

Inference in Randomized Studies with Informative Censoring and Discrete Time-to-Event Endpoints

Daniel Scharfstein,^{1,*} James M. Robins,^{2,3} Wesley Eddings,¹ and Andrea Rotnitzky²

¹Department of Biostatistics, Johns Hopkins School of Hygiene and Public Health,
Baltimore, Maryland 21025, U.S.A.

Departments of ²Biostatistics and ³Epidemiology, Harvard School of Public Health,
Boston, Massachusetts 02115, U.S.A.

* *email*: dscharf@jhsph.edu

SUMMARY. In this article, we present a method for estimating and comparing the treatment-specific distributions of a discrete time-to-event variable from right-censored data. Our method allows for (1) adjustment for informative censoring due to measured prognostic factors for time to event and censoring and (2) quantification of the sensitivity of the inference to residual dependence between time to event and censoring due to unmeasured factors. We develop our approach in the context of a randomized trial for the treatment of chronic schizophrenia. We perform a simulation study to assess the practical performance of our methodology.

KEY WORDS: Coarsening at random; Competing risks; Curse of dimensionality; Inverse probability of censoring weighted estimation; Kaplan–Meier estimator; Sequential ignorability of censoring.

1. Introduction

In the early 1990s, the JANSSEN Research Foundation supported a United States/Canadian randomized clinical trial to evaluate the efficacy of different drug regimes in the treatment of chronic schizophrenia (Chouinard et al., 1993; Marder and Meibach, 1994). In this trial, subjects were randomly assigned to one of six treatment groups (placebo, haloperidol [20 mg], or risperidone at one of four levels [2, 6, 10, and 16 mg]) and then followed for a total of 8 weeks. To simplify our presentation, we focus on two of the treatment groups: placebo and 2 mg risperidone. Measurements of the Positive and Negative Symptom Scale (PANSS) were scheduled to be collected at weeks 0, 1, 2, 4, 6, and 8, where 0 refers to baseline. One of the major endpoints of the study was week of first clinical improvement, defined to be at least a 20% reduction from baseline in PANSS score. An analysis of this endpoint is complicated because the study suffered from severe and potentially informative dropout (i.e., censoring). Of the 88 placebo subjects with complete baseline information, 51 (58.0%) dropped out prior to observation of the main endpoint and end of study. Of these subjects, 42 dropped out because of “inadequate treatment response” (ITR), while the other 9 dropped out for other reasons, including abnormal lab results, adverse experiences, intercurrent illness, loss to follow-up, withdrawal of consent, and uncooperativeness. Of the eighty-four 2-mg risperidone subjects with complete baseline information, 33 (39.3%) were censored—25 due to ITR and 8 due to other reasons. Table 1 presents the treatment-specific number of subjects who were censored (ITR in parentheses), experienced first clinical improvement, and remained at risk for improvement, stratified by week of clinic visit.

Using the combined U.S. and Canadian sample, Figure 1 presents the treatment-specific Kaplan–Meier (1958) estimates of the cumulative distribution functions of week of first clinical improvement. The associated log-rank statistic for the test of no treatment difference is -1.3386 (p -value = 0.1807). Under the assumption of noninformative censoring (i.e., equality between the net and cause-specific hazards of time to first improvement; Fleming and Harrington, 1991, pp. 26–27), we have only mild evidence that 2 mg risperidone yields stochastically shorter times to first clinical improvement than placebo. However, the assumption of noninformative censoring is dubious. Since ITR is a major reason for dropout, the Kaplan–Meier estimator will tend to underestimate the length of time to first clinical improvement. The extent of underestimation will depend on the censoring rate and the degree of negative dependence between censoring and time to first clinical improvement. The data in Table 1 suggest that the underestimation may be larger in the placebo group than in the 2-mg risperidone arm. If so, the log-rank statistic and associated p -value will be conservative, i.e., the inappropriate assumption of noninformative censoring may be masking statistically significant treatment effects. Thus, it is critical that the treatment comparisons be adjusted for informative censoring.

The purpose of this article is to provide a methodology for constructing estimates of the treatment-specific marginal distributions of time to first clinical improvement that (1) appropriately adjusts for informative censoring due to measured time-independent and -dependent prognostic factors for censoring and time to improvement and (2) simultaneously quantifies the sensitivity of the inference to nonidentified residual

Table 1
Treatment-specific number of subjects who were censored (ITR in parantheses), improved, and remained at risk for improvement, stratified by week of clinical visit

Treatment		Week						
		0	1	2	4	6	8	
Placebo	At risk	88	86	57	37	22	14	
	Dropped Out	2 (1)	16 (12)	13 (12)	14 (12)	6 (5)		
	Improved		13	7	1	2	3	
2 mg Risperidone	At risk	84	84	61	39	23	16	
	Dropped Out	0	6 (5)	11 (11)	8 (7)	5 (5)		
	Improved		17	11	8	2	3	

dependence between censoring and time to improvement due to unmeasured factors. The ideas in this article generalize the work of Robins (1987, 1993), Robins and Rotnitzky (1992), Satten, Datta, and Robins (2000), and Robins and Finkelstein (2000), who addressed (1), and the nonparametric sensitivity analysis ideas of Fisher and Kanarek (1974), Slud and Rubinstein (1983), Klein and Moeschberger (1988), Klein et al. (1992), and Zheng and Klein (1994, 1995), who addressed (2) in the absence of prognostic factors.

The article is organized as follows. In Section 2, we describe in greater detail the data structure encountered in the JANSSEN trial. In Section 3, we show that the distribution of week of first clinical improvement is not identified without further restrictions. In Section 4, we propose a class of restrictions that are sufficient for identification. In Section 5, we discuss issues of testing and estimation. In Section 6, we present the results of a simulation study. In Section 7, we apply our methodology to the JANSSEN clinical trial. The final section is devoted to a discussion.

2. Data Structure

For each treatment group, we assume the study design calls for $M + 1$ clinic visits. In the JANSSEN study, $M = 5$. To ease exposition, we suppose that these visits were to occur at weeks $t = 0, 1, \dots, M$. Let T be the week of first clinical improvement (defined as a 20% reduction from baseline in PANSS score). If a subject does not improve through week M , we arbitrarily define $T = \tau$, where τ is some fixed whole number greater than M . So T takes values in the set $\{1, 2, \dots, M, \tau\}$. Let the censoring time C denote the last visit time prior to dropout and $C = M$ if a subject does not drop out. Therefore, C is an integer between 0 and M and is observed on all subjects. Let B be the vector of baseline covariates, which in this study includes age at entry, race, gender, age at onset of psychological symptoms, age of first psychiatric hospitalization, total duration of hospitalization prior to entry into the study, and PANSS score measured prior to baseline. Let P_t be the PANSS score at time t and $\bar{V}_t = (B', P_0, \dots, P_t)'$ for $t = 0, 1, \dots, M$.

