Recall that $\beta_0$ and $\lambda_{k0}(\cdot)$ are the true values and functions of $\beta$ and $\lambda_k(\cdot)$, and $M_{ik}(\cdot)$ are iid copies of $M_k(\cdot)$. The proof is essentially same as that of Andersen and Gill (1982). Write

$$U_h(\beta_0) = \frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{K} \int_{0}^{T_k} [h_{ik}(t) - \mu_{k,h}(t)]dM_{ik}(t) - \frac{1}{n} \sum_{k=1}^{K} \sum_{i=1}^{n} \int_{0}^{T_k} [h_k(t; \beta_0) - \mu_{k,h}(t)]dM_{ik}(t)$$

(S1.1)

The first term of the right hand side is the sum of $n$ random vectors which are iid copies of $\sum_{k=1}^{K} \sum_{i=1}^{n} \int_{0}^{T_k} [h_{ik}(t) - \mu_{k,h}(t)]dM_{ik}(t)$ and is thus asymptotically normal at the rate $n^{1/2}$ with asymptotic variance $V_h$. For every $1 \leq k \leq K$,

$$\sum_{i=1}^{n} \int_{0}^{T_k} [h_k(t; \beta_0) - \mu_{k,h}(t)]dM_{ik}(t) \overset{\otimes 2}{=} o(n)$$

Therefore the second term of the right hand side of (S1.1) is $o_P(n^{1/2})$. As a result,

$$n^{-1/2} U_h(\beta_0) \rightarrow N(0, V_h).$$

(S1.2)

It follows from the law of large numbers that

$$\frac{1}{n} \frac{\partial}{\partial \beta} U_h(\beta) \bigg|_{\beta=\beta_0} = -\frac{1}{n} \sum_{k=1}^{K} \sum_{i=1}^{n} \int_{0}^{T_k} \frac{\sum_{j=1}^{n} (h_{jk}(t) - \tilde{h}_k(t; \beta_0))}{\sum_{j=1}^{n} e^{\beta Z_{jk}Y_{jk}(t)}} \times (Z'_{jk} - \bar{Z}'_{k}(t; \beta_0)) e^{\beta Z_{jk}Y_{jk}(t)}dN_{ik}(t)$$

$$\rightarrow -\sum_{k=1}^{K} E(\xi_{k,h}^{\xi_{k,h}}) = -A_h$$

(S1.3)
in probability as \( n \to \infty \). The above convergence can be shown to hold uniformly over \( \{ \beta : \| \beta - \beta_0 \| \leq \epsilon_n \} \) for any sequence of \( \epsilon_n \downarrow 0 \). Since \( A_h \) is assumed nondegenerate, let \( a = \inf \{ \| A_h x \| / \| x \| : x \in \mathbb{R}^p \} \). Then \( a > 0 \). Let \( B \) be the ball in \( \mathbb{R}^p \) centered at \( \beta_0 \) with radius \( \epsilon \), where \( \epsilon > 0 \) is small but fixed, and let \( D_n = \{ (1/n)U_h(x) : x \in B \} \) be the image of \( B \) for the continuous mapping \((1/n)U_h(\cdot)\). With probability tending to 1, for any two \( p \)-vectors \( x_1 \) and \( x_2 \) in \( B \),

\[
1/n\|U_h(x_1) - U_h(x_2)\| > (a/2)\|x_1 - x_2\|.
\]

It implies that, with probability tending to 1, \((1/n)U_h(\beta_0)\) is a homeomorphism from \( B \) to \( D_n \) and, moreover, \( D_n \) contains a ball centered at \((1/n)U_h(\beta_0)\) with radius \( a/2 \). Then, with probability tending to 1, this ball contains 0 since \((S1.2)\) implies \((1/n)U_h(\beta_0) = o_p(1)\). This proves that, with probability tending to 1, there exists a zero solution of the equation \( U_h(\beta_0) = 0 \) in any small but fixed neighborhood of \( \beta_0 \). The consistency follows. Then, \((S1.2)\) and \((S1.3)\) together ensures asymptotic normality and the asymptotic variance is given by the sandwich formula. The proof is complete.

**S2 Proof of the Proposition**

The proof is divided into five steps. Step 1: Introducing some notations. Let \( L_0^2 \) be the space of all \( p \)-dimensional random vectors measurable to the \( \sigma \)-algebra generated by \( \{(Y_k, \delta_k), k = 1, \ldots, K, Z\} \) and with zero conditional mean given \( Z \) and finite variance. Define an inner product to be the sum of component-wise covariances so that \( L_0^2 \) can be verified to be a Hilbert space. Let \( S_k = \{ \eta : \eta \in L_0^2, E[\eta|Z] = 0, \eta \in \sigma(Y_k, \delta_k, Z) \} \). Denote by \( M_k \) the closure of \( \{ \eta : \eta = \int_0^\infty (h(t, Z) - \mu_k(t))dM_k(t) \} \) for all \( p \)-dimensional continuous functions \( h \) such that \( \eta \in L_0^2 \) where \( \mu_k = E(h(t, Z)|Y_k = t, \delta_k = 1) \). It is seen that \( M_k \) and \( S_k \) are both closed linear subspaces of \( L_0^2 \) and that \( M_k \subseteq S_k \). Set \( M = M_1 + \cdots + M_K \) and \( M_k = M_1 + \cdots + M_{k-1} + M_{k+1} + \cdots + M_K \). Then \( S \) and \( S_k \) are likewise defined. To avoid trivialities, we assume throughout the paper \( M_k, S_k, M_k + M_k, \) and \( S_k + S_k \) are closed and \( M_k \cap M_k = \{0\} = S_k \cap S_k \) for every \( k = 1, \ldots, K \).

