

The Graft Versus Leukemia Effect after Bone Marrow Transplantation: A Case Study Using Structural Nested Failure Time Models

Niels Keiding,^{1,*} Marusca Filiberti,¹ Sille Esbjerg,¹
James M. Robins,² and Niels Jacobsen³

¹Department of Biostatistics, University of Copenhagen

²Departments of Epidemiology and Biostatistics, Harvard School of Public Health

³Department of Haematology, Rigshospitalet, University of Copenhagen

* *email:* N.Keiding@biostat.ku.dk

SUMMARY. Over the last decade, J. M. Robins has developed a set of tools for assessing, from observational data, the causal effects of a time-dependent treatment or exposure in the presence of time-dependent covariates that may be simultaneously confounders and intermediate variables. This report concerns a case study of the application of one these techniques, G-estimation using structural nested failure time models, to the problem of assessing the effect of graft versus host disease on leukemia relapse after bone marrow transplantation.

KEY WORDS: Accelerated failure time; Causal inference; Competing risks; Cox regression model; Counterfactuals; Survival synthesis; Time-dependent confounder.

1. Introduction

Patients receiving bone marrow transplant (BMT) as a treatment for leukemia frequently develop graft versus host disease (GvHD) wherein the transplanted (grafted) immune cells attack the host tissues. As a consequence, on biological grounds, one might expect that the development of GvHD would increase the risk of patients dying in remission while possibly decreasing the risk of leukemic relapse (because the graft's immune cells kill both normal and leukemic host cells). These problems have constituted a challenge to biostatisticians for the last two decades and, at the same time, have inspired methodological work in survival analysis and related areas.

The goal of this paper is to estimate the overall effect of the exposure variable graft versus host disease (GvHD) on time to leukemic relapse with death in remission treated as a competing risk. (Note: We are thus not interested here in whether BMT as such is useful against leukemia.) To accomplish this goal, we must overcome two different methodological challenges. The first challenge is the difficult problem of how to appropriately adjust for the competing risk, death in remission, when the primary endpoint is leukemic relapse. The second challenge is a consequence of the fact that the time-dependent covariate, infection with cytomegalovirus (CMV), is an independent prognostic factor for relapse that both (A1) predicts the subsequent development of the exposure GvHD and (A2) is predicted by past exposure history.

These relationships are represented in Figure 1. In standard epidemiologic terminology, a prognostic factor, such as CMV, that satisfies (A1) is referred to as a confounding fac-

tor or confounder, and the standard procedure in analytical epidemiology is to adjust (usually by stratification or regression) for confounders in order to avoid spurious associations between exposure and outcome generated only by the association of both of these to the confounder. On the other hand, a prognostic factor that satisfies (A2) is referred to as an intermediate variable on the causal pathway from the exposure GvHD to the outcome relapse when the arrows from GvHD to CMV and from CMV to relapse in Figure 1 represent causal relationships. It is important not to adjust for intermediate variables in the analysis.

We term a time-dependent independent prognostic factor satisfying (A1) and (A2) a time-dependent confounder and note that these can only exist when the exposure variable (here GvHD) is time dependent.

The fact that the CMV variable may be both a confounder and an intermediate variable thus presents the following analytic challenge: If we fail to adjust for CMV in the analysis, then, even under the null hypothesis of no effect of GvHD on relapse, we will find an artifactual noncausal association since GvHD and relapse are correlated with the confounder CMV. On the other hand, if we do adjust for CMV, say, by stratification, we will dilute a true causal effect of GvHD on relapse because we will have removed, by conditioning, the part of the effect that is mediated by causal pathways involving CMV. The G-estimation procedure described in this paper has been specifically designed to control for confounding by intermediate variables.

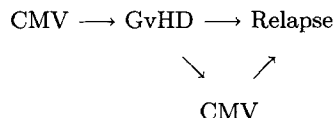


Figure 1. Basic structure of graft versus leukemia effect possibly confounded by posttransplant cytomegalovirus infection. CMV may be a confounder for the causal effect of GvHD on relapse since there are transition arrows from CMV to both GvHD and relapse. Also, transitions to CMV may be predicted by past GvHD history which would imply that the size of the arrow representing the GvHD-to-relapse transition may not have a direct causal interpretation.

