Published in final edited form as:

Stat Med. 2012 October 30; 31(24): 2805–2821. doi:10.1002/sim.5456.

Analysis of multi-stage treatments for recurrent diseases

Xuelin Huang^{a,*} and Jing Ning^b

^aDepartment of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX 77230

^bDepartment of Biostatistics, School of Public Health, The University of Texas at Houston Health Science Center, Houston, TX 77230

Abstract

Patients with a non-curable disease such as many types of cancer usually go through the process of initial treatment, a various number of disease recurrences and salvage treatments, and eventually death. The analysis of the effects of initial and salvage treatments on overall survival is not trivial. One may try to use disease recurrences and salvage treatments as time-dependent covariates. This can estimate the treatment effects conditional on disease progress history, but not the marginal effects of treatments on overall survival. Nevertheless, such marginal effect estimates are critical for treatment decision-making. Our approach to address this issue is that, at any treatment stage, for each patient, we compute a potential survival time if he/she had received the optimal subsequent treatments, and use this potential survival time to do comparison between initial treatment groups. This potential survival time is assumed to follow an accelerated failure time model at each treatment stage, and calculated by backward induction, starting from the last stage of treatment. By doing that, the effects on survival of different treatments at each stage can be consistently estimated and fairly compared. Under suitable conditions, these estimated effects have a causal interpretation. The proposed model and estimation method are evaluated by simulation studies, and illustrated using the motivating, real data set that describes initial and salvage treatments for patients with soft tissue sarcoma.

Keywords

backward induction; causal inference; coarsening at random; optimal treatment sequence; recurrent events; survival analysis

1. Introduction

Many diseases remain incurable, including most types of cancer. After the initial treatment, a recurrence of such diseases is common, perhaps after a few years or even within a few months. Disease recurrence can substantially increase the risk of death. Patients may also die without disease recurrences. If they die of causes unrelated to the disease of interest, then it is appropriate to view their time to disease-specific survival as censored. However, in many situations, their deaths without disease recurrences are still related to the disease of interest. For instance, the disease symptoms for some patients do not disappear after their initial treatment, so these patients never achieve any kind of “cure”, and they die of the disease without “recurrence”. Another scenario is that when patients die of treatment-related

complications. For example, cancer patients who receive chemotherapy may die from infections due to their low immunity response secondary to chemotherapy. A common approach to the analysis of such data is to estimate the distribution of disease-free survival, i.e., the time to the first disease recurrence or death, whichever occurs first. However, this approach equates the first disease recurrence with death, which may not appropriately reflect the research goal. In many situations, the recurring disease is treatable, and the length of survival after the first disease recurrence can be substantial. When the clinical research interest is to prolong overall survival, an analytic approach that uses disease-free survival can give misleading results.

The analysis of the effect of the initial treatment on overall survival is complicated by the fact that patients receive different treatments after their disease recurrences. For an easy description, we use salvage treatment to mean the treatment after disease recurrence. If the salvage treatments are ignored in the analysis of the effects of initial treatments on overall survival, then the estimates can be biased since salvage treatments also affect overall survival. If we use a Cox [1] proportional hazards model with disease recurrence and salvage treatment as time-dependent covariates, then the estimates of the effects of baseline covariates (including initial treatment indicators) on overall survival do not have a marginal interpretation [2]. This is because disease recurrence is an intermediate outcome of the initial treatments. When this intermediate outcome is used as an explanatory variable (predictor) in the model, the estimation for initial treatment effects on survival is distorted. An accelerated failure time (AFT) model [3] with the same time-dependent covariates as above has the same problem. A lack of appropriate methods for analyzing such data is the main reason that disease-free survival is widely used in medical research, although overall survival is often the ultimate treatment goal.

A good number of publications proposed marginal models for jointly modeling recurrent events and a terminal event such as death [4, 5, 6, 7, 8]. A copula approach was proposed to model the first disease recurrence and death, without considering covariates [9]. Since death censors disease recurrence, but not vice versa, these variables have been called semi-competing risks. Some regression models have been provided for semi-competing risk data [10, 11]. There are also many papers in the literature using conditional approaches to jointly model disease recurrences and death by assuming that these events from the same subject share a common frailty term [12, 13, 14]. All of these models, however, face the same dilemma as mentioned above for the Cox models and AFT models, i.e., whether or not to include disease recurrence and salvage treatment as time-dependent covariates in the model for time to death. If not to include them, the estimation for initial treatment effects on survival is biased since salvage treatments also have an impact on survival. If including them, the effects of initial treatment on survival may be masked because an intermediate outcome (disease recurrence) is included.

Our approach to deal with the above problem is to use backward induction method to identify the optimal treatment sequences. Backward induction is one of the main methods for mathematical dynamic programming optimization [15]. We use the situation of only one disease recurrence for an easy illustration. Denote the time from the initial treatment to disease recurrence by R , the time from salvage treatment to death by S , and the time from the initial treatment to death by T . Assuming that no time elapses from disease recurrence to salvage treatment, we have $T = R + S$. An AFT model for S is assumed to estimate the effects of salvage treatments on S . Then, for each patient with a disease recurrence, a hypothetical overall survival time T^* is computed for the situation if the patient had received the optimal salvage treatment. For those patients who actually received the optimal salvage treatment, they have $T = T^*$. For those patients who did not receive the optimal salvage treatment, they have $T^* > T$. Some patients may die without disease recurrences. For them

we let $T^* = T$. After these preparations, we assume T^* to follow another AFT model. Then the effects of initial treatments on overall survival can be estimated. The two AFT models for S and T^* are joint because the results from the former are used in the building of the latter. The above counterfactual setting for T^* is adopted from causal inference framework. The estimated marginal effects of initial treatments on overall survival can have a causal interpretation under suitable assumptions, which will be specified later.

The above R and S are sometimes called gap times. The estimation of the distribution of the second gap time S is difficult even in the case of independent censoring, as illustrated by Lin, Sun and Ying [16]. They proposed a nonparametric estimator for gap time distributions, which was shown to be equivalent as an inverse-probability-of-censoring-weighted (IPCW) estimator. In this article, we also use the IPCW estimation method. The consistency and optimality properties of the IPCW estimation method in many situations have been shown by Robins and Rotnitzky [17].

We describe our joint accelerated failure time model in Section 2.1, give its estimation method and property in Section 2.2. We describe simulation studies in Section 3. In Section 4, we apply the method to a study of patients with soft tissue sarcoma. We close with some discussion in Section 5, and provide mathematical details in the Appendix.

2. Joint accelerated failure time models

We now describe the backward induction method for identifying the optimal treatment sequences. Suppose there are two options for the initial treatment: A and B . A direct comparison between the effects of A and B on survival is difficult because the survival time is also affected by subsequent salvage treatments after disease recurrences. To compare A with B , a sensible approach is to use the survival times patients would have if they receive the same optimal subsequent treatment sequence. Under this consideration, the optimization over treatment sequences is done in two stages in a backward fashion. First, we use an AFT model to identify the optimal salvage treatment. To simplify the illustration, here in this article, we assume the optimal subsequent treatment sequence is the same for the two initial treatments A and B . In the second step, with another AFT model, the optimal initial treatment is identified. The combination of these two steps gives the optimal treatment sequence. In the second step, the result from the first step is used to compute a hypothetical survival time for each subject, under a potentially counterfactual scenario that this subject had received the optimal salvage treatment. That is how the dependent variable of the second AFT model is constructed. It is also the reason that the construction of AFT models has to be done in a backward fashion, starting with the last treatment stage. The AFT model for the last stage does not need the computation of hypothetical survival times as above. Therefore, it can be done first. In the Section 2.1 below, we illustrate this whole backward induction procedure for identifying optimal treatment sequence.

Suppose there are K stages. To implement the above comparison approach, we start with the last stage K . We first identify the optimal treatment for the K -th stage. Then we compare the treatment options for stage $K - 1$ using the above approach. Continue in this manner until we have completed the same procedure for all the stages. Then the optimal treatment sequence is identified. To simplify the notation and description, hereafter we assume $K = 2$. The methods presented below can be easily extended to situations with $K > 2$ stages. Practical values for K are about 2 to 5 in cancer studies. A cancer patient cannot survive too many disease recurrences. The number of disease recurrences are different from patient to patient. Some subjects may not experience any disease recurrence before death. They may die of treatment toxicities, or they may never achieve any disease-free period and die rapidly of the disease.

2.1. Joint models

Denote by V be a covariate vector observed prior to the initial treatment P . For patients with a recurrence time $R < T$ and $T = R + S$, let Y be summary statistics of all the information observed up to time R , and Q be the salvage treatment. For simplicity, we assume both treatments P and Q take values 0 or 1.