Step 2. Defining the score \( \sum_{l=1}^K \xi_{l,h^*} \) through alternating projection. Denote the projection operator in \( L_0^2 \) by \( \Pi \). Write

\[
\Sigma_h = \left[ \sum_{k=1}^K E(\xi_{k,h}^2) \right]^{-1} \sum_{k=1}^K E\left( \xi_{k,h} \left\{ \Pi \left( \sum_{l=1}^K \xi_{l,h} | M_k \right) \right\} \right) \left[ \sum_{k=1}^K E(\xi_{k,h}^2) \right]^{-1}.
\]

Let \( h^* \) satisfy

\[
\Pi\left( \sum_{l=1}^K \xi_{l,h^*} | M_k \right) = \xi_k, \quad k = 1, \ldots, K. \tag{S2.1}
\]

Then,

\[
\Sigma_{h^*} = \left[ \sum_{k=1}^K E(\xi_{k,h^*}^2) \right]^{-1} = \left[ \text{var} \left( \sum_{k=1}^K \xi_{k,h^*} \right) \right]^{-1}.
\]
The existence of the solution of (S2.1) can be argued as follows. Let $\Xi_k \equiv \xi_k - \Pi(\xi_k|\tilde{M}_k) + \Pi(\Pi(\xi_k|\tilde{M}_k)|\tilde{M}_k) - \Pi(\Pi(\Pi(\xi_k|\tilde{M}_k)|\tilde{M}_k)|\tilde{M}_k) + \cdots$. The convergence of the series follows from, e.g., Theorem 2 of Chapter A.4 of Bickel et al. (1993, pp.438) and $\Xi_k$ is an element of $\mathcal{M}$. Furthermore, $\Pi(\Xi_k|\tilde{M}_k) = \xi_k$, $\Pi(\Xi_k|\tilde{M}_k) = 0$ and, therefore, $\Pi(\Xi_k|\tilde{M}_k) = 0$ for $l \neq k$ since $\mathcal{M}_l \subseteq \mathcal{M}_k$ for $l \neq k$. Thus, $\Pi(\sum_{i=1}^{K} \Xi_i|\tilde{M}_k) = \xi_k$, for $1 \leq k \leq K$. The existence is established. The uniqueness is argued as follows. Let $\sum_{k=1}^{K} \xi_{k|h} = 0$ be the difference of any two solutions. Then, $\Pi(\sum_{i=1}^{K} \xi_{i|h}|\tilde{M}_k) = 0$ for all $k = 1, \ldots, K$. This implies $\sum_{k=1}^{K} \xi_{k|h} \perp \mathcal{M}_k$. Since $\sum_{k=1}^{K} \xi_{k|h} \in \mathcal{M}$, it follows that $\sum_{k=1}^{K} \xi_{k|h} = 0$.

Step 3. Martingale representations of the projections of $\sum_{i=1}^{K} \xi_{i|h}$. Let $M^g_k(t) = (1 - \delta_k)I(Y_k \leq t) - \int_{0}^{t} \lambda_{C_k}(s, Z)Y_k(s)ds$ where $\lambda_{C_k}(s, Z)$ is the true conditional hazard of $C_k$ given $Z = z$. It follows from the counting process martingale representation of random variables with zero mean and finite second moment that

$$
\Pi(\sum_{i=1}^{K} \xi_{i|h} | S_k) = E(\sum_{i=1}^{K} \xi_{i|h} | Y_k, \delta_k, Z) = \int_{0}^{\tau_k} \tilde{h}_k(t, Z)dM_k(t) + \int_{0}^{\tau_k} \tilde{a}_k(t)dM_k(t) + \int_{0}^{\tau_k} \tilde{G}_k(t, Z)dM^g_k(t) \quad (S2.2)
$$

for some measurable functions $\tilde{h}_k$, $\tilde{a}_k$ and $\tilde{G}_k$, $1 \leq k \leq K$, where $\tilde{h}_k$ satisfies $E(\tilde{h}_k(t, Z)|Y_k = t, \delta_k = 1) = 0$ and $\tilde{a}_k$ is a non-random function. The last two terms of (S2.2) are orthogonal to each other and both are orthogonal to $\mathcal{M}_k$ while the first is an element of $\mathcal{M}_k$. Combining (S2.2) with (S2.1), it follows that $\tilde{h}_k(t, Z) = Z_k(t) - \mu_k(t)$.

