Supplementary Document for
Marginal Structural Cox Models with Case-Cohort Sampling

Hana Lee[1], Michael G. Hudgens[2], Jianwen Cai[2], and Stephen R. Cole[2]

[1] *Brown University and* [2] *University of North Carolina at Chapel Hill*

This supplementary material contains three main sections: S1 provides detailed proofs corresponding to consistency and asymptotic normality of MSCM parameter estimators along with consistency of the cumulative baseline hazard estimator proposed in Section 5.1 of the main text; S2 describes (i) how a MSCM can easily be fit via inverse probability weighting for either the full cohort or case-cohort setting using standard survival analysis software, such as R or SAS, and (ii) additional simulation study results including performance of the proposed baseline cumulative hazard estimator; S3 provides a summary of notation introduced in the main text and the supplement. For clarity, we display theorems shown in the main text again in this document. Note that some equation numbers presented in this document are the same as the equation numbers in the main text. Therefore equation numbers from the main text are denoted by *.

S1. Proofs for Theorems 3.1 - 3.6 and the consistency of the proposed cumulative baseline hazard estimator.

This section consists of two main subsections: S1-1 provides proofs corresponding to consistency of MSCM parameter estimators $\hat{\beta}, \tilde{\beta}$, and β^* along with a proof for the consistency of the proposed baseline cumulative hazard estimator $\tilde{\Lambda}_{\hat{W}}(\cdot)$; S1-2 presents proofs for asymptotic normality results.

S1-1. Consistency Proofs

Recall that we define $\hat{\beta}$, $\tilde{\beta}$, and β^* to be solutions to $\partial l(\beta, 1; \hat{W})/\partial\beta = 0$, $\partial\tilde{l}(\beta, 1; \hat{W})/\partial\beta = 0$, and $\partial l^*(\beta, 1; \hat{W})/\partial\beta = 0$, respectively in the main text.

Consider the following processes

$$X(\beta,t;W) = n^{-1}\{l(\beta,t;W) - l(\beta_0,t;W)\} \tag{1.1}$$
$$= n^{-1}\sum_{i=1}^{n}\int_0^t W_i(u)\Big[(\beta-\beta_0)'A_i(u) - \log\frac{\sum_{l=1}^{n} W_l(u)Y_l(u)r\{\beta' A_l(u)\}}{\sum_{l=1}^{n} W_l(u)Y_l(u)r\{\beta_0' A_l(u)\}}\Big]dN_i(u),$$
$$\tilde{X}(\beta,t;W) = n^{-1}\{\tilde{l}(\beta,t;W) - \tilde{l}(\beta_0,t;W)\} \tag{1.2}$$
$$= n^{-1}\sum_{i\in C}\int_0^t W_i(u)\Big[(\beta-\beta_0)'A_i(u) - \log\frac{\sum_{l\in\tilde{C}} W_l(u)Y_l(u)r\{\beta' A_l(u)\}}{\sum_{l\in\tilde{C}} W_l(u)Y_l(u)r\{\beta_0' A_l(u)\}}\Big]dN_i(u)$$

corresponding to (2.7*) and (2.8*), respectively, in the main text. We will first show that $X(\beta,t;\hat{W})$ and (1.1) are asymptotically equivalent, and so are $\tilde{X}(\beta,t;\hat{W})$ and (1.2). Thus, further technical developments will be made based on (1.1) and (1.2). We then show that (1.1) and (1.2) at $t=1$ converge in probability to functions of β which are concave with a unique maximum β_0 under certain conditions. Using the same argument as in Andersen and Gill (1982), it follows that $\hat{\beta}\to_p \beta_0$ and $\tilde{\beta}\to_p \beta_0$. That $\beta^*\to_p\beta_0$ can be shown analogously by using $X^*(\beta,t;W) = n^{-1}\{l^*(\beta,t;W) - l^*(\beta_0,t;W)\}$. Asymptotic normality of $\hat{\beta}$ and $\tilde{\beta}$ will be shown via asymptotic normality of score statistics corresponding to (2.7*) and (2.8*).

Theorem 3.1. (Consistency of full cohort MSCM estimator $\hat{\beta}$) Under conditions A-F, $\hat{\beta}\to_p\beta_0$.

Proof. Consider (1.1) and its compensator counterpart $K(\beta,t;W)$ which is

$$K(\beta,t;W) = n^{-1}\sum_{i=1}^{n}\int_0^t W_i(u)\Big[(\beta-\beta_0)'A_i(u) - \log\{\frac{S^{(0)}_{W_{(1)}}(\beta,u)}{S^{(0)}_{W_{(1)}}(\beta_0,u)}\}\Big]\lambda_i(u)du$$

where $\lambda_i(t)$ is given as in (2.11*). We start by showing that

$$|\{X(\beta,t;\hat{W}) - K(\beta,t;\hat{W})\} - \{X(\beta,t;W) - K(\beta,t;W)\}| \to_p 0 \tag{1.3}$$

so that we can consider the asymptotic behavior of $X(\beta,t;W) - K(\beta,t;W)$ instead of $X(\beta,t;\hat{W}) - K(\beta,t;\hat{W})$ to prove consistency of $\hat{\beta}$. To prove (1.3), first note the term $|\{X(\beta,t;\hat{W})-K(\beta,t;\hat{W})\}-\{X(\beta,t;W)-K(\beta,t;W)\}|$ in (1.3) equals

$$\Big|n^{-1}\sum_{i=1}^{n}\int_0^1\Big[\hat{W}_i(u)(\beta-\beta_0)'A_i(u)-\hat{W}_i(u)\log\Big\{\frac{S^{(0)}_{\hat{W}_{(1)}}(\beta,u)}{S^{(0)}_{\hat{W}_{(1)}}(\beta_0,u)}\Big\}\Big]dM_i(u)$$
$$-n^{-1}\sum_{i=1}^{n}\int_0^1\Big[W_i(u)(\beta-\beta_0)'A_i(u)-W_i(u)\log\Big\{\frac{S^{(0)}_{W_{(1)}}(\beta,u)}{S^{(0)}_{W_{(1)}}(\beta_0,u)}\Big\}\Big]dM_i(u)\Big|.$$

Replacing $W_i(u)$ in front of $\log\{S^{(0)}_{W_{(1)}}(\beta,u)/S^{(0)}_{W_{(1)}}(\beta_0,u)\}$ with $W_i(u)-\hat{W}_i(u)+\hat{W}_i(u)$ and rearranging terms yields

$$\Big|n^{-1}\sum_{i=1}^{n}\int_0^1\{\hat{W}_i(u)-W_i(u)\}(\beta-\beta_0)'A_i(u)dM_i(u) \tag{1.4}$$
$$-n^{-1}\sum_{i=1}^{n}\int_0^1\{\hat{W}_i(u)-W_i(u)\}\log\Big\{\frac{S^{(0)}_{W_{(1)}}(\beta,u)}{S^{(0)}_{W_{(1)}}(\beta_0,u)}\Big\}dM_i(u)$$
$$-n^{-1}\sum_{i=1}^{n}\int_0^1\hat{W}_i(u)\log\Big\{\frac{S^{(0)}_{\hat{W}_{(1)}}(\beta,u)}{S^{(0)}_{\hat{W}_{(1)}}(\beta_0,u)}\Big/\frac{S^{(0)}_{W_{(1)}}(\beta,u)}{S^{(0)}_{W_{(1)}}(\beta_0,u)}\Big\}dM_i(u)\Big|.$$

Each term in (1.4) is a local square integrable martingale since $g(W_i(\cdot),A_i(\cdot))$ is predictable for any continuous function $g(\cdot)$ due to predictableness of $W_i(\cdot)$ and $A_i(\cdot)$. Because $\hat{W}_i(\cdot)$ is also bounded and predictable, the same argument can be made for $g_1(\hat{W}_i(\cdot),A_i(\cdot))$ and $g_2(W_i(\cdot),\hat{W}_i(\cdot))$ for any continuous functions $g_1(\cdot)$ and $g_2(\cdot)$. We will show that the variation process of each martingale in (1.4) converges in probability to zero, thus proving (1.3).

Let $B_1(\beta,t)$ be the variation process of the first martingale in (1.4). Then

$$B_1(\beta,t)=n^{-2}\sum_{i=1}^{n}\int_0^t\{\hat{W}_i(u)-W_i(u)\}^2(\beta-\beta_0)'A_i(u)^{\otimes 2}(\beta-\beta_0)\lambda_i(u)du$$
$$=n^{-2}\sum_{i=1}^{n}\int_0^t\{\hat{W}_i(u)-W_i(u)\}^2(\beta-\beta_0)'Y_i(u)r^{(2)}\{\beta_0'A_i(u)\}(\beta-\beta_0)\lambda_0(u)du$$
$$\le n^{-1}\int_0^t\mathsf{M}^2_{\hat{W}}(\beta-\beta_0)'\Big[n^{-1}\sum_{i=1}^{n}Y_i(u)r^{(2)}\{\beta_0'A_i(u)\}\Big](\beta-\beta_0)\lambda_0(u)du$$
$$=n^{-1}\mathsf{M}^2_{\hat{W}}\int_0^t(\beta-\beta_0)'S^{(2)}(\beta_0,u)(\beta-\beta_0)\lambda_0(u)du$$

which converges in probability to zero due to conditions A, B, D, and F. The second equality is owing to (2.11*), and the inequality comes from replacing $\{\hat{W}_i(u) - W_i(u)\}^2$ by its supremum value $\mathsf{M}^2_{\hat{W}}$ shown in condition A of the main text. Let $B_2(\beta, t)$ be the variation process of the second martingale term in (1.4). Then

$$\begin{aligned}
B_2(\beta,t) =& n^{-2}\sum_{i=1}^n \int_0^t \left\{\hat{W}_i(u) - W_i(u)\right\}^2 \left\{\log S^{(0)}_{W_{(1)}}(\beta,u) - \log S^{(0)}_{W_{(1)}}(\beta_0,u)\right\}^2 \lambda_i(u)du \\
\le& n^{-1}\int_0^t \mathsf{M}^2_{\hat{W}} \left\{\log S^{(0)}_{W_{(1)}}(\beta,u) - \log S^{(0)}_{W_{(1)}}(\beta_0,u)\right\}^2 S^{(0)}(\beta_0,u)\lambda_0(u)du
\end{aligned}$$

which converges to zero due to conditions A, B, D, and F. Lastly, let the variation process of the third martingale term in (1.4) be $B_3(\beta, t)$. Then

$$\begin{aligned}
B_3(\beta,t) =& n^{-2}\sum_{i=1}^n \int_0^t \hat{W}_i(u)^2 \Big[\{\log S^{(0)}_{\hat{W}_{(1)}}(\beta,u) - \log S^{(0)}_{W_{(1)}}(\beta,u)\} \\
&- \{\log S^{(0)}_{\hat{W}_{(1)}}(\beta_0,u) - \log S^{(0)}_{W_{(1)}}(\beta_0,u)\}\Big]^2 \lambda_i(u)du \\
=& n^{-1}\int_0^1 \Big[\{\log S^{(0)}_{\hat{W}_{(1)}}(\beta,u) - \log S^{(0)}_{W_{(1)}}(\beta,u)\} \\
&- \{\log S^{(0)}_{\hat{W}_{(1)}}(\beta_0,u) - \log S^{(0)}_{W_{(1)}}(\beta_0,u)\}\Big]^2 S^{(0)}_{\hat{W}_{(2)}}(\beta_0,u)\lambda_0(u)du \\
\le& n^{-1}\int_0^1 \Big[\sup_{\beta,u} |\log S^{(0)}_{\hat{W}_{(1)}}(\beta,u) - \log S^{(0)}_{W_{(1)}}(\beta,u)|^2 \\
&+ 2\sup_{\beta,u} |\log S^{(0)}_{\hat{W}_{(1)}}(\beta,u) - \log S^{(0)}_{W_{(1)}}(\beta,u)| \sup_u |\log S^{(0)}_{\hat{W}_{(1)}}(\beta_0,u) - \log S^{(0)}_{W_{(1)}}(\beta_0,u)| \\
&+ \sup_u |\log S^{(0)}_{\hat{W}_{(1)}}(\beta_0,u) - \log S^{(0)}_{W_{(1)}}(\beta_0,u)|^2\Big] S^{(0)}_{\hat{W}_{(2)}}(\beta_0,u)\lambda_0(u)du
\end{aligned}$$

which converges in probability to zero due to conditions A, B, D, F, and by the continuous mapping theorem. It follows that $X(\beta, t; \hat{W}) - K(\beta, t; \hat{W})$ and $X(\beta, t; W) - K(\beta, t; W)$ in (1.3) are asymptotically equivalent processes. Thereby we proceed to describe asymptotic behavior of the process $X(\beta, t; W) - K(\beta, t; W)$. Hereafter for notational convenience we suppress W when writing $X(\beta, t; W)$ and $K(\beta, t; W)$.

Now consider $X(\beta, t) - K(\beta, t)$, which equals to

$$n^{-1}\sum_{i=1}^n \int_0^t W_i(u)\Big[(\beta-\beta_0)'A_i(u) - \log\Big\{\frac{S^{(0)}_{W_{(1)}}(\beta,u)}{S^{(0)}_{W_{(1)}}(\beta_0,u)}\Big\}\Big]dM_i(u),$$

which is a martingale. After some calculation, it can be shown that its variation process $B(\beta, t)$ can be simplified as

$$n^{-1} \int_0^1 \Big[(\beta - \beta_0)' S^{(2)}_{W_{(2)}}(\beta_0, u)(\beta - \beta_0) - 2(\beta - \beta_0)' S^{(1)}_{W_{(2)}}(\beta_0, u) \log \Big\{ \frac{S^{(0)}_{W_{(1)}}(\beta, u)}{S^{(0)}_{W_{(1)}}(\beta_0, u)} \Big\} + \left\{ \log \Big(\frac{S^{(0)}_{W_{(1)}}(\beta, u)}{S^{(0)}_{W_{(1)}}(\beta_0, u)} \Big) \right\}^2 S^{(0)}_{W_{(2)}}(\beta_0, u) \Big] \lambda_0(u) du \tag{1.5}$$

where each term inside the integral converges in probability to a function of finite quantities $s^{(j)}_{W_{(k)}}$ on $\beta \in \mathcal{B}$ in view of conditions D and F. Therefore, (1.5) converges in probability to zero. It follows that $X(\beta, t)$ and $K(\beta, t)$ converge in probability to the same limit by the Lenglart inequality, i.e., that $\mathrm{pr}[\sup_{t,\beta} ||X(\beta, t) - K(\beta, t) > \eta||] \le \delta/\eta^2 + \mathrm{pr}[B(\beta, 1) > \delta]$ for all $\delta, \eta > 0$. Therefore, to investigate asymptotic properties of $X(\beta, 1)$, consider asymptotic properties of $K(\beta, 1)$ instead:

$$K(\beta, 1) \to_p \int_0^1 \Big[(\beta - \beta_0)' s^{(1)}_{W_{(1)}}(\beta_0, u) - \log \Big\{ \frac{s^{(0)}_{W_{(1)}}(\beta, u)}{s^{(0)}_{W_{(1)}}(\beta_0, u)} \Big\} s^{(0)}_{W_{(1)}}(\beta_0, u) \Big] \lambda_0(u) du$$

by (2.11*). Let $K_l(\beta, 1)$ be the limiting quantity shown in the above. Then

$$\frac{\partial K_l(\beta, 1)}{\partial \beta} = \int_0^1 \Big[s^{(1)}_{W_{(1)}}(\beta_0, u) - \frac{s^{(1)}_{W_{(1)}}(\beta, u)}{s^{(0)}_{W_{(1)}}(\beta, u)} s^{(0)}_{W_{(1)}}(\beta_0, u) \Big] \lambda_0(u) du$$

which is zero at $\beta = \beta_0$. In addition, $\partial^2 K_l(\beta, 1)/\partial \beta^2$ is

$$\begin{aligned} & - \int_0^1 \Big[\frac{s^{(2)}_{W_{(1)}}(\beta, u) s^{(0)}_{W_{(1)}}(\beta, u) - s^{(1)}_{W_{(1)}}(\beta, u)^{\otimes 2}}{s^{(0)}_{W_{(1)}}(\beta, u)^2} \Big] s^{(0)}_{W_{(1)}}(\beta_0, u) \lambda_0(u) du \\ &= - \int_0^1 v_{W_{(1)}}(\beta, u) s^{(0)}_{W_{(1)}}(\beta_0, u) \lambda_0(u) du \end{aligned}$$

which equals to $-\Sigma_{W_{(1)}}$ and is negative definite when $\beta = \beta_0$ based on condition F. Therefore $K(\beta, 1)$ converges to a concave function having unique maximum at β_0. This enables us to make use of Theorem II.1 in Andersen and Gill (1982) that proves in probability convergence of $X(\beta,1)$ to the same concave function of β as does $K(\beta,1)$, with a unique maximum at $\beta=\beta_0$. Then $\hat{\beta} \to_p \beta_0$. □