The important conceptual analysis of the biostatistical methodology of competing risks by Prentice et al. (1978) and Kalbfleisch and Prentice (1980, Chapter 7) was, to a large extent, motivated by the obviously unacceptable assumption of independent competing risks of relapse and death in remission after BMT. The classical calculation of cause-specific failure in competing risks analysis uses a Kaplan–Meier estimator counting all other deaths as censorings. The Kaplan–Meier estimator is consistent for the time to relapse distribution that would be observed with death from other causes eliminated under the assumption that time to relapse is independently censored by death from competing risks. Pepe and Mori (1993) contrasted this calculation with the two other possible definitions of probability of cause-specific failure: the marginal probability, i.e., cause-specific failure ‘in this world’ where other causes may be at work, and the conditional probability, calculated in this world conditionally on the other causes not actually occurring. In this paper, following Robins and Rotnitzky (1992) and Robins et al. (1992), we attempt to calculate the time to relapse distribution that would be observed were death from other causes eliminated. However, we do not assume independent censoring of time to relapse by death from competing risks. Rather, we assume that (a) time to competing risk and time to relapse are dependent given baseline covariates but (b) the cause-specific intensity (hazard) of the competing risk at t given baseline covariates and the history of the measured time-dependent covariates up to t does not further depend on (the possibly unobserved) time to relapse.

More elaborate multistate models may also be formulated along the lines of statistical models based on counting processes (see Andersen et al., 1993, Section IV.4, for a general survey). This allows combination of transition intensities to transition probabilities (survival synthesis) (see Klein, Keiding, and Copelan, 1993; Dabrowska, Sun, and Horowitz, 1994). Arjas and Eerola (1993) embedded these techniques in a framework of dynamic probabilistic causality as documented by Eerola (1994) on the same data as used here. Estimation of effects using the method of dynamic probabilistic causality is equivalent to effect estimation by the G-computation algorithm of Robins (1986, 1987), who noted that the estimation of causal effects from sparse multivariate data with the G-computation algorithm suffered from inherent robustness difficulties because, in the presence of time-dependent covariates that are simultaneously confounders and intermediate variables, the parameters of models of the transition

intensities lack a causal interpretation; only the counterfactual synthesized (i.e., integrated) transition probabilities can be causally interpreted.

To overcome these robustness problems, Robins (1989, 1992, 1993) proposed a new class of counterfactual causal models, the structural nested failure time models (SNFT) whose parameters have a direct causal interpretation. In this paper, we use SNFT models to attempt to estimate the effect of GvHD on relapse, considering death in remission as (potentially informative) censoring and CMV infection as both a potential time-dependent confounder and a potential intermediate variable.

2. Data and Descriptive Analyses

Our case study is a reanalysis of a series of 163 consecutive acute myeloid leukemia (AML) and acute lymphoid leukemia (ALL) patients previously published by Jacobsen et al. (1987, 1990) (see also Andersen et al., 1993, Examples I.3.14 and VII.2.18; Eerola, 1994, Section 4.3.1). Within the follow-up period of 2–7 years after BMT, 36 patients relapsed and 45 died in remission (i.e., before relapse of the leukemia). In a first phase, we considered the situation in Figure 1, where arrows represent transitions between states and transitions to death and end-to-follow-up have been temporarily suppressed.

Our interest will be to assess the overall (net) effect of the (time-dependent) exposure GvHD on relapse, taking into account the influence of the time-dependent covariate CMV, which may serve both as confounder and as intermediate variable. (The biological problem is somewhat simplified compared to current BMT literature, where it is customary to distinguish between acute and chronic GvHD.)

There are two complications due to censoring. First, patients may die in remission and, second, patients may be still alive and in remission (leukemia-free) at end of follow-up. We consider these two censoring complications in detail later.