We consider the observed data for an individual to be $O = (X = \min(T, C), \Delta = I(T \leq C), \bar{V}_X)$. Our goal is to use n i.i.d. copies of O , $O = \{O_i = (X_i, \Delta_i, \bar{V}_i, X_i) : i = 1, \dots, n\}$, to draw inference about the distribution of T . Let $\lambda(t) = P(T = t | T \geq t)$, $S(t) = P(T > t) = \prod_{s=1}^t \{1 - \lambda(s)\}$, $F(t) = 1 - S(t)$, and $f(t)$ denote the discrete hazard function, survivor function, cumulative distribution function, and

probability mass functions for $t = 1, 2, \dots, M$. Note that we are artificially discarding information that is available on C once T is observed to occur. A separate report will address whether and how this information might be used to improve inference about the distribution of T .

3. Identifiability

In the JANSSEN study, the law, F_O , of the observed data O is not sufficient to identify the distribution of T . To see this, note that this distribution can be written as a mixture of conditional distributions of T for subjects who are observed to improve and those who are censored, stratified by censoring occasion and recorded past data. Specifically,

$$\begin{aligned}
 f(T) = & f(T | \Delta = 1)P(\Delta = 1) \\
 & + f(T | X = M, \Delta = 0)P(X = M, \Delta = 0) \\
 & + \sum_{t=0}^{M-1} \left\{ \int f(T | X = t, \Delta = 0, \bar{V}_t) \right. \\
 & \quad \left. \times f(\bar{V}_t | X = t, \Delta = 0) d\mu(\bar{V}_t) \right\} \\
 & \times P(X = t, \Delta = 0). \tag{1}
 \end{aligned}$$

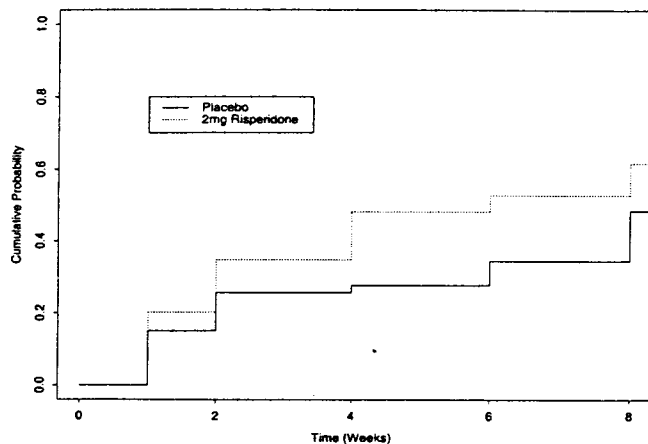


Figure 1. Treatment-specific Kaplan-Meier estimates for the cumulative distribution function of time to first clinical improvement.

In this mixture, the only components not identified from the law of the observed data are $f(T | X = t, \Delta = 0, \bar{V}_t)$, $t = 0, \dots, M-1$. This implies that $f(T)$ is not identified (Robins, 1987). In the absence of auxiliaries, many authors have addressed the nonidentifiability issue (cf., Crowder, 1996, 1997; Bedford and Meilijson, 1997).

4. Model

We can identify the distribution of T by assuming relationships between the nonidentified components of (1) and identifiable distributions. Our approach will be to model the relationship between $f(T | X = t, \Delta = 0, \bar{V}_t)$ and $f(T | X > t, \bar{V}_t)$. In particular, we impose a model for the full data $(T, \Delta, \bar{V}_T, (1-\Delta)C)$, which specifies that, for $t = 0, \dots, M-1$,

$$f(T | X = t, \Delta = 0, \bar{V}_t) = f(T | X > t, \bar{V}_t) \frac{\exp\{q_t(\bar{V}_t, T)\}}{E(\exp\{q_t(\bar{V}_t, T)\} | X > t, \bar{V}_t)}, \quad (2)$$

where $q_t(\bar{V}_t, T)$ are known functions of t , \bar{V}_t , and T . Implicit in equation (2) is the assumption that the supports of $f(T | X = t, \Delta = 0, \bar{V}_t)$ and $f(T | X > t, \bar{V}_t)$ are identical for all \bar{V}_t . This is equivalent to assuming that, at each time t , $t < M$, the chance that a subject at risk for censoring at time t remains uncensored is positive within all levels of \bar{V}_t and $T (T > t)$. In addition, note that we divide by the conditional expectation on the right-hand side of (2) to guarantee that $f(T | X = t, \Delta = 0, \bar{V}_t)$ is a proper conditional probability mass function for $t = 0, \dots, M-1$. The proof that (2) identifies $S(t)$ is provided in the Appendix.

Using Bayes' rule, it can be shown that model (2) is equivalent to the model

$$\text{logit } \lambda_C^\dagger(t | \bar{V}_t, T, T > t) = h_t(\bar{V}_t) + q_t(\bar{V}_t, T), \quad (3)$$

where $\lambda_C^\dagger(t | \cdot, T > t) = P(C = t | C \geq t, \cdot, T > t)$ is the conditional cause-specific hazard for dropout at time t given \cdot , $q_t(\bar{V}_t, T)$ are known functions of t , \bar{V}_t , and T and $h_t(\bar{V}_t)$ are unrestricted, unknown functions of t and \bar{V}_t . The above support condition is the same as assuming that, for all $t (t < M)$, \bar{V}_t , and $T (T > t)$, $\lambda_C^\dagger(t | \bar{V}_t, T, T > t) < 1$. From (3), we see that specifying the functions q_t , $t = 0, \dots, M-1$, is equivalent to quantifying for those who remain at risk after visit t , the dependence, on a log odds ratio scale, of censoring between visits t and $t+1$ on T after adjusting for all the recorded prognostic factors up to time t . To emphasize this important property, we refer to the functions q_t as censoring bias functions. The censoring bias functions essentially determine how T enters into the logistic regression model for the cause-specific probabilities of censoring. Note that $q_t = 0$ (i.e., $\lambda_C^\dagger(t | \bar{V}_t, T, T > t) = \lambda_C^\dagger(t | \bar{V}_t, T > t)$) for all $t = 0, \dots, M-1$ is equivalent to the assumptions of no unmeasured confounders for censoring (Robins, 1987) and sequential ignorability of censoring (Robins and Finkelstein, 2000). This assumption is a generalization of coarsening at random (Hietjan and Rubin, 1991; Jacobsen and Keiding, 1995; Gill, van der Laan, and Robins, 1997) and indicates that there is no residual dependence between T and C -after adjusting for recorded prognostic factors.

The following three points, proved using essentially the same techniques as in the proof of Theorem 3 of Rotintzky,

Robins, and Scharfstein (1998), summarize the critical features of model (2, 3):

- (a) Specification of the functions q_t , $t = 0, \dots, M-1$, identifies the distribution of T .
- (b) The functions q_t are not identified because all choices of these functions are compatible with the law of the observed data. Thus, no statistical test can reject any specific choice of q_t .
- (c) By specifying the functions q_t , we do not place any restrictions on the law of the observed data. Thus, restriction (2, 3) with specified q_t 's forms a nonparametric identified (NPI) model for O .