As described in Section 3 of the main text, consistency of $\tilde{\beta}$ can be shown using similar arguments as in Theorem 3.1. We start from showing that $\tilde{X}(\beta,t)$ converges in probability to $K(\beta,t)$. Then the same argument as in the proof of Theorem 3.1 can be made. $|\tilde{X}(\beta,t)-K(\beta,t)|$ will be decomposed into two terms, $|X(\beta,t)-K(\beta,t)|$ plus a term that is asymptotically negligible.

Theorem 3.2. (Consistency of case-cohort MSCM estimator $\tilde{\beta}$) Under conditions A-G, $\tilde{\beta} \to_p \beta_0$.

Proof. First, $|\tilde{X}(\beta,t)-K(\beta,t)|$ can be rewritten as

$$\begin{aligned}
&\Big| n^{-1}\int_0^t \sum_{i=1}^n W_i(u)(\beta-\beta_0)' A_i(u) dM_i(u) \\
&\quad - n^{-1}\int_0^t \sum_{i=1}^n W_i(u) \log\Big\{ \frac{\tilde{S}^{(0)}_{W_{(1)}}(\beta,u)}{\tilde{S}^{(0)}_{W_{(1)}}(\beta_0,u)} \Big\} dN_i(u) \\
&\quad + n^{-1}\int_0^t \sum_{i=1}^n W_i(u) \log\Big\{ \frac{S^{(0)}_{W_{(1)}}(\beta,u)}{S^{(0)}_{W_{(1)}}(\beta_0,u)} \Big\} \lambda_i(u) du \Big| \\
\le & |X(\beta,t)-K(\beta,t)| \\
&\quad + \Big| n^{-1}\int_0^t \sum_{i=1}^n W_i(u) \Big\{ \log\Big(\frac{\tilde{S}^{(0)}_{W_{(1)}}(\beta,u)}{\tilde{S}^{(0)}_{W_{(1)}}(\beta_0,u)} \Big) - \log\Big(\frac{S^{(0)}_{W_{(1)}}(\beta,u)}{S^{(0)}_{W_{(1)}}(\beta_0,u)} \Big) \Big\} dN_i(u) \Big|.
\end{aligned}$$

We have shown that $|X(\beta,t)-K(\beta,t)| \to_p 0$. The remaining term can be decomposed as

$$\begin{aligned}
&\Big| n^{-1}\int_0^t \Big[\sum_{i=1}^n W_i(u) \Big\{ \log\Big(\frac{\tilde{S}^{(0)}_{W_{(1)}}(\beta,u)}{\tilde{S}^{(0)}_{W_{(1)}}(\beta_0,u)} \Big) - \log\Big(\frac{S^{(0)}_{W_{(1)}}(\beta,u)}{S^{(0)}_{W_{(1)}}(\beta_0,u)} \Big) \Big\} dM_i(u) \Big] \qquad (1.6) \\
&+ n^{-1}\int_0^t \sum_{i=1}^n \Big[W_i(u) \Big\{ \log\Big(\frac{\tilde{S}^{(0)}_{W_{(1)}}(\beta,u)}{\tilde{S}^{(0)}_{W_{(1)}}(\beta_0,u)} \Big) - \log\Big(\frac{S^{(0)}_{W_{(1)}}(\beta,u)}{S^{(0)}_{W_{(1)}}(\beta_0,u)} \Big) \Big\} \lambda_i(u) du \Big] \Big|.
\end{aligned}$$

Then the second term in (1.6) can easily be shown to converge in probability to zero in view of conditions C, D, F, and G-3. Also the martingale in (1.6) converges in probability to zero because its variation process is

$$\begin{aligned}
&\Big|n^{-2}\int_0^t \sum_{i=1}^n W_i(u)^2 \Big[\Big\{ \log \tilde{S}^{(0)}_{W_{(1)}}(\beta,u) - \log S^{(0)}_{W_{(1)}}(\beta,u)\Big\} \\
&\quad - \Big\{ \log \tilde{S}^{(0)}_{W_{(1)}}(\beta_0,u) - \log S^{(0)}_{W_{(1)}}(\beta_0,u)\Big\}\Big]^2 \lambda_i(u)du\Big| \\
\leq&\Big|n^{-1}\int_0^t \Big[\sup_{\beta,u} |\log \tilde{S}^{(0)}_{W_{(1)}}(\beta,u) - \log S^{(0)}_{W_{(1)}}(\beta,u)| \\
&\quad + \sup_u |\log \tilde{S}^{(0)}_{W_{(1)}}(\beta_0,u) - \log S^{(0)}_{W_{(1)}}(\beta_0,u)|\Big]^2 S^{(0)}_{W_{(2)}}(\beta_0,u)\lambda_0(u)du\Big|
\end{aligned}$$

which converges in probability to zero, again by (2.11*) with conditions C, D, F, and G-3. Note that sum of supremums in the integrand (which can be taken outside the integral) converges in probability to zero by conditions D and G-3. □

Theorem 3.3. (Asymptotic equivalence between two case-cohort MSCM estimators) Under conditions A-G, $\tilde{\beta} - \beta^* \to_p 0$.

Proof. We sketch a proof of Theorem 3.3. Consider the following process

$$X^*(\beta,t) = n^{-1}\{l^*(\beta,t) - l^*(\beta_0,t)\}.$$

Then $X^*(\beta,t) = n^{-1}\{\tilde{l}(\beta,t) - \tilde{l}(\beta_0,t)\} + o_p(1)$ because $n^{-1}l^*(\beta,t) = n^{-1}\tilde{l}(\beta,t) + o_p(1)$. Therefore, $X^*(\beta,t)$ and $\tilde{X}(\beta,t)$ are asymptotically equivalent processes and we can repeat the proof of Theorem 3.2 using $X^*(\beta,t)$ instead of $\tilde{X}(\beta,t)$. □

Lastly, we show consistency of the MSCM cumulative baseline hazard estimator presented in Section 5.1 of the main text: Under conditions A-G, $\sup_{t\in[0,1]} |\tilde{\Lambda}_{\hat{W}}(\tilde{\beta},t) - \Lambda_0(t)| \to_p 0$, where $\Lambda_0(t) = \Lambda_0(\beta_0,t)$.

Proof. Recall:

$$\tilde{\Lambda}_{\hat{W}}(\tilde{\beta},t) = \tilde{n}n^{-1}\int_0^t \Big[\sum_{i\in\tilde{C}} \hat{W}_i(u)Y_i(u)r\{\tilde{\beta}' A_i(u)\}\Big]^{-1} \sum_{i=1}^n \hat{W}_i(u)dN_i(u).$$

Note that $\tilde{\Lambda}_{\hat{W}}(\tilde{\beta},t) - \Lambda_0(t)$ equals

$$\sum_{i=1}^n \int_0^t n^{-1}\frac{\hat{W}_i(u)}{\tilde{S}^{(0)}_{\hat{W}_{(1)}}(\tilde{\beta},u)}dM_i(u) + \int_0^t \frac{S^{(0)}_{\hat{W}_{(1)}}(\beta_0,u) - \tilde{S}^{(0)}_{\hat{W}_{(1)}}(\tilde{\beta},u)}{\tilde{S}^{(0)}_{\hat{W}_{(1)}}(\tilde{\beta},u)}\lambda_0(u)du \quad (1.7)$$

in view of (2.11*). The first term of (1.7) is a local square integrable martingale because the term $n^{-1}\hat{W}_i(u)/\tilde{S}^{(0)}_{\hat{W}_{(1)}}(\tilde{\beta},u)$ is a bounded predictable process based on conditions A, B, F, and G. The variation process is $n^{-1}\int_0^t S^{(0)}_{\hat{W}_{(1)}}(\tilde{\beta},u)/\{\tilde{S}^{(0)}_{\hat{W}_{(1)}}(\tilde{\beta},u)\}^2$ $\lambda_0(u)du$ which converges in probability to zero due to conditions A, B, D, F, and G, and thus the first term of (1.7) converges in probability to zero. The second term in (1.7) can be written as

$$\int_0^t \frac{S^{(0)}_{\hat{W}_{(1)}}(\beta_0,u)-\tilde{S}^{(0)}_{\hat{W}_{(1)}}(\beta_0,u)+\tilde{S}^{(0)}_{\hat{W}_{(1)}}(\beta_0,u)-\tilde{S}^{(0)}_{\hat{W}_{(1)}}(\tilde{\beta},u)}{\tilde{S}^{(0)}_{\hat{W}_{(1)}}(\tilde{\beta},u)}\lambda_0(u)du. \qquad (1.8)$$

Then uniform consistency of $\hat{W}$, consistency of $\tilde{\beta}$ along with conditions C, D, F, and G imply in probability convergence of (1.8) to zero. □

S1-2. Asymptotic Normality of Marginal Structural Cox Model Estimators

In this section we show that the full cohort and the case-cohort MSCM estimators are asymptotically normally distributed. We start from showing that the full and the case-cohort score statistics are asymptotically normal. Recall that we referred to the score process under the full cohort setting as *the full cohort MSCM score process*, and the score process under the case-cohort setting as *the case-cohort cohort MSCM score process.*

Theorem 3.4. (Asymptotic normality of the full cohort MSCM score statistic) Under conditions A-F,

$$n^{-1/2}U(\beta_0,1)\to_d N(0,\Sigma_U)$$

where $\Sigma_U=\Sigma_{W_{(2)}}+\Delta_{W_{(1)},W_{(2)}}$ with

$$\Delta_{W_{(1)},W_{(2)}}=\int_0^1\{e_{W_{(2)}}(\beta_0,u)-e_{W_{(1)}}(\beta_0,u)\}^{\otimes 2}s^{(0)}_{W_{(2)}}(\beta_0,u)\lambda_0(u)du. \qquad (1.9)$$

Proof. Let $U(\beta_0,t)$ be the full cohort MSCM score process at time t. Then

$$n^{-1/2}U(\beta_0,t)=n^{-1/2}\partial l(\beta,t)/\partial\beta\Big|_{\beta=\beta_0} \qquad (1.10)$$

$$= n^{-1/2} \sum_{i=1}^{n} \int_0^t W_i(u) \Big[A_i(u) - \frac{S^{(1)}_{W_{(1)}}(\beta_0, u)}{S^{(0)}_{W_{(1)}}(\beta_0, u)} \Big] dN_i(u)$$

$$= n^{-1/2} \sum_{i=1}^{n} \int_0^t W_i(u) \Big[A_i(u) - E_{W_{(1)}}(\beta_0, u) \Big] dM_i(u).$$

The third equality follows from (2.10*) and the fact that

$$n^{-1/2} \sum_{i=1}^{n} \int_0^t W_i(u)[A_i(u) - E_{W_{(1)}}(\beta_0, u)]\lambda_i(u) du = 0, \tag{1.11}$$

based on (2.11*). Similar to the proof of Theorem 3.1, we start by showing that $|n^{-1/2}U(\beta, t; \hat{W}) - n^{-1/2}U(\beta, t; W)| \to_p 0$ so that the rest of the arguments can be made based on $n^{-1/2}U(\beta, t; W)$. It can be seen that

$$\begin{aligned} & |n^{-1/2}U(\beta, t; \hat{W}) - n^{-1/2}U(\beta, t; W)| \\ =& n^{-1/2} \sum_{i=1}^{n} \int_0^t \Big[\{\hat{W}_i(u) - W_i(u)\} A_i(u) dM_i(u) \qquad (1.12) \\ & - n^{-1/2} \sum_{i=1}^{n} \int_0^t \{\hat{W}_i(u) E_{\hat{W}_{(1)}}(\beta, u) - W_i(u) E_{W_{(1)}}(\beta, u)\} \Big], \end{aligned}$$

which is a sum of two local square integrable martingales. The variation process of the first term in (1.12) is given by

$$\begin{aligned} & n^{-1} \sum_{i=1}^{n} \int_0^t \{\hat{W}_i(u) - W_i(u)\}^2 A_i^{\otimes 2}(u) \lambda_i(u) du \\ =& n^{-1} \sum_{i=1}^{n} \int_0^t \{\hat{W}_i(u) - W_i(u)\}^2 Y_i(u) r^{(2)}\{\beta_0' A_i(t)\} \lambda_0(u) du \\ \leq & \int_0^t \mathsf{M}^2_{\hat{W}} \Big[n^{-1} \sum_{i=1}^{n} Y_i(u) r^{(2)}\{\beta_0' A_i(u)\} \Big] \lambda_0(u) du \\ =& \mathsf{M}^2_{\hat{W}} \int_0^t S^{(2)}(\beta_0, u) \lambda_0(u) du, \end{aligned}$$

which converges in probability to zero by conditions A, C, D, E, and F. The second term in (1.12) can be rewritten as

$$\begin{aligned} & n^{-1/2} \sum_{i=1}^{n} \int_0^t \{\hat{W}_i(u) - W_i(u)\} E_{\hat{W}_{(1)}}(\beta, u) \qquad (1.13) \\ & \qquad + W_i(u)\{E_{\hat{W}_{(1)}}(\beta, u) - E_{W_{(1)}}(\beta, u)\} dM_i(u) \Big], \end{aligned}$$

which is, again a sum of two local square integrable martingales. The variation process of the first term in (1.13) is

$$\begin{aligned} &n^{-1}\sum_{i=1}^{n}\int_0^t \{\hat{W}_i(u) - W_i(u)\}^2 E_{\hat{W}_{(1)}}(\beta,u)^{\otimes 2}\lambda_i(u)du \\ \le& \mathsf{M}_{\hat{W}}^2 \int_0^t E_{\hat{W}_{(1)}}(\beta,u)^{\otimes 2} S^{(0)}(\beta_0,u)\lambda_0(u)du, \end{aligned}$$

which converges in probability to zero in view of conditions A, C, D, E, and F. The variation process of the second term in (1.13) can be written as

$$\begin{aligned} &n^{-1}\sum_{i=1}^{n}\int_0^t W_i^2(u)\{E_{\hat{W}_{(1)}}(\beta,u) - E_{W_{(1)}}(\beta,u)\}^{\otimes 2}\lambda_i(u)du \\ \le& \int_0^t \mathsf{M}_1^2\{E_{\hat{W}_{(1)}}(\beta,u) - E_{W_{(1)}}(\beta,u)\}^{\otimes 2} S^{(0)}(\beta_0,u)\lambda_0(u)du. \end{aligned} \tag{1.14}$$