The precise definitions of the variables chosen here are

$$\text{GvHD}(t) = I\{\text{acute or chronic graft versus host disease has occurred by time } t\}$$

and

$$\text{CMV}(t) = I\{\text{CMV infection has occurred by time } t\}.$$

As a first step, Cox regression models with time-dependent covariates were fitted to the four transitions indicated by Figure 1. In addition to the time-dependent covariates indicated by the figure, fixed-time covariates suggested by the earlier analyses cited above were assessed.

The first Cox regression analysis shows that, after accounting for donor age, CMV immunity at transplantation, and the relapse/remission stage of the patient at transplantation, GvHD is associated with a significant ($P < 0.01$) reduction in the risk of relapse (estimated relative risk = 0.33). In this analysis, censoring by death and end of follow-up is treated in standard fashion, i.e., subjects who die or reach end-of-follow-up by time t do not contribute to the estimation of the hazard at times subsequent to t .

The basic difficulty of this analysis, which may at first sight seem perfectly standard and reliable, is that the apparent treatment effect of $\text{GvHD}(t)$, as expressed by the relative risk estimate 0.33, may be confounded by time-dependent confounders such as CMV history. As discussed in Section

1, CMV history is a time-dependent confounder, conditional on baseline covariates, if both (i) the transition intensity to GvHD at time t depends on CMV history before t and (ii) the transition intensity to relapse at t depends on past CMV history conditional on past GvHD history. Furthermore, the regression coefficient for GvHD(t) in an intensity model for relapse that adjusts for both baseline covariates and past CMV history will still fail to have a causal interpretation as the overall (net) effect of GvHD on relapse if, conditional on baseline covariates, the transition intensity to CMV at time t depends on past GvHD history, as would be the case if CMV were an intermediate variable on the causal pathway from GvHD to relapse.

To test condition (i), we next fit a Cox model for the onset of GvHD. After correction for patient CMV immunity at transplantation, mismatch score, and relapse/remission at transplantation, this analysis shows that CMV is associated with at least marginally significantly ($P = 0.07$) increased risk of later development of GvHD (estimated relative risk = 1.9), confirming condition (i).

To test condition (ii), we added to the previous Cox model for the onset of relapse the additional time-dependent covariate CMV(t). The resulting analysis indicates that CMV(t) is associated with an at least marginally significant ($P = 0.06$) reduction in the relapse intensity (estimated relative risk = 0.47), confirming condition (ii). Although in this analysis GvHD(t) is associated with a significant ($P = 0.02$) reduction in relapse intensity (estimated relative risk = 0.37), nonetheless, as discussed above, this relative risk may fail to have a causal interpretation if GvHD affects the risk of transitions to CMV.

The fit of the fourth Cox model, for transition to CMV, shows that, in addition to donor age and donor CMV immunity at transplantation, GvHD is indeed associated with significantly ($P < 0.01$) increased risk of later development of CMV (estimated relative risk = 1.9).

3. Hypothesis of No Unmeasured Confounders; Structural Nested Failure Time Models

Disregarding temporarily the complications from censoring due to death in remission and to end of follow-up, we now introduce the key concepts in Robins's approach (cf., Robins et al., 1992; Robins, 1998; and several other expositions by Robins and colleagues).

A generally acknowledged basis for drawing inference from clinical trials is for the investigator to randomly assign treatment to the study subjects. This insures both that there is no systematic confounding by unmeasured factors and that there is a well-defined probability measure describing the noise against which to compare a possible signal. It is well known that the power of the trial may be enhanced by performing stratified randomization, and it is also perfectly permissible to let the randomization deviate from 1:1 depending on the strata (or time-fixed covariates) available before randomization, as long as proper account is taken of these strata and covariates in the analysis.

We are here interested in problems where some of the covariates influencing treatment decisions are time-dependent: Only after the registration of some intermediate response does it become relevant to institute a particular treatment. Robins's idea here is to point out that valid inference may still

be drawn using the randomization paradigm. Randomization is performed along the way, depending on explicitly observed time-fixed and time-dependent covariates. The principles are still that the investigator randomly assigns the treatment (so that no unknown factors influence treatment decision) and that randomization allows 'like to be compared to like,' provided proper account of the covariates is taken in the analysis.

