{\gamma}) - \boldsymbol{Q}_n^{GL}(\boldsymbol{\gamma}_0) - n^{1/2}\boldsymbol{\Lambda}_0^{GL}(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0)||}{1+n^{1/2}||\boldsymbol{\gamma}-\boldsymbol{\gamma}_0||} = o(1)$$
$$\sup_{||\boldsymbol{\gamma}-\boldsymbol{\gamma}_0||\le c_n} \frac{||\boldsymbol{Q}_n^{H}(\boldsymbol{\gamma}) - \boldsymbol{Q}_n^{H}(\boldsymbol{\gamma}_0) - n^{1/2}\boldsymbol{\Lambda}_0^{H}(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0)||}{1+n^{1/2}||\boldsymbol{\gamma}-\boldsymbol{\gamma}_0||} = o(1),$$

whenever c_n *converges to 0 almost surely.*

As before, we will prove this lemma later. By strong consistency of $\hat{\boldsymbol{\alpha}}$ and Lemma 2,

$$\boldsymbol{G}_n(\hat{\boldsymbol{\alpha}}) = \boldsymbol{G}_n(\boldsymbol{\alpha}_0) + n^{1/2}\dot{\boldsymbol{G}}_n(\boldsymbol{\alpha}_0)(\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0) + o_p(1). \tag{5}$$

Note that by Lemma 2, $n^{-1/2}\boldsymbol{Q}_n^{GL}(\hat{\boldsymbol{\gamma}}^{GL}) = o(1)$ and $n^{-1/2}\boldsymbol{Q}_n^{H}(\hat{\boldsymbol{\gamma}}^{H}) = o(1)$ (Lin, Wei and Ying. 1998; Ghosh, 2000, Chapter 6). Then by using arguments in Ying (1993),

$$\boldsymbol{Q}_n^{GL}(\hat{\boldsymbol{\gamma}}^{GL}) = \boldsymbol{Q}_n^{GL}(\boldsymbol{\gamma}_0) + n^{1/2}\boldsymbol{\Lambda}_0^{GL}(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0) + o_p(1),$$
$$\boldsymbol{Q}_n^{H}(\hat{\boldsymbol{\gamma}}^{H}) = \boldsymbol{Q}_n^{H}(\boldsymbol{\gamma}_0) + n^{1/2}\boldsymbol{\Lambda}_0^{H}(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0) + o_p(1),$$

and by consistency of $\hat{\boldsymbol{\gamma}}^{GL}$ and $\hat{\boldsymbol{\gamma}}^{H}$,

$$n^{1/2}(\hat{\boldsymbol{\gamma}}^{GL}-\boldsymbol{\gamma}_0) = -(\boldsymbol{\Lambda}_0^{GL})^{-1}\boldsymbol{Q}_n^{GL}(\boldsymbol{\gamma}_0) + o_p(1),$$
$$n^{1/2}(\hat{\boldsymbol{\gamma}}^{H}-\boldsymbol{\gamma}_0) = -(\boldsymbol{\Lambda}_0^{H})^{-1}\boldsymbol{Q}_n^{H}(\boldsymbol{\gamma}_0) + o_p(1),$$

Then by Slusky's theorem,

$$n^{1/2}(\hat{\boldsymbol{\gamma}}^{GL} - \boldsymbol{\gamma}_0) \xrightarrow{d} N(0, (\boldsymbol{\Lambda}_0^{GL})^{-1}\boldsymbol{\Omega}_0^{GL}((\boldsymbol{\Lambda}_0^{GL})^{-1})^T,$$
$$n^{1/2}(\hat{\boldsymbol{\gamma}}^{H} - \boldsymbol{\gamma}_0) \xrightarrow{d} N(0, (\boldsymbol{\Lambda}_0^{H})^{-1}\boldsymbol{\Omega}_0^{H}((\boldsymbol{\Lambda}_0^{H})^{-1})^T.$$

where $\xrightarrow{d}$ denotes convergence in distribution.