The above points tell us that our model (2, 3) has the same NPI characteristic as the models proposed for nonparametric sensitivity analysis in the competing risks literature without auxiliaries, i.e., the censoring bias functions are not identified, but once they are specified, the model is nonparametric for the observed data and the distribution of T is identified. Thus, following the lead of this literature, we suggest drawing inference by performing a sensitivity analysis in which the censoring bias functions are varied over a plausible range (see Section 5 for details of inference).

To facilitate our sensitivity analysis, it will be useful to parameterize the censoring bias functions by a parameter vector α , where $\alpha = 0$ if and only if sequential ignorability holds. Further, it is important that α be chosen so that it is interpretable by subject matter experts. The parameterization that we use in the simulation and analysis of the JANSSEN data is

$$q_t(\bar{V}_t, T; \alpha) = \alpha\{T - (t+1)\} \quad (4)$$

for $t = 0, \dots, M-1$. The censoring bias parameter α is interpreted as the conditional log odds ratio of dropping out between times t and $t+1$ per unit increase in T for subjects who have not improved by time t and have the same covariate history \bar{V}_t . So $\alpha > 0$ (< 0) implies that subjects who would improve later or never are more (less) likely to be censored at any time t than subjects who would improve earlier. Note that this censoring bias function implies that a subject who would never improve has $\exp[\alpha\{\tau - (t+1)\}]$ times the odds of dropping out between t and $t+1$ as compared with a subject with the same covariate history but who improves at time $(t+1)$. As mentioned above, the choice of τ is arbitrary, and sensitivity analysis should be performed with respect to τ as well. The choice of τ indicates the degree to which never improvers are assumed to be different from improvers with respect to their risk of dropping out.

When prognostic factor data are not recorded, our model becomes a new NPI model for the competing risks setting without auxiliaries, given by

$$f(T | X = t, \Delta = 0) = f(T | X > t) \frac{\exp\{q_t(T)\}}{E(\exp\{q_t(T)\} | X > t)}, \quad (5)$$

or equivalently,

$$\text{logit } \lambda_C^\dagger(t | T, T > t) = h_t + q_t(T), \quad (6)$$

where h_t are unknown constants and $q_t(T)$ is a known function of t and T . When $q_t(T) = 0$ for all t , restriction (5, 6) is identical to the assumption of noninformative censoring.

Point (b) above tells us that the assumption of noninformative censoring is not identifiable. Crowder (1996, 1997) prove that, for discrete survival data, the assumption of independent censoring is sometimes testable. This does not conflict with our result because independent censoring implies noninformative censoring but not vice versa. In addition, Crowder's test requires that T and C can be observed to occur simultaneously, which is ruled out in our formulation of the problem.

Thus, we suggest that sensitivity analyses be performed with respect to q_t . Letting $q_t(T)$ equal the right-hand side of (4) and varying α (and τ) over a plausible range is a convenient way of conducting such a sensitivity analysis. The advantage of our NPI model over the other NPI models for the competing risks setting without auxiliaries is that ours can naturally be extended to settings in which data have been collected on both time-independent and -dependent prognostic factors for T and C .

Nonparametric estimation of the distribution of T in model (2, 3) with fixed censoring bias functions will require nonparametric estimators for the functions $h_t(\bar{V}_t)$. When \bar{V}_t is high dimensional, nonparametric estimation is infeasible in moderate-sized datasets because the number of subjects at any given level of \bar{V}_t is too small to yield reasonable estimators. One way around this problem is to assume that the functions $h_t(\bar{V}_t)$ follow a parametric model, i.e.,

$$h_t(\bar{V}_t) = h_t(\bar{V}_t; \gamma^*), \tag{7}$$

where $h_t(\bar{V}_t; \gamma)$ is a known function and γ is an unknown finite dimensional parameter vector with true value γ^* . If model (7) is not a saturated model for the functions $h_t(\bar{V}_t)$ (which will be required when \bar{V}_t is high dimensional), then the model defined by the specification of the functions q_t and the restrictions (2, 3) and (7) is no longer a nonparametric model for O . Thus, in principle, given (7), we can test our specification of q_t . However, since q_t is not identified when the model for h_t is saturated, our ability to test our specification of q_t requires that we be certain that our unsaturated model $h_t(\bar{V}_t; \gamma)$ is correct. Since we will never be certain, we continue to recommend sensitivity analysis rather than testing q_t . However, since inference can be sensitive to the misspecification of the model for h_t , we recommend choosing as flexible a parametric model for h_t as possible. Given q_t , it is theoretically possible to test for misspecification of (7). However, with sample sizes found in practice, our ability to detect misspecification of such flexible models will be quite limited. In a simulation study in Section 6, we report on the impact of incorrect specification of (7).

5. Estimation and Testing

5.1 One-Sample Estimation

In this subsection, we show how to estimate the distribution of T under our model, characterized by (2, 3) and (7). Define $\lambda = (\lambda(1), \dots, \lambda(M))'$, $S = (S(1), \dots, S(M))'$, and $q = (q_0, \dots, q_{M-1})$. Parameters that are superscripted by $*$ denote the true value of the parameters. Parameter subscripts on expectations indicate the distribution under which the expectation is taken and expectations without subscripts indicate expectations with respect to the true law generating the data.

We estimate S^* and γ^* using an extension of the methods proposed in Rotnitzky et al. (1998). For each fixed q , our

estimators, \hat{S} and $\hat{\gamma}$, are the solution to the estimating equation $\sum_{i=1}^n U(O_i; S, \gamma; q) = 0$, where $U(O; S, \gamma) = (W_1(O; S, \gamma; q), W_2(O; S, \gamma; q), \dots, W_M(O; S, \gamma; q), Z(O; S, \gamma; q))'$,

$$\begin{aligned} &W_t(O; S; \gamma; q) \\ &= \frac{I(C \geq \min(T, M))}{\prod_{s=0}^{\min(T, M)-1} \{1 - \lambda_C^\dagger(s | \bar{V}_s, T, T > s; \gamma, q)\}} \\ &\quad \times \{I(T > t) - S(t)\}, \\ &Z(O; \gamma; q) \\ &= \sum_{t=0}^{M-1} I(C \geq t)I(T > t) \\ &\quad \times \left[I(C = t) \right. \\ &\quad \left. - I(C \geq \min(T, M))\lambda_C^\dagger(t | \bar{V}_t, T, T > t; \gamma, q) \right. \\ &\quad \left. \div \prod_{s=t}^{\min(T, M)-1} \{1 - \lambda_C^\dagger(s | \bar{V}_t, T, T > s; \gamma, q)\} \right] \\ &\quad \times \phi_t(\bar{V}_t), \end{aligned} \tag{8}$$

$\lambda_C^\dagger(t | \bar{V}_t, T, T > t; \gamma, q) = \exp\{h_t(\bar{V}_t; \gamma) + q_t(\bar{V}_t, T)\} / [1 + \exp\{h_t(\bar{V}_t; \gamma) + q_t(\bar{V}_t, T)\}]$, and $\phi_t(\bar{V}_t)$ is a vector function of \bar{V}_t of the same dimension as γ . The solution $\hat{S}(t)$ is a proper survivor function since it lies in the interval $[0, 1]$ and it is nonincreasing in t .