In our example we do not have a treatment to be manipulated by the investigator. Our treatment is GvHD and 'nature' decides when to switch this on. The idea is now to mimic our observational study of nature's treatment experiment by a (sequential) clinical trial of the kind just outlined. Indeed, we assume that, based on values of covariates known to us, nature decides for each patient, at each time t , essentially by randomization, whether to initiate 'treatment' (GvHD). That the treatment initiation is conditionally independent of patient characteristics, given the values of the measured covariates, is Robins's fundamental assumption of no unmeasured confounders, which, in the present version of Robins's theory, is formalized as follows.

As mentioned at the beginning of this section, assume for the moment that no censoring occurs either by death or end to follow-up so that the subject's time to relapse T is always observed. Consider the times T_{G+} and T_{G-} of reaching the endpoint (relapse) given that the treatment (GvHD) is always switched on (+) or always switched off (-). T_{G+} and T_{G-} are referred to as counterfactual variables since, e.g., T_{G-} is the time to relapse if, possibly contrary to fact, GvHD is eliminated. Note that T_{G+} and T_{G-} can never both be observed. Indeed, neither will be observed with the exception of subjects who either develop GvHD at time zero (so that $T = T_{G+}$) or never develop GvHD (so that $T = T_{G-}$). The assumption of no unmeasured confounders in this situation states that, at each time t , the onset of treatment at t should be conditionally independent of T_{G+} and T_{G-} given the covariate history at t . If $\lambda_G(t | B)$ is the intensity of onset of treatment at t given B , this means that we require

$$\lambda_G(t | \mathcal{H}(t)) = \lambda_G(t | \mathcal{H}(t), T_{G-}) = \lambda_G(t | \mathcal{H}(t), T_{G+}), \quad (1)$$

where $\mathcal{H}(t)$ denotes the covariate history at $t-$. Robins further defined the sharp null hypothesis $T = T_{G-} = T_{G+}$ of no treatment effect, with T the actually realized time to the endpoint, so that a concrete test may be performed by testing whether $\lambda_G(t | \mathcal{H}(t)) = \lambda_G(t | \mathcal{H}(t), T)$.

To further estimate the treatment effect, a structural nested failure time model is postulated. Let G be the time of onset of treatment and $\Gamma(t) \equiv I\{G < t\}$. The effect of treatment is described by the one-dimensional parameter ψ_0 , which accelerates the counterfactual time T_{G-} to the endpoint under no treatment as follows, where T is the realized time to the endpoint: $T_{G-} = T(\psi_0)$, where we define

$$T(\psi) = \int_0^T e^{\psi\Gamma(t)} dt = \begin{cases} T & \text{if } G \geq T \\ G + (T - G)e^{\psi} & \text{if } G < T. \end{cases}$$

It is seen that, for $\psi_0 > 0$, $T_{G-} > T$, i.e., the treatment is harmful; for $\psi_0 < 0$, $T_{G-} < T$, i.e., the treatment is beneficial; and for $\psi_0 = 0$, $T_{G-} = T$, i.e., the treatment does neither harm nor good. John P. Klein has pointed out that this model was discussed by Fisher and Kanarek (1974).

Obviously, if $T < G$, then the endpoint has occurred under no treatment, so by definition $T(\psi) = T$. This simple structural nested failure time model implies that T_{G-} is a deterministic function of T and G . Robins (1989) referred to such a model as a rank-preserving structural failure time model (RPSFTM) and pointed out that an RPSFTM is equivalent to the strong version of the time-dependent accelerated failure time model described by Cox and Oakes (1984). Robins et al. (1992) and Robins (1998) described how the assumption of a deterministic relationship of T_{G-} with T and G can be relaxed.

