BIOMETRICS 57, 343–347
June 2001

Discussion of the Frangakis and Rubin Article

James Robins,[1,2,*] Andrea Rotnitzky,[1] and Marco Bonetti[1]

[1]Department of Biostatistics and [2]Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts 02115, U.S.A.
* *email:* robins@hsph.harvard.edu

1. Introduction

In clinical survival studies conducted in the United States, rich data are frequently available on variables recording time to drop out and the evolution over time of a patient's clinical signs, symptoms, and laboratory measurements. Motivated by a Greek study, Frangakis and Rubin (FR) consider estimation of a marginal survival curve under a double sampling design from severely impoverished reduced data that included none of the aforementioned variables due to confidentiality restrictions. To make our discussion relevant to settings with and without confidentiality restrictions, we shall consider survival curve estimation both from rich data that includes the aforementioned variables and from reduced data that does not. Let T, L, and C be continuous failure, dropout, and administrative censoring times, respectively, with time measured from date of enrollment. The goal is to estimate the cumulative net (marginal) hazard of failure $\Lambda(t) = \int_0^t \lambda_T(u)du$ and the survival function $S(t) = e^{-\Lambda(t)}$ under a double sampling design in which a subset of the dropouts is followed up in a second-phase sample. In their analysis, FR assumed (a) all dropouts had the same chance of being selected into the second-stage sample, (b) all dropouts selected into the second phase had their censoring indicator $\Delta \equiv \delta(T < C)$ and their minimum $X = \min(T, C)$ of censoring and failure successfully ascertained, and (c) C was independent of underlying variables such as T and L. It was necessary for FR to impose these assumptions because FR's estimator of $\Lambda(t)$ is inconsistent unless (a)–(c) hold.

In practice, one or more of assumptions (a)–(c) may often fail to hold. For instance, in the first paragraph of their Section 5, FR noted it will often happen that a subset of the dropouts pursued in the second phase will fail to have (X, Δ) ascertained, violating assumption (b). FR recommend that members of this subset be treated in the analysis as having been administratively censored. However, were this recommendation to be followed, FR's survival estimator would be inconsistent and could be severely biased if the number of second-stage subjects who do not have (X, Δ) ascertained is large. Assumption (a) will be false if a potentially more efficient design has been employed in which subjects who dropped out early are oversampled in the second phase. Assumption (c) will be false when there are secular trends in the distribution of T, as was the case during the 1980s and 1990s for the survival time T of AIDS patients. If only assumption (a) were false, then, as FR note, identification of $\Lambda(t)$ could be restored by addition of the known second-phase sampling probabilities to FR's reduced data. However, when assumptions (b) and/or (c) are false, both additional assumptions and rich data are required to restore identification.

In this discussion, we make the following points. First, we show in Section 4 that, if assumptions (a)–(c) hold, then (i), when only the reduced data are available for analysis, FR's estimator is algebraically identical to the efficient inverse probability weighted (IPW) estimator, but (ii) when rich data are available, FR's estimator is inefficient; for this case, we provide in Section 8 a locally semiparametric efficient (LSE) estimator of $S(t)$ that exploits the information in the rich data. Second, we consider the more realistic setting in which (i) FR's estimator is inconsistent because one or more of assumptions (a)–(c) fail to hold and (ii) rich data are available. In Sections 6 and 7, we derive doubly robust LSE survival curve estimators under the assumption that the missingness process is ignorable.

In summary, we describe aspects of a powerful methodology for the analysis of doubly sampled censored survival data that can resolve the three principal problems left open by FR: (i) how to construct locally LSE estimators of $S(t)$ when FR's assumptions (a)–(c) hold and, as would typically be the case in the United States, rich data are available for analysis; (ii) how to construct LSE doubly robust estimators of $S(t)$ when FR's assumptions (b) and/or (c) fail but missingness remains ignorable; and (iii) how to conduct a sensitivity analysis when missingness may be nonignorable. This methodology is in effect a subset of the general theory developed in Robins and Rotnitzky (1992), Robins (1993a,b), Robins, Rotnitzky, and Scharfstein (1999), and Scharfstein, Rotnitzky, and Robins (1999a) for the analysis of semi- and nonparametric right-censored data models specialized to the case of doubly sampled censored survival data. Because of space limitations, our resolution of problem (iii) will be described elsewhere.