Proof of Lemma 2.

$$\begin{pmatrix} \boldsymbol{G}_n(\boldsymbol{\gamma}) \\ S_n(\eta^{tr}, \boldsymbol{\alpha}) \\ U_n^{GL}(\beta^{tr}, \boldsymbol{\alpha}) \end{pmatrix} = \begin{pmatrix} \boldsymbol{G}_n(\boldsymbol{\gamma}_0) \\ S_n(\eta_0^{tr}, \boldsymbol{\alpha}_0) \\ U_n^{GL}(\beta_0^{tr}, \boldsymbol{\alpha}_0) \end{pmatrix} +$$
$$n^{-1/2}\sum_{i=1}^{n} \begin{pmatrix} \Psi_i(\boldsymbol{\alpha}) \\ \int_{-\infty}^{\infty}\{Z_i - \bar{Z}^{(1)}(t;\eta^{tr}, \boldsymbol{\alpha})\}dN_{1i}^{w}(t;\eta^{tr}, \boldsymbol{\alpha}) \\ \int_{-\infty}^{\infty}\{Z_i - \bar{Z}^{(2)}(t;\beta^{tr}, \boldsymbol{\alpha})\}[dN_{2i}^{w}(t;\beta^{tr}, \boldsymbol{\alpha}) - w_i(\boldsymbol{\alpha})I\{\tilde{D}_i^{**}(\beta^{tr}) \geq t\}r_0(t;\theta^{tr}, \eta^{tr})]dt \end{pmatrix}$$
$$-n^{-1/2}\sum_{i=1}^{n} \begin{pmatrix} \Psi_i(\boldsymbol{\alpha}_0) \\ \int_{-\infty}^{\infty}\{Z_i - \bar{Z}^{(1)}(t;\eta_0^{tr}, \boldsymbol{\alpha}_0)\}dN_{1i}^{w}(t;\eta_0^{tr}, \boldsymbol{\alpha}_0) \\ \int_{-\infty}^{\infty}\{Z_i - \bar{Z}^{(2)}(t;\beta_0^{tr}, \boldsymbol{\alpha}_0)\}[dN_{2i}^{w}(t;\beta_0^{tr}, \boldsymbol{\alpha}_0) - w_i(\boldsymbol{\alpha}_0)I\{\tilde{D}_i^{**}(\beta_0^{tr}) \geq t\}r_0(t;\theta_0^{tr}, \eta_0^{tr})]dt \end{pmatrix}$$
$$+n^{-1/2}\sum_{i=1}^{n} \begin{pmatrix} 0 \\ 0 \\ \int_{-\infty}^{\infty}[\{Z_i - \bar{Z}^{(2)}(t;\beta^{tr}, \boldsymbol{\alpha})\}I\{\tilde{D}_i^{**}(\beta^{tr}) \geq t\}w_i(\boldsymbol{\alpha})r_0(t;\theta^{tr}, \eta^{tr}) \\ -\{Z_i - \bar{Z}^{(2)}(t;\beta_0^{tr}, \boldsymbol{\alpha}_0)\}I\{\tilde{D}_i^{**}(\beta_0^{tr}) \geq t\}w_i(\boldsymbol{\alpha}_0)r_0(t;\theta_0^{tr}, \eta_0^{tr})]dt \end{pmatrix}.$$

M-estimation theory Stefanski and Boos (2002) implies that

$$\boldsymbol{G}_n(\boldsymbol{\alpha}) = \boldsymbol{G}_n(\boldsymbol{\alpha}_0) + n^{1/2}\dot{\boldsymbol{G}}_n(\boldsymbol{\alpha}_0)(\boldsymbol{\alpha} - \boldsymbol{\alpha}_0) + o_p(n^{1/2}||\boldsymbol{\alpha} - \boldsymbol{\alpha}_0||), \tag{6}$$

where $\dot{\boldsymbol{G}}_n(\boldsymbol{\alpha}_0) = [\partial \boldsymbol{G}_n/\partial \boldsymbol{\alpha}]_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0}$. From arguments in Chapter 6 of Ghosh (2000), Theorem 1 and Corollary 1 of Ying (1993) imply that

$$S_n(\eta^{tr}, \boldsymbol{\alpha}) = S_n(\eta_0^{tr}, \boldsymbol{\alpha}_0) + A_1^* n^{1/2}(\boldsymbol{\gamma}^{sub_1} - \boldsymbol{\gamma}_0^{sub_1}) + o_p(1 + n^{1/2}||\boldsymbol{\gamma}^{sub_1} - \boldsymbol{\gamma}_0^{sub_1}||), \quad (7)$$