Later we will show that $|E_{\hat{W}_{(1)}}(\beta,u) - E_{W_{(1)}}(\beta,u)| \to_p 0$ uniformly in β and t. For now, assume that

$$\sup_{(\beta,t)\in\mathcal{B}\times[0,1]} |E_{\hat{W}_{(1)}}(\beta,t) - E_{W_{(1)}}(\beta,t)| \to_p 0 \tag{1.15}$$

holds. Then it can be shown that (1.14) is less than equal to

$$\begin{aligned} &\mathsf{M}_1^2\int_0^t \sup_{\beta,u}|E_{\hat{W}_{(1)}}(\beta,u) - E_{W_{(1)}}(\beta,u)|^{\otimes 2}\{S^{(0)}(\beta_0,u) - s^{(0)}(\beta_0,u) + s^{(0)}(\beta_0,u)\}\lambda_0(u)du \\ =& \mathsf{M}_1^2 \sup_{\beta,u}|E_{\hat{W}_{(1)}}(\beta,u) - E_{W_{(1)}}(\beta,u)|^{\otimes 2}\int_0^t \{S^{(0)}(\beta_0,u) - s^{(0)}(\beta_0,u)\}\lambda_0(u)du \\ &+ \mathsf{M}_1^2 \sup_{\beta,u}|E_{\hat{W}_{(1)}}(\beta,u) - E_{W_{(1)}}(\beta,u)|^{\otimes 2}\int_0^t s^{(0)}(\beta_0,u)\lambda_0(u)du \\ \le& \mathsf{M}_1^2 \sup_{\beta,u}|E_{\hat{W}_{(1)}}(\beta,u) - E_{W_{(1)}}(\beta,u)|^{\otimes 2}\sup_u |S^{(0)}(\beta_0,u) - s^{(0)}(\beta_0,u)|\int_0^t \lambda_0(u)du \\ &+ \mathsf{M}_1^2 \sup_{\beta,u}|E_{\hat{W}_{(1)}}(\beta,u) - E_{W_{(1)}}(\beta,u)|^{\otimes 2}\int_0^t s^{(0)}(\beta_0,u)\lambda_0(u)du, \end{aligned}$$

which converges uniformly in β and t based on (1.15) and conditions B, C, D, and F. To show that (1.15) is satisfied, let's rewrite $|E_{\hat{W}_{(1)}}(\beta,t) - E_{W_{(1)}}(\beta,t)|$ as

$$
\begin{aligned}
&|E_{\hat{W}_{(1)}}(\beta,t) - e_{W_{(1)}}(\beta,t) - E_{W_{(1)}}(\beta,t) + e_{W_{(1)}}(\beta,t)| \\
\leq&|E_{\hat{W}_{(1)}}(\beta,t) - e_{W_{(1)}}(\beta,t)| + |E_{W_{(1)}}(\beta,t) - e_{W_{(1)}}(\beta,t)| \\
=&\Big|\frac{S^{(1)}_{\hat{W}_{(1)}}(\beta,t)s^{(0)}_{W_{(1)}}(\beta,t) - s^{(1)}_{W_{(1)}}(\beta,t)S^{(0)}_{\hat{W}_{(1)}}(\beta,t)}{S^{(0)}_{\hat{W}_{(1)}}(\beta,t)s^{(0)}_{W_{(1)}}(\beta,t)}\Big| \qquad (1.16)\\
&+\Big|\frac{S^{(1)}_{W_{(1)}}(\beta,t)s^{(0)}_{W_{(1)}}(\beta,t) - s^{(1)}_{W_{(1)}}(\beta,t)S^{(0)}_{W_{(1)}}(\beta,t)}{S^{(0)}_{W_{(1)}}(\beta,t)s^{(0)}_{W_{(1)}}(\beta,t)}\Big|. \qquad (1.17)
\end{aligned}
$$

First, we can show that (1.16) converges in probability to zero uniformly in (β,t); because by conditions A, D(ii), and condition F (i.e., $s^{(0)}_{W_{(1)}}(\beta,t)$ is bounded away from zero), there exists an integer N_0 such that when $n > N_0$, $S^{(0)}_{\hat{W}_{(1)}}(\beta,t)$ is bounded away from 0. Therefore when $n > N_0$, $\min_{\beta,t}|S^{(0)}_{\hat{W}_{(1)}}(\beta,t)s^{(0)}_{W_{(1)}}(\beta,t)|$ exists and we will denote it by M. Then when $n > N_0$,

$$
\begin{aligned}
&\Big|\frac{S^{(1)}_{\hat{W}_{(1)}}(\beta,t)s^{(0)}_{W_{(1)}}(\beta,t) - s^{(1)}_{W_{(1)}}(\beta,t)S^{(0)}_{\hat{W}_{(1)}}(\beta,t)}{S^{(0)}_{\hat{W}_{(1)}}(\beta,t)s^{(0)}_{W_{(1)}}(\beta,t)}\Big| \\
\leq&\frac{1}{M}\Big|S^{(1)}_{\hat{W}_{(1)}}(\beta,t)s^{(0)}_{W_{(1)}}(\beta,t) - s^{(1)}_{W_{(1)}}(\beta,t)S^{(0)}_{\hat{W}_{(1)}}(\beta,t)\Big| \\
\leq&\frac{1}{M}\Big|S^{(1)}_{\hat{W}_{(1)}}(\beta,t)s^{(0)}_{W_{(1)}}(\beta,t) - s^{(1)}_{W_{(1)}}(\beta,t)s^{(0)}_{W_{(1)}}(\beta,t)\Big| + \frac{1}{M}\Big|s^{(1)}_{W_{(1)}}(\beta,t)S^{(0)}_{\hat{W}_{(1)}}(\beta,t) - s^{(1)}_{W_{(1)}}(\beta,t)s^{(0)}_{W_{(1)}}(\beta,t)\Big| \\
\leq&\frac{1}{M}\Big|s^{(0)}_{W_{(1)}}(\beta,t)\Big|\Big|S^{(1)}_{\hat{W}_{(1)}}(\beta,t) - s^{(1)}_{W_{(1)}}(\beta,t)\Big| + \frac{1}{M}\Big|s^{(1)}_{W_{(1)}}(\beta,t)\Big|\Big|S^{(0)}_{\hat{W}_{(1)}}(\beta,t) - s^{(0)}_{W_{(1)}}(\beta,t)\Big| \\
&\to_p 0 \quad \text{uniformly in } (\beta,t) \in \mathcal{B}\times[0,1],
\end{aligned}
$$

because it is clear that $\sup_{\beta,t}|S^{(j)}_{\hat{W}_{(1)}}(\beta,t) - s^{(j)}_{W_{(1)}}(\beta,t)| \to_p 0$ from condition A and D(ii). A similar argument can be made to show that (1.17) uniformly converges in probability to zero. This enables us to establish asymptotic results using true IPWs, instead of using estimated weights, and therefore we will proceed with the MSCM score function evaluated at the true weights.

Set $H_i(t) = n^{-1/2}W_i(t)[A_i(t) - E_{W_{(1)}}(\beta_0,t)]$ for $i = 1,...,n$. This is a locally bounded predictable process. Therefore, (1.10) is a local square integrable martingale. To apply the martingale central limit theorem to the local square integrable martingale, we show that $n^{-1/2}U(\beta_0,1)=\sum_{i=1}^n\int_0^1 H_i(t)dM_i(t)$ satisfies (i) $\int_0^1\sum_{i=1}^n H_{ij}(t)^2 I\{|H_{ij}(t)|>\epsilon\}\lambda_i(t)dt\to_p 0$ for any $\epsilon>0$ (the Lindeberg condition), and that (ii) the variation process of (1.10) evaluated at $t=1$ converges in probability to a finite quantity. Condition (i) is satisfied because of condition E. To see if condition (ii) is satisfied, consider the variation process of $n^{-1/2}U(\beta_0,1)$,

$$\begin{aligned}
&\sum_{i=1}^n\int_0^1 H_i(u)^{\otimes 2}\lambda_i(u)du\\
&=\int_0^1 n^{-1}\sum_{i=1}^n W_i(u)^2\Big[A_i(u)-E_{W_{(1)}}(\beta_0,u)\Big]^{\otimes 2}\lambda_i(u)du\\
&=\int_0^1 n^{-1}\sum_{i=1}^n\Big[W_i(u)^2Y_i(u)r^{(2)}\{\beta_0'A_i(u)\}-2W_i(u)^2Y_i(u)r^{(1)}\{\beta_0'A_i(u)\}\{E_{W_{(1)}}(\beta_0,u)\}'\\
&\quad+W_i(u)^2Y_i(u)r\{\beta_0'A_i(u)\}E_{W_{(1)}}(\beta_0,u)^{\otimes 2}\Big]\lambda_0(u)du\\
&=\int_0^1\Big[S^{(2)}_{W_{(2)}}(\beta_0,u)-2S^{(1)}_{W_{(2)}}(\beta_0,u)\{E_{W_{(1)}}(\beta_0,u)\}'+S^{(0)}_{W_{(2)}}(\beta_0,u)E_{W_{(1)}}(\beta_0,u)^{\otimes 2}\Big]\lambda_0(u)du\\
&=\int_0^1\Big[\frac{S^{(2)}_{W_{(2)}}(\beta_0,u)}{S^{(0)}_{W_{(2)}}(\beta_0,u)}-2\frac{S^{(1)}_{W_{(2)}}(\beta_0,u)}{S^{(0)}_{W_{(2)}}(\beta_0,u)}\Big\{\frac{S^{(1)}_{W_{(1)}}(\beta_0,u)}{S^{(0)}_{W_{(1)}}(\beta_0,u)}\Big\}'+\Big\{\frac{S^{(1)}_{W_{(1)}}(\beta_0,u)}{S^{(0)}_{W_{(1)}}(\beta_0,u)}\Big\}^{\otimes 2}\Big]S^{(0)}_{W_{(2)}}(\beta_0,u)\lambda_0(u)du\\
&=\int_0^1\Big[\Big\{\frac{S^{(2)}_{W_{(2)}}(\beta_0,u)}{S^{(0)}_{W_{(2)}}(\beta_0,u)}-\Big(\frac{S^{(1)}_{W_{(2)}}(\beta_0,u)}{S^{(0)}_{W_{(2)}}(\beta_0,u)}\Big)^{\otimes 2}\Big\}+\Big\{\Big(\frac{S^{(1)}_{W_{(2)}}(\beta_0,u)}{S^{(0)}_{W_{(2)}}(\beta_0,u)}\Big)^{\otimes 2}\\
&\quad-2\frac{S^{(1)}_{W_{(2)}}(\beta_0,u)}{S^{(0)}_{W_{(2)}}(\beta_0,u)}\Big(\frac{S^{(1)}_{W_{(1)}}(\beta_0,u)}{S^{(0)}_{W_{(1)}}(\beta_0,u)}\Big)'+\Big(\frac{S^{(1)}_{W_{(1)}}(\beta_0,u)}{S^{(0)}_{W_{(1)}}(\beta_0,u)}\Big)^{\otimes 2}\Big\}\Big]S^{(0)}_{W_{(2)}}(\beta_0,u)\lambda_0(u)du\\
&=\int_0^1\Big[V_{W_{(2)}}(\beta_0,u)+\{E_{W_{(2)}}(\beta_0,u)-E_{W_{(1)}}(\beta_0,u)\}^{\otimes 2}\Big]S^{(0)}_{W_{(2)}}(\beta_0,u)\lambda_0(u)du.
\end{aligned}$$

Finally we can see that the variation process of $n^{-1/2}U(\beta_0,1)$ converges in probability to

$$\Sigma_{W_{(2)}}+\Delta_{W_{(1)},W_{(2)}}\equiv\Sigma_U \tag{1.18}$$

where $\Delta_{W_{(1)},W_{(2)}}$ is given in (1.9). Based on conditions C and F, (1.18) is a finite quantity. Therefore, the full cohort MSCM score statistic converges in distribution to a Gaussian process with mean zero and the limiting variance-covariance process Σ_U by the martingale central limit theorem. When $W_i(t)\equiv 1$

for all $i = 1, ..., n$ and $t \in [0, 1]$, $\Delta_{W_{(1)}, W_{(2)}}$ becomes zero and (1.18) equals to Σ which is the asymptotic variance of the score process under the full cohort. □

Before we prove Theorem 3.5, we present the following Proposition which is taken from Self and Prentice (1998):

Proposition 1. *(Self and Prentice, 1988) Let* $\mathbf{X}_n = (X_{1n}, ..., X_{nn})$ *and* $\boldsymbol{\delta}_n = (\delta_{1n}, ..., \delta_{nn})$ *be independent random variables such that:*
(I) δ_n *is a vector of* $\tilde{n}$ *ones and* $n - \tilde{n}$ *zeros, each possible configuration of zeros and ones is equally likely and* $\tilde{n}/n \rightarrow_p \alpha \in (0, 1)$.
(II) For some scalar functions of $\mathbf{X}_n$, $f_{in}(\mathbf{X}_n)$, *and for any* $\epsilon > 0$,

$$n^{-1} \sum_{i=1}^{n} [f_{in}(\mathbf{X}_n) - f_{\cdot n}(\mathbf{X}_n)]^2 I\{|f_{in}(\mathbf{X}_n) - f_{\cdot n}(\mathbf{X}_n)| > n^{1/2}\epsilon\} \rightarrow_p 0,$$

and $\mathbf{S}^2_{f_n} \rightarrow_p \sigma^2_f > 0$, *where* $f_{\cdot n}(\mathbf{X}_n) = n^{-1} \sum_{i=1}^{n} f_{in}(\mathbf{X}_n)$ *and*

$$\mathbf{S}^2_{f_n} = n^{-1} \sum_{i=1}^{n} [f_{in}(\mathbf{X}_n) - f_{\cdot n}(\mathbf{X}_n)]^2.$$

(III) The scalar functions of $\mathbf{X}_n$, $g_n(\mathbf{X}_n)$, *converge in distribution to a Gaussian random variable with mean zero and variance* σ^2_g.