Since, in the absence of censoring, T (always) and G (when $T > G$) are observable, then $T(\psi)$ is observable for each given value of ψ , which means that one may test the hypothesis that a test value ψ equals the true but unknown ψ_0 by testing whether

$$\lambda_G(t | \mathcal{H}_t) = \lambda_G \{t | \mathcal{H}_t, T(\psi)\} \tag{2}$$

for each value of ψ . Those ψ compatible with the hypothesis at some significance level α form a $(1 - \alpha)$ confidence interval for ψ_0 , and a point estimate may be obtained as that value of ψ that yields the least significant value of the relevant test statistic. Robins (1998) described how the consequences of violations of our assumption of no unmeasured confounders could be explored through a sensitivity analysis.

In our application, the treatment is GvHD and the endpoint is relapse. We shall fit a Cox regression model to the endpoint GvHD using relevant covariates, and our estimates and test statistics will be based on the usual score equations. But before we can proceed to the actual inference, we need to handle the two censoring complications.

4. Allowing for Censoring Due to End of Follow-Up: The Potential Censoring Time

Still disregarding censoring due to the competing risk death in remission, we now turn to censoring due to end of follow-up, for the moment pretending that all events before end of follow-up are observable. We further assume (as is justified for our data, where everybody was censored at a fixed calendar date) that the potential censoring time C from study start is well defined and known also for those patients who relapsed (or died) before then. Specifically, it is the difference between the fixed (nonrandom) end-of-follow-up date and the calendar date of study entry. The objective of what follows is to replace, as in standard survival analysis, the sometimes unobservable time T of relapse by the observable $T \wedge C$ and to elaborate the time accelerations correspondingly. Define, for $\psi \geq 0$, $C_t(\psi) = C$ for $t \leq G$ while, for $\psi \leq 0$, $C_t(\psi) = t + (C - t)e^\psi$ for $t \leq G$.

The motivation for the asymmetric definition is that the time accelerations must not allow violation of the inherent predictability, and therefore some additional censoring may be necessary. Note we need not worry about defining $C_t(\psi)$ for subjects with $T < G$ since such subjects do not contribute to the risk set at t in our Cox model for G .

Since C is known at the time $t = 0$ and thus is in $\mathcal{H}(t)$ for all t , it follows that, if $\psi = \psi_0$, equation (2) is true with $T(\psi)$ replaced by any fixed function of $T(\psi)$ and C . The obvious choice would be $X_{1t}(\psi) = \min(T(\psi), C_t(\psi))$, but we shall report separate results for three additional functions which also, contrary to $T(\psi)$, depend only on what happened before end

of follow-up since, if $t \leq G$, (i) $C_t(\psi)$ is observed and (ii) $T > C$ implies $T(\psi) \geq C_t(\psi)$: $X_{2t}(\psi) = I\{T(\psi) < C_t(\psi)\}X_{1t}(\psi)$, $X_{3t}(\psi) = I\{T(\psi) > C_t(\psi)\}X_{1t}(\psi)$, and $X_{4t}(\psi) = I\{T(\psi) < C_t(\psi)\}$.

In particular, $X_{1t}(0) = \min(T, C)$, independent of t , so we shall use the notation $X = \min(T, C) = T \wedge C$.

5. Correction for Censoring by Competing Risk

In our BMT analysis, the primary endpoint is relapse, but some patients die in remission prior to relapse, and the risk factors relevant for death in remission may well overlap those for relapse. Following the approach explained in the appendices of Robins et al. (1992), we handle this complication by assigning each patient a weight W (to be specified below), where $W = 0$ if the patient died in remission and $W = 1/(\text{the conditional probability of not dying in remission by time } X[= T \wedge C])$ otherwise. These weights are then multiplied onto the estimating equations.

We pause to remark that this idea of inverse probability of censoring weighted (IPCW) estimating equations is analogous to the Horvitz–Thompson device in sampling theory. The rationale is to compensate for the attrition, here due to death in remission, by letting each observed relapse [or censoring] represent not only that particular relapse but also those patients who died before relapse. The innovation by Robins and colleagues was to use this approach to adjust for dependent censoring by replacing the usual Kaplan–Meier estimate of the probability of not dying in remission by time T by an estimate of $W^{-1} = \exp[-\int_0^T \lambda_D\{u | \bar{H}^*(u)\}du]$ based on the fitting of a time-dependent Cox model for the intensity $\lambda_D\{u | \bar{H}^*(u)\}du$ of death at u among those with $X > u$ conditional on the history $\bar{H}^*(u)$ of fixed and time-dependent covariates $H(u)$ and of treatment up to u . Specifically, this approach adequately corrects for dependent censoring under the assumption that $\lambda_D\{u | \bar{H}^*(u), X\} = \lambda_D\{u | \bar{H}^*(u)\}$.