Under mild regularity conditions, we can show that $(\hat{S}', \hat{\gamma}')$ are consistent and asymptotically normal. Note that $U(O; S, \gamma; q)$ is an unbiased estimating function, i.e., $E_{S, \gamma, q}\{U(O; S, \gamma; q)\} = 0$. This follows from the key identity,

$$\begin{aligned} &E_{\gamma, q} \left(I(C \geq \min(T, M)) \right. \\ &\quad \left. \div \prod_{s=t}^{\min(T, M)-1} \{1 - \lambda_C^\dagger(s | \bar{V}_t, T, T > s; \gamma, q)\} \right) \\ &\quad \left. C \geq t, T > t, \bar{V}_t, T \right) = 1 \end{aligned}$$

for all t . The asymptotic variance of $(\hat{S}', \hat{\gamma}')$ is given by the sandwich variance formula,

$$\begin{aligned} \Sigma(q) &= E \left\{ \frac{\partial U(O; S^*, \gamma^*; q)}{\partial (S', \gamma)'} \right\}^{-1} \\ &\quad \times E \{ U(O; S^*, \gamma^*; q) U(O; S^*, \gamma^*; q)' \} \\ &\quad \times E \left\{ \frac{\partial U(O; S^*, \gamma^*; q)}{\partial (S', \gamma)'} \right\}^{-1}, \end{aligned} \tag{10}$$

and it can be consistently estimated by replacing, in (10), S^* and γ^* by \hat{S} and $\hat{\gamma}$ and the expectations by their empiricals. We refer to this estimator as $\hat{\Sigma}$. Let $\Sigma_{\hat{S}}(q)$ be the associated asymptotic covariance matrix for $\hat{S}(q)$ and let $\hat{\Sigma}_{\hat{S}}(q)$ be the corresponding estimator.

Now we can estimate λ^* by $\hat{\lambda}$, where $\hat{\lambda}(t) = \{\hat{S}(t-1) - \hat{S}(t)\}/\hat{S}(t-1)$ and $S(0) \equiv 1$. By a simple application of the multivariate delta method, we can show that $\hat{\lambda}$ will be asymptotically normal with mean zero and variance $\Sigma_{\hat{\lambda}}(q) = G(S^*)\Sigma_{\hat{S}}(q)G(S^*)$, where $G(S)$ is an $M \times M$ matrix with the t th row zero except for $S(t)/S(t-1)^2$ and $-1/S(t-1)$ in the $(t-1)$ th and t th entries, respectively. We estimate $\Sigma_{\hat{\lambda}}(q)$ by $\hat{\Sigma}_{\hat{\lambda}}(q) = G(\hat{S})\hat{\Sigma}_{\hat{S}}(q)G(\hat{S})$.

The efficiency of the above estimators (i.e., when restriction (7) is not saturated) will depend on the choice of ϕ_t . The optimal choice of ϕ_t is the solution to a complicated integral equation. A reasonably efficient compromise choice of ϕ_t is to use the one that would yield the MLE of γ^* when sequential ignorability of censoring holds. Now, under sequential ignorability (i.e., $q_t = 0$ for all t), the maximum likelihood estimator (MLE) of γ^* , $\hat{\gamma}^{MLE}$, solves

$$\sum_{i=1}^n \sum_{t=0}^{M-1} I(C_i \geq t, T_i > t) \times \left[I(C_i = t) - \frac{\exp\{h_t(\bar{V}_{i,t}; \gamma)\}}{1 + \exp\{h_t(\bar{V}_{i,t}; \gamma)\}} \right] \times \frac{\partial h_t(\bar{V}_{i,t}; \gamma)}{\partial \gamma} = 0. \tag{11}$$

When $q_t = 0$ for all t , some algebra shows that the solution to equation (9) will equal $\hat{\gamma}^{MLE}$ when $\phi_t(\bar{V}_t)$ in (9) is equal to

$$\phi_t^*(\bar{V}_t) = - \sum_{s=0}^{t-1} \lambda_C^\dagger(s | \bar{V}_s, T, T > s; \hat{\gamma}^{MLE}, 0) \times \frac{\partial h_s(\bar{V}_s; \hat{\gamma}^{MLE})}{\partial \gamma} + \left\{ 1 - \lambda_C^\dagger(t | \bar{V}_t, T, T > t; \hat{\gamma}^{MLE}, 0) \right\} \times \frac{\partial h_t(\bar{V}_t; \hat{\gamma}^{MLE})}{\partial \gamma}.$$

This approach constitutes a reasonable compromise because Robins and Rotnitzky (1992) have shown that, under sequential ignorability, the resulting estimator for S^* is nearly efficient (provided the dimension of γ is reasonably large). Thus, local to sequential ignorability, setting $\phi_t(\bar{V}_t)$ equal to $\phi_t^*(\bar{V}_t)$ in (9) should yield a reasonably efficient estimator of S^* .

It is interesting to note that, in the special case in which no covariates \bar{V}_t are available, our estimators for S and λ when $q_t = 0$ for all t (i.e., noninformative censoring) reduce to the Kaplan–Meier and Nelson–Aalen estimators, respectively. Furthermore, we can show, provided that there is an observed improvement at all potential support points of the distribution of T , that, as α in (4) goes to $\pm\infty$ (regardless of τ), our estimator attains the empirical version of the bounds derived by Bedford and Meilijson (1997). At $\alpha = -\infty$, the lower bound corresponds to the assumption that improvement occurs immediately after censoring, and at $\alpha = +\infty$, the upper bound is equivalent to the assumption that censored subjects never improve. When time to event and censoring have discrete distributions, these bounds are sharp and improve upon those derived by Peterson (1976).

5.2 Two-Sample Testing

To test the null hypothesis of no treatment difference, we will use a discrete time analog of Gill’s (1980) test statistics based on the integrated weighted differences in hazard functions. Letting the superscript 1 denote the 2-mg risperidone group and 0 the placebo group, our test statistic is of the form

$$\hat{G}_K = (n^1 + n^0)^{-1/2} \sum_{t=1}^M \hat{K}(t) \{ \hat{\lambda}^1(t) - \hat{\lambda}^0(t) \} = (n^1 + n^0)^{-1/2} \hat{K}' \{ \hat{\lambda}^1 - \hat{\lambda}^0 \},$$

where $\hat{K}(t)$ is a random, time-specific weight function such that $\hat{K}(t)/(n^1 + n^0)$ converges uniformly in probability to a deterministic weight function $K(t)$ and $\hat{K} = (\hat{K}(1), \dots, \hat{K}(M))'$. By Slutsky’s theorem, it is easy to show that, under the null hypothesis of no treatment difference, \hat{G}_K will be asymptotically normal with mean zero and variance $K' \{ \Sigma_{\hat{\lambda}^1}/p^1 + \Sigma_{\hat{\lambda}^0}/p^0 \} K$, where $0 < p^1 < 1$ is the limit of $n^1/(n^1 + n^0)$, $p^0 = 1 - p^1$, and $K = (K(1), \dots, K(M))'$. The asymptotic variance of \hat{G}_K can be consistently estimated by $(n^1 + n^0)^{-1} \hat{K}' \{ \hat{\Sigma}_{\hat{\lambda}^1}/n^1 + \hat{\Sigma}_{\hat{\lambda}^0}/n^0 \} \hat{K}$. So we know that, under the null hypothesis,