We consider potential outcomes. Suppressing the dependence on covariates, denote by $R(0)$ ($R(1)$) the time to disease recurrence under treatment $P = 0$ ($P = 1$). Also denote by $Y(0)$ ($Y(1)$) the covariate at the disease recurrence time under treatment $P = 0$ ($P = 1$). Let $S(0, 0)$ be the time from disease recurrence to death under treatments $P = 0$ and $Q = 0$. Let $T(0, 0)$ be the survival time with $P = 0$, and if disease recurs, using $Q = 0$. Other potential outcomes $S(j, k)$ and $T(j, k)$ for $j, k = 0, 1$ are defined similarly. Denote all these potential outcomes for a subject by $\mathbf{O} = \{R(j), Y(k), S(j, k), T(j, k), j, k = 0, 1\}$. We try not to use the potential outcome notation when it is not necessary. For example, we denote by S_i the actually observed time from disease recurrence to death for subject i , and similarly for other variables.

Throughout this article, we make the following assumptions. (1) No unmeasured confounders: at any time, conditional on a patient's observed treatment and health history, the treatment assignment for this patient is independent of his/her potential outcomes. That is, $P \perp \mathbf{O} | V$ and $Q \perp \mathbf{O} | (P, R, Y)$. This assumption is also called sequential ignorability [18]. (2) Consistency: the potential outcome under any particular treatment sequence corresponds to the actual outcome if that treatment sequence is followed. (3) Independence between subjects.

As explained earlier, the procedure of backward induction for identifying the optimal treatment sequence starts with the last treatment stage. For the case of only two stages of treatments, we assume the following model for the potential outcome S ,

$$S_i(P_i, Q_i) = \exp(Y_i' \lambda + Q_i' \theta + \epsilon_i) \triangleq \exp(X_i' \beta + \epsilon_i), \quad (1)$$

where $Y_i = Y_i(P_i)$, $X_i = (Y_i', Q_i')'$, and $\beta = (\lambda', \theta)'$, and ϵ_i is assumed to be independent of X_i . Note that P_i can be included as a component of Y_i in the above model. We also assume $\theta \geq 0$. In this setting, if $\theta > 0$ then the optimal salvage treatment is $Q_i = 1$; otherwise the optimal salvage treatment is $Q_i = 0$. Denote $b = I(\theta > 0)$, then the optimal salvage treatment is $Q_i = b$.

In the above model (1), to keep things simple, the interaction between the initial and salvage treatments is not included, but this is easy to be incorporated. When applying the proposed method to real data sets, certainly it is important and wise to include the interaction.

For a subject who has received $P_i = p_i$ and $Q_i = q_i$, we can compute the potential time from disease recurrence to death under the optimal salvage treatment $Q_i = b$ as follows.

$$\begin{aligned} S_i^* \triangleq S(p_i, b) &= \exp(Y_i' \lambda + b' \theta + \epsilon_i) \\ &= \exp(Y_i' \lambda + q_i' \theta + \epsilon_i - q_i' \theta + (\theta)_+) \\ &= S_i \exp(-q_i' \theta + (\theta)_+), \end{aligned} \quad (2)$$

where $(\theta)_+ = |\theta| I(\theta > 0)$. Note that if $q_i = b$, then the above equation gives $S_i^* = S_i$. That is to say, the consistency assumption specified above is satisfied.

The above formulation of AFT model adopts the idea of the structural nested failure time (SNFT) model proposed by Robins and Greenland [19], but it is formulated in a different way. Applying their model to our setting, S_i^* can be negative (the term in the bracket in their equation (1b)). Our model does not have this problem.

After the optimal salvage treatment is identified, we consider the initial treatment. For all the patients (with or without disease recurrences), denote $\delta_{Ri} = I(R_i < T_i)$, then apparently $T_i = \delta_{Ri}(R_i + S_i) + (1 - \delta_{Ri})T_i$. For a subject who has received $P_i = p_i$, his/her potential survival time with the optimal salvage treatment as the treatment option in case of disease recurrence is as follows,

$$T_i^* \triangleq T(p_i, b) = \delta_{Ri}(R_i + S_i^*) + (1 - \delta_{Ri})T_i. \quad (3)$$

Then we assume the following model,

$$T_i^* = \exp(Z_i' \gamma + \xi_i), \quad (4)$$

where $Z_i = (V_i', P_i')'$. The parameter γ has an interpretation as the effects of baseline covariates (including the initial treatment) on overall survival if the optimal salvage treatment would be given in case of disease recurrence. The component of γ corresponding the initial treatments determines which initial treatment is the optimal choice.

The proposed method does not require either the salvage treatment or the initial treatment to be randomized. However, we make the assumption that there are no unmeasured confounders, and require that all potential confounders are appropriately included in the covariates Y or Z in the above models for salvage and initial treatments (in equations (1) and (4) respectively).

By the above formulation, we have an AFT model for the effect of salvage treatment on survival. We also have an AFT model for the effect of initial treatment on survival under a hypothetical scenario that every patient would receive his/her optimal salvage treatment option. The two models are joint in the sense that the results from the model for salvage treatment are used in the building of the model for initial treatment. The formulation of the above two AFT models is in spirit similar as the marginal modeling approach proposed by Wei, Lin and Weissfeld [20]. They considered marginal models for the elapsed time from initial treatment to first recurrence, the elapsed time from initial treatment to second recurrence, etc. They did not consider treatments after disease recurrences. By their formulation, it is not convenient to incorporate the effects of such salvage treatments. Our approach is to use marginal models for time from first-line treatment to death, time from second-line treatment to death, etc. By this formulation, the effects of each treatment (initial or salvage) on survival can be easily modeled. More over, this approach is sensible because the initial treatment may have effects beyond the first disease recurrence. The comparison between the aforementioned two approaches is illustrated by Figure 1. Common advantages of both approaches include: (1) the convenience for practical data analysis; and (2) easy, direct and useful parameter interpretation. A common disadvantage of both approaches is that they are somewhat “unnatural” in the sense the underlying data generation mechanism is not specified. This causes some difficulty for simulation studies. However, in the next section, we provide an easy way for generating data that satisfy the proposed joint AFT model. A natural way to specify the data generation mechanism is to use a regression model for R , and another regression model for S . However, the problem for such a modeling approach is that the resulting model for $T = R + S$ will usually be very complicated, especially when extending to scenarios with multiple disease recurrences. Consequently this

“natural” modeling approach is inconvenient for drawing inference on the overall survival time T . We will illustrate this point in more detail in the Discussion section.

The survival time used above can be replaced by, say, the quality-adjusted survival time, to penalize more times of disease recurrences and treatments. To ensure that the mean survival time is finite, we should consider survival time within, say, $L = 10$ years. Here L is chosen for convenience, but should be slightly less than the observed maximum follow-up in the data set. This restriction would not affect much the analysis results, but give us convenience when fitting the AFT models below. For patients with $T > L$, we redefine for them $T = L$. By doing that, the probability of censoring used in the next subsection is always larger than a small positive constant, so that its inverse is bounded above.

2.2. Estimation

Denote $\mu_\epsilon = E\epsilon_j$ and $\mu_\xi = E\xi_j$. Assume that $\epsilon_j, j = 1, \dots, n$ are independent and identically distributed with an unknown distribution, and similarly for $\xi_j, j = 1, \dots, n$. Furthermore, ϵ_j and ξ_j can be correlated, but ϵ_j is independent of ξ_j for any j .

For convenience, assume that patients $i = 1, \dots, n_1$ had observed disease recurrences ($\delta_{Ri} = 1$), and the remaining patients did not have observed disease recurrences ($\delta_{Ri} = 0$). For uncensored data, the classical least-squares principle is a natural approach to estimate β by minimizing the object function

$$\sum_{i=1}^{n_1} \{\log S_i - X_i' \beta - \mu_\xi\}^2. \quad (5)$$

Then, using the estimated parameters from above to compute T_i^* as done in (3) for each patient, and estimate γ by minimizing the following function.

$$\sum_{i=1}^n \{\log T_i^* - Z_i' \gamma - \mu_\epsilon\}^2, \quad (6)$$

The complications arise from the censored observations. With censoring time C_i , denote $\delta_i = I(T_i \leq C_i)$ and $\tilde{T}_i = \min\{T_i, C_i\}$. Note that for patients $i = 1, \dots, n_1$, T_i and S_i are either both observed (when $\delta_i = 1$) or both censored (when $\delta_i = 0$). For these patients, let $\tilde{S}_i = \min\{S_i, \max(C_i - R_i, 0)\}$. The inverse probability of censoring weighted method can be applied to modify the object functions (5) and (6), as shown hereafter, to accommodate right-censored data [21, 22]. The details of the independent censoring assumption are deferred to the end of this section.