In our situation, we estimated W for each patient by fitting a Cox regression model with the endpoint death in remission and with fixed and time-dependent covariates selected on the basis of this and earlier analyses of the data: After correction for the fixed-time covariates relapse/remission status at transplantation, potential censoring time, patient CMV immunity at transplantation, and mismatch score, death in remission depended strongly on both GvHD(t) (estimated relative risk 5.6) and CMV(t) (estimated relative risk 3.3). From this Cox regression model, an estimate of the probability of not having died in remission by time $X_i = T_i \wedge C_i$ may be directly derived for each patient i from the estimated regression coefficients and the estimated underlying hazard.

6. The Final Analysis

We now return to the general description in Section 3. Making the assumption of no unmeasured confounders, we want to test conditional independence, given covariate history \mathcal{H}_t at t , between occurrence of GvHD at t and the counterfactual time T_{G-} to relapse if GvHD were suppressed. This is done by fitting a Cox regression model for time to GvHD with covariates motivated in part by earlier analyses (patient CMV immunity at transplantation (yes/no), transplantation during relapse or remission, mismatch (yes/no), potential censoring time C , CMV(t)) combined with one of $A_{it}(\psi) = X_{it}(\psi)\tau W$

with $\tau = I\{\text{not died in remission}\}$, $i = 1, \dots, 4$, where the $X_{it}(\psi)$ are specified toward the end of Section 4.

For each of the four score functions $A_{it}(\psi)$, $i = 1, \dots, 4$, the sharp null hypothesis of no treatment effect ($\psi = 0$) is tested by performing the score test for the hypothesis that the coefficient of $A_{it}(0)$ vanishes. However, the variance in the score test must be based on the so-called robust variance (Lin and Wei, 1989). The usual variance of the Cox partial likelihood score can no longer be used because the contributions to the partial likelihood score are no longer uncorrelated for two distinct reasons. First, the weights W and therefore the time-dependent covariate $A_{it}(\psi)$ depend on a subject's treatment and covariate history beyond t , disrupting the martingale structure of the Cox partial likelihood score. Second, the weights W are based on an estimate that depends on the data. However, it can be shown that tests and intervals based on a score test that uses a robust variance will be conservative (Robins, 1998), i.e., in large samples, nominal 95% confidence intervals are guaranteed to exclude $\psi_0 = 0$ when true no more than 5% of the time.

Repeating the analysis for a grid of values of ψ (step size 0.05), approximate 95% confidence intervals for ψ_0 are obtained as the set of ψ for which the hypothesis of vanishing coefficient of $A_{it}(\psi)$ is accepted at the 5% level. The results are given in Table 1.

It is seen that none of the score functions yields a clearly significant effect of GvHD on relapse, although the intervals based on $X_{4t}(\psi)$ are the narrowest, a fact that is theoretically predicted by asymptotic calculations in an unpublished UCLA thesis of Marshall Joffe. The excessive length of even the confidence intervals based on $X_{4t}(\psi)$ has four causes. First, the data set is small with few events. Second, our SNFTM model is a semiparametric model rather than a parametric model for distribution of the data. This insures a great deal of robustness at the expense of some efficiency. Third, the 95% intervals reported in Table 1 based on score tests using a robust variance are conservative. Robins (1998) described how to calculate correct, nonconservative intervals that are guaranteed to be narrower. However, because of the considerable programming burden associated with calculating these intervals, we have not reported them here. Fourth, none of the four score functions that we used are expected to be fully efficient as none of the four approximates the efficient score function described in Appendix 4 of Robins (1993). In future work, we plan to try to compute the efficient score and to determine whether this shrinks the size of our confidence interval.