$$\hat{L}_K = \frac{\hat{K}' \{ \hat{\lambda}^1 - \hat{\lambda}^0 \}}{(\hat{K}' (\hat{\Sigma}_{\hat{\lambda}^1}/n^1 + \hat{\Sigma}_{\hat{\lambda}^0}/n^0) \hat{K})^{1/2}} \tag{12}$$

will be asymptotically normal with mean zero and variance one. Thus, we reject the null hypothesis (at level 0.05) when $|\hat{Z}_k|$ is greater than 1.96. A negative (positive) value of \hat{L}_K indicates evidence in favor of treatment group 0 (1). For the JANSSEN data analysis, we chose

$$\hat{K}(t) = \frac{Y^1(t)Y^0(t)}{Y^1(t) + Y^0(t)}, \tag{13}$$

where $Y^j(t)$ is the number of subjects in treatment group j who are at risk for censoring or first clinical improvement at time j . We chose this weight function because, in the absence of auxiliaries, \hat{L}_K reduces to the classic log-rank statistic.

6. Simulation Study

To assess the practical performance of our estimator of the distribution of T , we conducted a simulation study. In the study, we let $M = 5$, $n = 100$, and assumed that the PANSS scores for an individual followed a linear, random effects model with a random intercept and fixed slope. Specifically, we assumed that $P_t = \beta_0 + \beta_1 t + \epsilon_t$, where β_0 was normally distributed with mean 100 and standard deviation 10, $\beta_1 = -5$, and $\epsilon_0, \dots, \epsilon_5$ were independently, normally distributed errors with mean 0 and standard deviation 9.5. Under this assumption, $S^*(1) = 0.8860$, $S^*(2) = 0.7286$, $S^*(3) = 0.5381$, $S^*(4) = 0.3453$, and $S^*(5) = 0.1860$. No baseline factors were simulated. For the censoring mechanism, we assumed that

$$h_t(\bar{V}_t; \gamma) = \begin{cases} \gamma_{00} + \gamma_0 V_{10} & t = 0 \\ \gamma_{0t} + \gamma_1 V_{1t} + \gamma_2 V_{2t} & t = 1, \dots, 4, \end{cases} \tag{14}$$

where V_{1t} is equal to zero if $P_t \leq 90$, one if $90 < P_t \leq 110$, and two otherwise; V_{2t} is equal to zero if $(P_t - P_0)/P_0 \leq -0.05$, one if $-0.05 < (P_t - P_0)/P_0 \leq 0.1$, and two otherwise; $\gamma_{00}^* = -5.0$, $\gamma_{01}^* = -4.5$, $\gamma_{02}^* = -3.5$, $\gamma_{03}^* = -2.5$, $\gamma_{04}^* = -1.5$, $\gamma_0^* = 0.25$, $\gamma_1^* = 0.0$, and $\gamma_2^* = 0.5$. In addition, we parameterized the

Table 2
Results of simulation study

α	τ		Model A		Model B		Model C	
			Bias	Coverage (%)	Bias	Coverage (%)	Bias	Coverage (%)
-0.5	6	S(1)	-0.0076	86.6	-0.0098	85.0	-0.0101	89.6
		S(2)	-0.0288	88.2	-0.0333	71.8	-0.0336	83.6
		S(3)	-0.0616	75.4	-0.0703	36.8	-0.0636	72.0
		S(4)	-0.0887	54.2	-0.1031	14.4	-0.0920	54.4
		S(5)	-0.0835	47.0	-0.0957	8.2	-0.0874	42.8
	8	S(1)	-0.0080	86.4	-0.0099	89.0	-0.0104	89.6
		S(2)	-0.0297	88.2	-0.0337	83.8	-0.0344	83.6
		S(3)	-0.0640	73.6	-0.0717	68.2	-0.0659	70.8
		S(4)	-0.0960	47.4	-0.1077	39.2	-0.0991	45.2
		S(5)	-0.1047	18.2	-0.1129	33.7	-0.1075	15.8
0.0	6, 8	S(1)	-0.0011	86.8	-0.0033	89.6	-0.0035	87.6
		S(2)	-0.0102	92.2	-0.0137	91.0	-0.0153	87.8
		S(3)	-0.0285	90.2	-0.0361	85.0	-0.0314	86.0
		S(4)	-0.0485	82.8	-0.0638	72.8	-0.0530	80.4
		S(5)	-0.0517	81.8	-0.0685	67.5	-0.0575	77.0
0.5	6	S(1)	0.0024	83.6	0.0001	84.6	0.0000	87.2
		S(2)	0.0023	90.6	0.0000	88.6	-0.0028	88.8
		S(3)	0.0009	92.2	-0.0033	89.8	-0.0026	90.6
		S(4)	-0.0002	91.0	-0.0119	92.2	-0.0064	89.0
		S(5)	-0.0031	89.6	-0.0193	90.6	-0.0116	91.2
	8	S(1)	0.0029	83.6	0.0005	84.6	0.0005	87.0
		S(2)	0.0048	90.6	0.0025	87.8	-0.0005	88.4
		S(3)	0.0103	90.0	0.0071	87.8	0.0061	89.4
		S(4)	0.0279	83.6	0.0191	84.1	0.0204	85.4
		S(5)	0.0627	57.2	0.0505	61.7	0.0512	65.4
1.0	6	S(1)	0.0034	83.6	0.0010	84.6	0.0009	86.9
		S(2)	0.0070	90.4	0.0050	88.2	0.0019	88.2
		S(3)	0.0159	88.4	0.0178	86.6	0.0123	87.8
		S(4)	0.0343	79.4	0.0272	81.4	0.0279	82.3
		S(5)	0.0459	69.8	0.0334	72.9	0.0352	75.9
	8	S(1)	0.0037	83.6	0.0011	85.1	0.0011	86.9
		S(2)	0.0082	90.2	0.0059	88.2	0.0030	88.1
		S(3)	0.0217	86.2	0.0193	86.1	0.0178	87.8
		S(4)	0.0575	65.4	0.0521	67.3	0.0508	65.7
		S(5)	0.1194	14.8	0.1124	19.1	0.1083	23.3
1.5	6	S(1)	0.0037	83.6	0.0011	84.9	0.0012	86.7
		S(2)	0.0083	90.2	0.0061	88.2	0.0032	88.5
		S(3)	0.0212	86.6	0.0193	86.1	0.0176	87.5
		S(4)	0.0516	68.7	0.0465	70.5	0.0455	69.8
		S(5)	0.0814	40.9	0.0721	49.0	0.0695	51.5
	8	S(1)	0.0038	83.6	0.0021	85.3	0.0013	86.7
		S(2)	0.0087	90.2	0.0064	88.1	0.0038	88.3
		S(3)	0.0234	86.4	0.0211	86.1	0.0198	87.1
		S(4)	0.0629	59.0	0.0580	62.0	0.0572	62.0
		S(5)	0.1325	9.4	0.1267	12.5	0.1231	14.9

censoring bias function via (4) and assumed that the data were generated under $\alpha = 0.5$ and $\tau = 6$. With these additional assumptions, 26% of subjects drop out prior to observation of T and end of study. Our simulation study was based on 500 replications.