where $\boldsymbol{\gamma}^{sub_1} = (\boldsymbol{\alpha}^T, \eta^{tr})^T$, $\boldsymbol{\gamma}_0^{sub_1} = (\boldsymbol{\alpha}_0^T, \eta_0^{tr})^T$ and

$$A_1^* = \begin{pmatrix} \boldsymbol{L}_2 & E_1 \end{pmatrix}. \quad (8)$$

Then

$$\begin{aligned}
&\sum_{i=1}^{n} \int_{-\infty}^{\infty} \{Z_i - \bar{Z}^{(2)}(t; \beta^{tr}, \boldsymbol{\alpha})\}[dN_{2i}^w(t; \beta^{tr}, \boldsymbol{\alpha}) - w_i(\boldsymbol{\alpha}) I\{\tilde{D}_i^{**}(\beta^{tr}) \geq t\} r_0(t; \theta^{tr}, \eta^{tr})]dt \\
&= \sum_{i=1}^{n} \int_{-\infty}^{\infty} \{Z_i - \bar{Z}^{(2)}(t; \beta_0^{tr}, \boldsymbol{\alpha}_0)\}[dN_{2i}^w(t; \beta_0^{tr}, \boldsymbol{\alpha}_0) - w_i(\boldsymbol{\alpha}_0) I\{\tilde{D}_i^{**}(\beta_0^{tr}) \geq t\} r_0(t; \theta_0^{tr}, \eta_0^{tr})]dt \\
&+ (\boldsymbol{\alpha} - \boldsymbol{\alpha}_0) \\
&\times \frac{\partial}{\partial \boldsymbol{\alpha}} \sum_{i=1}^{n} \int_{-\infty}^{\infty} \{Z_i - \bar{Z}^{(2)}(t; \beta^{tr}, \boldsymbol{\alpha})\}[dN_{2i}^w(t; \beta^{tr}, \boldsymbol{\alpha}) - w_i(\boldsymbol{\alpha}) I\{\tilde{D}_i^{**}(\beta^{tr}) \geq t\} r_0(t; \theta^{tr}, \eta^{tr})]dt \Big|_{\substack{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0, \\ \beta^{tr}=\beta_0^{tr}}} \\
&+ (\eta^{tr} - \eta_0^{tr}) \times \\
&+ \frac{\partial}{\partial \eta^{tr}} \sum_{i=1}^{n} \int_{-\infty}^{\infty} \{Z_i - \bar{Z}^{(2)}(t; \beta^{tr}, \boldsymbol{\alpha})\}[dN_{2i}^w(t; \beta^{tr}, \boldsymbol{\alpha}) - w_i(\boldsymbol{\alpha}) I\{\tilde{D}_i^{**}(\beta^{tr}) \geq t\} r_0(t; \theta^{tr}, \eta^{tr})]dt \Big|_{\substack{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0, \\ \beta^{tr}=\beta_0^{tr}}} \\
&+ (\theta^{tr} - \theta_0^{tr}) \times \\
&+ \frac{\partial}{\partial \theta^{tr}} \sum_{i=1}^{n} \int_{-\infty}^{\infty} \{Z_i - \bar{Z}^{(2)}(t; \beta^{tr}, \boldsymbol{\alpha})\}[dN_{2i}^w(t; \beta^{tr}, \boldsymbol{\alpha}) - w_i(\boldsymbol{\alpha}) I\{\tilde{D}_i^{**}(\beta^{tr}) \geq t\} r_0(t; \theta^{tr}, \eta^{tr})]dt \Big|_{\substack{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0, \\ \beta^{tr}=\beta_0^{tr}}}.
\end{aligned}$$

Note that $\mathcal{B}$, $\mathcal{N}_0$ and $\mathcal{N}_1$ are neighborhoods of $\boldsymbol{\alpha}_0$, η_0^{tr} and β_0^{tr}, respectively. Then for