In order to successfully exploit our general theory in the context of doubly sampled data, an additional problem must be faced. Specifically, in this context, the distribution of the censoring variable has both discrete and continuous components; as a result, none of the estimators of $S(t)$ previously proposed in the aforementioned papers are directly applicable. In Section 6 and the Appendix, we provide a survival estimator that allows for a mixed discrete and continuous censoring distribution as in Satten, Datta, and Robins (unpublished manuscript).

2. Representation of the Rich Data as Right-Censored Full Data

One key insight underlying the construction of LSE estimators that are consistent under much less stringent assumptions than FR's (a)–(c) is that the observed rich data can be represented as "full data" that is right censored by a censoring variable Q whose distribution function is discontinuous, with an atom of positive mass at the random dropout time L. To specify the full data, suppose that, contrary to fact, administrative censoring was absent and data on T were obtained through double sampling for 100% of dropouts; then, for all subjects, we would observe full data $F = (T, L, \bar{Z}(L))$, where, by convention, $L = T$ if a subject never drops out and $\bar{Z}(u) = \{Z(x); 0 \leq x \leq u\}$ denote the data on all other possibly time-dependent covariates that would be available on a subject up to time u. In the actual study, there is administrative censoring and incomplete double sampling of dropouts. Hence, as in FR, we let $S = 1$ for a dropout selected for the second-phase sample and $S = 0$ otherwise. Let $\bar{V}(u) = (\delta(L < u), \delta(L < u)L, \bar{Z}\{\min(u, L)\})$ denote the covariate information actually recorded up to time u on a subject and whether and when the subject dropped out before u. We now construct a new censoring variable Q given by $Q = L$ if $L < \min(T, C)$ and $S = 0$ and $Q = C$ otherwise. Let $\tilde{X} = \min(T, Q)$ and $\tilde{\Delta} = \delta(\tilde{X} = T)$. Thus, $\tilde{\Delta} = 1$ if and only if a subject is actually observed to fail. Note that, when $\tilde{\Delta} = 0$, the censoring time Q is the dropout time L whenever the subject's dropout time is observed and the subject was not selected for follow-up in the second stage, i.e., when $S = 0$; otherwise, Q is the administrative censoring time C. Thus, in the setting considered by FR but with rich data available for analysis, the observed data can be represented as n i.i.d. copies of $O = (\tilde{X}, \tilde{\Delta}, \bar{V}(\tilde{X}))$. The important point here is that this representation shows that the observed rich data O is the full data F right censored by Q.

REMARK. Data on C may also be recorded on all subjects. This will have no bearing on our inferences concerning $\Lambda(t)$ if the known probability of selection into the second-stage sample does not depend on C (Robins, 1993a, Appendix 4). If second stage selection depends on C, then inference on $\Lambda(t)$ will require that one use the general theory of estimation in multiphase designs described in Scharfstein, Rotnitzky, and Robins (1999b, p. 1145) with a small correction printed in Robins (2000, Section 4) rather than the methods used in this discussion.