Let $h_n(\mathbf{X}_n, \boldsymbol{\delta}_n) = n^{1/2}[\tilde{n}^{-1} \sum_{i=1}^{n} \delta_{in} f_{in}(\mathbf{X}_n) - f_{\cdot n}(\mathbf{X}_n)]$, *then* $\{g_n(\mathbf{X}_n), h_n(\mathbf{X}_n, \boldsymbol{\delta}_n)\}$ *converge in distribution to a bivariate Gaussian random variable with mean zero and covariance matrix given by*

$$\begin{bmatrix} \sigma^2_g & \mathbf{0} \\ \mathbf{0} & (1-\alpha)\alpha^{-1}\sigma^2_f \end{bmatrix}.$$

Theorem 3.5. (Asymptotic normality of the case-cohort MSCM score statistic) Under conditions A-G,

$$n^{-1/2}\tilde{U}(\beta_0, 1) \rightarrow_d N(\mathbf{0}, \Sigma_{\tilde{U}})$$

where $\Sigma_{\tilde{U}} = \Sigma_U + \Delta_\alpha$,

$$\Delta_\alpha = \int_0^1 \int_0^1 G(\beta_0, x, v)\lambda_0(x)\lambda_0(v)dxdv, \quad (1.19)$$

and $G(\beta_0, x, v)$ is given by

$$G(\beta_0, x, v) = (1-\alpha)\alpha^{-1}\Big[h^{(1)}(\beta_0, x, v) - e_{W_{(1)}}(\beta_0, x)h^{(2)}(\beta_0, x, v)' \qquad (1.20)$$
$$- h^{(2)}(\beta_0, v, x)e_{W_{(1)}}(\beta_0, v)' + e_{W_{(1)}}(\beta_0, x)e_{W_{(1)}}(\beta_0, v)'h^{(0)}(\beta_0, x, v)\Big],$$

where

$$h^{(0)}(\beta, x, v) = q^{(0)}(\beta, x, v) - s^{(0)}_{W_{(1)}}(\beta, x)s^{(0)}_{W_{(1)}}(\beta, v)$$
$$h^{(1)}(\beta, x, v) = q^{(1)}(\beta, x, v) - s^{(1)}_{W_{(1)}}(\beta, x)s^{(1)}_{W_{(1)}}(\beta, v)'$$
$$h^{(2)}(\beta, x, v) = q^{(2)}(\beta, x, v) - s^{(0)}_{W_{(1)}}(\beta, x)s^{(1)}_{W_{(1)}}(\beta, v).$$

Proof. The score process corresponding to (2.8*), which will be referred to as *case-cohort MSCM score process*, is defined by

$$n^{-1/2}\tilde{U}(\beta_0, t) = n^{-1/2}\partial\tilde{l}(\beta, t)/\partial\beta\Big|_{\beta=\beta_0} \qquad (1.21)$$
$$= n^{-1/2}\sum_{i=1}^{n}\int_0^t W_i(u)\Big[A_i(u) - \tilde{E}_{W_{(1)}}(\beta_0, u)\Big]dN_i(u).$$

Replacing $\tilde{E}_{W_{(1)}}(\beta_0, u)$ in (1.21) with $E_{W_{(1)}}(\beta_0, u) + \tilde{E}_{W_{(1)}}(\beta_0, u) - E_{W_{(1)}}(\beta_0, u)$, we obtain

$$n^{-1/2}\sum_{i=1}^{n}\int_0^t W_i(u)\Big[A_i(u) - E_{W_{(1)}}(\beta_0, u)\Big]dM_i(u) \qquad (1.22)$$
$$- \int_0^t D_n(u)\lambda_0(u)du$$
$$- \int_0^t D_n(u)\{S^{(0)}_{W_{(1)}}(\beta_0, u)/\tilde{S}^{(0)}_{W_{(1)}}(\beta_0, u) - 1\}\lambda_0(u)du$$
$$+ \int_0^t n^{1/2}\{E_{W_{(1)}}(\beta_0, u) - e_{W_{(1)}}(\beta_0, u)\}\{\tilde{S}^{(0)}_{W_{(1)}}(\beta_0, u) - S^{(0)}_{W_{(1)}}(\beta_0, u)\}$$
$$\times S^{(0)}_{W_{(1)}}(\beta_0, u)/\tilde{S}^{(0)}_{W_{(1)}}(\beta_0, u)\lambda_0(u)du$$
$$- n^{-1/2}\sum_{i=1}^{n}\int_0^t W_i(u)\Big[\tilde{E}_{W_{(1)}}(\beta_0, u) - E_{W_{(1)}}(\beta_0, u)\Big]dM_i(u),$$

where

$$D_n(t) = n^{1/2}\Big[\Big\{\tilde{S}^{(1)}_{W_{(1)}}(\beta_0, t) - S^{(1)}_{W_{(1)}}(\beta_0, t)\Big\} - e_{W_{(1)}}(\beta_0, t)\Big\{\tilde{S}^{(0)}_{W_{(1)}}(\beta_0, t)$$
$$- S^{(0)}_{W_{(1)}}(\beta_0, t)\Big\}\Big]S^{(0)}_{W_{(1)}}(\beta_0, t).$$

To begin with, we first show how (1.21) can be rewritten as (1.22). Replacing $\tilde{E}_{W_{(1)}}(\beta_0, u)$ in (1.21) with $E_{W_{(1)}}(\beta_0, u) + \tilde{E}_{W_{(1)}}(\beta_0, u) - E_{W_{(1)}}(\beta_0, u)$, we obtain

$$
\begin{aligned}
n^{-1/2}\tilde{U}(\beta_0, t) = {} & n^{-1/2}\sum_{i=1}^{n}\int_0^t W_i(u)\Big[A_i(u) - E_{W_{(1)}}(\beta_0, u)\Big]dM_i(u) \\
& - n^{1/2}\int_0^t \Big[\tilde{E}_{W_{(1)}}(\beta_0, u) - E_{W_{(1)}}(\beta_0, u)\Big]S^{(0)}_{W_{(1)}}(\beta_0, u)\lambda_0(u)du \\
& - n^{-1/2}\sum_{i=1}^{n}\int_0^t W_i(u)\Big[\tilde{E}_{W_{(1)}}(\beta_0, u) - E_{W_{(1)}}(\beta_0, u)\Big]dM_i(u).
\end{aligned} \tag{1.23}
$$

We can rewrite $n^{1/2}[E_{W_{(1)}}(\beta_0, u) - E_{W_{(1)}}(\beta_0, u)\}]S^{(0)}_{W_{(1)}}(\beta_0, u)$ in the second term of (1.23) as follows:

$$
\begin{aligned}
& n^{1/2}[\tilde{E}_{W_{(1)}}(\beta_0, t) - E_{W_{(1)}}(\beta_0, t)]S^{(0)}_{W_{(1)}}(\beta_0, t) \\
= & n^{1/2}\Big[\frac{\tilde{S}^{(1)}_{W_{(1)}}(\beta_0, t)}{\tilde{S}^{(0)}_{W_{(1)}}(\beta_0, t)} - \frac{S^{(1)}_{W_{(1)}}(\beta_0, t)}{S^{(0)}_{W_{(1)}}(\beta_0, t)}\Big]S^{(0)}_{W_{(1)}}(\beta_0, t) \\
= & n^{1/2}\Big[\Big\{\frac{\tilde{S}^{(1)}_{W_{(1)}}(\beta_0, t)}{\tilde{S}^{(0)}_{W_{(1)}}(\beta_0, t)} - \frac{S^{(1)}_{W_{(1)}}(\beta_0, t)}{\tilde{S}^{(0)}_{W_{(1)}}(\beta_0, t)}\Big\} \\
& \quad + \Big\{\frac{S^{(1)}_{W_{(1)}}(\beta_0, t)}{\tilde{S}^{(0)}_{W_{(1)}}(\beta_0, t)} - \frac{S^{(1)}_{W_{(1)}}(\beta_0, t)}{S^{(0)}_{W_{(1)}}(\beta_0, t)}\Big\}\Big]S^{(0)}_{W_{(1)}}(\beta_0, t) \\
= & n^{1/2}\Big[\frac{1}{\tilde{S}^{(0)}_{W_{(1)}}(\beta_0, t)}\Big\{\tilde{S}^{(1)}_{W_{(1)}}(\beta_0, t) - S^{(1)}_{W_{(1)}}(\beta_0, t)\Big\} \\
& \quad + \frac{S^{(1)}_{W_{(1)}}(\beta_0, t)}{\tilde{S}^{(0)}_{W_{(1)}}(\beta_0, t)S^{(0)}_{W_{(1)}}(\beta_0, t)}\Big\{S^{(0)}_{W_{(1)}}(\beta_0, t) - \tilde{S}^{(0)}_{W_{(1)}}(\beta_0, t)\Big\}\Big]S^{(0)}_{W_{(1)}}(\beta_0, t) \\
= & n^{1/2}\Big[\Big\{\tilde{S}^{(1)}_{W_{(1)}}(\beta_0, t) - S^{(1)}_{W_{(1)}}(\beta_0, t)\Big\} - E_{W_{(1)}}(\beta_0, t)\Big\{\tilde{S}^{(0)}_{W_{(1)}}(\beta_0, t) - S^{(0)}_{W_{(1)}}(\beta_0, t)\Big\}\Big] \\
& \quad \times S^{(0)}_{W_{(1)}}(\beta_0, t)/\tilde{S}^{(0)}_{W_{(1)}}(\beta_0, t) \\
= & D_n(t) + D_n(t)\{S^{(0)}_{W_{(1)}}(\beta_0, t)/\tilde{S}^{(0)}_{W_{(1)}}(\beta_0, t) - 1\} \\
& - n^{1/2}\{E_{W_{(1)}}(\beta_0, u) - e_{W_{(1)}}(\beta_0, u)\}\{\tilde{S}^{(0)}_{W_{(1)}}(\beta_0, u) - S^{(0)}_{W_{(1)}}(\beta_0, u)\}S^{(0)}_{W_{(1)}}(\beta_0, u)/\tilde{S}^{(0)}_{W_{(1)}}(\beta_0, u),
\end{aligned}
$$

which leads to the form given by (1.22). Then the fourth term in (1.22) can be shown to converge in probability to zero uniformly in t because its integrand converges to zero uniformly in t in view of the stability conditions D, F, and G-3, combined with the Slutsky's theorem. The fifth term in (1.22) is a local square integrable martingale with the variation process

$$\int_0^1 \left[\tilde{E}_{W_{(1)}}(\beta_0, u) - E_{W_{(1)}}(\beta_0, u)\right]^{\otimes 2} S^{(0)}_{W_{(2)}}(\beta_0, u)\lambda_0(u)du$$

which converges in probability to zero by conditions C, D, F, and G-3. Therefore, if we can show that the first term in (1.22) and $D_n(u)$ converge jointly in distribution to independent Gaussian random variables then it implies that $D_n(u)$ converges in distribution to a Gaussian (the joint in distribution convergence will be shown through Proposition 1 in the main text). This further implies that the third term in (1.22) converges in probability to zero and that the first two terms in (1.22) converge jointly in distribution to independent Gaussian random variables. Then we can claim that the limiting covariance function of the case-cohort MSCM score process is given by the sum of each of the limiting covariances. The proof for Theorem 3.5 is lengthy thus we break it into three parts: Part 1) shows how Proposition 1 can be used to prove Theorem 3.5. Part 2) justifies the application of Proposition 1 by showing that conditions (I) to (III) in the Proposition are met. Part 3) shows detailed calculations to obtain the limiting covariance function of the MSCM case-cohort score process.

Part 1) Application of Proposition 1 Consider application of Proposition 1 to $D_n(t)$. In particular, X_{in} represents $\{W_i(u), Y_i(u), N_i(u), A_i(u); u \in [0, 1]\}$, $f_{in}(\mathbf{X}_n)$ represents a linear combination of elements of $W_i(t)Y_i(t)r\{\beta_0' A_i(t)\}$ and $W_i(t)Y_i(t)r^{(1)}\{\beta_0' A_i(t)\}$. The explicit form of $f_{in}(\mathbf{X}_n)$ will be presented below.

Our goal is to show that the difference of the first two terms in (1.22), which is given by

$$\begin{aligned}
&n^{-1/2}\sum_{i=1}^n \int_0^t W_i(u)[A_i(u) - E_{W_{(1)}}(\beta_0, u)]dM_i(u) - \int_0^t D_n(u)\lambda_0(u)du \\
=&B_n(t) - \int_0^t D_n(u)\lambda_0(u)du \\
=&B_n(t) - C_n(t),
\end{aligned}$$

converges in distribution to a finite dimensional Gaussian random variable where $B_n(\cdot)$, $C_n(\cdot)$, and $D_n(\cdot)$ are defined by

$$B_n(t) = n^{-1/2}\sum_{i=1}^{n}\int_0^t W_i(u)[A_i(u) - E_{W_{(1)}}(\beta_0,u)]dM_i(u), \tag{1.24}$$

$$C_n(t) = \int_0^t D_n(u)\lambda_0(u)du, \quad \text{and} \tag{1.25}$$

$$\begin{aligned} D_n(u) = n^{1/2}\Big[&\{\tilde{S}^{(1)}_{W_{(1)}}(\beta_0,u) - S^{(1)}_{W_{(1)}}(\beta_0,u)\} - e_{W_{(1)}}(\beta_0,u) \\ &\times\{\tilde{S}^{(0)}_{W_{(1)}}(\beta_0,u) - S^{(0)}_{W_{(1)}}(\beta_0,u)\}\Big] S^{(0)}_{W_{(1)}}(\beta_0,u). \end{aligned} \tag{1.26}$$

Let $g_n(\mathbf{X}_n)$ be a linear combination of elements of the MSCM full cohort score process (B_n), i.e., for any constants c_j ($j = 1,..,p$),

$$g_n(\mathbf{X}_n) = n^{-1/2}\sum_{i=1}^{n}\sum_{j=1}^{p} c_j \int_0^t W_i(u)[A_{i,j}(u) - E_{W_{(1)},j}(\beta_0,u)]dM_i(u)$$

where the subscript j denotes the jth component of a vector. Also, let $h_n(\mathbf{X}_n,\delta_n)$ be a linear combination of elements of D_n, i.e., for any constants $d_j (j = 1,..,p)$, $f_{in}(\mathbf{X}_n)$ is given by

$$\begin{aligned} f_{in}(\mathbf{X}_n) = \sum_{j=1}^{p} d_j\Big[&W_i(u_j)Y_i(u_j)r_j^{(1)}\{\beta_0' A_i(u_j)\} \\ &- e_{W(1),j}(\beta_0,u_j)W_i(u_j)Y_i(u_j)r\{\beta_0' A_i(u_j)\}\Big]. \end{aligned} \tag{1.27}$$

Then (1.27) leads to the desired form of $h_n(\mathbf{X}_n,\delta_n)$:

$$\begin{aligned} h_n(\mathbf{X}_n,\delta_n) =& n^{1/2}[\tilde{n}^{-1}\sum_{i=1}^{n}\delta_{in} f_{in}(\mathbf{X}_n) - f_{\cdot n}(\mathbf{X}_n)] \\ =& n^{1/2}\Big[\sum_{j=1}^{p} d_j\{\tilde{S}^{(1)}_{W_{(1)},j}(\beta_0,u_j) - e_{W_{(1)},j}(\beta_0,u_j)\tilde{S}^{(0)}_{W_{(1)}}(\beta_0,u_j)\} \\ &- \sum_{j=1}^{p} d_j\{S^{(1)}_{W_{(1)},j}(\beta_0,u_j) - e_{W_{(1)},j}(\beta_0,u_j)S^{(0)}_{W_{(1)}}(\beta_0,u_j)\}\Big] \\ =& n^{1/2}\sum_{j=1}^{p} d_j\Big[\{\tilde{S}^{(1)}_{W_{(1)},j}(\beta_0,u_j) - S^{(1)}_{W_{(1)},j}(\beta_0,u_j)\} \\ &- e_{W_{(1)},j}(\beta_0,u_j)\{\tilde{S}^{(0)}_{W_{(1)}}(\beta_0,u_j) - S^{(0)}_{W_{(1)}}(\beta_0,u_j)\}\Big], \end{aligned}$$

which is a linear combination of elements of D_n where each jth component can be evaluated at possibly different time points u_j, i.e., $h_n(\mathbf{X}_n,\delta_n) = \sum_{j=1}^{p} d_j D_{n,j}(u_j)$.