Table 1
Tests for $\psi_0 = 0$ and confidence intervals for ψ_0 for four score functions

Score function based on	P for $\psi_0 = 0$	95% confidence interval for ψ_0	Estimate ψ_0
$X_{1t}(\psi)$	0.58	$[-4.35, \infty)$	-1.55
$X_{2t}(\psi)$	0.74	$[-2.80, 1.90]$	-0.70
$X_{3t}(\psi)$	0.67	$(-\infty, \infty)$	-1.10
$X_{4t}(\psi)$	0.93	$[-2.60, 0.90]$	-0.80

Table 2
Estimated regression coefficients in Cox regression of occurrence of GvHD

Covariate	$\hat{\beta}$	s.d. ($\hat{\beta}$)	$\hat{\beta}/\text{s.d.}(\hat{\beta})$	$\exp(\hat{\beta})$
Transplantation during				
Relapse	1.015	0.345	2.942	2.759
Remission	0	—	—	—
C_i (in days)	-0.000100	0.000215	-0.465	1.000
Patient CMV immunity at transplantation				
Yes	0.810	0.261	3.103	2.249
No	0	—	—	—
Mismatch				
Yes	0.882	0.280	3.150	2.416
No	0	—	—	—
CMV(t)				
Yes at t	0.630	0.350	1.800	1.877
No at t	0	—	—	—
$A_{4t}(0)$	-0.0087	0.105	-0.0829	0.991

Our experience from other problems where the data sets were larger is that reasonable efficiency may be obtained using the score functions X_{1t}, \dots, X_{4t} . Despite the poor efficiency, it may be noted that the general tendency is consistent: a negative estimate of ψ_0 . Taking the apparently most efficient score function $A_{4t}(\psi)$ as an example, $\hat{\psi}_0 = -0.80$ and the Cox regression results for $A_{4t}(0)$ are given in Table 2.

As explained in Section 3, a negative ψ_0 means a beneficial effect of the 'treatment,' and one way of expressing this (Robins et al., 1992) is to remark that, under the structural nested failure time model, the relative increase in lifetime by getting GvHD immediately versus never getting it is

$$(T_{G+} - T_{G-})/T_{G-} = e^{-\psi_0} - 1 \approx 2.2 - 1 = 1.2,$$

or an increase of about 120% in relapse-free time following GvHD. Even though this conclusion is far from statistically significant, it does agree qualitatively with the earlier analyses of these data by Jacobsen et al. (1990) and Eerola (1994).

ACKNOWLEDGEMENTS

This study was initiated while the second author was at the Department of Biostatistics, University of Copenhagen, under a grant from the University of Palermo. The study was sponsored by contract 2R01 CA54706-04A1 from the National Cancer Institute.

RÉSUMÉ

Dans la dernière décennie, J. M. Robins a développé un ensemble d'outils pour, à partir de données d'observation, juger de la responsabilité causale d'un traitement ou d'une exposition dépendante du temps en présence de covariables dépendantes du temps qui peuvent être à la fois des facteurs de confusion et des variables intermédiaires. Cet article présente

une étude de cas de l'une de ces techniques, la G -estimation à partir de modèles de survie emboîtés, appliquée à l'analyse de l'effet de la réaction du greffon contre l'hôte sur les rechutes de leucémie après une greffe de moelle.