3. Weakening FR's Assumptions (a) and (c)

We now provide additional plausible modeling assumptions under which the observed rich data O is ignorable right-censored data. FR's assumptions (a)–(c) define a restrictive special case of this model. Specifically, we assume that the cause-specific hazard of censoring at time u given the full data F depends only on the observed past, i.e., $\lambda_Q(u \mid F) = \lambda_Q[u \mid \bar{V}(u)]$. Separating the discrete and continuous components of $\lambda_Q(u \mid F)$, this assumption says that (i) when $u \neq L$, the conditional cause-specific hazard of C, $\lambda_Q(u \mid F) = \lim_{h \to 0} \text{pr}(u \leq Q < u + h \mid \tilde{X} \geq u, u \neq L, F)/h$, depends on the full data F only on the observed past $\bar{V}(u)$, and (ii) when $u = L$, the known conditional probability of being selected into the second-phase study, $\lambda_Q(u \mid F) = \text{pr}(S = 0 \mid \tilde{X} \geq u, L = u, F)$, may depend on the observed past $\bar{V}(u)$. Robins and Rotnitzky (1992) noted that this assumption is equivalent to the assumption that the full data F are coarsened at random (CAR) (Heitjan and Rubin, 1991; Gill, Van der Laan, and Robins, 1997), i.e., the density $f_O(O \mid F)$ of the observed data O given the full data F is a function of F only through O. Missingness is ignorable if the data are CAR and, as we shall assume, the parameters of $f_O(O \mid F)$ are distinct from those of the marginal distribution of F.

We refer to the model defined by assumptions (a)–(c) and the reduced data as FR's model. It then follows that FR's model is the special case of our model in which, (1) by assumption (c), when $u \neq L$, the conditional cause-specific hazard of C does not depend on the past $\bar{V}(u)$, i.e., $\lambda_Q(u \mid \bar{V}(u)) = \lambda_Q(u)$, (2) by assumption (a), when $u = L$, the chance of a dropout not being selected for the second phase is a known constant $1 - \omega$, i.e., $\lambda_Q(u \mid \bar{V}(u)) = 1 - \omega$, and (3) the only data available for analysis are n i.i.d. copies of the reduced data $O_{\text{red}} = (I, R^{\text{obs}}, IX, I\Delta)$, where $\Delta \equiv \delta(T < C), X = \min(T, C), R^{\text{obs}} \equiv 1 - \delta(L < X)$ is the indicator that a subject was not observed to drop out during the study and $I \equiv R^{\text{obs}} + (1 - R^{\text{obs}})S$ is the indicator that (X, Δ) is recorded.

4. FR's Estimator Is an IPW Estimator

We now show that FR's model is itself an ignorable semiparametric missing data model and that FR's estimator of $\Lambda(t)$, which is the nonparametric maximum likelihood estimator (NPMLE) in this model, is algebraically identical to the efficient inverse probability weighted estimator. Specifically, FR's model is a missing data model with I as the missing data indicator and $F_{\text{red}} = (X, \Delta, R^{\text{obs}})$ rather than F acting as the full data since, when $I = 1$, O_{red} is equivalent to observing the full data F_{red}. FR's model also implies that (i) the data are CAR and thus ignorable in the model with full data F_{red} and observed data O_{red} since the conditional probability of nonresponse ($I = 0$) given F_{red} depends on F_{red} only through the reduced data $O_{\text{red}} = (I, R^{\text{obs}})$ and (ii) the joint law of F_{red} is completely unrestricted. Now, the cumulative hazard $\Lambda(t)$ is a functional of the distribution of the full data F_{red} since, as FR note, $\Lambda(t) = \text{E}(\int_0^t dN^*(u)/\text{pr}(X > u)) \equiv \text{E}(\int_0^t \delta(T = u, X \geq u)/\text{E}\{\delta(X \geq u)\})$. But Robins and Rotnitzky (1992) and Rotnitzky and Robins (1995) show that the NPMLE of a functional of an unrestricted full data distribution in a CAR model is the efficient IPW estimator $\Sigma_i I_i \hat{\pi}_i^{-1} \{\int_0^t dN_i^*(u)/\Sigma_i I_i \hat{\pi}_i^{-1} \delta(X_i \geq u)\}$ obtained by replacing all expectations of incompletely observed variables in the functional by weighted averages with inverse probability weights $I_i \hat{\pi}_i^{-1}$, where $\hat{\pi}_i$ is the empirical conditional probability of observing full data F_{red} on subject i, i.e., $\hat{\pi}_i = 1$ for nondropout i, and for dropout i, $\hat{\pi}_i$ is the empirical proportion $\hat{\omega}$ of the dropouts selected in the second-phase sampling. Since FR's estimator is the NPMLE in this model, it has to coincide with the efficient IPW estimator.