To fit the joint model specified by (1) to (4), first we use the information of patients with disease recurrences and find the roots of the following equation $U_1(\lambda, \theta)$, denoted as $\hat{\lambda}$, and $\hat{\theta}$,

$$U_1(\lambda, \theta) = \sum_{i=1}^n \frac{\delta_{Ri} \delta_i (X_i - \bar{X})}{\hat{G}_i(T_i)} (\log S_i - Y_i' \lambda - Q_i' \theta) = 0, \quad (7)$$

where $X_i = (Y_i', Q_i')'$, $\bar{X} = \sum_{i=1}^n \frac{\delta_{Ri} \delta_i X_i}{\hat{G}_i(T_i)} / \sum_{i=1}^n \frac{\delta_{Ri} \delta_i}{\hat{G}_i(T_i)}$, and $\hat{G}_i(t) = \Pr(C > t)$ is the survival function of censoring time C_i . This gives a closed-form solution for $\beta = (\lambda', \theta')'$ as below

$$\widehat{\beta} = \left\{ \sum_{i=1}^n \frac{\delta_{R_i} \delta_i (X_i - \bar{X}) X_i'}{\widehat{G}_i(T_i)} \right\}^{-1} \sum_{i=1}^n \frac{\delta_{R_i} \delta_i (X_i - \bar{X}) \log S_i}{\widehat{G}_i(T_i)}.$$

The use of $G_i(T_i)$ in equation (7) is based on the method used in equation (6) of Lin, Sun and Ying [16].

Denote $\widehat{S}_i^* = S_i \exp(-Q_i' \widehat{\theta} + (\widehat{\theta})_+)$ and $\widehat{T}_i^* = \delta_{R_i} (R_i + \widehat{S}_i^*) + (1 - \delta_{R_i}) T_i$. After this preparation, we get the estimating equations for parameter γ as follows,

$$U_2(\gamma) = \sum_{i=1}^n \frac{\delta_i (Z_i - \bar{Z})}{\widehat{G}_i(T_i)} (\log \widehat{T}_i^* - Z_i' \gamma), \quad (8)$$

where $\bar{Z} = \sum_{i=1}^n \frac{\delta_i Z_i}{\widehat{G}_i(T_i)} / \sum_{i=1}^n \frac{\delta_i}{\widehat{G}_i(T_i)}$. We obtain a closed-form solution for the estimator of γ from the equation $U_2(\gamma) = 0$:

$$\widehat{\gamma} = \left\{ \sum_{i=1}^n \frac{\delta_i (Z_i - \bar{Z}) Z_i'}{\widehat{G}_i(T_i)} \right\}^{-1} \sum_{i=1}^n \frac{\delta_i (Z_i - \bar{Z}) \log \widehat{T}_i^*}{\widehat{G}_i(T_i)}. \quad (9)$$

In general, the survival functions for the censoring times are unknown but can be consistently estimated by nonparametric models or semiparametric models under reasonable assumptions. If we assume censoring time C is independent of T , R and covariates X , and $G_X(\cdot) = G(\cdot)$, then we can use the Kaplan-Meier estimator to estimate G . More generally, we can assume coarsening at random, which means that, given the full data, the censoring event defining the observed data depends on only the observed part of the data [23, 24]. In our situation, let D represent all the other data, and $\bar{D}(t)$ represents the information about all the data (R , S , T , X , Z and Q , excluding C) up to the time t , and t the end of the study. Then the assumption is specified as follows,

$$h(t | \bar{D}(\tau)) = h(t | \bar{D}(t)), \quad (10)$$

where h is the hazard function of C . To estimate the distribution of C , we assume a Cox [1] proportional hazards model for the hazard function of C , as shown below,

$$h_i(t | \bar{D}_i(t)) = h_0(t) \exp(W_i'(t) \psi), \quad (11)$$

where $W_i(t)$ is a summary statistic from $\bar{D}_i(t)$, ψ is an unknown parameter, and $h_0(\cdot)$ is an arbitrary positive hazard function. From this model, we can estimate the probability

$G_i(t) = \Pr(C_i \geq t | \bar{D}_i(t))$ by plugging in the partial likelihood estimator $\widehat{\psi}$ and the Breslow [25] estimator for $\widehat{h}_0(t)$. The resulting estimator is then

$\widehat{G}_i(t) = \exp\left\{-\int_0^t \widehat{h}_0(s) \exp(W_i'(s) \widehat{\psi}) ds\right\}$. Given the consistent estimator of the survival

function for the censoring, we have asymptotic unbiased estimating equations for γ and β in (8) and (7).

Assume γ_0 and β_0 to be the true values of γ and β , respectively. We can prove that the estimating equations $U_2(\gamma)$ and $U_1(\beta)$ yield consistent and weak convergent estimators, $\widehat{\gamma}$ and $\widehat{\beta}$, under the following regularity conditions,

(C.1) Z and X are bounded, if there exists a constant vector d_1 such that $d_1'Z=0$ then $d_1=0$, and similarly for X ;

(C.2) There exists a constant $\rho > 0$ such that $Pr(R < T) > \rho$;

(C.3) There exists a constant $\sigma_c > 0$ such that $Pr(C > T) > \sigma_c$;

(C.4) $\Gamma_\gamma = E \left\{ \frac{\delta_i (Z_i - \bar{Z}) Z_i'}{G_i(T_i)} \right\}$ and $\Gamma_\beta = E \left\{ \frac{\delta_{R_i} \delta_i (X_i - \bar{X}) X_i'}{G_i(T_i)} \right\}$ exist (i.e., expectations $< \infty$) and are nonsingular.

(C.5) The asymptotic variance-covariance matrices of two estimating equations, Σ_γ and Σ_β , exist.

Theorem 1. If regularity conditions C.1–C.5 are satisfied, $\widehat{\gamma}$ and $\widehat{\beta}$ are consistent estimators of γ and β , respectively. Moreover, $\sqrt{n}(\widehat{\gamma} - \gamma_0) \rightarrow MVN\left(0, \Gamma_\gamma^{-1} \sum_\gamma (\Gamma_\gamma^{-1})'\right)$ and $\sqrt{n}(\widehat{\beta} - \beta_0) \rightarrow MVN\left(0, \Gamma_\beta^{-1} \sum_\beta (\Gamma_\beta^{-1})'\right)$, where Γ_γ and Γ_β are expectations of the Jacobian matrices. A detailed proof for Theorem 1 and the “sandwich” variance-covariance formulas for $\widehat{\gamma}$ and $\widehat{\beta}$ are provided in the Appendix.

The above estimation procedure can certainly be improved or replaced by more efficient approaches [26, 27, 28, 29, 30]. To extend these methods to our joint AFT models for sequential time intervals, the complication, again, is that a simple independent censoring time C in this situation will induce dependent censoring time for, say, S . This because the censoring time for S is $C - R$, which is correlated with S due to the correlation between R and S . These extensions warrant future research.

3. Simulation studies

We conducted simulation studies to evaluate the proposed estimators for regression coefficients in the joint models. Specifically, we aimed to evaluate the small sample accuracy and precision of our estimators and associated inferences. Each study comprised 500 runs. A sample size of 100 or 200 was used.

Let A_{i1} and A_{i2} be the initial and salvage treatment indicators with $P(A_{i1} = 1) = 0.5$, and independently $P(A_{i2} = 1) = 0.5$. Let Z_{i1} be a continuous baseline covariate with a Uniform(0,1) distribution, Z_{i2} be a binary covariate at disease recurrence with a $P(Z_{i2} = 1) = 0.6$. For each patient, let $\log T_{0i}$ have a uniform(−0.5,0.5) distribution, and $T_i = T_{0i} \exp(Z_{i1} + 0.5 A_{i1})$. Assume 75% of the patients have a disease recurrence time $R_i \sim \text{Uniform}(0.1 T_i, 0.9 T_i)$. The rest 25% patients do not have a disease recurrence. For the 75% of patients with a disease recurrence, let $S_{0i} = T_i - R_i$ and $S_i = S_{0i} \exp(-0.329 + 0.5 Z_{i2} A_{i2})$. This gives $\beta = (0.5, -1)$ for the parameter in model (1). By this setting, the optimal salvage treatment is $A_2 = 0$. Under this optimal salvage treatment, a patient's survival after disease recurrence would be $S_i^* = S_{0i} \exp(-0.329 + 0.5 Z_{i2})$, and the overall survival time would be $T_i^* = R_i + S_i^*$. Note that

$E[\exp(-0.329 + 0.5 Z_{i2})] = 1$. with this fact, it can be verified that $E(T_i^*|Z_{i1}, A_{i1}) = E(T_i|Z_{i1}, A_{i1})$ which is equal to $E(T_{0i}) \exp(Z_{i1} + 0.5 A_{i1})$. This shows that the model in (4) is satisfied with respect to its mean structure, and the parameter $\gamma = (1, 0.5)$.