$\boldsymbol{\alpha} \in \mathcal{B}$, $\beta^{tr} \in \mathcal{N}_1$, by using arguments from Ying (1993),

$$n^{-1/2} \sum_{i=1}^{n} \int_{-\infty}^{\infty} \{Z_i - \bar{Z}^{(2)}(t; \beta^{tr}, \boldsymbol{\alpha})\}[dN_{2i}^{w}(t; \beta^{tr}, \boldsymbol{\alpha}) - w_i(\boldsymbol{\alpha}) I\{\tilde{D}_i^{**}(\beta^{tr}) \geq t\} r_0(t; \theta^{tr}, \eta^{tr})]dt$$
$$= n^{-1/2} \sum_{i=1}^{n} \int_{-\infty}^{\infty} \{Z_i - \bar{Z}^{(2)}(t; \beta_0^{tr}, \boldsymbol{\alpha}_0)\}[dN_{2i}^{w}(t; \beta_0^{tr}, \boldsymbol{\alpha}_0) - w_i(\boldsymbol{\alpha}_0) I\{\tilde{D}_i^{**}(\beta_0^{tr}) \geq t\} r_0(t; \theta_0^{tr}, \eta_0^{tr})]dt$$
$$+ o_p(1). \tag{9}$$

By multivariate Taylor expansion,

$$n^{-1/2} \sum_{i=1}^{n} \int_{-\infty}^{\infty} [\{Z_i - \bar{Z}^{(2)}(t; \beta^{tr}, \boldsymbol{\alpha})\} I\{\tilde{D}_i^{**}(\beta^{tr}) \geq t\} w_i(\boldsymbol{\alpha}) r_0(t; \theta^{tr}, \eta^{tr}) - \{Z_i - \bar{Z}^{(2)}(t; \beta_0^{tr}, \boldsymbol{\alpha}_0)\}$$
$$\times I\{\tilde{D}_i^{**}(\beta_0^{tr}) \geq t\} w_i(\boldsymbol{\alpha}_0) r_0(t; \theta_0^{tr}, \eta_0^{tr})]dt.$$

$$= n^{-1/2} \sum_{i=1}^{n} \Big[\int_{-\infty}^{\infty} (\boldsymbol{\alpha} - \boldsymbol{\alpha}_0) \frac{\partial}{\partial \boldsymbol{\alpha}} \{Z_i - \bar{Z}^{(2)}(t; \beta^{tr}, \boldsymbol{\alpha})\} I\{\tilde{D}_i^{**}(\beta^{tr}) \geq t\} w_i(\boldsymbol{\alpha}) r_0(t; \theta^{tr}, \eta^{tr}) dt \Big|_{\boldsymbol{\alpha} = \boldsymbol{\alpha}_0, \beta^{tr} = \beta_0^{tr}}$$
$$+ \int_{-\infty}^{\infty} (\theta^{tr} - \theta_0^{tr}) \frac{\partial}{\partial \theta^{tr}} \{Z_i - \bar{Z}^{(2)}(t; \beta^{tr}, \boldsymbol{\alpha})\} I\{\tilde{D}_i^{**}(\beta^{tr}) \geq t\} w_i(\boldsymbol{\alpha}) r_0(t; \theta^{tr}, \eta^{tr}) dt \Big|_{\boldsymbol{\alpha} = \boldsymbol{\alpha}_0, \beta^{tr} = \beta_0^{tr}}$$
$$+ \int_{-\infty}^{\infty} (\eta^{tr} - \eta_0^{tr}) \frac{\partial}{\partial \eta^{tr}} \{Z_i - \bar{Z}^{(2)}(t; \beta^{tr}, \boldsymbol{\alpha})\} I\{\tilde{D}_i^{**}(\beta^{tr}) \geq t\} w_i(\boldsymbol{\alpha}) r_0(t; \theta^{tr}, \eta^{tr}) dt \Big|_{\boldsymbol{\alpha} = \boldsymbol{\alpha}_0, \beta^{tr} = \beta_0^{tr}}$$
$$+ o_p(||n^{1/2}(\boldsymbol{\gamma}^{sub_2} - \boldsymbol{\gamma}_0^{sub_2})||), \tag{10}$$

where $\boldsymbol{\gamma}^{sub_2} = (\boldsymbol{\alpha}^T, (\beta^{tr})^T)^T$ and $\boldsymbol{\gamma}_0^{sub_2} = (\boldsymbol{\alpha}_0^T, (\beta_0^{tr})^T)^T$. Combining (9) and (10) yields that