5. Relaxing FR's Assumption (b)

We now return to the general problem of estimation with rich data O. We wish to consider inference that is valid when assumption (b) fails, i.e., when there are nonrespondents in the second-phase sample. To do so, we augment the observed rich data O by J, where $J = 1$ for dropouts who have (X, Δ) ascertained and $J = 0$ otherwise. We redefine the censoring time Q to be L if $L < X$ and either (i) $S = 0$ or (ii) $S = 1$ and $J = 0$; otherwise, $Q = C$. Hence, a dropout is regarded as censored at his/her observed dropout time if the dropout is either not selected into the second-stage sample, i.e., if $S = 0$, or if he/she is selected but fails to respond, i.e., $S = 1$ and $J = 0$. Thus, the discrete hazard of censoring at the dropout time $u = L$ is now $\lambda_Q(u \mid F) = \text{pr}(S = 1 \mid F, L = u)\text{pr}(J = 0 \mid S = 1, F, L = u) + \text{pr}(S = 0 \mid F, L = u)$. As in Section 3, we consider the possibility that assumption (a) fails because selection into the second-phase sampling depends on the observed past, i.e., we make the less restrictive assumption (a′) that $\text{pr}(S = 1 \mid F, L = u) = \text{pr}(S = 1 \mid L = u, \bar{V}(u))$ and this probability is known by design. We replace assumption (b) by the less restrictive assumption (b′) that the unknown conditional probability $\text{pr}(J = 0 \mid S = 1, F, L = u)$ of being a nonrespondent depends only on the observed past, i.e., it equals $\text{pr}(J = 0 \mid S = 1, \bar{V}(u), L = u)$. Finally, as in Section 3, we relax (c) and allow the possibility of dependent censoring by making the assumption (c′) that, when $u \neq L$, $\lambda_Q(u \mid F) = \lambda_Q(u \mid \bar{V}(u))$. Note that, with the augmented data, assumptions (a′)–(c′) are equivalent to the CAR assumption $\lambda_Q(u \mid F) = \lambda_Q(u \mid \bar{V}(u))$.

6. Estimation Under the Less Restrictive Ignorable Assumptions (a′)–(c′)

The Nelson–Aalen estimator of $\Lambda(t)$ based on the data O is $\hat{\Lambda}_{NA}(t) = \int_0^t \Sigma_{i=1}^n \{d\tilde{N}_i(u)\}/\{\Sigma_{i=1}^n \tilde{Y}_i(u)\}$, where $d\tilde{N}(u) = \delta(\tilde{\Delta} = 1, \tilde{X} = u)$, and $\tilde{Y}(u) = \delta(\tilde{X} \geq u)$ is the at-risk indicator at u. In our setting, $\hat{\Lambda}_{NA}(t)$ will generally be inconsistent for $\Lambda(t)$ because censoring by Q may be dependent. Therefore, we propose estimating $\Lambda(t)$ by the inverse probability of censoring weighted Nelson–Aalen (IPCW NA) estimator of Robins (1993b). The IPCW NA estimator is defined as $\hat{\Lambda}(t) = \int_0^t \Sigma_{i=1}^n \{\hat{K}_i^{-1}(u)d\tilde{N}_i(u)\}/\{\Sigma_{i=1}^n \hat{K}_i^{-1}(u)\tilde{Y}_i(u)\}$, where $\hat{K}(u)$ is an efficient estimator of the conditional probability $K(u) = \text{pr}(Q > u \mid F)$ of remaining uncensored up to u under (a′)–(c′) based on the following model $\lambda_Q[u \mid \bar{V}(u); \eta]$ for $\lambda_Q[u \mid \bar{V}(u)]$, where $\eta = (\eta_1, \eta_2, \eta_3)$:

$$\lambda_Q[u \mid \bar{V}(u); \eta_1] = \lambda_{0Q}(u)\exp(\alpha' W(u)), \qquad u \neq L, \tag{1}$$

$$\text{logit pr}(J = 0 \mid S = 1, \bar{V}(u); \eta_2) = \psi_0 + \psi_1' W^*(u), \qquad u = L \tag{2}$$

$$\text{logit pr}(S = 1 \mid L = u, \bar{V}(u); \eta_3) = \text{logit pr}(S = 1 \mid L = u, \bar{V}(u)) + \eta_3' W^{**}(u), \qquad u = L. \tag{3}$$