Step 4. Constructing a parametric submodel. Let $\beta$ be in a small but fixed neighborhood of $\beta_0$. Let

$$
\lambda_k(t; \beta) = \lambda_{k0}(t)e^{(\beta - \beta_0)\midt - \mu_k(t) + \tilde{a}_k(t)} \quad \text{and} \quad \lambda_{C_k|Z}(t, Z; \beta) = \lambda_{C_k|Z}(t, Z)e^{(\beta - \beta_0)\tilde{G}_k(t, Z)}.
$$

Define

$$
f_k(y, d|Z; \beta) = e^{d\beta_k}Z_k \lambda_k^d(y|\beta)e^{-\int_{0}^{\tau_k} e^{\beta_k \lambda_k(t; \beta)dt}} \lambda_{C_k|Z}(y, Z; \beta)e^{-\int_{0}^{\tau_k} \lambda_{C_k|Z}(t, Z; \beta)dt},
$$

where $z = (z_1, \ldots, z_K)$ and $d$ takes value 0 or 1. If a parametric family, with parameter $\beta$, has (conditional) marginal densities as $f_k$, then the family is a parametric submodel since the expression of $f_k$ fulfills the requirement of proportional hazards in (1). Such a family of densities is constructed in the following.

Let $u_k(\beta) = f_k(Y_k, \delta_k|Z, \beta)/f_k(Y_k, \delta_k|Z, \beta_0) - 1$. Then $u_k(\beta) \in \mathcal{S}_k$ and $u_k(\beta_0) = 0$. Let

$$
v_k(\beta) = u_k(\beta) - \Pi(u_k(\beta)|\tilde{S}_k) + \Pi(\Pi(u_k(\beta)|\tilde{S}_k)|\tilde{S}_k) - \Pi(\Pi(\Pi(u_k(\beta)|\tilde{S}_k)|\tilde{S}_k)|\tilde{S}_k) + \cdots.
$$

Theorem 2 of A.4 of Bickel et al. (1993) ensures the convergence of the series and that

$$
v_k(\beta) \in \mathcal{S}, \quad \Pi(v_k(\beta)|\tilde{S}_k) = u_k(\beta) \quad \text{and} \quad \Pi(v_k(\beta)|\tilde{S}_k) = 0. \quad (S2.3)
$$
Let $v(\beta) = 1 + \sum_{k=1}^{K} v_k(\beta)$ and
\[ f(y_1, \delta_1, ..., y_K, \delta_K|z; \beta) = v(\beta) f_0(y_1, \delta_1, ..., y_K, \delta_K|z; \beta_0) \]
where $f_0$ denotes the true conditional density of $(Y_1, \delta_1, ..., Y_K, \delta_K)$ given $Z$. Notice that $f$ is a (conditional) density since $E(v(\beta)|Z) = 1$ and $v(\beta) \geq 0$ for $\beta$ in a small neighborhood of $\beta_0$. Observe that $f_k(y_k, \delta_k|z; \beta_0)$ are the true conditional marginal densities. Write
\[ f(y_1, \delta_1, ..., y_K, \delta_K|z; \beta) = f_k(y_k, \delta_k|z; \beta_0) \times \frac{v(\beta) f_0(y_1, \delta_1, ..., y_K, \delta_K|z; \beta_0)}{f_k(y_k, \delta_k|z; \beta_0)}. \]
Then, the log of the marginal density of $f$ is
\[ \log f_k(Y_k, \delta_k|z; \beta_0) + \log E(v(\beta)|Y_k, \delta_k, Z) = \log f_k(Y_k, \delta_k|z; \beta_0) + \log[1 + \Pi(v(\beta) - 1|S_k)] = \log f_k(Y_k, \delta_k|z; \beta_0) + \log(1 + u_k(\beta)). \]
Thus $f$ as a parametric family of densities is indeed a parametric submodel with parameter $\beta$.

Step 5. Verifying that the score of the parametric submodel is $\sum_{i=1}^{K} \xi_i h_i$. Observe that $v(\beta_0) = 1$ since $v_k(\beta_0) = u_k(\beta_0) = 0$. Moreover, $\frac{\partial}{\partial \beta} u_k(\beta)|_{\beta = \beta_0}$ is the same as (S2.2). The score of the parametric family $f$ at $\beta = \beta_0$ is $\frac{\partial}{\partial \beta} \log v(\beta)|_{\beta = \beta_0} = \frac{\partial}{\partial \beta} v(\beta)|_{\beta = \beta_0}$. It follows from (S2.3) that
\[ \Pi(\frac{\partial}{\partial \beta} v(\beta)|_{\beta = \beta_0} |S_k) = \Pi(\frac{\partial}{\partial \beta} v_k(\beta)|_{\beta = \beta_0} |S_k) = \frac{\partial}{\partial \beta} u_k(\beta)|_{\beta = \beta_0} = \Pi(\sum_{i=1}^{K} \xi_i h_i |S_k). \]
The uniqueness of the alternating projection solution then implies that the score of the parametric submodel $f$ at $\beta = \beta_0$ is $\frac{\partial}{\partial \beta} v(\beta)|_{\beta = \beta_0} = \sum_{i=1}^{K} \xi_i h_i$. The proof is complete.