$$U_n^{GL}(\beta^{tr}, \boldsymbol{\alpha}) = U_n^{GL}(\beta_0^{tr}, \boldsymbol{\alpha}_0) + o_p(1) + A_2^* n^{1/2}(\boldsymbol{\gamma}^{sub_2} - \boldsymbol{\gamma}_0^{sub_2}) + o_p(||n^{1/2}(\boldsymbol{\gamma}^{sub_2} - \boldsymbol{\gamma}_0^{sub_2})||), \tag{11}$$

where $A_2^* = (\boldsymbol{L}_3 \ E_2 \ E_3)$. Combining results from (6) - (11), when c_n converges to 0 in probability,

$$\sup_{||\boldsymbol{\gamma}-\boldsymbol{\gamma}_0||\leq c_n} \frac{||\boldsymbol{Q}_n^{GL}(\boldsymbol{\gamma}) - \boldsymbol{Q}^{GL}(\boldsymbol{\gamma}_0) - n^{1/2}\boldsymbol{\Lambda}_0^{GL}(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0)||}{1+n^{1/2}||\boldsymbol{\gamma}-\boldsymbol{\gamma}_0||} = o_p(1).$$

By Lemma 2 of Honoré and Powell (1994),

$$\sup_{\boldsymbol{\alpha}\in\mathcal{B},\beta^{tr}\in\mathcal{N}_1} \frac{||U_n^H(\beta^{tr},\boldsymbol{\alpha}) - U_n^H(\beta_0^{tr},\boldsymbol{\alpha}_0) - n^{1/2}\lambda(\beta^{tr},\boldsymbol{\alpha})||}{1+n^{1/2}||\lambda(\beta^{tr},\boldsymbol{\alpha})||} = o_p(1).$$

Then Taylor expansion provides the following expression :

$$U_n^H(\beta^{tr},\boldsymbol{\alpha}) = U_n^H(\beta_0^{tr},\boldsymbol{\alpha}_0) + n^{1/2}A_3^*(\boldsymbol{\gamma}^{sub_2}-\boldsymbol{\gamma}_0^{sub_2}) + o_p(1+n^{1/2}||\boldsymbol{\gamma}^{sub_2}-\boldsymbol{\gamma}_0^{sub_2}||), \tag{12}$$

where $A_3^* = (\boldsymbol{L}_4\ E_4\ E_5)$. Combining results from (6), (7), (8) and (12), when c_n converges to 0 in probability,

$$\sup_{||\boldsymbol{\gamma}-\boldsymbol{\gamma}_0||\leq c_n} \frac{||\boldsymbol{Q}_n^{H}(\boldsymbol{\gamma}) - \boldsymbol{Q}^{H}(\boldsymbol{\gamma}_0) - n^{1/2}\boldsymbol{\Lambda}_0^{H}(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0)||}{1+n^{1/2}||\boldsymbol{\gamma}-\boldsymbol{\gamma}_0||} = o_p(1).$$

The first part of the lemma is proved. The second part of the lemma follows immediately from Chapter 6 in Ghosh (2000). □

5. Proof of Theorem 4:

The proof of Theorem 4 is almost identical to Cho, Hu and Ghosh (2018). Let $\boldsymbol{Q}_n(\boldsymbol{\gamma})$ be either $\boldsymbol{Q}_n^{GL}(\boldsymbol{\gamma})$ or $\boldsymbol{Q}_n^{H}(\boldsymbol{\gamma})$. Let $\hat{\boldsymbol{\gamma}}$ be either $\hat{\boldsymbol{\gamma}}^{GL}$ or $\hat{\boldsymbol{\gamma}}^{H}$. Let $\boldsymbol{\gamma}^*$ be either solution of $\boldsymbol{\gamma}^{GL*}$ or $\boldsymbol{\gamma}^{H*}$. Recall that

$$\boldsymbol{Q}_n(\boldsymbol{\gamma}) = -n^{-1/2}\sum_{i=1}^{n}\boldsymbol{P}_iA_i \tag{13}$$

where $\boldsymbol{P}_i$ is the empirical influence function for the limiting distribution of $\boldsymbol{Q}_n(\boldsymbol{\gamma}_0)$ (Peng and Fine, 2006) and A_i are standard normal random variables. $\boldsymbol{\gamma}^*$ is a solution of (13). In the proof of asymptotic normality we can show that