Assume that $f_{in}(\mathbf{X}_n)$ and $g_n(\mathbf{X}_n)$ satisfy conditions (I) to (III) stated in Proposition 1, which will be shown later in the Part 2. Then by varying c_j and d_j, we can show that any chosen elements of B_n and D_n jointly converge in distribution to an independent bivariate Gaussian process by application of Proposition 1. For example, consider $c_1 = d_1 = 1$ and $c_2 = ... = c_p = d_2 = ... = d_p = 0$. Then Proposition 1 states that the first element of B_n and the first element of D_n converge jointly in distribution to an independent bivariate Gaussian. In iterative fashion, we can show that jth element of B_n and kth element of D_n converge in distribution to an independent bivariate Gaussian for all combinations of $(j, k) \in [1, 2, ..., p] \times [1, 2, ..., p]$. Therefore, B_n and D_n converge in distribution to independent processes. We have shown that B_n, the MSCM full cohort score process, converges in distribution to a Gaussian process. Therefore, what we have left to show is that D_n converges in distribution to a Gaussian process (and later to show that C_n converges in distribution to a Gaussian process).

In the above arguments we have shown that, for any d_j $(j = 1, ..., p)$, $\sum_{j=1}^{p} d_j D_{n,j}$ converges in distribution to a univariate Gaussian because $f_{in}(\mathbf{X}_n)$ satisfies conditions in Proposition 1 for any d_j (which, as we mentioned above, will be shown in the Part 2). Therefore, it follows that D_n converges in distribution to a multidimensional mean zero Gaussian random variable by the Cramer-Wold device. As in Self and Prentice (1988), the fact that linear functionals of the Gaussian processes are Gaussian combined with the fact that $\lambda_0(\cdot)$ is absolutely continuous with respect to the Lebesque measure leads to that C_n converges to a Gaussian random variable, say C. Then it follows that $B_n - C_n$ converges to a mean zero Gaussian random variable with covariance $\Sigma_U + \Delta_\alpha$, as the limiting covariance of C_n will be shown to equal Δ_α later in the Part 3.

In the next two parts, we verify that $f_{in}(\mathbf{X}_n)$ and $g_n(\mathbf{X}_n)$ satisfy conditions in Proposition 1, and show the explicit form of limiting covariance structure of C_n respectively.

Part 2) Conditions in Proposition 1 Condition (I) in Proposition 1 is satisfied by condition G-1(i) in the main text and the fact that the subcohort is selected by the simple random sampling without replacement. The first subcondition of condition (II) in Proposition 1 follows from the inequality used by Andersen and Gill (1982) and Self and Prentice (1988),

$$|a-b|^2 I\{|a-b|>\epsilon\} \leq 4|a|^2 I\{|a|>\epsilon/2\} + 4|b|^2 I\{|b|>\epsilon/2\}, \tag{1.28}$$

by letting $n^{-1/2} f_{in}(\mathbf{X}_n)$ be a and $n^{-1/2} f_{\cdot n}(\mathbf{X}_n)$ be b, combined with conditions D, F, and G-1(ii). Recall that condition (II) of Proposition 1 has the following two subconditions:
For any $\epsilon > 0$,

$$n^{-1} \sum_{i=1}^{n} [f_{in}(\mathbf{X}_n) - f_{\cdot n}(\mathbf{X}_n)]^2 I_{\{|f_{in}(\mathbf{X}_n) - f_{\cdot n}(\mathbf{X}_n)| > n^{1/2}\epsilon\}} \rightarrow_p 0, \text{ and} \tag{1.29}$$

$$\mathbf{S}_{f_n}^2 = n^{-1} \sum_{i=1}^{n} [f_{in}(\mathbf{X}_n) - f_{\cdot n}(\mathbf{X}_n)]^2 \rightarrow_p \sigma_f. \tag{1.30}$$

To show (1.29) based on the inequality (1.28), we need to show that for any $\epsilon > 0$,

$$n^{-1} \sum_{i=1}^{n} |f_{in}(\mathbf{X}_n)|^2 I\{|f_{in}(\mathbf{X}_n)| > n^{1/2}\epsilon/2\} \rightarrow_p 0, \quad \text{and} \tag{1.31}$$

$$n^{-1} |f_{\cdot n}(\mathbf{X}_n)|^2 I\{|f_{\cdot n}(\mathbf{X}_n)| > n^{1/2}\epsilon/2\} \rightarrow_p 0. \tag{1.32}$$

To show (1.31), recall condition G-1(ii) in the main text: For any $\epsilon > 0$

$$\sup_t n^{-1} \sum_{i=1}^{n} W_i(t)^2 Y_i(t) r\{\beta_0' A_i(t)\}^2 \times I\{n^{-1/2} W_i(t) Y_i(t) r\{\beta_0' A_i(t)\} > \epsilon\} \rightarrow_p 0, \tag{1.33}$$

$$\sup_t n^{-1} \sum_{i=1}^{n} W_i(t)^2 Y_i(t) ||r^{(1)}\{\beta_0' A_i(t)\}||^2 \times I\{n^{-1/2} W_i(t) Y_i(t) ||r^{(1)}\{\beta_0' A_i(t)\}|| > \epsilon\} \rightarrow_p 0, \tag{1.34}$$

where (1.33) implies

$$\sup_t n^{-1} \sum_{i=1}^{n} W_i(t)^2 Y_i(t) r\{\beta_0' A_i(t)\}^2 ||e_{W_{(1)}}(\beta_0, t)||^2 \times I\{n^{-1/2} W_i(t) Y_i(t) r\{\beta_0' A_i(t)\} ||e_{W_{(1)}}(\beta_0, t)|| > \epsilon\} \rightarrow_p 0. \tag{1.35}$$

It can be shown that (1.34) and (1.35) imply (1.31), by repeatedly applying

(1.28). Also, we can rewrite

$$f_{\cdot n}(\mathbf{X}_n) = \sum_{j=1}^{p} d_j [S^{(1)}_{W_{(1)},j}(\beta_0, t) - e_{W_{(1)},j}(\beta_0, t) S^{(0)}_{W_{(1)}}(\beta_0, t)].$$

Then (1.32) can immediately be seen by the stability property implied by condition D in the main text.

To show (1.30), note that $\mathbf{S}^2_{f_n}$ can be rewritten as

$$\begin{aligned} \mathbf{S}^2_{f_n} &= n^{-1} \sum_{i=1}^{n} [f_{in}(\mathbf{X}_n) - f_{\cdot n}(\mathbf{X}_n)]^2 \\ &= n^{-1} \sum_{i=1}^{n} f_{in}(\mathbf{X}_n)^2 - \{f_{\cdot n}(\mathbf{X}_n)\}^2. \end{aligned} \tag{1.36}$$

For notational and calculational convenience, let

$$f_{in}(\mathbf{X}_n) = \sum_{j=1}^{p} d_j (a_j - b_j)$$

by letting

$$\begin{aligned} a_j &= W_i(u_j) Y_i(u_j) r^{(1)}_j \{\beta_0' A_i(u_j)\}, \quad \text{and} \\ b_j &= e_{W(1),j}(\beta_0, u_j) W_i(u_j) Y_i(u_j) r\{\beta_0' A_i(u_j)\} \end{aligned}$$

then calculate the form of each term in (1.36). First term in (1.36) can be written as follows:

$$\begin{aligned} n^{-1} \sum_{i=1}^{n} f_{in}(\mathbf{X}_n)^2 &= n^{-1} \sum_{i=1}^{n} \Big[\sum_{j=1}^{p} d_j (a_j - b_j) \Big]^2 \\ =& n^{-1} \sum_{i=1}^{n} \Big[\sum_{j=1}^{p} d_j^2 (a_j - b_j)^2 + 2 \sum_{j<k} d_j d_k (a_j - b_j)(a_k - b_k) \Big] \\ =& n^{-1} \sum_{i=1}^{n} \Big[\sum_{j=1}^{p} d_j^2 \Big\{ W_i(u_j)^2 Y_i(u_j) r^{(1)}_j \{\beta_0' A_i(u_j)\}^2 \\ &\quad - 2 W_i(u_j) Y_i(u_j) r^{(1)}_j \{\beta_0' A_i(u_j)\} e_{W(1),j}(\beta_0, u_j) W_i(u_j) Y_i(u_j) r\{\beta_0' A_i(u_j)\} \\ &\quad + e_{W(1),j}(\beta_0, u_j)^2 W_i(u_j)^2 Y_i(u_j) r\{\beta_0' A_i(u_j)\}^2 \Big\} \\ &+ 2 \sum_{j<k} d_j d_k \Big\{ W_i(u_j) Y_i(u_j) r^{(1)}_j \{\beta_0' A_i(u_j)\} W_i(u_k) Y_i(u_k) r^{(1)}_k \{\beta_0' A_i(u_k)\} \\ &\quad - W_i(u_j) Y_i(u_j) r^{(1)}_j \{\beta_0' A_i(u_j)\} e_{W(1),k}(\beta_0, u_k) W_i(u_k) Y_i(u_k) r\{\beta_0' A_i(u_k)\} \end{aligned}$$

$$- e_{W(1),j}(\beta_0, u_j)W_i(u_j)Y_i(u_j)r\{\beta_0' A_i(u_j)\}W_i(u_k)Y_i(u_k)r_k^{(1)}\{\beta_0' A_i(u_k)\}$$
$$+ e_{W(1),j}(\beta_0, u_j)W_i(u_j)Y_i(u_j)r\{\beta_0' A_i(u_j)\}e_{W(1),k}(\beta_0, u_k)W_i(u_k)Y_i(u_k)r\{\beta_0' A_i(u_k)\}\Big\}\Big]$$

Then using $Q^{(j)}(j = 0, 1, 2)$ notation defined in condition G-2, the above equation can be abbreviated as

$$\begin{aligned}
n^{-1}\sum_{i=1}^{n} f_{in}(\mathbf{X}_n)^2 = &\sum_{j=1}^{p} d_j^2 \Big\{ Q^{(1)}_{(j,j)}(\beta_0, u_j, u_j) - 2e_{W(1),j}(\beta_0, u_j)Q^{(2)}_j(\beta_0, u_j, u_j) \\
&\quad + e_{W(1),j}(\beta_0, u_j)^2 Q^{(0)}(\beta_0, u_j, u_j)\Big\} \\
&+2\sum_{j<k} d_j d_k \Big\{ Q^{(1)}_{(j,k)}(\beta_0, u_j, u_k) - e_{W(1),k}(\beta_0, u_k)Q^{(2)}_j(\beta_0, u_k, u_j) \\
&\quad - e_{W(1),j}(\beta_0, u_j)Q^{(2)}_k(\beta_0, u_j, u_k) \\
&\quad + e_{W(1),j}(\beta_0, u_j)e_{W(1),k}(\beta_0, u_k)Q^{(0)}(\beta_0, u_j, u_k)\Big\}.
\end{aligned}$$

Now it can be seen that the above equation converges in probability to a fixed quantity in view of stability properties of $Q^{(\cdot)}$ stated in condition G-2 in the main text. The convergence of $f_{\cdot n}(\mathbf{X}_n)$ can be shown using the same manner as the above. In particular, let

$$f_{\cdot n}(\mathbf{X}_n) = \sum_{j=1}^{p} d_j(a_j - b_j)$$

where

$$\begin{aligned}
a_j &= S^{(1)}_{W_{(1),j}}(\beta_0, u_j), \quad \text{and} \\
b_j &= e_{W(1),j}(\beta_0, u_j)S^{(0)}_{W_{(1)}}(\beta_0, u_j),
\end{aligned}$$

then

$$\begin{aligned}
\{f_{\cdot n}(\mathbf{X}_n)\}^2 = &\sum_{j=1}^{p} d_j^2 \Big\{ S^{(1)}_{W_{(1),j}}(\beta_0, u_j)^2 - 2S^{(1)}_{W_{(1),j}}(\beta_0, u_j)e_{W(1),j}(\beta_0, u_j)S^{(0)}_{W_{(1)}}(\beta_0, u_j) \\
&\quad + e_{W(1),j}(\beta_0, u_j)^2 S^{(0)}_{W_{(1)}}(\beta_0, u_j)^2\Big\} \\
&+ 2\sum_{j<k} d_j d_k \Big\{ S^{(1)}_{W_{(1),j}}(\beta_0, u_k)S^{(1)}_{W_{(1),k}}(\beta_0, u_k) \\
&\quad - S^{(1)}_{W_{(1),j}}(\beta_0, u_j)e_{W(1),k}(\beta_0, u_k)S^{(0)}_{W_{(1)}}(\beta_0, u_k) \\
&\quad - e_{W(1),j}(\beta_0, u_j)S^{(0)}_{W_{(1)}}(\beta_0, u_j)S^{(0)}_{W_{(1)}}(\beta_0, u_k)
\end{aligned}$$

$$+ e_{W(1),j}(\beta_0, u_j) S^{(0)}_{W_{(1)}}(\beta_0, u_j) e_{W(1),k}(\beta_0, u_k) S^{(0)}_{W_{(1)}}(\beta_0, u_k) \Big\}.$$

Then without further calculation, it can be seen that the above equation also converges to a fixed quantity by conditions D, F, and G-2 in the main text, and therefore we prove that (1.30) holds.

Lastly, $g_n(\mathbf{X}_n)$ represents linear combinations of elements of the full cohort MSCM score process all evaluated at a finite number of fixed time points in $[0, 1]$. It can easily be seen that, for any such $g_n(\mathbf{X}_n)$, condition (III) of the Proposition 1 is satisfied due to the convergence of the full cohort MSCM score process to a Gaussian process with mean zero and finite covariance function.