The censoring times were independently generated from exponential distribution $\exp(\tau_1)$, where the parameter τ_1 controlled the censoring percentages for disease recurrence ($C_1\%$) and death ($C_2\%$). In addition to the scenarios with covariate-independent censoring, we also studied scenarios in which the censoring variables were generated from a Cox regression model $h_i(t) = \tau_2 t \exp(-Z_{i1})$, where the parameter τ_2 controlled the degree of censoring. In these scenarios, the Cox proportional hazards models were used to model the covariate-dependent survival distributions for the censoring.

The simulation results are presented in Table 1. The proposed estimators performed well in all the scenarios with various degrees of censoring. The empirical biases were very small, the estimated asymptotic standard errors match their empirical counterparts well, and the coverage probabilities of the 95% confidence intervals were close to the nominal level.

4. Application to a soft tissue sarcoma study

The primary treatment for patients with soft tissue sarcoma is surgical resection of the tumor. The role of systemic chemotherapy for these patients, given as adjuvant treatment to prevent distant tumor recurrence, remains controversial [31, 32]. To elucidate the effect of chemotherapy on disease recurrence and survival, a cohort of 674 patients was identified from the prospectively collected tumor registries of two comprehensive cancer centers. This homogeneous population consisted of all the patients presenting with primary, stage III, soft tissue sarcoma of an extremity, who were treated between 1984 and 1999 [32]. All the patients received definitive surgical resection. Medical records were examined retrospectively for verification of known prognostic factors, treatment sequences and survival outcomes. All clinically important variables were included in the data set.

The most important prognostic factors for patients with soft tissue sarcoma are tumor size, grade and anatomic relationship to the deep fascia. The American Joint Committee on Cancer [33] has defined soft tissue sarcomas that are large (> 5 cm), high-grade, and which occur deep in an extremity as stage III tumors. Patients with stage III soft tissue sarcomas are known to be at significant risk for distant recurrence and subsequent sarcoma-related death. In addition to primary tumor factors, age, gender, pathologic resection margins, anatomic tumor location and histologic subtype have been reported to be important clinical factors associated with patient outcome [34, 35, 36].

Of the 674 patients described in the data set, 336 received chemotherapy as a component of their initial treatment regimen. Of these, 179 patients had at least one recurrence of the disease and 59 received salvage chemotherapy. The mean (median) time to first distant recurrence for these 179 patients was 1.74 (1.12) years. Among the 338 patients who did not receive chemotherapy as a component of their initial treatment, 171 had at least one recurrence of disease and 42 received salvage chemotherapy. The mean (median) time to first recurrence for these 171 patients was 1.73 (0.82) years. More details are provided in Table 2. Since, as mentioned earlier, the effects of chemotherapy was not clear, the decisions to receive chemotherapy or not in initial and/or treatments were largely based on treating physicians' personal beliefs.

Figure 1 shows the Kaplan-Meier estimators [37] for the distributions of disease recurrence and survival times, stratified by initial and salvage chemotherapy treatments. Death due to the disease or treatment complications is the event of interest. Death due to unrelated causes or remaining alive at the end of study is censored. In the plot of recurrence time distribution

(Fig. 1a), the 13 deaths due to treatment complications are also treated as disease recurrence (8 with and 5 without initial chemotherapy).

Examining the plot in Figure 1(a), we can see that the two survival functions cross at about the second year. So it appears that chemotherapy as part of the initial treatment benefited patients for a period of time, but that the benefit disappeared after about two years. The log-rank test comparing these two groups gives a p -value of 0.73. When the survival curves cross with each other, the log-rank test and the Cox proportional hazard models are not the best methods to detect the differences. The accelerated failure time (AFT) models provide a good alternative analytic approach. Using this approach, the mean lifetime after treatment is used as the criterion for treatment comparisons; this is a sensible criterion. It is difficult to use a single AFT model to fully describe the initial and salvage treatments and the multivariate outcomes (disease recurrence time and survival time). The proposed joint models in Section 2 are useful in this situation.

Some important confounding factors for these comparisons are patient age, tumor size and histologic subtype. Older patients were much less likely to receive chemotherapy because they have lower tolerance for the drugs toxicity. Patients with a larger tumor size, and patients malignant fibrous histiocytoma are more likely to receive chemotherapy (results not shown). Overall speaking, patients with worse prognosis probably were more likely to be offered chemotherapy. Due to these important confounding factors, a direct comparison based on Figure 1 would be misleading. It is important to adjust the effects of the covariates.

To apply the method in Section 2, we used a Cox model to compute the probabilities of (not) being censored. Disease recurrence and salvage chemotherapy were used as time-dependent covariates. To satisfy the condition (C.3) in Theorem 1, we truncated survival times at 15 years. That is to say, if $T > 15$, then we set $T = 15$ and $\delta = 1$, regardless of the actual values of T and δ . By doing this, the estimated $\widehat{G}_i(T_i)$ in the denominators of the estimating equations in Section 2 is always bounded away from zero.

Using joint models from Section 2, first, we considered the analysis on the time from recurrence to death using the data for the 350 patients who had at least one disease recurrence; we solved the estimating equation in (7). The results are listed in Table 3. The effect of salvage chemotherapy on survival was very small (mean lifetime ratio = 0.99, $p = 0.95$). The ratios of remaining lifetime after disease recurrence were: 0.50 ($p < 0.01$) for patients with tumors proximal vs. distal to their breast chests, 0.53 ($p < 0.01$) for patients who were 60 years or older vs. 40 years or younger, and 1.61 for liposarcoma vs. malignant fibrous histiocytoma ($p = 0.04$).

Next, we analyzed the effect of the initial chemotherapy and other factors on overall survival using the joint models from section 2. Data from all 674 patients were used. As indicated in (8), for the 350 patients who had at least one recurrence, their observed survival times were adjusted by the estimated effect of the salvage chemotherapy. A hypothetical survival time was constructed for each one of them as described in §2.2. For the 324 patients without disease recurrence, their original survival times were used. The solution of the resulting estimating equation is given by (9). The results, listed in Table 4, show that the initial chemotherapy appeared to be beneficial, extending the patients' overall survival times by 32% ($p = 0.04$), after the effects of the salvage treatment and baseline covariates were adjusted. In addition to the significant effects of patient age and tumor site, we found that the size and histology type of the tumor had a significant effect on a patient's survival.

It is always risky to make causal inference based on observational studies since the assumption of “no unmeasured confounders” may not be satisfied. We conducted a

sensitivity analysis for the results in the Table 4, using the method proposed by Lin, Psaty and Kronmal [38]. Suppose there is a unmeasured confounder Z_u . Its mean difference between groups with and without initial chemotherapy is d . Its regression coefficient in the AFT model (4) is κ . That is to say, the true model is $T^* = \exp(\gamma'Z + \kappa Z_u + \xi)$, where T^* is the survival time under optimal salvage treatment, and Z includes all the variables listed in Table 4. When there are multiple unmeasured confounders, we may assume their combination as Z_u . Under different scenarios of d and κ , the resulting estimates for the effect of initial chemotherapy on survival are listed in Table 5. We can see that if a large Z_u value prolongs survival ($\kappa > 0$), and the mean of Z_u is higher in the group with initial chemotherapy than the group without initial chemotherapy ($d > 0$), then the results in Table 4 overestimates the benefit of initial chemotherapy. This is reflected by that the estimated mean life time ratios in the lower half part of Table 5 are all lower than 1.32, the estimated mean life time ratios of initial chemotherapy versus no initial chemotherapy. On the other hand, similarly from the upper half part of Table 5, we can see that if a large Z_u value hurts survival ($\kappa < 0$), and still $d > 0$, then the result of the benefit of initial chemotherapy reported in Table 4 is underestimated. The magnitude of bias in each case of κ and d combination can be seen from Table 5.