Here, $W(u)$, $W^*(u)$, and $W^{**}(u)$ are known vector functions of $\bar{V}(u)$ chosen by the analyst; $\alpha, \eta_2 = (\psi_0, \psi_1')'$, and η_3 are parameter vectors to be estimated; $\lambda_{0Q}(u)$ is an unknown baseline hazard function; and η_1 is the infinite dimensional parameter $(\lambda_{0Q}(u), \alpha)$. We provide the formula for $\hat{K}(u)$ in the Appendix. Note the logit of the known second-phase sampling probabilities are entered as an offset in model (3). Hence, the true value of η_3 is known to be zero. However, we ignore this knowledge and estimate η_3 by maximum likelihood. We do so because, under CAR, one never decreases and usually increases the asymptotic efficiency of our IPCW estimator $\hat{\Lambda}(t)$ by replacing known values of parameters in any model $\lambda_Q[u \mid \bar{V}(u); \eta]$ for censoring by efficient estimates. Indeed, even when assumptions (a)–(c) hold so that FR's estimator based on the reduced data O_{red} is also consistent, the asymptotic variance of the IPCW NA estimator $\hat{\Lambda}(t)$ never exceeds and is often much less than that of FR's estimator. However, $\hat{\Lambda}(t)$ is not fully semiparametric efficient unless the functions $W(u)$, $W^*(u)$, and $W^{**}(u)$ are chosen optimally, as discussed in the following paragraph. Robins (1993b) showed that the IPCW estimator $\hat{\Lambda}(t)$ is consistent and asymptotically normal (CAN) when CAR holds and the model $\lambda_Q[u \mid \bar{V}(u); \eta]$ is correct.

7. Locally Semiparametric Efficient Doubly Robust Estimators Under the CAR Assumptions (a′)–(c′)

Unfortunately, $\hat{\Lambda}(t)$ will be inconsistent if $\lambda_Q[u \mid \bar{V}(u); \eta]$ is misspecified. When assumption (b) and/or (c) may not hold, one cannot be certain that the model $\lambda_Q[u \mid \bar{V}(u); \eta]$ is correctly specified. Because of this uncertainty, one might choose to specify a parametric model $f(F; \theta)$ for density of the full data F and then estimate both θ and the full data functional $\Lambda(t)$ based on the data O with parametric Bayes, parametric maximum likelihood, and/or parametric multiple imputation estimators (Rubin, 1987). However, if the model $f(F; \theta)$ is misspecified, these estimators of $\Lambda(t)$ are inconsistent and can be severely biased when the number of nonrespondents in the second-stage sample is large (Scharfstein, Rotnitzky, and Robins, 1999b, Section 3.2.6; Robins and Wang, 2000). Hence, the best that can be hoped for is to find a locally semiparametric efficient (LSE) doubly robust estimator. An estimator is doubly robust if it is CAN under the assumption of CAR when either (but not necessarily both) of the models $f(F; \theta)$ or $\lambda_Q(u \mid \bar{V}(u); \eta)$ is correct. A doubly robust estimator is LSE if it is the asymptotically most efficient doubly robust estimator of $\Lambda(t)$ when both models $f(F; \theta)$ and $\lambda_Q(u \mid \bar{V}(u); \eta)$ happen to be correct. Thus, in CAR models, it is best to simultaneously model the censoring (missingness) mechanism and the law of the full data and then calculate a LSE doubly robust estimator (Robins, 2000; Robins, Rotnitzky, and van der Laan, 2000, Section 7; Scharfstein, Rotnitzky, and Robins, 2000b, Section 3.2). In the Appendix, we show that there exist random functions $W_{\text{opt}}(u)$, $W_{\text{opt}}^*(u)$, $W_{\text{opt}}^{**}(u)$ such that the associated IPCW estimator $\hat{\Lambda}_{\text{opt}}(t)$ is doubly robust and LSE.