Since the true form of h_t and true values of α and τ are unknown, we analyzed the simulated data using various α ($\alpha = -0.5, 0.0, 0.5, 1.0, 1.5$), τ ($\tau = 6, 8$), and three forms for

h_t , two correctly specified and one misspecified. The first form (model A) was similar to (14) except that γ_1 was excluded, the second form (model B) was similar to (14) except that γ_2 was excluded, and the third form (model C) was the same as (14). Models A and C are correctly specified because, in the true data generating process, $\gamma_1^* = 0$. Model B is misspecified since $\gamma_2^* \neq 0$. Models A and B are submodels of model C. For each model, we report the bias for estimation of S^* as well as

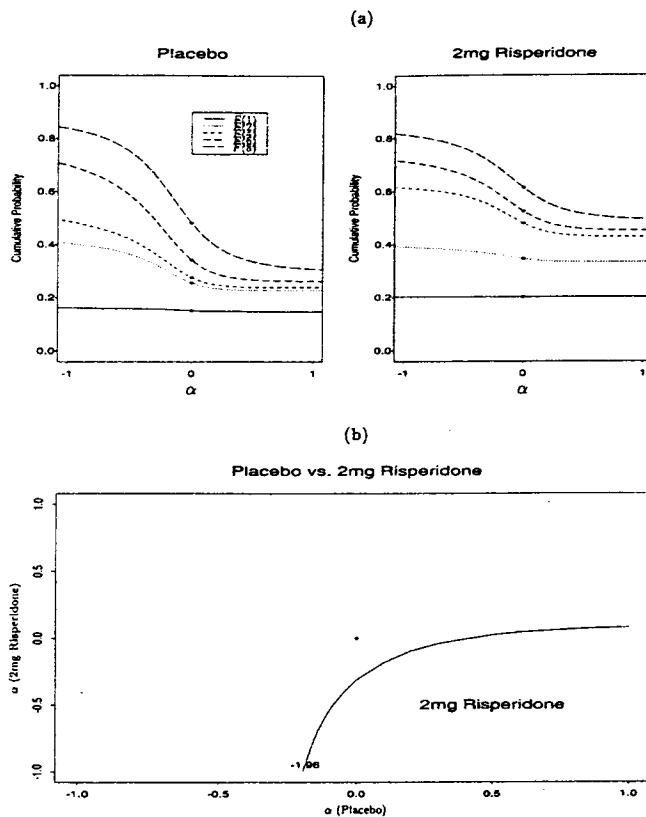


Figure 2. a. Treatment-specific cumulative distribution function estimates as a function of α , for $t = 1, 2, 4, 6, 8$, under model (5, 6). b. Contour plot of Z_K statistic as a function of treatment-specific censoring bias parameters (α) under model (5, 6).

the coverage rate of a 90% pointwise confidence interval for S^* . Each confidence interval was formed by exponentiating a 90% symmetric Wald-type confidence interval for the $\log(S^*(t))$, which was derived by using the delta method. Specifically, our confidence interval for $S^*(t)$ is of the form

$$\left[\exp \left\{ \log \{ \hat{S}(t) \} - 1.645 \sqrt{e_t' \hat{\Sigma}_{\hat{S}}(q) e_t} / \hat{S}(t) \right\}, \right. \\ \left. \exp \left\{ \log \{ \hat{S}(t) \} + 1.645 \sqrt{e_t' \hat{\Sigma}_{\hat{S}}(q) e_t} / \hat{S}(t) \right\} \right],$$

where e_t is an $M \times 1$ vector with one as the t th entry and zeros otherwise.

Table 2 reports the results of the simulation. The salient features of this simulation are that, when α , τ , and h_t are correctly specified, our estimator performs well in terms of bias and coverage. Coverages worsen as α and τ deviate from the truth. This is overwhelmingly due to the worsening of bias. The results also worsen as a function of t . For given α and τ , the results for the misspecified model for h_t did not differ appreciably from those of the other models. This latter result is not guaranteed by our methodology but holds in this simulation since V_{1t} and V_{2t} are highly correlated. In general, misspecification of h_t can induce bias and poor coverage.

7. Analysis of JANSSEN Trial

In the JANSSEN trial, the five visits occurred at weeks 1, 2, 4, 6, and 8, so the support points of T are 1, 2, 4, 6, 8, and τ . We conducted two sensitivity analyses. In the first, we ignored the auxiliaries. In the second, we chose $V_t = P_t$. In both analyses, we parameterized the censoring bias functions using (4) and varied α and τ . Our results were relatively insensitive to the choice of τ , so we present the results for $\tau = 10$. We ranged α from -1 to 1 in increments of 0.1 . Note that our *a priori* belief is that α is positive. For illustration, however, we will carry out a sensitivity analysis over the entire range. Both the data analysis and the simulation were conducted in MATLAB on a SPARC Ultra 10 workstation with 512 megabytes of RAM. The CPU time for the sensitivity analysis with $V_t = P_t$ was 89.61 and 78.21 seconds for the placebo and 2-mg risperidone arms, respectively. The CPU time for the sensitivity analysis without auxiliaries was appreciably shorter.

For the sensitivity analysis without auxiliaries, we used the nonparametric model (5, 6). In Figure 2a, we present the treatment-specific cumulative distribution function (CDF) estimates, $\hat{F}(t)$, as a function of α for $t = 1, 2, 4, 6, 8$. At $\alpha = 0$, $\hat{F}(t)$ is the Kaplan-Meier estimate presented in Figure 1. As α goes to $\pm\infty$, Bedford and Meilijson's (1997) bounds would be obtained since there is at least one event (improvement) at each of the potential event times. Treatment groups are compared in Figure 2b. On the x -axis, we have varying levels of censoring bias for the placebo arm, and on the y -axis, we have varying levels of censoring bias for the 2-mg risperidone arm. For each combination of censoring biases, we perform a test (at the 0.05 level) of the null hypothesis of no treatment difference. The graph is a contour plot of the L_K -statistic (12) with weight function (13) as a function of the two levels of censoring biases. The line(s) in each plot represent the combinations that lead to a L_K -statistic of ± 1.96 . The areas of the plots are labeled to indicate the favored treatment. Unlabeled areas indicate that there is not enough evidence to draw a conclusion. The point (0, 0) represents the noninformative censoring comparison. Our *a priori* belief is that the true combination of censoring biases should lie in the upper right quadrant of the contour plot. Here we see that only differential and implausible censoring biases would result in a statistically significant treatment effect.