5. Discussion

It is ethically imperative that cancer patients whose initial treatments fail be given subsequent salvage therapy. The treatment sequence we considered, the initial treatment followed by disease recurrence and then salvage treatment, is very common in cancer treatment. The ultimate goal of treatment is to prolong patient survival. Consequently, it is very important to assess the effect of the treatment sequence on patient survival in this type of observational setting. Despite the importance of this type of assessment, a satisfactory method for it has been lacking. The methods most commonly used for this type of data generally either examine the overall survival, without accounting for the effects of salvage treatments, or use disease-free survival to compare the treatment effects. Results from the former method are potentially biased because of an imbalance between the numbers of patients receiving each type of salvage treatments. The latter method ignores the data that represent the time after disease recurrence. This is not ideal when the recurrence is treatable and the outcome of interest is overall survival. This article provides a solution for this dilemma.

By using backward induction, we solve two easy AFT equations to find out the optimal 2-stage treatment sequences. One might wonder why we do not do the modeling in a forward fashion. We use the following example (with some notation different from that used in earlier sections) to illustrate the difficulty of such a “natural” approach. To do a forward modeling, assume the time from initial treatment to disease recurrence $R = \exp(Z_1'\alpha + \epsilon_1)$, and time from disease recurrence to death $S = \exp(Z_2'\beta + \epsilon_2)$, where Z_1 denotes baseline covariates (including indicators for initial treatments), and Z_2 is the covariates at the time of disease recurrence (including indicators for salvage treatments). Since Z_2 should include all important information at the time of disease recurrence, suppose R is a component of Z_2 . For easy notation, we rewrite the original Z_2 as $(Z_2', R)'$, and the original β as (β', γ) . Then we have that

$$\begin{aligned} T &= R + S \\ &= \exp(Z_1'\alpha + \epsilon_1) + \exp(Z_2'\beta + R\gamma + \epsilon_2) \\ &= \exp(Z_1'\alpha + \epsilon_1) + \exp(Z_2'\beta + \gamma \exp(Z_1'\alpha + \epsilon_1) + \epsilon_2). \end{aligned}$$

It can be seen that the effects of initial treatment (a component of Z_1) on the final survival time T is in a complicated model, which is difficult to solve. Due to this reason, we use backward induction to directly build a model for the effects of initial treatment on T . As the number of stages increases, the advantages of the backward induction modeling approach get even more clear.

In this article, the data we use can be from observational studies, randomized clinical trials, or their combinations. Certainly, the gold standard for the identification of optimal treatment sequences is through randomized clinical trials. However, most clinical trials compare treatments at a single stage. The organization of a trial evaluating multi-stage treatments is extremely difficult and costly due to the long follow-up time and large sample size needed. Current and future work to identify the optimal multi-stage treatment sequences thus depend heavily on data from observational studies and combinations of clinical trials. Even if a clinical trial of multi-stage regimes can be organized, the proposed method is still useful. This is because it can be at least used to eliminate treatment sequences that are apparently inferior and select a few good candidates of treatment sequences for the multi-stage trial. Without such an elimination procedure, there would be too many treatment sequences to be tested, and some patients will unnecessarily receive inferior treatment sequences. Simply put, there is no reason not to take advantage of available data.

A major assumption of the proposed method is that there are no unmeasured confounders (also called sequential ignorability). This assumption is commonly made by many other researchers working on observational studies [18, 39]. In the recent decades of computerized information system, with extensive health information electronically recorded, this assumption should be reasonable in clinical studies because all the information physicians based on to make treatment decisions should have been recorded. However, such an assumption might be too strong for observational epidemiology studies. For any study, to apply the proposed method, careful sensitivity analyses are always recommended.

We have assumed AFT models for the effects of the initial and salvage treatments. It is worth considering other models in order to check the sensitivity of the identified optimal regimes on model choices, and achieve reliable results. The key criterion for a model to be suitable for our purpose is that it must directly address the final outcome of interest, namely, the survival time. The Cox proportional hazards model is not a good choice because it works on the hazard function rather than directly on the survival time. This feature of the Cox model makes it convenient to handle censored observations, but difficult to be used for the construction of a counterfactual survival time as done in Section 2. The application of other models to the setting of multi-stage treatment optimization, and the goodness-of-fit testing for these models in the new setting, are both of interest for future research.

Another important aspect of using the proposed models to make treatment decisions is the problem of variable selection. The number of covariates can be large, even larger than the number of subjects in the study. Beside this, there is another special features for this classical problem in our setting. That is, a variable may appear in multiple equations. For example, a baseline covariate may appear in both equations (1) and (4). This is in contrast to the variable selection in a single-stage treatment setting, where each covariate appears only in one equation. Therefore, for the proposed modeling approach, it is an open question on how to do the variable selection in a joint fashion rather than separately.

Acknowledgments

This research was supported by the U.S.A. National Institutes of Health grants 5P50 CA100632, 5PO1 CA055164 and 1PO1 CA108631-01.

Contract/grant sponsor: U.S.A. National Institutes of Health grants 5P50 CA100632, 5PO1 CA055164 and 1PO1 CA108631-01.

A. Appendix

For simplicity, we derive the asymptotic properties of the estimated coefficients under the assumptions that the censoring time is independent of the covariates ($G_i(R_i) = G(R_i)$), that $\mu_\epsilon = \mu_\xi = 0$, and that the arguments can be easily generalized to a setting with covariate-dependent censoring and non-zero means.

A.1. Consistency of $\widehat{\beta}$

We derive the expectation of the following quantity to be zero

$$E \left\{ \frac{\delta_{R_i} \delta_i X_i}{G(T_i)} (\log S_i - X_i' \beta) \right\} = E \left[E \left\{ \frac{\delta_{R_i} \delta_i X_i}{G(T_i)} (\log S_i - X_i' \beta) | T_i, S_i, X_i \right\} \right] = 0$$

where the inner side of the double expectation is taken with respect to the censoring time C . Thus, $U_1(\beta)$ is an asymptotic unbiased estimating equation of β . For convenience, we denote $a^{\otimes 2} = aa'$ for any vector a . Since the first derivative of $U_1(\beta)$ with respect to β ,

$$\Gamma_n(\beta) = dU_1(\beta) / d\beta = - \sum_{i=1}^n \frac{\delta_{R_i} \delta_i X_i^{\otimes 2}}{\widehat{G}(T_i)},$$

is negative semi-definite, the estimating equation $U_1(\beta)$ has a unique solution $\widehat{\beta}$. With probability one, the quantity $n^{-1}(\beta' - \beta_0') U_1(\beta)$ converges to

$$\int Pr(R < T | x) \{x'(\beta - \beta_0)\}^2 dH(x),$$

where $H(x)$ is the distribution function of covariates X . Then the consistency of $\widehat{\beta}$ follows from the fact that the above limit is non-negative and is zero if and only if $\beta = \beta_0$.

A.2. Asymptotic distribution of $\widehat{\beta}$

We establish the asymptotic distribution of $\widehat{\beta}$ as follows. First, we derive the asymptotic distribution of $U_1(\beta_0)$. The estimating equation $U_1(\beta_0)$ can be rewritten as

$$U_1(\beta_0) = \sum_{i=1}^n \frac{\delta_{R_i} \delta_i X_i}{G(T_i)} (\log S_i - X_i' \beta_0) + \sum_{i=1}^n \frac{\delta_{R_i} \delta_i X_i \{G(T_i) - \widehat{G}(T_i)\}}{G(T_i) \widehat{G}(T_i)} (\log S_i - X_i' \beta_0).$$

Given the asymptotic martingale expression of $\widehat{G}(t)$,

$$\sqrt{n} \{ \widehat{G}(t) - G(t) \} / \widehat{G}(t) = - \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^\infty \frac{I(s \leq t) dM_i^c(s)}{G(s) S_T(s)} + o_p(1),$$

we have

$$\begin{aligned} & \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\delta_{R_i} \delta_i X_i \{G(T_i) - \widehat{G}(T_i)\}}{G(T_i) \widehat{G}(T_i)} (\log S_i - X_i' \beta_0) \\ &= \frac{1}{n} \sum_{i=1}^n \left\{ \frac{1}{\sqrt{n}} \sum_{j=1}^n \int_0^\infty \frac{I(s \leq T_i) dM_j^c(s)}{G(s) S_T(s)} \right\} \frac{\delta_{R_i} \delta_i X_i (\log S_i - X_i' \beta_0)}{G(T_i)} + o_p(1), \end{aligned}$$

where $S_T(s) = P(T \geq s)$, and $M_i^c(s) = I(T_i \leq s, \delta_i = 0) - \int_0^s I(\tilde{T}_i \geq u) d\Lambda_c(u)$ and $\Lambda_c(u)$, respectively, are the martingale and cumulative hazard function of the censoring times.