8. A Rich Data Estimator That Improves on FR's Estimator Under FR's Assumptions

Under FR's assumptions (a)–(c), $\hat{\Lambda}_{\text{opt}}(t)$ is guaranteed to be CAN since these assumptions imply that our censoring model $\lambda_Q[u \mid \bar{V}(u); \eta]$ is correctly specified with $\alpha = n_3 = \psi_1 = 0$ and $\psi_0 = -\infty$. Further, $\hat{\Lambda}_{\text{opt}}(t)$ has asymptotic variance that is always less than or equal to that of FR's estimator. In fact, $\hat{\Lambda}_{\text{opt}}(t)$ remains LSE under FR's assumptions (a)–(c) since these assumptions are just further *a priori* restrictions on the ignorable censoring model $\lambda_Q[u \mid \bar{V}(u); \eta]$ and, in an ignorable model, knowledge of the censoring mechanism has no effect on efficiency.

References

Gill, R. D., van der Laan, M. J., and Robins, J. M. (1997). Coarsening at random: Characterizations, conjectures and counterexamples. *Proceedings of the First Seattle Symposium on Survival Analysis,* 255–294. New York: Springer-Verlag.

Heitjan, D. F. and Rubin, D. B. (1991). Ignorability and coarse data. *The Annals of Statistics* **19,** 2244–2253.

Robins, J. M. (1993a). Analytic methods for estimating HIV treatment and cofactor effects. In *Methodological Issues of AIDS Mental Health Research,* D. G. Ostrow and R. Kessler (eds), 213–290. New York: Plenum.

Robins, J. M. (1993b). Information recovery and bias adjustment in proportional hazards regression analysis of randomized trials using surrogate markers. *Proceedings of the Biopharmaceutical Section, American Statistical Association,* 24–33. Washington, D.C.: American Statistical Association.

Robins, J. M. (2000). Robust estimation in sequentially ignorable missing data and causal inference models. *Proceedings of the 1999 Joint Statistical Meetings,* Washington, D.C.: American Statistical Association in press.

Robins, J. M. and Rotnitzky, A. (1992). Recovery of information and adjustment for dependent censoring using surrogate markers. In *AIDS Epidemiology—Methodological Issues,* N. Jewell, K. Dietz, and V. Farewell (eds), 297–331. Boston: Birkhäuser.

Robins, J. M. and Wang, N. (2000). Inference for imputation estimators. *Biometrika* **87,** 113–124.

Robins, J. M., Rotnitzky, A., and Scharfstein, D. O. (1999). Sensitivity analysis for selection bias and unmeasured confounding in missing data and causal inference models. In *Statistical Models in Epidemiology,* E. Halloran (ed), 1–94. New York: Springer-Verlag.

Robins, J. M., Rotnitzky, A., and van der Laan, M. J. (2000). Discussion of "On Profile Likelihood" by Murphy and van der Vaart. *Journal of the American Statistical Association* **95,** 477–482.

Rotnitzky, A. and Robins, J. M. (1995). Semiparametric regression estimation in the presence of dependent censoring. *Biometrika* **82,** 805–820.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys.* New York: Wiley.

Scharfstein, D. O., Rotnitzky, A., and Robins, J. M. (1999a). Adjusting for non-ignorable drop-out using semiparametric non-response models. *Journal of the American Statistical Association* **94,** 1096–1120.

Scharfstein, D. O., Rotnitzky, A., and Robins, J. M. (1999b). Rejoinder to "Adjusting for Non-ignorable Drop-out Using Semiparametric Non-response Models." *Journal of the American Statistical Association* **94,** 1135–1146.