For the second sensitivity analysis, we fit the semiparametric model (2, 3) with additional restriction (14), where $V_{1t} = j - 1$ and $V_{2t} = j - 1$ if P_t and $(P_t - P_0)/P_0$ are the j th tertile of P_t and $(P_t - P_0)/P_0$, respectively. The tertiles were formed by considering all subjects in the trial who were at risk for censoring at time t . At $\alpha = 0$, this model says that the log odds of censoring at time t is linearly related to history of PANSS scores through the tertile of the current PANSS score and the tertile of the percentage change between the current and baseline PANSS score. Small sample sizes precluded use of richer models. In Figure 3a, we present the treatment-specific cumulative distribution function estimates as a function of α for $t = 1, 2, 4, 6, 8$. For reference, the solid points at $\alpha = 0$ are the Kaplan-Meier estimator. For each treatment group, we see that, under sequential ignorability ($\alpha = 0$) with high-dimensional prognostic factors, the CDF estimator differs from Kaplan-Meier. In Figure 4, we present a blow up of this comparison. As expected, the larger

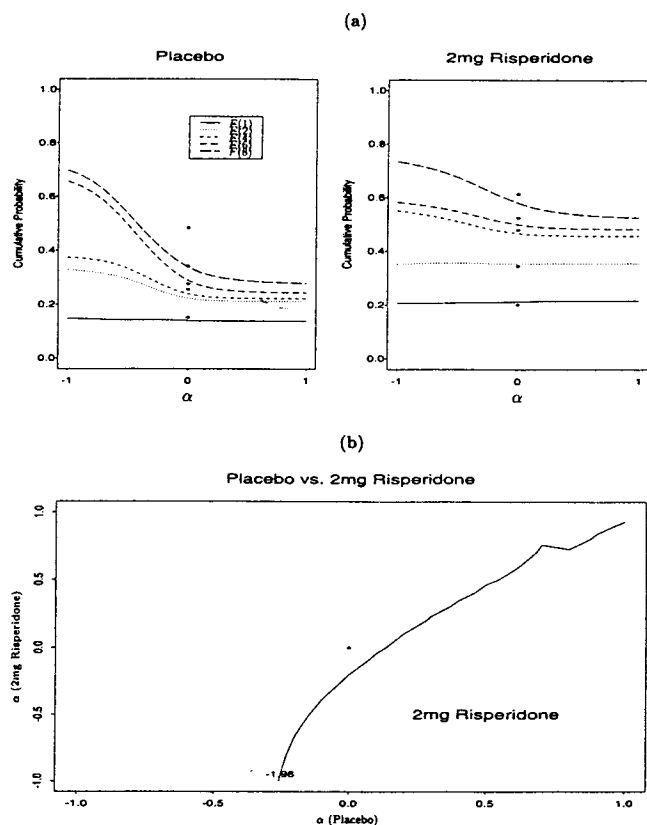


Figure 3. a. Treatment-specific cumulative distribution function estimates as a function of α , for $t = 1, 2, 4, 6, 8$, under model (2, 3) and (14). b. Contour plot of Z_K statistic as a function of treatment-specific censoring bias parameters (α) under model (2, 3) and (14).

correction occurs in the placebo group. Under the assumption of sequential ignorability in both treatment groups, $Z_k = -1.80$ (p -value = 0.07). Under this assumption, the evidence in favor of 2 mg risperidone is stronger than under noninformative censoring. In Figure 3b, we present the analog of the contour plot in Figure 2b. Again, we expect the true combination of censoring bias to lie in the upper right quadrant. Here, small, quite plausible degrees of differential selection bias result in rejection of the null hypothesis of no treatment difference at the 0.05 level.

8. Discussion

In this article, we presented a method for estimation and comparison of treatment arm-specific marginal survival curves of a time-to-event variable from right-censored data. Our method allows adjustment for informative censoring due to measured prognostic factors for time to event and censoring while simultaneously quantifying the sensitivity of the inference to non-identifiable assumptions concerning the residual dependence between time to event and censoring due to unmeasured factors. We focused on studies with discrete time-to-event outcomes. For future work, we plan to extend our results to the continuous time-to-event setting.

Investigators may feel dissatisfied with our sensitivity analysis approach because (1) a single inference (e.g., based on

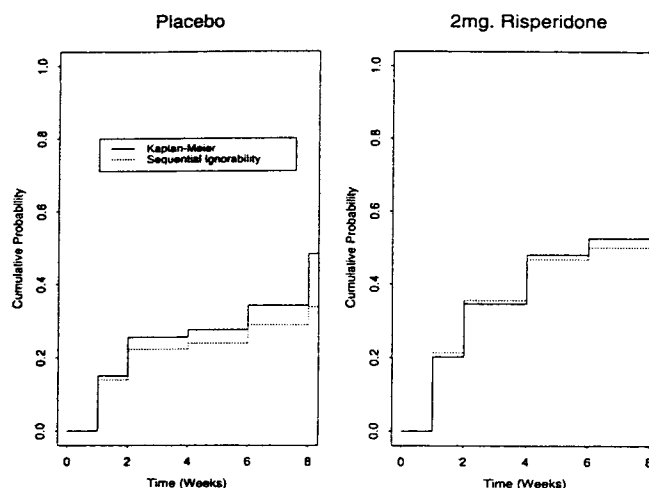


Figure 4. Treatment-specific estimates of the cumulative distribution under noninformative censoring versus sequential ignorability.

noninformative censoring or sequential ignorability of censoring) is not drawn and (2) there are no limits to the assessment of sensitivity (i.e., one can consider an infinite number of forms for q_t and h_t). One possible solution is to successfully follow up a random sample of subjects who dropped out. In this case, q_t is nonparametrically identified. An alternative approach would be to forgo the sensitivity analysis and just report Bedford and Meilijson's (1997) bounds. However, in situations where there is a substantial amount of censoring, they will be too wide to be of substantive use. When additional data are not available, our approach represents a compromise between presenting a single inference and bounds since it quantifies inference along a user-defined smooth path between the two. (For a discussion on choosing a plausible range for the censoring bias parameter/function, see Section 7.2.3 of Scharfstein, Rotnitzky, and Robins (1999).) Finally, for the situation where a decision must be made, we are in the process of developing a semiparametric Bayesian approach that allows experts to specify their prior beliefs about the censoring bias function q_t .

ACKNOWLEDGEMENTS

This research was partially supported by National Institute of Health grants 1-R29-GM48704-04, 5R01A132475, R01-CA74112, 1-R01-MH56639-01A1, and 1-R01-DA10184-01A2. This work was partly conducted while Professor Andrea Rotnitzky was visiting the University Di Tella, Buenos Aires, Argentina.

RÉSUMÉ

Dans cet article, nous présentons une méthode pour estimer et comparer, selon le traitement, les distributions d'une variable discrète associée au délai d'apparition d'un événement pour des données censurées à droite. Notre méthode permet d'une part d'ajuster pour de la censure informative due à des facteurs pronostiques mesurés du délai d'apparition de

l'événement et du temps de censure. Elle permet d'autre part de quantifier la sensibilité de l'inférence à la dépendance résiduelle entre le délai de l'apparition de l'événement et la censure due à des facteurs non mesurés. Nous développons notre approche dans le contexte d'un essai randomisé pour le traitement de la schizophrénie chronique. Nous réalisons une étude de simulation pour évaluer la performance pratique de notre méthode.

