In order to stress that dM depends on μ, Λ_0 and that $g(h)$ depends on μ, we will now and then denote these quantaties with dM_{μ,Λ_0} and g_μ, respectively. Here $\int g(h)(t)dM(t)$ equals the projection of $\int h(t,\bar{Z}(t-))dM(t)$ onto the tangent space $\{\int g(t)dM(t) : g\}$ of Λ_0. In other words, the full data estimating functions are of the form $\int h(t,\bar{Z}(t-))dM(t)$ for h chosen so that it is orthogonal to the tangent space of Λ_0. Thus, each h yields an estimating function for $\mu = \beta$ indexed by a nuisance parameter λ_0. The optimal estimating function in the full data model in which one observes n i.i.d. observations of X is obtained by selecting $h(t,\bar{Z}(t-)) = W(t)$, which corresponds with the full data efficient score $S^F_{eff}(X \mid F_X)$. Our general choice (3.6) $IC_0(Y \mid G, D_h)$ is given by

$$IC_{01}(Y \mid G, D_h(\cdot \mid \mu, \Lambda_0)) = D_h(X \mid \mu, \Lambda_0)\frac{\Delta}{\bar{G}(T \mid X)},$$

where $\Delta = I(C \geq T)$ and $\bar{G}(t \mid X) = P(C \geq t \mid X)$. Because D_h is an integral (sum) of unbiased estimating functions, an alternative choice of $IC_0(Y \mid G, D_h)$ is given by

$$IC_{02}(Y \mid G, D_h(\cdot \mid \mu, \Lambda_0)) = \int \{h(t,\bar{Z}(t-)) - g_\mu(h)(t)\}\frac{I(C \geq t)}{\bar{G}(t \mid X)}dM_{\mu,\Lambda_0}(t).$$

We have the following lemma (as in Robins, 1993a).

Lemma 3.1 *If $D(X)I(\bar{G}(T \mid X) > 0) = D(X)$ F_X-a.e., then*

$$E(IC_{01}(Y \mid G, D) \mid X) = D(X) \;\; F_X\text{-a.e.}$$

If $\int h(t,\bar{Z}(t-))I(\bar{G}(t \mid X) > 0)dM(t) = \int h(t,\bar{Z}(t-))dM(t)$ a.e., then

$$E(IC_{02}(Y \mid G, D) \mid X) = D(X) \;\; F_X\text{-a.e.}$$

Proof. We have

$$\begin{aligned} E\left(\frac{D(X)I(T \leq C)}{\bar{G}(T \mid X)} \mid X\right) &= I(\bar{G}(T \mid X) > 0)D(X) \\ E\left(IC_{02}(Y \mid G, D) \mid X\right) &= \int h(t,\bar{Z}(t-))I(\bar{G}(t \mid X) > 0)dM(t). \square \end{aligned}$$

Let $IC_0(Y \mid G, D_h(\cdot \mid \mu, \Lambda_0))$ denote one of these two choices of estimating functions for μ with nuisance parameters Λ_0, G. Given estimators $G_n, \Lambda_{0,n}$ of G, Λ_0, each of these observed data estimating functions yields an estimating equation for $\mu = \beta$:

$$0 = \sum_{i=1}^n IC_0(Y_i \mid G_n, D_h(\cdot \mid \mu, \Lambda_{0,n})).$$

As discussed in Section 3.1, if we assume a multiplicative intensity model for $A(t) = I(C \leq t)$ w.r.t. $\mathcal{F}(t) = (\bar{A}(t), \bar{X}(\min(t,C)))$, then Coxph() can be used to obtain an estimate of G. We also need a reasonable estimator of Λ_0. Since our estimating function is orthogonal to the nuisance tangent space in the observed data model $\mathcal{M}(G)$ with G known, which thus includes the tangent space generated by Λ_0, the influence curve of μ_n is not affected by the first-order behavior of $\Lambda_{0,n}$ (except that it needs to be consistent at an appropriate rate). Therefore, it suffices to construct an ad hoc estimator of Λ_0. Since $E(dN(t)) = EE(dN(t) \mid \bar{Z}(t-)) = \lambda_0(t)E(Y(t)\exp(\beta W(t)))$ it follows that

$$\Lambda_0(t) = \Lambda_0(t \mid \beta) \equiv \int_0^t \frac{E(dN(t))}{E(Y(t)\exp(\beta W(t)))}.$$

For general β, we denoted the right-hand side of the last equation with $\Lambda_0(t \mid \beta)$, while at the true β it equals $\Lambda_0(t)$. Now, note that

$$\begin{aligned} E(dN(t)) &= E\left(dN(t)\frac{I(C > t)}{\bar{G}(t \mid X)}\right), \\ E(Y(t)\exp(\beta W(t))) &= E\left(Y(t)\exp(\beta W(t))\frac{I(C > t)}{\bar{G}(t \mid X)}\right). \end{aligned}$$

This suggests the following estimator of $\Lambda_0(t \mid \beta)$:

$$\Lambda_{0,n}(t \mid \beta) = \frac{1}{n}\sum_{i=1}^n \int \frac{dN_i(t)I(C_i > t)/\bar{G}_n(t \mid X_i)}{\frac{1}{n}\sum_{i=1}^n Y_i(t)\exp(\beta W_i(t))I(C_i > t)/\bar{G}_n(t \mid X_i)}.$$

Substitution of $\Lambda_{0,n}(t \mid \beta)$ for Λ_0 in our estimating function yields the following estimating equation for β:

$$0 = \sum_{i=1}^n IC_0(Y_i \mid G_n, D_h(\cdot \mid \beta, \Lambda_{0,n}(\cdot \mid \beta))). \tag{3.10}$$

We now reparametrize the full data estimating function so that it has a variation-independent nuisance parameter

$$D_h^r(X \mid \beta, \rho) = D_h(X \mid \beta, \Lambda_0(\beta)),$$

where ρ denotes the additional parameters beyond β identifying $\Lambda_0(\cdot \mid \beta)$. We denote this reparametrized class of full data structure estimating functions with $D_h(x \mid \beta, \rho)$ again.

We will now provide a sensible data-adaptive choice for the full data index h. If $\lambda_C(t \mid X) = \lambda_C(t \mid Z)$ so that censoring is explained by the covariates in our multiplicative intensity model, then a sensible estimating function is the one corresponding with the score of the partial likelihood for β and Λ_0 ignoring $V_2(t)$ which is given by (Andersen, Borgan, Gill, and Keiding, 1993)

$$\int \left\{W(t) - \frac{E(W(t)Y(t)I(C > t)\exp(\beta W(t)))}{E(Y(t)I(C > t)\exp(\beta W(t)))}\right\} I(C > t)dM(t). \tag{3.11}$$