Part 3) Limiting Covariance function Now we need to show the limiting covariance function of C_n. First we will show the limiting covariance function of D_n. Let $h_n(\mathbf{X}_n, \delta_n) = D_{n,j}(u_j) + D_{n,k}(u_k)$ (i.e., let $d_j = d_k = 1$ and $d_l = 0$ for all $l \neq j$ in $\sum_{j=1}^{p} d_j D_{n,j}(u_j)$). Covariance between $D_{n,j}(u_j)$ and $D_{n,k}(u_k)$ is given by

$$\begin{aligned} &\mathrm{Cov}(D_{n,j}(u_j), D_{n,k}(u_k)) \\ &= \Big\{ \mathrm{Var}(h_n(\mathbf{X}_n, \delta_n)) - \mathrm{Var}(D_{n,j}(u_j)) - \mathrm{Var}(D_{n,k}(u_k)) \Big\} / 2. \end{aligned} \tag{1.37}$$

Then the limiting values of (1.37) will lead to the (j, k)th components of the limiting covariance, i.e.,

$$\begin{aligned} &\lim_{n\to\infty} \mathrm{Cov}(D_{n,j}(u_j), D_{n,k}(u_k)) \\ &= \lim_{n\to\infty} \Big\{ \mathrm{Var}(h_n(\mathbf{X}_n, \delta_n)) - \mathrm{Var}(D_{n,j}(u_j)) - \mathrm{Var}(D_{n,k}(u_k)) \Big\} / 2. \end{aligned} \tag{1.38}$$

By Proposition 1, we can obtain limiting values of $\mathrm{Var}(h_n(\mathbf{X}_n, \delta_n))$, $\mathrm{Var}(D_{n,j}(u_j))$ and $\mathrm{Var}(D_{n,k}(u_k))$ using sample covariances calculated based on corresponding $f_{in}(\mathbf{X}_n)$ equipped with conditions G-2 and G-3 in the main text. Note that condition G-2 ensures the convergence of the finite sample covariance function to that of the limiting distribution. For notational convenience, let

$$\begin{aligned} \mathbf{F}_{in,j}(\mathbf{X}_n) = \Big[& W_i(u_j) Y_i(u_j) r_j^{(1)} \{\beta_0' A_i(u_j)\} \\ & - e_{W(1),j}(\beta_0, u_j) W_i(u_j) Y_i(u_j) r\{\beta_0' A_i(u_j)\} \Big], \quad \text{and} \end{aligned}$$

$$\mathbf{F}_{\cdot n,j}(\mathbf{X}_n) = n^{-1} \sum_{i=1}^{n} \mathbf{F}_{in,j}(\mathbf{X}_n); \qquad j = 1, ..., p.$$

Now, straightforward calculation based on Proposition 1 yields that

$$
\begin{aligned}
&\lim_{n\to\infty}\Big\{\mathrm{Var}(h_n(\mathbf{X}_n,\delta_n)) - \mathrm{Var}(D_{n,j}(u_j)) - \mathrm{Var}(D_{n,k}(u_k))\Big\} \\
=&(1-\alpha)\alpha^{-1}\lim_{n\to\infty} n^{-1}\sum_{i=1}^{n}\Big[\mathbf{F}_{in,j}(\mathbf{X}_n) + \mathbf{F}_{in,k}(\mathbf{X}_n) - \{\mathbf{F}_{\cdot n,j}(\mathbf{X}_n) + \mathbf{F}_{\cdot n,k}(\mathbf{X}_n)\}\Big]^2 \\
&-(1-\alpha)\alpha^{-1}\lim_{n\to\infty} n^{-1}\sum_{i=1}^{n}\Big[\mathbf{F}_{in,j}(\mathbf{X}_n) - \mathbf{F}_{\cdot n,j}(\mathbf{X}_n)\Big]^2 \\
&-(1-\alpha)\alpha^{-1}\lim_{n\to\infty} n^{-1}\sum_{i=1}^{n}\Big[\mathbf{F}_{in,k}(\mathbf{X}_n) - \mathbf{F}_{\cdot n,k}(\mathbf{X}_n)\Big]^2 \\
=&(1-\alpha)\alpha^{-1}\Big\{\lim_{n\to\infty} n^{-1}\sum_{i=1}^{n}\Big[\mathbf{F}_{in,j}(\mathbf{X}_n) + \mathbf{F}_{in,k}(\mathbf{X}_n) - \{\mathbf{F}_{\cdot n,j}(\mathbf{X}_n) + \mathbf{F}_{\cdot n,k}(\mathbf{X}_n)\}\Big]^2 \\
&\qquad -\lim_{n\to\infty} n^{-1}\sum_{i=1}^{n}\Big[\mathbf{F}_{in,j}(\mathbf{X}_n) - \mathbf{F}_{\cdot n,j}(\mathbf{X}_n)\Big]^2 \\
&\qquad -\lim_{n\to\infty} n^{-1}\sum_{i=1}^{n}\Big[\mathbf{F}_{in,k}(\mathbf{X}_n) - \mathbf{F}_{\cdot n,k}(\mathbf{X}_n)\Big]^2\Big\}.
\end{aligned}
$$

The whole term after $(1-\alpha)\alpha^{-1}$ can be simplified as follows:

$$
\begin{aligned}
&\lim_{n\to\infty} n^{-1}\sum_{i=1}^{n}\Big[\{\mathbf{F}_{in,j}(\mathbf{X}_n)^2 + 2\mathbf{F}_{in,j}(\mathbf{X}_n)\mathbf{F}_{in,k}(\mathbf{X}_n) + \mathbf{F}_{in,k}(\mathbf{X}_n)^2\} \\
&\qquad\qquad - 2\{\mathbf{F}_{in,j}(\mathbf{X}_n) + \mathbf{F}_{in,k}(\mathbf{X}_n)\}\{\mathbf{F}_{\cdot n,j}(\mathbf{X}_n) + \mathbf{F}_{\cdot n,k}(\mathbf{X}_n)\} \\
&\qquad\qquad + \{\mathbf{F}_{\cdot n,j}(\mathbf{X}_n)^2 + 2\mathbf{F}_{\cdot n,j}(\mathbf{X}_n)\mathbf{F}_{\cdot n,k}(\mathbf{X}_n) + \mathbf{F}_{\cdot n,k}(\mathbf{X}_n)^2\}\Big] \\
&-\lim_{n\to\infty} n^{-1}\sum_{i=1}^{n}\Big[\mathbf{F}_{in,j}(\mathbf{X}_n)^2 - 2\mathbf{F}_{in,j}(\mathbf{X}_n)\mathbf{F}_{\cdot n,j}(\mathbf{X}_n) + \mathbf{F}_{\cdot n,j}(\mathbf{X}_n)^2\Big] \\
&-\lim_{n\to\infty} n^{-1}\sum_{i=1}^{n}\Big[\mathbf{F}_{in,k}(\mathbf{X}_n)^2 - 2\mathbf{F}_{in,k}(\mathbf{X}_n)\mathbf{F}_{\cdot n,k}(\mathbf{X}_n) + \mathbf{F}_{\cdot n,k}(\mathbf{X}_n)^2\Big] \\
=&\lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} 2\Big[\mathbf{F}_{in,j}(\mathbf{X}_n)\mathbf{F}_{in,k}(\mathbf{X}_n) - \mathbf{F}_{in,j}(\mathbf{X}_n)\mathbf{F}_{\cdot n,k}(\mathbf{X}_n) - \mathbf{F}_{in,k}(\mathbf{X}_n)\mathbf{F}_{\cdot n,j}(\mathbf{X}_n) \\
&\qquad\qquad + \mathbf{F}_{\cdot n,j}(\mathbf{X}_n)\mathbf{F}_{\cdot n,k}(\mathbf{X}_n)\Big] \\
=&\lim_{n\to\infty} 2\Big[n^{-1}\sum_{i=1}^{n}\mathbf{F}_{in,j}(\mathbf{X}_n)\mathbf{F}_{in,k}(\mathbf{X}_n) - \mathbf{F}_{\cdot n,j}(\mathbf{X}_n)\mathbf{F}_{\cdot n,k}(\mathbf{X}_n)\Big],
\end{aligned}
$$

where the first term inside the bracket is given by

$$
\begin{aligned}
&n^{-1}\sum_{i=1}^{n}\mathbf{F}_{in,j}(\mathbf{X}_n)\mathbf{F}_{in,k}(\mathbf{X}_n)\\
=&n^{-1}\sum_{i=1}^{n}\Big[W_i(u_j)Y_i(u_j)r_j^{(1)}\{\beta_0'A_i(u_j)\}W_i(u_k)Y_i(u_k)r_k^{(1)}\{\beta_0'A_i(u_k)\}\\
&\qquad - e_{W(1),j}(\beta_0,u_j)W_i(u_j)Y_i(u_j)r\{\beta_0'A_i(u_j)\}W_i(u_k)Y_i(u_k)r_k^{(1)}\{\beta_0'A_i(u_k)\}\\
&\qquad - W_i(u_j)Y_i(u_j)r_j^{(1)}\{\beta_0'A_i(u_j)\}W_i(u_k)Y_i(u_k)r\{\beta_0'A_i(u_k)\}e_{W(1),k}(\beta_0,u_k)\\
&\qquad + e_{W(1),j}(\beta_0,u_j)W_i(u_j)Y_i(u_j)r\{\beta_0'A_i(u_j)\}W_i(u_k)Y_i(u_k)r\{\beta_0'A_i(u_k)\}e_{W(1),k}(\beta_0,u_k)\Big]\\
&= Q_{(j,k)}^{(1)}(\beta_0,u_j,u_k) - e_{W(1),j}(\beta_0,u_j)Q_k^{(2)}(\beta_0,u_j,u_k)\\
&\qquad - Q_j^{(2)}(\beta_0,u_k,u_j)e_{W(1),k}(\beta_0,u_k) + e_{W(1),j}(\beta_0,u_j)Q^{(0)}(\beta_0,u_j,u_k)e_{W(1),k}(\beta_0,u_k),
\end{aligned}
$$

and the second term inside the bracket is given by

$$
\begin{aligned}
\mathbf{F}_{\cdot n,j}(\mathbf{X}_n) - \mathbf{F}_{\cdot n,k}(\mathbf{X}_n) =&S_{W_{(1),j}}^{(1)}(\beta_0,u_j)S_{W_{(1),k}}^{(1)}(\beta_0,u_k)\\
&- e_{W(1),j}(\beta_0,u_j)S_{W_{(1)}}^{(0)}(\beta_0,u_j)S_{W_{(1),k}}^{(1)}(\beta_0,u_k)\\
&- S_{W_{(1),j}}^{(1)}(\beta_0,u_j)e_{W(1),k}(\beta_0,u_k)S_{W_{(1)}}^{(0)}(\beta_0,u_k)\\
&+ e_{W(1),j}(\beta_0,u_j)S_{W_{(1)}}^{(0)}(\beta_0,u_j)S_{W_{(1)}}^{(0)}(\beta_0,u_k)e_{W(1),k}(\beta_0,u_k).
\end{aligned}
$$

Then $\lim_{n\to\infty} 2\Big[n^{-1}\sum_{i=1}^{n}\mathbf{F}_{in,j}(\mathbf{X}_n)\mathbf{F}_{in,k}(\mathbf{X}_n) - \mathbf{F}_{\cdot n,j}(\mathbf{X}_n)\mathbf{F}_{\cdot n,k}(\mathbf{X}_n)\Big]$ can be rewritten as

$$
\begin{aligned}
&\lim_{n\to\infty} 2\Big[\Big\{Q_{(j,k)}^{(1)}(\beta_0,u_j,u_k) - S_{W_{(1),j}}^{(1)}(\beta_0,u_j)S_{W_{(1),k}}^{(1)}(\beta_0,u_k)\Big\}\\
&\qquad - e_{W(1),j}(\beta_0,u_j)\Big\{Q_k^{(2)}(\beta_0,u_j,u_k) - S_{W_{(1)}}^{(0)}(\beta_0,u_j)S_{W_{(1),k}}^{(1)}(\beta_0,u_k)\Big\}\\
&\qquad - \Big\{Q_j^{(2)}(\beta_0,u_k,u_j) - S_{W_{(1)}}^{(0)}(\beta_0,u_k)S_{W_{(1),j}}^{(1)}(\beta_0,u_j)\Big\}e_{W(1),k}(\beta_0,u_k)\\
&\qquad + e_{W(1),j}(\beta_0,u_j)\Big\{Q^{(0)}(\beta_0,u_j,u_k) - S_{W_{(1)}}^{(0)}(\beta_0,u_j)S_{W_{(1)}}^{(0)}(\beta_0,u_k)\Big\}e_{W(1),k}(\beta_0,u_k)\Big]\\
=&\lim_{n\to\infty} 2\Big[H_{(j,k)}^{(1)}(\beta_0,u_j,u_k) - e_{W(1),j}(\beta_0,u_j)H_k^{(2)}(\beta_0,u_j,u_k)\\
&\qquad - H_j^{(2)}(\beta_0,u_k,u_j)e_{W(1),k}(\beta_0,u_k) + e_{W(1),j}(\beta_0,u_j)H^{(0)}(\beta_0,u_j,u_k)e_{W(1),k}(\beta_0,u_k)\Big],
\end{aligned}
$$

where

$$
\begin{aligned}
H^{(0)}(\beta,x,v) &= Q^{(0)}(\beta,x,v) - S_{W_{(1)}}^{(0)}(\beta,x)S_{W_{(1)}}^{(0)}(\beta,v)\\
H^{(1)}(\beta,x,v) &= Q^{(1)}(\beta,x,v) - S_{W_{(1)}}^{(1)}(\beta,x)S_{W_{(1)}}^{(1)}(\beta,v)'
\end{aligned}
$$

$$H^{(2)}(\beta, x, v) = Q^{(2)}(\beta, x, v) - S^{(0)}_{W_{(1)}}(\beta, x) S^{(1)}_{W_{(1)}}(\beta, v).$$

It follows that (1.38) is the (j, k)th element of $G(\beta_0, u_j, u_k)$ in view of convergence property implied by conditions D and G-2. Then it can be seen that the limiting covariance function of D_n is given by G, and therefore we complete showing the in distribution convergence of (1.26) to a Gaussian random variable. By applying the basic properties of covariance matrix, we obtain the limiting covariance function of C_n given by Δ_α. This completes the proof of Theorem 3.5. □

Before presenting proof for the main result of this paper, asymptotic normality of $\tilde{\beta}$, let

$$\tilde{\mathcal{I}}(\beta, t) = -\partial^2 \tilde{l}(\beta, t)/\partial\beta^2, \quad \text{and}$$
$$\mathcal{I}(\beta, t) = -\partial^2 l(\beta, t)/\partial\beta^2.$$

In the proof of Theorem 3.6 we consider asymptotic properties of $\tilde{\mathcal{I}}$ instead of $\mathcal{I}$ because the two processes converge in probability to the same quantity. To see this, note

$$\begin{aligned}
&\sup_{\beta,t} |n^{-1}\{\mathcal{I}(\beta, t) - \tilde{\mathcal{I}}(\beta, t)\}| \\
\le& n^{-1} \sum_{i=1}^{n} \int_0^1 \sup_{\beta,u} |W_i(u)\{\tilde{V}_{W_{(1)}}(\beta, u) - V_{W_{(1)}}(\beta, u)\}| dN_i(u) \\
\le& \mathsf{M}_1 \int_0^1 \sup_{\beta,u} |\{\tilde{V}_{W_{(1)}}(\beta, u) - V_{W_{(1)}}(\beta, u)\}| n^{-1} \sum_{i=1}^{n} dN_i(u) \to_p 0
\end{aligned}$$

for any $(\beta, t) \in \mathcal{B} \times [0, 1]$ due to conditions B, D, F, and G-3 in the main text, by the continuous mapping theorem, and the fact that the total number of jumps are bounded by n.