Scharfstein, D. O., Robins, J. M., Rotnitzky, A., and Eddings, W. (2001). Causal inference in randomized studies with informative censoring and discrete time-to-event outcomes. *Biometrics,* in press.

Received April 2000. *Revised October* 2000.
Accepted October 2000.

Appendix

Formula for $\hat{K}(u)$

In this Appendix, we provide a formula for $\hat{K}(u)$ that takes into account that Q has a mixed continuous and discrete distribution,

$$\begin{aligned}\hat{K}(u) &= \hat{K}(u,\hat{\eta}) \\ &= \{\text{expit}\{\text{logit pr}(S=1 \mid L, \bar{V}(L)) + \hat{\eta}_3' W^{**}(L)\} \\ &\quad \times [1 - \text{expit}\{\hat{\psi}_0 + \hat{\psi}_1' W^*(L)\}]\}^{\delta(L<u)} \\ &\quad \times \exp\left\{-\int_0^u d\hat{\Lambda}_{0Q}(v)\exp(\hat{\alpha}' W(v))\right\},\end{aligned}$$

where $\text{expit}(x) = e^x/(1+e^x)$;

$$d\hat{\Lambda}_{0Q}(v) = \frac{\sum_{i=1}^{n} \delta(L_i \neq v)\delta(Q_i = v, \tilde{X}_i \geq v)}{\sum_{i=1}^{n} \exp(\hat{\alpha}' W_i(v))\delta(\tilde{X}_i \geq v)},$$

$\hat{\alpha}$ is the Cox partial likelihood estimator of α in the Cox model (1) for the cause-specific hazard of C for $C < T$, $\hat{\eta}_3$ is the MLE of η_3 solving

$$0 = \sum_i \{S_i - \text{pr}(S = 1 \mid \bar{V}_i(L_i); \eta_3)\}\delta(L_i \leq \tilde{X}_i) W_i^{**}(L_i),$$

and $\hat{\eta}_2 = (\hat{\psi}_0, \hat{\psi}_1')'$. To compute $\hat{\Lambda}(t)$, $W(u)$ needs only to be evaluated at the observed administrative censoring times $C \leq \min(T,t)$, while $W^{**}(u)$ and $W^*(u)$ need to be evaluated at observed dropout times $L < t$.

Formula for $W_{\text{opt}}(u)$, $W^*_{\text{opt}}(u)$, $W^{**}_{\text{opt}}(u)$

Let $\hat{\eta}_0$ and $\hat{\Lambda}_0(t)$ equal $\hat{\eta}$ and $\hat{\Lambda}(t)$ as defined previously and let $\hat{\theta}$ be the MLE of θ based on data O in model $f(F:\theta)$. Define, for any function $r(v,\eta)$,

$$\begin{aligned}U_{\theta\eta}(u,r) &= \int_u^t K^{-1}(v;\eta)\{d\tilde{N}(v) - \lambda_T(v;\theta)Y(v)dv\}/r(v,\eta) \\ H_{\theta\eta}(u,r) &= \text{E}_{\theta\eta}[U_{\theta\eta}(u,r) \mid \bar{V}(u), Y(u)=1, u \neq L], \\ H^*_{\theta\eta}(u,r) &= \text{E}_{\theta\eta}[U_{\theta\eta}(u,r) \mid \bar{V}(u), Y(u)=1, u = L, S=1],\end{aligned}$$

and

$$H^{**}_{\theta\eta}(u,r) = \text{E}_{\theta\eta}[U_{\theta\eta}(u,r) \mid \bar{V}(u), Y(u) = 1, u = L].$$