Denote the expectation of $I(s \leq T_i) \delta_{R_i} \delta_i X_i (\log S_i - X_i' \beta_0) / G(T_i)$. Using the fact that

$\frac{1}{n} \sum_{i=1}^n \frac{I(s \leq T_i) \delta_{R_i} \delta_i X_i (\log S_i - X_i' \beta_0)}{G(T_i)}$ converges to $D(s)$, we have

$$\frac{1}{\sqrt{n}} U_1(\beta_0) = \frac{1}{\sqrt{n}} \left\{ \sum_{i=1}^n \frac{\delta_{R_i} \delta_i X_i}{G(T_i)} (\log S_i - X_i' \beta_0) + \int_0^\infty \frac{D(s)}{G(s) S_T(s)} dM_i^c(s) \right\} + o_p(1).$$

Therefore $1/\sqrt{n} U_1(\beta_0)$ converges weakly to a normal distribution with zero mean and variance-covariance matrix $\sum_{\beta} = E\{A_i^{\otimes 2}\}$, where

$$A_i = \frac{\delta_{R_i} \delta_i X_i}{G(T_i)} (\log S_i - X_i' \beta_0) + \int_0^\infty \frac{D(s)}{G(s) S_T(s)} dM_i^c(s).$$

Then by Taylor series expansion,

$$n^{-1/2} U_1(\widehat{\beta}) = n^{-1/2} U_1(\beta_0) + \frac{1}{n} \Gamma_n(\beta_0) \sqrt{n}(\widehat{\beta} - \beta_0) + o_p(1),$$

$\sqrt{n}(\widehat{\beta} - \beta_0)$ converges weakly to a normal distribution with mean zero and variance-covariance matrix $\Gamma_{\beta}^{-1} \sum_{\beta} (\Gamma_{\beta}^{-1})'$, in which Γ_{β} is the expectation of the Jacobian matrices $\Gamma_n(\beta)$.

A.3. Asymptotic distribution of $\widehat{\gamma}$

The consistency of $\widehat{\gamma}$ can be shown by the same arguments in the proof of the consistency of $\widehat{\beta}$. We outline only the main steps of the proof for the asymptotic distribution of $\widehat{\gamma}$.

Assume θ_0 , λ_0 and γ_0 to be the true values of three regression coefficient vectors, θ , λ and γ , respectively. First, we derive the weak convergence of the estimating equation $n^{-1/2} U_2(\gamma_0)$, which can be rewritten as

$$\begin{aligned}
n^{-1/2} U_2(\gamma_0) &= n^{-1/2} \sum_{i=1}^{n_1} \frac{\delta_i Z_i}{G(T_i)} \left\{ \log \left(R_i + S_i e^{-Q'_i \theta_0 + (\theta_0)_+} \right) - Z'_i \gamma_0 \right\} \\
&+ n^{-1/2} \sum_{i=n_1+1}^n \frac{\delta_i Z_i}{G(T_i)} \left\{ \log T_i - Z'_i \gamma_0 \right\} \\
&+ n^{-1/2} \sum_{i=1}^{n_1} \frac{\delta_i Z_i}{\widehat{G}(T_i)} \left\{ \log \left(R_i + S_i e^{-Q'_i \widehat{\theta} + (\widehat{\theta})_+} \right) - \log \left(R_i + S_i e^{-Q'_i \theta_0 + (\theta_0)_+} \right) \right\} \quad (1) \\
&+ n^{-1/2} \sum_{i=1}^{n_1} \delta_i Z_i \log \left(R_i + S_i e^{-Q'_i \theta_0 + (\theta_0)_+} \right) \frac{G(T_i) - \widehat{G}(T_i)}{\widehat{G}(T_i) G(T_i)} \\
&+ n^{-1/2} \sum_{i=n_1+1}^n \frac{\delta_i Z_i \{G(T_i) - \widehat{G}(T_i)\}}{G(T_i) \widehat{G}(T_i)} (\log T_i - Z'_i \gamma_0).
\end{aligned}$$

It follows from the asymptotic i.i.d. expression of $\widehat{\theta}$ that

$$\begin{aligned}
& n^{-1/2} \sum_{i=1}^{n_1} \frac{\delta_i Z_i}{\widehat{G}(T_i)} \left\{ \log \left(R_i + S_i e^{-Q'_i \widehat{\theta} + (\widehat{\theta})_+} \right) - \log \left(R_i + S_i e^{-Q'_i \theta_0 + (\theta_0)_+} \right) \right\} \\
&= n^{-1/2} \sum_{i=1}^{n_1} \frac{\delta_{R_i} \delta_i Z_i (-Q'_i + \text{sign}(\theta_0)) S_i e^{-Q'_i \theta_0 + (\theta_0)_+}}{G(T_i) (R_i + S_i e^{-Q'_i \theta_0 + (\theta_0)_+})} (\widehat{\theta} - \theta_0) + o_p(1) \quad (2) \\
&= n^{-1/2} C_\theta \Gamma_\theta^{-1} \sum_{i=1}^{n_1} A_i + o_p(1),
\end{aligned}$$

where and $C_\theta = E \left\{ \frac{\delta_{R_i} \delta_i Z_i (-Q'_i + \text{sign}(\theta_0)) S_i e^{-Q'_i \theta_0 + (\theta_0)_+}}{G(T_i) (R_i + S_i e^{-Q'_i \theta_0 + (\theta_0)_+})} \right\}$, and Γ_θ^{-1} is a sub-matrix of Γ_β^{-1} corresponding to the parameter θ .

Using the asymptotic martingale expression of $\widehat{G}(t)$, we have

$$\begin{aligned}
& n^{-1/2} \sum_{i=1}^{n_1} \delta_i Z_i \log \left(R_i + S_i e^{-Q'_i \theta_0 + (\theta_0)_+} \right) \frac{G(T_i) - \widehat{G}(T_i)}{\widehat{G}(T_i) G(T_i)} \\
&= n^{-1/2} \sum_{i=1}^n \frac{\delta_{R_i} \delta_i Z_i \log \left(R_i + S_i e^{-Q'_i \theta_0 + (\theta_0)_+} \right)}{G(T_i)} \frac{1}{n} \sum_{j=1}^n \int_0^\infty \frac{I(s \leq T_i) dM_j^c(s)}{G(s) S_T(s)} + o_p(1) \quad (3) \\
&= n^{-1/2} \sum_{i=1}^n \int_0^\infty \frac{D_\theta(s) dM_i^c(s)}{G(s) S_T(s)} + o_p(1),
\end{aligned}$$

$$\begin{aligned}
& n^{-1/2} \sum_{i=n_1+1}^n \frac{\delta_i Z_i \{G(T_i) - \widehat{G}(T_i)\}}{G(T_i) \widehat{G}(T_i)} (\log T_i - Z'_i \gamma_0) \\
&= n^{-1/2} \sum_{i=1}^n \frac{(1 - \delta_{R_i}) \delta_i Z_i (\log T_i - Z'_i \gamma_0)}{G(T_i)} \frac{1}{n} \sum_{j=1}^n \int_0^\infty \frac{I(s \leq T_i) dM_j^c(s)}{G(s) S_T(s)} + o_p(1) \quad (4) \\
&= n^{-1/2} \sum_{j=1}^n \int_0^\infty \frac{D_\gamma(s) dM_j^c(s)}{G(s) S_T(s)} + o_p(1),
\end{aligned}$$

$$\begin{aligned}
D_\theta(s) &= E \left\{ \frac{\delta_{R_i} \delta_i Z_i \log \left(R_i + S_i e^{-Q'_i \theta_0 + (\theta_0)_+} \right) I(s \leq T_i)}{G(T_i)} \right\} \text{ and} \\
D_\gamma(s) &= E \left\{ \frac{(1 - \delta_{R_i}) \delta_i Z_i (\log T_i - Z'_i \gamma_0) I(s \leq T_i)}{G(T_i)} \right\}.
\end{aligned}$$