Theorem 3.6. (Asymptotic normality of $\tilde{\beta}$) Under conditions A-G,

$$n^{1/2}(\tilde{\beta} - \beta_0) \to_d N(\mathbf{0}, \Sigma^{-1}_{W_{(1)}} \Sigma_{\tilde{U}} \Sigma^{-1}_{W_{(1)}})$$

where $\Sigma_{\tilde{U}}$ is given in Theorem 3.5.

Proof. A Taylor expansion of the MSCM case-cohort score process around β_0

evaluated at $\tilde{\beta}$ and $t = 1$ gives

$$n^{-1/2}\tilde{U}(\beta_0, 1) = \Big\{ - n^{-1}\frac{\partial^2 \tilde{l}(\dot{\beta}, 1)}{\partial \beta^2} \Big\} n^{1/2}(\tilde{\beta} - \beta_0) \tag{1.39}$$

for any $\dot{\beta}$ on the line segment between $\tilde{\beta}$ and β_0. It is clear that we need to show (in probability) convergence of $-n^{-1}\partial^2 \tilde{l}(\dot{\beta}, 1)/\partial \beta^2$, for any $\dot{\beta}$ in between $\tilde{\beta}$ and β_0. Here, $\tilde{V}_{W_{(1)}} = \tilde{S}^{(2)}_{W_{(1)}}/\tilde{S}^{(0)}_{W_{(1)}} - (\tilde{S}^{(1)}_{W_{(1)}}/\tilde{S}^{(0)}_{W_{(1)}})^{\otimes 2}$. Therefore, it is sufficient to show that $n^{-1}\mathcal{I}(\beta, 1)$ converges in probability to a fixed matrix. Using (2.10*), decompose $n^{-1}\mathcal{I}(\beta_0, 1)$ by

$$n^{-1}\sum_{i=1}^{n}\int_0^1 W_i(u)\Big[\frac{S^{(2)}_{W_{(1)}}(\beta_0, u)S^{(0)}_{W_{(1)}}(\beta_0, u) - \{S^{(1)}_{W_{(1)}}(\beta_0, u)\}^{\otimes 2}}{S^{(0)}_{W_{(1)}}(\beta_0, u)^2}\Big] dM_i(u)$$

$$+ \int_0^1 \Big[\frac{S^{(2)}_{W_{(1)}}(\beta_0, u)S^{(0)}_{W_{(1)}}(\beta_0, u) - \{S^{(1)}_{W_{(1)}}(\beta_0, u)\}^{\otimes 2}}{S^{(0)}_{W_{(1)}}(\beta_0, u)^2}\Big] S^{(0)}_{W_{(1)}}(\beta_0, u)\lambda_0(u)du.$$

The elements of the first term are local square integrable martingale with the variation process for the (i, j) element equals

$$n^{-1}\int_0^1 \big\{V_{W_{(1)}}(\beta, u)\big\}^2_{ij} S^{(0)}_{W_{(2)}}(\beta_0, u)\lambda_0(u)du$$

which converges in probability to zero by virtue of the stability, regularity, and boundedness conditions A-F. It follows that

$$n^{-1}\mathcal{I}(\beta, 1) \to_p \int_0^1 v_{W_{(1)}}(\beta_0, u)s^{(0)}_{W_{(1)}}(\beta_0, u)\lambda_0(u)du = \Sigma_{W_{(1)}} \tag{1.40}$$

for any $\beta \in \mathcal{B}$, and therefore $n^{-1}\mathcal{I}(\dot{\beta}, 1) \to_p \Sigma_{W_{(1)}}$ for any $\dot{\beta}$ in between $\tilde{\beta}$ and β_0. Then Theorem 3.5 along with (1.39) complete the proof. In particular, the covariance matrix $\Sigma^{-1}_{W_{(1)}}\Sigma_{\tilde{U}}\Sigma^{-1}_{W_{(1)}}$ has a form

$$\Sigma^{-1}_{W_{(1)}}(\Sigma_U + \Delta_\alpha)\Sigma^{-1}_{W_{(1)}} = \Sigma^{-1}_{W_{(1)}}(\Sigma_{W_{(2)}} + \Delta_{W_{(1)}, W_{(2)}} + \Delta_\alpha)\Sigma^{-1}_{W_{(1)}}$$

where $\Sigma_U = \Sigma_{W_{(2)}} + \Delta_{W_{(1)}, W_{(2)}}$ as in Theorem 3.4 and the explicit form of Δ_α is given by (1.19). □

S2. Implementation and Simulation

We have shown that we can obtain a consistent and asymptotically normally distributed estimator of treatment effect in the case-cohort setting by fitting a MSCM via inverse probability weighting. This provides theoretical justification for simulation results shown in Cole et al. (2012). In this section we (i) describe how a MSCM can easily be fit via inverse probability weighting for either the full cohort or case-cohort setting using standard survival analysis software, such as R or SAS, and (ii) present additional results from the simulation study described in Section 4 of the main text, as well as results from a simulation study showing finite sample performance of the cumulative baseline hazard estimator proposed in Section 5.1 of the main text.

S2-1. Implementation

To fit a MSCM via inverse probability weighting for a full cohort, first create a data set in which each person-visit corresponds to one row. Specifically, let each row contain a subject identifier, visit (or date) information, treatment and time-varying confounder information at the corresponding visit/date time, and baseline covariates. Depending upon the user-defined models to estimate $W_i(t)$, the data set may be augmented by treatment/covariate histories in each row as well. For example, one might fit pooled logistic models to obtain the estimated probability of receiving treatment at time t by regressing the log-odds of receiving treatment $A(t)$ on prior treatment status (say, $A(t^-)$) alone (for the numerator in $W^T(t)$ (which is (2.5*) in the main text)), or with current covariate information $L(t)$ (for the denominator in $W^T(t)$) (Hernán, Brumback, and Robins (2001)). Analogously, the estimated probability of being uncensored at time t can be obtained by regressing the log-odds of being uncensored ($C(t) = 0$) on current treatment status ($A(t)$) alone, or with $L(t)$. For such models flexible functional forms (e.g., splines) are often used for continuous confounders (Cole and Hernán (2008); Cole et al. (2003, 2012)). Predicted values of the denominator and numerator probabilities in $W^T(t)$ and $W^C(t)$ (which is (2.6*) in the main text) can then be used to calculate $\hat{W}_i(t)$ for all participants $i = 1, ..., n$ and all study visit times t. Then $\hat{W}_i(t)$ needs to be added to the data set to fit the MSCM. Finally, the data set should be prepared in the counting process type format whereby each row contains the start and stop times corresponding to the previous and current visits, along with an event status indicator for the current visit. Then standard software can be used to fit the MSCM via inverse probability weighting. For instance, using the `survival` package in R (Therneau (2013)), the following code can be used:

```
coxph(Surv(start, stop, delta) ~ trt, weight=w, data=dataname)
```

where `delta` is the event indicator having value 1 if an event occurred at `stop` and 0 otherwise, `trt` indicates whether an individual received treatment (assuming treatment is a scalar) over the interval `(start, stop]`, and `w` is $\hat{W}_i(t)$. The same model can be fit in SAS by using the following code:

```
proc phreg data = dataname covout;
     model (start,stop)*delta(0)=trt;
     weight w;
run;
```

Fitting a MSCM in the case-cohort setting can be accomplished with some additional data modifications. First, prepare a reduced (case-cohort) data set including the randomly selected $\tilde{n}$ subcohort members and all cases. Just as in the full cohort data preparation, each row of the case-cohort data set should correspond to each person-visit record. Second, estimate the individual-time-specific weights $W_i(t)$ based on the user-specified model as before (e.g., logistic regression), except with individuals in the subcohort that are not cases weighted by inverse-probability-sampling weights $n/\tilde{n}$ (Cole et al. (2012)). For example, if `pi` is the subcohort fraction $\tilde{n}/n$ and `sub` is the subcohort indicator, we can use the following SAS code to estimate $W(t)$;

```
data casecohortdata;
     if delta=1 then w2=1;
     else if delta=0 and sub=1 then w2=1/pi;
run;
proc logistic data=casecohortdata;
     weight=w2;
     model trt= l1 l2;
     output out=outdata1 p=denom;
run;
```

where `l1` and `l2` are two (possibly time-varying) covariates included in the treatment model. The variable `denom` in the `outdata1` data set will contribute to the calculation of the denominator of $\hat{W}_i(t)$. Similarly, contributions to the numerator of $\hat{W}_i(t)$ can be obtained by using `model trt=` and `output out=outdata2 p=num` statement. Then we can merge `outdata1` with `outdata2` to create `w`.

After adding the estimated individual-time-specific weights $\hat{W}_i(t)$ to each person-visit row, modify each nonsubcohort case to contribute only one line of data with start time $t_j - \epsilon$ and stop time t_j where t_j is the event time for that individual and ϵ is chosen to be very small, for instance $\epsilon = 0.0001$. This insures that nonsubcohort cases appear only in the risk set when they fail. One should make sure that the start times for nonsubcohort cases are positive, such that $t_j - \epsilon > 0$ for your choice of ϵ. This modification of the data set for the nonsubcohort cases is sufficient to obtain β^*, and the same R/SAS code as above can be employed using the modified data set. Obtaining $\tilde{\beta}$ can be accomplished with an additional data step wherein a dummy variable is coded equal to a relatively small negative value (e.g., -20) for nonsubcohort cases and 0 otherwise (Therneau and Li (1999)). Then, $\tilde{\beta}$ can be obtained as follows in R:

```
coxph(Surv(start, stop, delta) ~ trt + offset(dummy), weight=w)
```

or in SAS:

```
proc phreg data = dataname covout;
     model (start,stop)*delta(0)=trt/offset=dummy;
     weight w;
run;
```

The `offset` term enforces a relative weight of $\exp(-20) < 10^{-8}$, assuming -20 is used for the dummy value, to the nonsubcohort cases so that they effectively do not contribute to the sum of the log (inside the integral) in (2.8*) in the main text. Therneau and Li (1999) suggested using -100 ($\exp(-100) < 10^{-40}$) for the dummy variable value, however, we found that sometimes the `coxph` function in R did not converge when dummy $= -100$; this convergence problem was observed when the event rate was very low, say 3-4%. Therefore, we recommend several dummy values be considered to ensure robustness of analysis results. The choice of dummy $= -20$ yielded reasonable analysis results under average event rate

$\geq$5% in our simulation study.

The proposed variance estimator (3.1*) in the main text requires computation of four components: $\hat{\Sigma}^{-1}_{W_{(1)}}$, $\hat{\Sigma}_{W_{(2)}}$, $\hat{\Delta}_{W_{(1)},W_{(2)}}$, and $\hat{\Delta}_{\alpha}$. The naive variance estimator obtained by fitting the Cox model with the `weight` option is the inverse of minus the second derivative of $\tilde{l}(\beta, 1)$ evaluated at $\tilde{\beta}$ (i.e., $\tilde{\mathcal{I}}^{-1}(\tilde{\beta}, 1)$, the inverse of the observed information matrix) which is n^{-1} times $\hat{\Sigma}^{-1}_{W_{(1)}}$. Therefore, $\hat{\Sigma}^{-1}_{W_{(1)}}$ can be obtained by multiplying n times the naive variance estimate. Likewise, $\hat{\Sigma}_{W_{(2)}}$ can be obtained by multiplying n^{-1} times the inverse of the naive variance estimate obtained by fitting the Cox model with the variable `weight` equal to the square of the original weight variable. Unfortunately, it does not seem that $\hat{\Delta}_{W_{(1)},W_{(2)}}$ and $\hat{\Delta}_{\alpha}$ can be obtained as simply as $\hat{\Sigma}_{W_{(1)}}$ or $\hat{\Sigma}_{W_{(2)}}$. One can create vectors/matrices of $\tilde{S}^{(j)}_{\hat{W}_k}(\tilde{\beta}, \cdot)$, and then calculate $\tilde{E}_{\hat{W}_k}(\tilde{\beta}, \cdot)$, $\tilde{Q}^{(j)}(\tilde{\beta}, \cdot)$, and $\tilde{H}^{(j)}(\tilde{\beta}, \cdot)$ to obtain $\hat{\Delta}_{W_{(1)},W_{(2)}}$ and $\hat{\Delta}_{\alpha}$. Alternatively, one may want to apply the LY estimator in practice (Cole et al. (2012)). The LY estimator appears to perform well empirically if we have moderate subcohort size and event rate (See Cole et al. (2012); Table 4.1 in the main text and Table 2.1 in S2-2 below), and is computationally straightforward to implement. The LY estimator associated with $\tilde{\beta}$ can be obtained by using the following R or SAS code:

```
coxph(Surv(start, stop, delta) ~ trt + offset(dummy)
                                 + cluster(id), weight=w)
proc phreg data = dataname covs(aggregate) covout;
     id id;
     model (start,stop)*delta(0)=trt/offset=dummy;
     weight w;
run;
```

The LY estimator corresponding to β^* can be obtained by deleting `offset(dummy)` or `/offset=dummy`.

S2-2. Simulation

As Table 4.1 in the main text shows simulation results under the null hypothesis only, we present additional simulation results obtained from the alternative hypothesis using the same scenarios as in Table 4.1. Under the alternative hypotheis, β_0 was set to $\log(1/2) \simeq -0.6931$ representing a scenario that treatment

Sub-cohort(%)	Event rate(%)	Estimator	Bias	ESE	ASE		Coverage	
					proposed	LY	proposed	LY
5	5	β^*	-0.11	0.53	0.65	0.46	0.97	0.92
		$\tilde{\beta}$	-0.23	0.78	0.98	0.36	0.97	0.77
	25	β^*	0.02	0.38	0.37	0.33	0.96	0.92
		$\tilde{\beta}$	-0.01	0.44	0.38	0.36	0.94	0.92
10	5	β^*	-0.06	0.45	0.49	0.42	0.97	0.95
		$\tilde{\beta}$	-0.08	0.47	0.50	0.36	0.97	0.89
	25	β^*	0.01	0.27	0.27	0.23	0.95	0.93
		$\tilde{\beta}$	-0.01	0.28	0.27	0.26	0.94	0.94
20	5	β^*	-0.04	0.41	0.43	0.40	0.97	0.96
		$\tilde{\beta}$	-0.05	0.41	0.43	0.37	0.97	0.94
	25	β^*	0.00	0.20	0.21	0.20	0.96	0.95
		$\tilde{\beta}$	-0.01	0.21	0.21	0.21	0.95	0.95

Table 2.1: Summary of simulation study under the alternative hypothesis. Bias denotes the empirical bias of the different estimators of β_0. ESE denotes the empirical standard errors. ASE denotes the average estimated standard errors and Coverage denotes the empirical coverage of 95% Wald-type confidence intervals using either (3.1*) or the LY variance estimator.

lowers the rate of having an event by half. See Section 4 of the main text for additional details regarding the simulation study. Results from simulations under the alternative hypothesis are summarized in Table 2.1. Similar to the null hypothesis setting, $\tilde{\beta}$ and β^* were nearly unbiased and the proposed variance estimator was usually less biased than the LY variance estimator when the subcohort fraction was only 5%, regardless of the event rate. Although not all simulation results are shown in Table 2.1, our numerical study indicated that both the proposed and the LY variance estimators were approximately unbiased when the subcohort fraction and event rate were both greater than 15%. Wald confidence intervals (CIs) using the LY variance estimator again tended to undercover when the subcohort fraction was 5%, and also when event rate was 5% and subcohort fraction was 10% if $\tilde{\beta}$ was used. Wald CIs using (3.1*) exhibited coverage close to the nominal level. As seen from Table 4.1 in the main text, $\tilde{\beta}$ and β^* along with the proposed variance estimator exhibited good finite sample properties for the scenarios considered, while performance of the LY variance estimator depended on subcohort size and event rate.