REFERENCES

- Bedford, T. and Meilijson, I. (1997). A characterization of marginal distributions of (possibly dependent) lifetime variables which right censor each other. *The Annals of Statistics* **25**, 1622–1645.
- Chouinard, G., Jones, B., Remington, G., Bloom, D., Addington, D., MacEwan, G. W., Labelle, A., Beauclair, L., and Arnott, W. (1993). A Canadian multicenter placebo-controlled study of fixed doses of risperidone and haloperidol in the treatment of chronic schizophrenia patients. *Journal of Clinical Psychopharmacology* **13**, 25–40.
- Crowder, M. (1996). On assessing independence of competing risks when failure times are discrete. *Lifetime Data Analysis* **2**, 195–209.
- Crowder, M. (1997). A test for independence of competing risks with discrete failure times. *Lifetime Data Analysis* **3**, 215–223.
- Fisher, L. and Kanarek, P. (1974). Presenting censored survival data when censoring and survival times may not be independent. In *Reliability and Biometry: Statistical Analysis of Lifelength*, F. Proschan and R. Serfling (eds), 303–326. Philadelphia: SIAM.
- Fleming, T. R. and Harrington, D. P. (1991). *Counting Processes and Survival Analysis*. New York: Wiley.
- Gill, R. D. (1980). *Censoring and Stochastic Integrals*. Mathematical Centre Tracts 124. Amsterdam: Mathematisch Centrum.
- Gill, R. D., van der Laan, M. J., and Robins, J. M. (1997). Coarsening at random: Characterizations, conjectures and counterexamples. In *Proceedings of the First Seattle Symposium on Survival Analysis*, D. Y. Lin and T. R. Fleming (eds), 255–294. New York: Springer.
- Heitjan, D. F. and Rubin, D. B. (1991). Ignorability and coarse data. *The Annals of Statistics* **19**, 2244–2253.
- Jacobsen, M. and Keiding, N. (1995). Coarsening at random in general sample spaces and random censoring in continuous time. *Annals of Statistics* **23**, 774–786.
- Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* **53**, 457–481.
- Klein, J. P. and Moeschberger, M. L. (1988). Bounds on net survival probabilities for dependent competing risks. *Biometrics* **44**, 529–538.
- Klein, J. P., Moeschberger, M. L., Li, Y. H., and Wang, S. T. (1992). Estimating random effects in the Framingham Heart Study (Disc: P118–120). In *Survival Analysis: State of the Art*, J. P. Klein and P. Goel (eds), 99–118. Norwell, Massachusetts: Kluwer Academic.
- Marder, S. R. and Meibach, R. C. (1994). Risperidone in the treatment of chronic schizophrenia. *American Journal of Psychiatry* **151**, 825–835.
- Peterson, A. V., Jr. (1976). Bounds for a joint distribution function with fixed sub-distribution functions: Application to competing risks. *Proceedings of the National Academy of Sciences* **73**, 11–13.
- Robins, J. M. (1987). A new approach to causal inference in mortality studies with sustained exposure periods—Application to control of the healthy worker survivor effect (addendum). *Computers and Mathematics with Applications* **14**, 917–921.
- Robins, J. M. (1993). Information recovery and bias adjustment in proportional hazards regression analysis of randomized trials using surrogate markers. In *ASA Proceedings of the Biopharmaceutical Section*, 24–33. Alexandria, Virginia: American Statistical Association.
- Robins, J. M. and Finkelstein, D. H. (2000). Correcting for noncompliance and dependent censoring in an AIDS clinical trial with inverse probability of censoring weighted (IPCW) log-rank tests. *Biometrics* **56**, 779–788.
- 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 and K. Dietz (eds), 297–331. Boston: Birkhäuser.
- Rotnitzky, A., Robins, J. M., and Scharfstein, D. O. (1998). Semiparametric regression for repeated outcomes with nonignorable nonresponse. *Journal of the American Statistical Association* **93**, 1321–1339.
- Satten, G. A., Datta, S., and Robins, J. M. (2000). An estimator for the survival function when data are subject to dependent censoring. *Lifetime Data Analysis* **56**, 779–788.
- Scharfstein, D. O., Rotnitzky, A., and Robins, J. M. (1999). Adjusting for nonignorable drop-out using semiparametric nonresponse models (with discussion). *Journal of the American Statistical Association* **94**, 1096–1146.
- Slud, E. V. and Rubinstein, L. V. (1983). Dependent competing risks and summary survival curves. *Biometrika* **70**, 643–649.
- Zheng, M. and Klein, J. P. (1994). A self-consistent estimator of marginal survival functions based on dependent competing risk data and an assumed copula. *Communications in Statistics, Part A—Theory and Methods* **23**, 2299–2311.
- Zheng, M. and Klein, J. P. (1995). Estimates of marginal survival for dependent competing risks based on an assumed copula. *Biometrika* **82**, 127–138.

Received March 2000. Revised December 2000.

Accepted December 2000.

APPENDIX

LEMMA 1: $f(T | X > t, \bar{V}_t)$ in (2) is identifiable from F_O .

Proof. We proceed via a reverse-time inductive proof. First, $f(T | X > M-1, \bar{V}_{M-1})$ is identified from F_O . Now, suppose that $f(T | X > t, \bar{V}_t)$ is identified. We need to show that $f(T | X > t-1, \bar{V}_{t-1})$ is identified. Note that

$$f(T | X > t-1, \bar{V}_{t-1})$$

$$\begin{aligned}
 &= \sum_{\delta=0}^1 f(T | X = t, \Delta = \delta, \bar{V}_{t-1}) \\
 &\quad \times P(X = t, \Delta = \delta | X > t-1, \bar{V}_{t-1}) \\
 &+ \left\{ \int f(T | X > t, \bar{V}_t) f(V_t | X > t, \bar{V}_{t-1}) d\mu(V_t) \right\} \\
 &\quad \times P(X > t | X > t-1, \bar{V}_{t-1}).
 \end{aligned}$$

Now, all components on the right side of the above equation are identified. It is clear that $P(X = t, \Delta = \delta | X > t-1, \bar{V}_{t-1})$, $\delta = 0, 1$, $f(V_t | X > t, \bar{V}_{t-1})$, and $P(X > t | X > t-1, \bar{V}_{t-1})$ are identifiable directly from $F_{\mathcal{O}}$. Further, $f(T | X = t, \Delta = 1, \bar{V}_{t-1})$ is trivially identified since $P(T = t | X = t, \Delta = 1, \bar{V}_{t-1} = \bar{v}_{t-1}) = 1$ for all \bar{v}_{t-1} , $f(T | X > t, \bar{V}_t)$ is identified by assumption, and $f(T | X = t, \Delta = 0, \bar{V}_{t-1})$ is identified via restriction (2).