Let $\hat{r}(v,\eta) = n^{-1}\Sigma_{i=1}^n K_i^{-1}(v,\eta)\tilde{Y}_i(v)$. Recursively for $j = 0,1,2,\ldots,$ define

$$\begin{aligned}W_{\hat{\theta}\hat{\eta}_j}(u) &= \left(W(u)', H_{\hat{\theta}\hat{\eta}_j}(u,\hat{r})\right)', \\ W^*_{\hat{\theta}\hat{\eta}_j}(u) &= \left(W^*(u)', H^*_{\hat{\theta}\hat{\eta}_j}(u,\hat{r})\right)',\end{aligned}$$

and

$$W^{**}_{\hat{\theta}\hat{\eta}_j}(u) = \left(1, H^{**}_{\hat{\theta}\hat{\eta}_j}(u,\hat{r})\right)',$$

Then $\hat{\eta}_{j+1}$ and $\hat{\Lambda}_{j+1}(t)$ are $\hat{\eta}$ and $\hat{\Lambda}(t)$ except with $W_{\hat{\theta}\hat{\eta}_j}$,

$W^*_{\hat\theta\hat\eta_j}$, $W^{**}_{\hat\theta\hat\eta_j}$ replacing W, W^*, W^{**}. Finally, W_{opt}, W^*_{opt}, W^{**}_{opt} and $\hat\Lambda_{\mathrm{opt}}(t)$ are the limits as $j \to \infty$ of $W_{\hat\theta\hat\eta_j}$, $W^*_{\hat\theta\hat\eta_j}$, $W^{**}_{\hat\theta\hat\eta_j}$, and $\hat\Lambda_j(t)$, respectively. They can be computed by iterating on j until convergence. Consider the union model that specifies that the data are CAR and either $f(F;\theta)$ or $\lambda_Q(u \mid \bar V(u);\eta)$ is correct. The cumulative hazard estimator $\hat\Lambda_{\mathrm{opt}}(t)$ is LSE in the union model at the intersection submodel in which both $f(F;\theta)$ and $\lambda_Q(u \mid \bar V(u);\eta)$ are correct. When model $\lambda_Q[u \mid \bar V(u);\eta]$ is correct, $\hat\eta_j$ is assymptotically equivalent to $\hat\eta_0$. When model $\lambda_Q[u \mid \bar V(u);\eta]$ is misspecified, $\hat\eta_j$ can differ substantially from $\hat\eta_0$. In that case, $\hat\Lambda_{\mathrm{opt}}(t)$, in contrast to $\hat\Lambda_j(t)$, remains CAN, provided model $f(F;\theta)$ is correct.

If either $\hat\theta$ or the required expectations are too difficult to compute, one can sacrifice double robustness and use the following approach. In $U_{\theta\eta}(u)$, replace $\lambda_T(v;\theta)$ by a preliminary estimate based on nonoptimal W, W^*, W^{**} and replace $\mathrm{E}_{\theta\eta}\{\delta(\tilde X > v)\}$ by a sample average to give $\hat U(u)$. $\hat U(u)$ can be regressed on functions of $\bar V(u)$ among subjects with $Y(u) = 1$ and $u \neq L$ to give $\hat H(u)$. $\hat U(L)$ can be regressed on functions of $\bar V(L)$ to give $\hat H^{**}(L)$, and $\hat U(L)$ can be regressed on functions of $\bar V(L)$ among subjects with $S = 1$ to give $\hat H^*(L)$. Then, in the model characterized by CAR and (1)–(3) being true, the associated estimator, say $\hat{\hat\Lambda}_{\mathrm{opt}}(t)$, will be (nearly) semiparametric efficient if the regressions models used to calculate $\hat H(u)$, $\hat H^*(u)$, and $\hat H^{**}(u)$ are (nearly) correct and will be CAN even if they are badly misspecified, provided the censoring model is correct.

Finally, we note that the "1" column did not have to be included in $W^{**}_{\theta\eta}(u)$ to insure either double robustness or local semiparametric efficiency. Rather, its inclusion was necessary to guarantee that $\hat\Lambda_{\mathrm{opt}}(t)$ is always at least as efficient as FR's estimator of $\Lambda(t)$ when FR's assumptions (a)–(c) hold.