Subcohort (%)	$\Lambda_0(t)$	$\beta_0 = 0$		$\beta_0 = \log(1/2)$	
		Bias	ESE	Bias	ESE
5	0.05	0.008	0.074	0.010	0.179
	0.10	0.030	0.834	0.063	2.627
	0.15	0.042	0.837	0.071	2.629
	0.20	0.055	0.844	0.081	2.630
10	0.05	0.002	0.023	0.001	0.011
	0.10	0.005	0.028	0.003	0.018
	0.15	0.008	0.037	0.006	0.029
	0.20	0.012	0.048	0.009	0.041
20	0.05	0.001	0.011	0.000	0.009
	0.10	0.002	0.016	0.001	0.014
	0.15	0.004	0.024	0.002	0.021
	0.20	0.005	0.032	0.003	0.029

Table 2.2: Summary of simulation study to evaluate performance of $\tilde{\Lambda}_{\hat{W}}(\tilde{\beta}, t)$ at event rate 25%. Bias denotes the empirical bias of $\tilde{\Lambda}_{\hat{W}}(\tilde{\beta}, t)$ at different failure times. ESE denotes the empirical standard error.

We also examined performance of the cumulative baseline hazard estimator proposed in Section 5.1 of the main text. We use the same simulation settings as in Section 4 of the main text at a fixed event rate 25%. Survival times when all individuals were untreated (i.e., potential outcome under no treatment) followed an exponential distribution with mean 1 as in Cole et al. (2012). Therefore the true cumulative baseline hazard function is given by $\Lambda_0(t) = t$. We evaluated the bias and standard error of the proposed estimator at time points $t = 0.05, 0.10, 0.15$, and 0.20 under the null and alternative simulations. Results are summarized in Table 2.2. For each time point, the bias and standard error of the proposed estimator decreased as the subcohort fraction increased, with negligible bias when the subcohort fraction equaled 20%.

S3. Notation

In this section we present a list of key notation used throughout the main text and this supplement.

1. $l(\beta, t; W)$: The log-WPPL under full cohort setting ((2.7*) in the main text).

2. $\tilde{l}(\beta, t; W)$: The log-WPPL under case-cohort setting (Self and Prentice (1988) type; (2.8*) in the main text).

3. $l^*(\beta, t; W)$: The log-WPPL under case-cohort setting (Prentice (1986) type, Cole et al. (2012); (2.9*) in the main text).

4. $\hat{\beta}$, $\tilde{\beta}$, and β^*: solutions to $\partial l(\beta, 1; \hat{W})/\partial\beta = 0$, $\partial\tilde{l}(\beta, 1; \hat{W})/\partial\beta = 0$, and $\partial l^*(\beta, 1; \hat{W})/\partial\beta = 0$, respectively. β_0 is the parameter of interest in the MSCM (See (2.4*) and (2.11*) in the main text).

5. $X(\beta, t, W) = n^{-1}\{l(\beta, t; W) - l(\beta_0, t; W)\}$
   $\tilde{X}(\beta, t, W) = n^{-1}\{\tilde{l}(\beta, t; W) - \tilde{l}(\beta_0, t; W)\}$
   $X^*(\beta, t, W) = n^{-1}\{l^*(\beta, t; W) - l^*(\beta_0, t; W)\}$
   W can be replaced to $\hat{W}$ if we plug-in estimator of IPWs into $X, \tilde{X}$, or X^*.

6. $K(\beta, t, W)$: compensator of $X(\beta, t, W)$, where $dN_i(u)$ inside the integral of $X(\beta, t, W)$ is replaced by $\lambda_i(u)du$. (See S1-1)

7. $S^{(j)}(\beta, t) = n^{-1}\sum_{i=1}^n Y_i(t)r^{(j)}\{\beta' A_i(t)\}; j = 0, 1, 2$, where their limits are defined by $s^{(j)}(\beta, t)$. (See Section 3 and condition D in the Appendix of the main text.)

8. $S^{(j)}_{W_{(k)}} = n^{-1}\sum_{i=1}^n W_i(t)^k Y_i(t)r^{(j)}\{\beta' A_i(t)\}$ for $j = 0, 1, 2$ and $k = 1, 2$, where their limits are defined by $s^{(j)}_{W_{(k)}}(\beta, t)$. (See Section 3 and condition D in the Appendix of the main text.)

9. $\tilde{S}^{(j)}_{W_{(k)}}(\beta, t) = \tilde{n}^{-1}\sum_{i\in\tilde{\mathcal{C}}} W_i(t)^k Y_i(t)r^{(j)}\{\beta' A_i(t)\}$ for $j = 0, 1, 2$ and $k = 1, 2$, where their limits are defined by $s^{(j)}_{W_{(k)}}(\beta, t)$. (See Section 3 and condition G-3 in the Appendix of the main text.)

10. $E = S^{(1)}/S^{(0)}$, with its limit $e = s^{(1)}/s^{(0)}$
    $E_{W_{(k)}} = S^{(1)}_{W_{(k)}}/S^{(0)}_{W_{(k)}}$, with their limits $e_{W_{(k)}}; k = 1, 2$.
    $\tilde{E}_{W_{(k)}} = \tilde{S}^{(1)}_{W_{(k)}}/\tilde{S}^{(0)}_{W_{(k)}}$, with their limits $e_{W_{(k)}}; k = 1, 2$.
    $V = S^{(2)}/S^{(0)} - E^{\otimes 2}$, with its limit $v = s^{(2)}/s^{(0)} - e^{\otimes 2}$.
    $V_{W_{(k)}} = S^{(2)}_{W_{(k)}}/S^{(0)}_{W_{(k)}} - E^{\otimes 2}_{W_{(k)}}$, with their limits $v_{W_{(k)}}; k = 1, 2$.
    $\Sigma = \int_0^1 v(\beta_0, t)s^{(0)}(\beta_0, t)\lambda_0(t)dt$
    $\Sigma_{W_{(k)}} = \int_0^1 v_{W_{(k)}}(\beta_0, t)s^{(0)}_{W_{(k)}}(\beta_0, t)\lambda_0(t)dt; \quad k = 1, 2$
    See Section 3 and conditions F and G in the Appendix of the main text.

11. $Q^{(0)}(\beta,t,u) = n^{-1}\sum_{i=1}^{n} W_i(t)Y_i(t)r\{\beta_0' A_i(t)\}W_i(u)Y_i(u)r\{\beta_0' A_i(u)\}$
$Q^{(1)}(\beta,t,u) = n^{-1}\sum_{i=1}^{n} W_i(t)Y_i(t)r^{(1)}\{\beta_0' A_i(t)\}W_i(u)Y_i(u)r^{(1)}\{\beta_0' A_i(u)\}'$
$Q^{(2)}(\beta,t,u) = n^{-1}\sum_{i=1}^{n} W_i(t)Y_i(t)r\{\beta_0' A_i(t)\}W_i(u)Y_i(u)r^{(1)}\{\beta_0' A_i(u)\}$,
where their limits are denoted by $q^{(j)}; j = 0, 1, 2$. $Q^{(j)}$ are covariance functions based on full cohort. Similary, $\tilde{Q}^{(j)}(\beta,t,u)$ are covariance functions based on subcohort members where their limits are also given by $q^{(j)}$. See Theorem 3.5 in Section 3 and conditions G-2 and G-3 in the Appendix of the main text.

12. $n^{-1/2}U(\beta_0,t) = n^{1/2}\partial l(\beta,t)/\partial\beta\big|_{\beta=\beta_0}$: full cohort MSCM Score process
$n^{-1/2}\tilde{U}(\beta_0,t) = n^{1/2}\partial\tilde{l}(\beta,t)/\partial\beta\big|_{\beta=\beta_0}$: case-cohort MSCM Score process, presented in Theorem 3.4 and 3.5 respectively.

13. (Theorem 3.4) $\Sigma_U = \Sigma_{W_{(2)}} + \Delta_{W_{(1)},W_{(2)}}$
$\Delta_{W_{(1)},W_{(2)}} = \int_0^1 \{e_{W_{(2)}}(\beta_0,u) - e_{W_{(1)}}(\beta_0,u)\}^{\otimes 2} s^{(0)}_{W_{(2)}}(\beta_0,u)\lambda_0(u)du$
$D_n(t) = n^{1/2}\Big[\Big\{\tilde{S}^{(1)}_{W_{(1)}}(\beta_0,t) - S^{(1)}_{W_{(1)}}(\beta_0,t)\Big\} - e_{W_{(1)}}(\beta_0,t)\Big\{\tilde{S}^{(0)}_{W_{(1)}}(\beta_0,t) - S^{(0)}_{W_{(1)}}(\beta_0,t)\Big\}\Big] S^{(0)}_{W_{(1)}}(\beta_0,t)$

14. (Theorem 3.5-3.6) $\Sigma_{\tilde{U}} = \Sigma_U + \Delta_\alpha$
$\Delta_\alpha = \int_0^1\int_0^1 G(\beta_0,x,v)\lambda_0(x)\lambda_0(v)dxdv$
$G(\beta_0,x,v) = (1-\alpha)\alpha^{-1}\Big[h^{(1)}(\beta_0,x,v) - e_{W_{(1)}}(\beta_0,x)h^{(2)}(\beta_0,x,v)' - h^{(2)}(\beta_0,v,x)e_{W_{(1)}}(\beta_0,v)' +$
$e_{W_{(1)}}(\beta_0,x)e_{W_{(1)}}(\beta_0,v)'h^{(0)}(\beta_0,x,v)\Big]$
$h^{(0)}(\beta,x,v) = q^{(0)}(\beta,x,v) - s^{(0)}_{W_{(1)}}(\beta,x)s^{(0)}_{W_{(1)}}(\beta,v)$
$h^{(1)}(\beta,x,v) = q^{(1)}(\beta,x,v) - s^{(1)}_{W_{(1)}}(\beta,x)s^{(1)}_{W_{(1)}}(\beta,v)'$
$h^{(2)}(\beta,x,v) = q^{(2)}(\beta,x,v) - s^{(0)}_{W_{(1)}}(\beta,x)s^{(1)}_{W_{(1)}}(\beta,v)$
$H^{(0)}(\beta,x,v) = Q^{(0)}(\beta,x,v) - S^{(0)}_{W_{(1)}}(\beta,x)S^{(0)}_{W_{(1)}}(\beta,v)$
$H^{(1)}(\beta,x,v) = Q^{(1)}(\beta,x,v) - S^{(1)}_{W_{(1)}}(\beta,x)S^{(1)}_{W_{(1)}}(\beta,v)'$
$H^{(2)}(\beta,x,v) = Q^{(2)}(\beta,x,v) - S^{(0)}_{W_{(1)}}(\beta,x)S^{(1)}_{W_{(1)}}(\beta,v)$
$n^{-1}\tilde{\mathcal{I}}(\beta,t) = -n^{-1}\partial^2\tilde{l}(\beta,t)/\partial\beta^2$
$n^{-1}\mathcal{I}(\beta,t) = -n^{-1}\partial^2 l(\beta,t)/\partial\beta^2$

15. Notation in S1 in the supplement:
$B_n(t) = n^{-1/2}\sum_{i=1}^{n}\int_0^t W_i(u)[A_i(u) - E_{W_{(1)}}(\beta_0,u)]dM_i(u)$

$C_n(t) = \int_0^t D_n(u)\lambda_0(u)du$

$$D_n(u) = n^{1/2}\Big[\{\tilde{S}^{(1)}_{W_{(1)}}(\beta_0,u) - S^{(1)}_{W_{(1)}}(\beta_0,u)\} - e_{W_{(1)}}(\beta_0,u)\{\tilde{S}^{(0)}_{W_{(1)}}(\beta_0,u) - S^{(0)}_{W_{(1)}}(\beta_0,u)\}\Big]S^{(0)}_{W_{(1)}}(\beta_0,u)$$

$g_n(\mathbf{X}_n) = n^{-1/2}\sum_{i=1}^n \sum_{j=1}^p c_j \int_0^t W_i(u)[A_{i,j}(u) - E_{W_{(1)},j}(\beta_0,u)]dM_i(u)$ where $c_j (j = 1, ..., p)$ can be any constant.

$f_{in}(\mathbf{X}_n) = \sum_{j=1}^p d_j \Big[W_i(u_j)Y_i(u_j)r_j^{(1)}\{\beta_0' A_i(u_j)\} - e_{W(1),j}(\beta_0,u_j)W_i(u_j)Y_i(u_j)r\{\beta_0' A_i(u_j)\}\Big]$ where $d_j (j = 1, ..., p)$ can be any constant.

$f_{\cdot n}(\mathbf{X}_n) = n^{-1}\sum_{i=1}^n f_{in}(\mathbf{X}_n)$

$\mathbf{S}^2_{f_n} = n^{-1}\sum_{i=1}^n [f_{in}(\mathbf{X}_n) - f_{\cdot n}(\mathbf{X}_n)]^2$

$h_n(\mathbf{X}_n, \delta_n) = n^{1/2}[\tilde{n}^{-1}\sum_{i=1}^n \delta_{in} f_{in}(\mathbf{X}_n) - f_{\cdot n}(\mathbf{X}_n)] = \sum_{j=1}^p d_j D_{n,j}(u_j)$

$\boldsymbol{\delta}_n = (\delta_{1n}, ..., \delta_{nn})$: vector of $\tilde{n}$ ones and $n - \tilde{n}$ zeros representing subcohort membership.

$\mathbf{F}_{in,j}(\mathbf{X}_n) = \Big[W_i(u_j)Y_i(u_j)r_j^{(1)}\{\beta_0' A_i(u_j)\} - e_{W(1),j}(\beta_0,u_j)W_i(u_j)Y_i(u_j)r\{\beta_0' A_i(u_j)\}\Big]$

$\mathbf{F}_{\cdot n,j}(\mathbf{X}_n) = n^{-1}\sum_{i=1}^n \mathbf{F}_{in,j}(\mathbf{X}_n); \quad j = 1, ..., p$