$$\boldsymbol{Q}_n(\boldsymbol{\gamma}^*) = \boldsymbol{Q}_n(\hat{\boldsymbol{\gamma}}) + n^{1/2}\boldsymbol{\Lambda}_0(\boldsymbol{\gamma}^* - \hat{\boldsymbol{\gamma}}) + o_p(1)$$
$$n^{1/2}(\boldsymbol{\gamma}^* - \hat{\boldsymbol{\gamma}}) = -\boldsymbol{\Lambda}_0^{-1} n^{-1/2} \sum_{i=1}^{n} \boldsymbol{P}_i A_i + o_p(1).$$

where $\boldsymbol{\Lambda}_0$ is either $\boldsymbol{\Lambda}_0^{GL}$ or $\boldsymbol{\Lambda}_0^{H}$. Since the observed data are independent and identically distributed, given observed data, the only random term is standard normal random variable. Then given the observed data, $n^{1/2}(\boldsymbol{\gamma}^* - \hat{\boldsymbol{\gamma}})$ is asymptotic normal distribution with mean 0 and same covariance matrix as $n^{1/2}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0)$ Hence given observed data, conditional distribution of $n^{1/2}(\boldsymbol{\gamma}^* - \hat{\boldsymbol{\gamma}})$ is asymptotically equivalent to unconditional distribution of $n^{1/2}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0)$.

References

[1] Cho, Y., Hu, C. and Ghosh, D. (2018). Covariate adjustment using propensity scores for dependent censoring problems in the accelerated failure time model. *Statistics in Medicine* 37 (3): 390-404. doi: 10.1002/sim.7513

[2] Ferguson, T. S. (1996). *A Course in Large Sample Theory*, London: Chapman and Hall.

[3] Ghosh, D. and Lin, D. Y. (2003). Semiparametric analysis of recurrent events data in the presence of dependent censoring. *Biometrics* 59 (4), 877-885. doi: 10.1111/j.0006-341X.2003.00102.x.

[4] Honoré, B. E. and Powell, J. L. (1994). Pairwise difference estimators of censored and truncated regression models. *Journal of Econometrics* 64 (1-2): 241-278. doi : 10.1016/0304-4076(94)90065-5

[5] Hsieh, J.-J., Ding, A. A. and Wang, W. (2011). Regression analysis for recurrent events data under dependent censoring. *Biometrics* 67 (3): 719-729. doi: 10.1111/j.1541-0420.2010.01497.x.

[6] Lin, D. Y., Wei, L. J., and Ying, Z. (1998). Accelerated failure time models for counting processes. *Biometrika*, 85 (3): 605-618. doi : 10.1093/biomet/85.3.605

[7] Peng, L. and Fine, J. P. (2006). Rank estimation of accelerated lifetime models with dependent censoring. *Journal of the American Statistical Association* 101 (475): 1085-1093. doi: 10.1198/016214506000000131.

[8] Pollard, D. (1990). Empirical processes: theory and applications. In: *NSF-CBMS regional conference series in probability and statistics*, Institute of Mathematical Statistics and the American Statistical Association. pp. i-86.

[9] Shorack, G. R. and Wellner, J. A. (1986) *Empirical Processes with Applications to Statistics.* John Wiley and Sons.

[10] Stefanski, L. A. and Boos, D. D. (2002). The calculus of M-estimation. *The American Statistician*, 56 (1): 29-38. doi : 10.1198/000313002753631330.

[11] Ying, Z. (1993). A large sample study of rank estimation for censored regression data, *The Annals of Statistics*, 21 (1): 76-99.

[12] Zhu Y. *Data-adaptive approaches to modeling propensity scores in causal inference problems.* Ph.D. Thesis, The Pennsylvania State University, 2013.

[13] Zhu, Y., Ghosh, D., Mitra, N., and Mukherjee, B. (2014). A data-adaptive strategy for inverse weighted estimation of causal effects, *Health Services Outcomes and Research Methodology*, 14 (3): 69-91. doi: 10.1007/s10742-014-0124-y

[14] Ghosh, D. (2000). Nonparametric and semiparametric analysis of recurrent events in the presence of terminal events and dependent censoring, Ph.D. Thesis. Seattle, WA: University of Washington.