REFERENCES

- Andersen, P. K., Borgan, Ø., Gill, R. D., and Keiding, N. (1993). *Statistical Models Based on Counting Processes*. New York: Springer.
- Arjas, E. and Eerola, M. (1993). On predictive causality in longitudinal studies. *Journal of Statistical Planning and Inference* **34**, 361–384.
- Cox, D. R. and Oakes, D. (1984). *Analysis of Survival Data*. London: Chapman and Hall.
- Dabrowska, D. M., Sun, G.-W., and Horowitz, M. M. (1994). Cox regression in a Markov renewal model: An application to the analysis of bone marrow transplant data. *Journal of the American Statistical Association* **89**, 867–877.
- Eerola, M. (1994). *Probabilistic Causality in Longitudinal Studies*, Volume 92, *Lecture Notes in Statistics*. Berlin: Springer.
- 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. J. Serfling (eds), 303–326. Philadelphia: SIAM.
- Jacobsen, N., Badsberg, J. H., Lönnqvist, B., et al., for the Nordic Bone Marrow Transplantation Group. (1987). Predictive factors for chronic graft-versus-host disease and leukaemic relapse after allogeneic bone marrow transplantation. In *Recent Advances and Future Directions in Bone Marrow Transplantation*, S. J. Baum, G. W. Santon, and F. Takaku (eds), 161–164. New York: Springer.
- Jacobsen, N., Badsberg, J. H., Lönnqvist, B., Ringdén, O., Volin, L., Rajantie, J., Nikoskelainen, J., and Keiding, N., for the Nordic Bone Marrow Transplantation Group. (1990). Graft-versus-leukaemia activity associated with CMV-seropositive donor, post-transplant CMV infection, young donor age and chronic graft-versus-host disease in bone marrow allograft recipients. *Bone Marrow Transplantation* **5**, 413–418.
- Kalbfleisch, J. D. and Prentice, R. L. (1980). *The Statistical Analysis of Failure Time Data*. New York: Wiley.
- Klein, J. P., Keiding, N., and Copelan, E. A. (1993). Plotting summary predictions in multistate survival models: Probabilities of relapse and death in remission for bone marrow transplantation patients. *Statistics in Medicine* **12**, 2315–2332.
- Lin, D. Y. and Wei, L. J. (1989). The robust inference for the Cox proportional hazards model. *Journal of the American Statistical Association* **84**, 1074–1078.
- Pepe, M. S. and Mori, M. (1993). Kaplan–Meier, marginal or conditional probability curves in summarizing competing risks failure time data? *Statistics in Medicine* **12**, 737–751.
- Prentice, R. L., Kalbfleisch, J. D., Peterson, Jr., A. V., Flournoy, N., Farewell, V. T., and Breslow, N. E. (1978). The analysis of failure times in the presence of competing risks. *Biometrics* **34**, 541–554.
- Robins, J. M. (1986). A new approach to causal inference in mortality studies with sustained exposure periods—Application to control of the healthy worker survivor effect. *Mathematical Modelling* **7**, 1393–1512.
- Robins, J. M. (1987). Addendum to “A new approach to causal inference in mortality studies with sustained exposure periods—Application to control of the healthy worker survivor effect.” *Computers and Mathematics with Applications* **14**, 923–945.
- Robins, J. M. (1989). The analysis of randomized and non-randomized AIDS treatment trials using a new approach to causal inference in longitudinal studies. In *Health Service Research Methodology: A Focus on AIDS*, L. Sechrest, H. Freeman, and A. Mulley A. (eds), 113–159. Washington, D.C.: NCHSR, U.S. Public Health Service.
- Robins, J. M. (1992). Estimation of the time-dependent accelerated failure time model in the presence of confounding factors. *Biometrika* **79**, 321–334.
- Robins, J. M. (1993). Analytic methods for estimating HIV-treatment and cofactor effects. In *Methodological Issues in AIDS Research*, D. G. Ostrow and R. C. Kessler (eds), 213–290. New York: Plenum.
- Robins, J. M. (1998). Structural nested failure time models. In *Encyclopedia of Biostatistics*, Volume 6, P. Armitage and T. E. Colton (eds), 4372–4389. Chichester: Wiley.
- Robins, J. M. and Rotnitzky, A. (1992). Recovery of information and adjustment for dependent censoring using surrogate markers. In *AIDS Epidemiology, Methodological Issues*, N. P. Jewell, K. Dietz, and V. T. Farewell (eds), 297–331. Boston: Birkhäuser.
- Robins, J. M., Blevins, D., Ritter, G., and Wulfsohn, M. (1992). G -estimation of the effect of prophylaxis therapy for *Pneumocystis carinii* pneumonia on the survival of AIDS patients. *Epidemiology* **3**, 319–336.

Received October 1997. Revised April 1998.

Accepted April 1998.