Using equations (1), (2), (3) and (4), we can derive the asymptotic representation of $U_2(\gamma_0)$:

$$\begin{aligned}
n^{-1/2} U_2(\gamma_0) &= n^{-1/2} \sum_{i=1}^n \frac{\delta_{R_i} \delta_i Z_i}{G(T_i)} \left\{ \log \left(R_i + S_i e^{-Q'_i \theta_0 + (\theta_0)_+} \right) - Z'_i \gamma_0 \right\} \\
&+ n^{-1/2} \sum_{i=n_1+1}^n \left\{ \frac{(1-\delta_{R_i}) \delta_i Z_i}{G(T_i)} (\log T_i - Z'_i \gamma_0) + C_\theta - \Gamma_\theta^{-1} A_i \right\} \\
&+ n^{-1/2} \sum_{i=1}^n \int_0^\infty \left[\frac{\{D_\theta(s) + D_\gamma(s)\} dM_i^c(s)}{G(s) S_T(s)} \right] + o_p(1) \\
&= n^{-1/2} \sum_{i=1}^{n_1} \varphi_i + o_p(1).
\end{aligned} \tag{5}$$

Given this i.i.d. representation, the weak convergence of $n^{-1/2} U_2(\gamma_0)$ follows from the standard central limit theorem. Then by Taylor series expansion,

$$\frac{1}{\sqrt{n}} U_2(\widehat{\gamma}) = \frac{1}{\sqrt{n}} U_2(\gamma_0) + \Gamma_\gamma \sqrt{n} (\widehat{\gamma} - \gamma_0) + o_p(1),$$

$\sqrt{n}(\widehat{\gamma} - \gamma_0)$ converges weakly to a normal distribution with mean zero and variance-covariance matrix $\Gamma_\gamma^{-1} \sum_\gamma \Gamma_\gamma^{-1}$, in which Σ_γ is the asymptotic variance-covariance matrix of $n^{-1/2} U_2(\gamma_0)$ and $\Gamma_\gamma = -E \left\{ \frac{\delta Z^{\otimes 2}}{G(T)} \right\}$. The matrix Σ_γ can be estimated by $n^{-1} \sum_{i=1}^n \widehat{\varphi}_i^{\otimes 2}$, where $\widehat{\varphi}_i$ can be obtained by replacing $G(s)$, $S_T(s)$, etc., by their respective estimators.

References

1. Cox D. Regression models and life tables (with discussion). *Journal of the Royal Statistical Society, Series B.* 1972; 34:187–220.
2. Kalbfleisch, JD.; Prentice, RL. *The Statistical Analysis of Failure Time Data.* Wiley; 2002.
3. Cox, D.; Oakes, D. *Analysis of Survival Data.* Chapman & Hall/CRC; 1984.
4. Cook RJ, Lawless JF. Marginal analysis of recurrent events and a terminating event. *Statistics in Medicine.* 1997; 16:911–924. [PubMed: 9160488]
5. Li Q, Lagakos S. Use of wei-lin-weissfeld method for the analysis of a recurring and a terminating event. *Statistics in Medicine.* 1997; 16:925–940. [PubMed: 9160489]
6. Ghosh D, Lin DY. Nonparametric analysis of recurrent events and death. *Biometrics.* 2000; 56:554–562. [PubMed: 10877316]
7. Ghosh D, Lin DY. Marginal regression models for recurrent and terminal events. *Statistica Sinica.* 2002; 12:663–688.
8. Huang Y, Wang MC. Frequency of recurrent events at failure time: Modeling and inference. *Journal of American Statistical Association.* 2003; 98:663–670.
9. Fine JP, Jiang H, Chappell R. On semi-competing risks data. *Biometrika.* 2001; 88:907–919.
10. Peng L, Fine JP. Rank estimation of accelerated lifetime models with dependent censoring. *Journal of the American Statistical Association.* 2006; 101:1085–1093.
11. Peng L, Fine JP. Regression modeling of semicompeting risks data. *Biometrics.* 2007; 63:96–108. [PubMed: 17447934]
12. Lancaster A, Intrator O. Panel data with survival: hospitalization of hiv-positive patients. *Journal of the American Statistical Association.* 1998; 93:46–53.
13. Liu L, Wolfe RA, Huang X. Shared frailty models for recurrent events and a terminal event. *Biometrics.* 2004; 60:747–756. [PubMed: 15339298]
14. Huang X, Liu L. A joint frailty model for survival and gap times between recurrent events. *Biometrics.* 2007; 63:389–397. [PubMed: 17688491]
15. Bellman, R. *Dynamic Programming.* Princeton University Press; Princeton, NJ: 1957.

16. Lin DY, Sun W, Ying Z. Nonparametric estimation of the gap time distributions for serial events with censored data. *Biometrika*. 1999; 86:59–70.
17. Robins, J.; Rotnitzky, A. Recovery of information and adjustment for dependent censoring using surrogate markers. Birkhäuser; Boston, MA: 1992.
18. Robins, J. Robust estimation in sequentially ignorable missing data and causal inference models. 1999. Proceedings of the American Statistical Association Section on Bayesian Statistical Science; 2000. p. 6-10.
19. Robins J, Greenland S. Adjusting for differential rates of pcp prophylaxis in high- versus low-dose azt treatment arms in an aids randomized trial. *Journal of the American Statistical Association*. 1994; 89:737–749.
20. Wei LJ, Lin DY, Weissfeld L. Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *Journal of the American Statistical Association*. 1989; 84:1065–1073.
21. Robins, JM.; Rotnitzky, A. Recovery of information and adjustment for dependent censoring using surrogate markers. In: Jewell, NP.; Dietz, K.; Farewell, VT., editors. *AIDS Epidemiology: Methodological Issues*. Birkhäuser; Boston: 1992. p. 297-331.
22. Robins, J. Information recovery and bias adjustment in proportional hazards regression analysis of randomized trials using surrogate markers. Proceedings of the Biopharmaceutical Section of the American Statistical Association; 1993. p. 33
23. Heitjan D, Rubin D. Ignorability and coarse data. *The Annals of Statistics*. 1991; 19:2244–2253.
24. Miloslavsky M, Keles S, ver der Laan M, Butler S. Recurrent events analysis in the presence of time-dependent covariates and dependent censoring. *Journal of the Royal Statistical Society, Series B*. 2004; 66:239–257.
25. Breslow N. Discussion of dr. cox's paper “regression models and life tables”. *Journal of the Royal Statistical Society, Series B*. 1972; 34:216–217.
26. Buckley J, James I. Linear regression with censored data. *Biometrika*. 1979; 66:429–436.
27. Tsiatis AA. Estimating regression parameters using linear rank tests for censored data. *Ann Statist*. 1990; 18:354–372.
28. Jin Z, Lin D, Wei L, Ying Z. Rank-based inference for the accelerated failure time models. *Biometrika*. 2003; 90:341–353.
29. Jin Z, Lin DY, Ying Z. On least-squares regression with censored data. *Biometrika*. 2006; 93:147–161.
30. Wahed A, Tsiatis AA. Semi-parametric efficient estimation of the survival distribution for treatment policies in two-stage randomization designs in clinical trials with censored data. *Biometrika*. 2006; 93:147–161.
31. Tierney JF. Adjuvant chemotherapy for localised resectable soft-tissue sarcoma of adults: Meta-analysis of individual data. sarcoma meta-analysis collaboration. *Lancet*. 1997; 350:1647–1654. [PubMed: 9400508]
32. Cormier JN, Huang X, Xing Y, Thall PF, Wang X, Benjamin RS, Pollock RE, Antonescu RG, Rand Maki C, Brennan MF, Pisters PWT. Cohort analysis of 674 patients with localized high-risk extremity soft tissue sarcoma treated at two cancer centers: Chemotherapy-associated outcomes. *Journal of Clinical Oncology*. 2004; 22:4567–4574. [PubMed: 15542808]
33. American Joint Committee on Cancer. *Cancer Staging Manual*. 6th edition. Springer; New York: 2002.
34. Coindre J, Terrier P, Bui N, Bonichon F, Collin F, LeDoussal V, Mandard A, Vilain M, Jacquemier J, Duplay H, et al. Prognostic factors in adult patients with locally controlled soft tissue sarcoma: A study of 546 patients from the French Federation of Cancer Centers Sarcoma Group. *Journal of Clinical Oncology*. Mar; 1996 14(3):869–877. [PubMed: 8622035]
35. Pisters PW, Leung DH, Woodruff J, Shi W, Brennan MF. Analysis of prognostic factors in 1,041 patients with localized soft tissue sarcomas of the extremities. *Journal of Clinical Oncology*. 1996; 14:1679–1689. [PubMed: 8622088]
36. Zagars GK, Ballo MT, Pisters PW, Pollock RE, Patel SR, Benjamin RS, Evans HL. Prognostic factors for patients with localized soft-tissue sarcoma treated with conservation surgery and radiation therapy: An analysis of 225 patients. *Cancer*. 2003; 97:2530–2543. [PubMed: 12733153]

37. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*. 1958; 53:457–481.
38. Lin DY, Psaty BM, Kronmal RA. Assessing the sensitivity of regression results to unmeasured confounders in observational studies. *Biometrics*. 1998; 54:948–963. [PubMed: 9750244]
39. Murphy SA. Optimal dynamic treatment regimes (with discussion). *Journal of the Royal Statistical Society, Series B*. 2003; 65:331–366.

\$watermark-text

\$watermark-text

\$watermark-text

A “natural” modeling approach

Wei, Lin, Weissfeld marginal model

Joint AFT model

$$\frac{\text{PSfrag replacements}}{\frac{ds}{d\phi_1}} \\ \phi_1$$

Figure 1.
Comparison of different modeling approaches for recurrent events.

Figure 2. Kaplan-Meier estimators for the distributions of (a) time from initial treatment to disease recurrence, and (b) time from disease recurrence to death, and (c) time from initial treatment to death.

Table 1

Simulation study: estimation of β and γ under the joint models

N	($C_1\%$, $C_2\%$)	Mean	Empirical SE	Asymptotic SE	95% C.P.
$C \sim \exp(\tau_i)$					
$\beta = (0.5, -1)$					
200	(10%, 45%)	(0.494, -0.981)	(0.143, 0.134)	(0.137, 0.135)	(0.934, 0.953)
400	(10%, 45%)	(0.496, -0.992)	(0.098, 0.095)	(0.097, 0.095)	(0.947, 0.940)
200	(20%, 60%)	(0.486, -0.960)	(0.158, 0.156)	(0.159, 0.158)	(0.954, 0.935)
400	(20%, 60%)	(0.488, -0.975)	(0.110, 0.113)	(0.113, 0.113)	(0.950, 0.939)
$\gamma = (1, 0.5)$					
200	(10%, 45%)	(0.994, 0.498)	(0.090, 0.051)	(0.087, 0.050)	(0.937, 0.943)
400	(10%, 45%)	(0.990, 0.498)	(0.060, 0.034)	(0.062, 0.035)	(0.941, 0.957)
200	(20%, 60%)	(0.981, 0.493)	(0.094, 0.056)	(0.099, 0.057)	(0.941, 0.955)
400	(20%, 60%)	(0.984, 0.494)	(0.069, 0.039)	(0.070, 0.040)	(0.946, 0.942)
$C: h_i(\theta) = \tau_2 t \exp(-Z_{it})$					
$\beta = (0.5, -1)$					
200	(10%, 45%)	(0.490, -0.974)	(0.142, 0.137)	(0.141, 0.140)	(0.951, 0.945)
400	(10%, 45%)	(0.493, -0.984)	(0.100, 0.098)	(0.100, 0.098)	(0.947, 0.946)
200	(20%, 60%)	(0.480, -0.946)	(0.166, 0.164)	(0.166, 0.165)	(0.941, 0.939)
400	(20%, 60%)	(0.481, -0.960)	(0.119, 0.120)	(0.119, 0.118)	(0.950, 0.925)
$\gamma = (1, 0.5)$					
200	(10%, 45%)	(0.984, 0.495)	(0.086, 0.051)	(0.090, 0.052)	(0.955, 0.942)
400	(10%, 45%)	(0.986, 0.497)	(0.063, 0.035)	(0.063, 0.036)	(0.939, 0.954)
200	(20%, 60%)	(0.965, 0.487)	(0.108, 0.061)	(0.111, 0.062)	(0.933, 0.934)
400	(20%, 60%)	(0.972, 0.491)	(0.078, 0.042)	(0.078, 0.043)	(0.934, 0.954)

Table 2

Distributions of initial and salvage chemotherapies and disease recurrence (DR) and death

Total	674 patients			
Initial chemotherapy	yes: 336		no: 338	
Number of patients with DR	179		171	
Mean (median) time to DR	1.74 (1.12)		1.73 (0.82)	
Salvage chemotherapy	yes: 59	no: 120	yes: 42	no: 129
Number of patients died	54	92	35	93
Mean (median) time from DR to death	1.20 (0.88)	1.65 (1.13)	1.51 (1.30)	1.27 (0.88)

Table 3

Joint AFT models: for the time from disease recurrence to death

Covariate	Estimate	Mean lifetime ratio	SE of estimate	P-value
Initial chemotherapy	0.26	1.29	0.17	0.12
Salvage chemotherapy	−0.01	0.99	0.16	0.95
Amputation	−0.35	0.70	0.26	0.18
Upper vs lower limb	0.19	1.21	0.28	0.50
Sex (male vs female)	−0.04	0.96	0.16	0.79
Proximal vs distal tumor site	−0.70	0.50	0.22	< 0.01
Positive vs negative margin	−0.26	0.77	0.25	0.29
Patient age 40–60 vs 40– years	−0.34	0.71	0.20	0.08
Patient age 60+ vs 40– years	−0.63	0.53	0.21	< 0.01
Tumor size 10–15 vs 10– cm	−0.26	0.77	0.18	0.15
Tumor size 15+ vs 10– cm	−0.36	0.70	0.21	0.09
Unclassified sarcoma vs MFH*	−0.40	0.67	0.33	0.23
Synovial sarcoma vs MFH	0.24	1.27	0.21	0.25
Liposarcoma vs MFH	0.48	1.61	0.23	0.04
Leiomyosarcoma vs MFH	−0.11	0.90	0.34	0.75
Other sarcoma vs MFH	−0.01	0.99	0.27	0.96

Remark: MFH=malignant fibrous histiocytoma

Table 4

Joint AFT models: for the overall survival time

Covariate	Estimate	Mean lifetime ratio	SE of estimate	P-value
Initial chemotherapy	0.28	1.32	0.13	0.04
Amputation	−0.01	0.99	0.22	0.98
Upper vs lower limb	−0.07	0.93	0.21	0.72
Sex (male vs female)	−0.16	0.85	0.12	0.21
Proximal vs distal tumor site	−0.42	0.65	0.20	0.03
Positive vs negative margin	−0.14	0.86	0.17	0.38
Patient age 40–60 vs 40– years	−0.33	0.72	0.18	0.05
Patient age 60+ vs 40– years	−0.43	0.65	0.19	0.02
Tumor size 10–15 vs 10– cm	−0.16	0.85	0.15	0.29
Tumor size 15+ vs 10– cm	−0.49	0.61	0.14	< 0.01
Unclassified sarcoma vs MFH*	−0.34	0.70	0.20	0.08
Synovial sarcoma vs MFH	0.09	1.10	0.19	0.62
Liposarcoma vs MFH	0.57	1.77	0.18	< 0.01
Leiomyosarcoma vs MFH	0.08	1.09	0.21	0.69
Other sarcoma vs MFH	−0.19	0.83	0.21	0.37

Remark: MFH=malignant fibrous histiocytoma

\$watermark-text

\$watermark-text

\$watermark-text

Table 5

The point estimates and 95% confidence intervals for the mean lifetime ratio associated with the initial chemotherapy with adjustment for an unmeasured confounder Z_u , with κ being the regression coefficient of Z_u in the AFT model (4), and d the mean difference of Z_u between groups with and without initial chemotherapy. When $\kappa = 0$ or $d = 0$, i.e., no unmeasured confounders, the estimate is 1.32 with 95% confidence intervals (1.03, 1.54).

κ	$\exp(\kappa)$	d			
		0.10	0.20	0.30	0.40
-0.40	0.67	1.38(1.07-1.78)	1.43(1.11-1.85)	1.49(1.16-1.92)	1.55(1.20-2.00)
-0.30	0.74	1.36(1.06-1.76)	1.40(1.09-1.81)	1.45(1.12-1.87)	1.49(1.16-1.92)
-0.20	0.82	1.35(1.05-1.74)	1.38(1.07-1.78)	1.40(1.09-1.81)	1.43(1.11-1.85)
-0.10	0.90	1.34(1.04-1.72)	1.35(1.05-1.74)	1.36(1.06-1.76)	1.38(1.07-1.78)
0.10	1.11	1.31(1.02-1.69)	1.30(1.01-1.67)	1.28(1.00-1.66)	1.27(0.99-1.64)
0.20	1.22	1.30(1.01-1.67)	1.27(0.99-1.64)	1.25(0.97-1.61)	1.22(0.95-1.58)
0.30	1.35	1.28(1.00-1.66)	1.24(0.97-1.61)	1.21(0.94-1.56)	1.17(0.91-1.51)
0.40	1.49	1.27(0.99-1.64)	1.22(0.95-1.58)	1.17(0.91-1.51)	1.13(0.87-1.45)