Published in final edited form as:

Stat Med. 2014 June 30; 33(14): 2363–2378. doi:10.1002/sim.6104.

Optimization of Individualized Dynamic Treatment Regimes for Recurrent Diseases

Xuelin Huang^{a,*}, Jing Ning^a, and Abdus S. Wahed^b

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

^bDepartment of Biostatistics, The University of Pittsburgh, Pittsburgh, PA 15260

Abstract

Patients with cancer or other recurrent diseases may undergo a long process of initial treatment, disease recurrences and salvage treatments. It is important to optimize the multi-stage treatment sequence in this process to maximally prolong patients' survival. Comparing disease-free survival for each treatment stage over-penalizes disease recurrences but under-penalizes treatment-related mortalities. Moreover, treatment regimes used in practice are dynamic, i.e., the choice of next treatment depends on a patient's responses to previous therapies. In this article, using accelerated failure time models, we develop a method to optimize such dynamic treatment regimes (DTRs). This method utilizes all the longitudinal data collected during the multi-stage process of disease recurrences and treatments, and identifies the optimal DTR for each individual patient by maximizing his/her expected overall survival. The application of this method is illustrated using data from a study of acute myeloid leukemia. The optimal treatment strategies for different patient subgroups are identified.

Keywords

Backward induction; Causal inference; Coarsening at random; Optimal treatment sequence; Recurrent events; Survival analysis

1. Introduction

Medical advancements have changed many types of cancer from a rapidly fatal disease to a chronic disease. Many people now live with cancer and have a reasonable quality of life. However, curing cancer remains a challenge. Cancer patients usually undergo a long process of initial treatment, disease recurrences and salvage treatments. The period of survival after a disease recurrence and salvage treatment can vary considerably across patients. Also, there are usually multiple treatment options that can be used for initial and salvage treatments. In this circumstance of multi-stage treatments, for both clinicians and researchers, an important goal is to determine the optimal treatment sequences that can maximally prolong a patient's

overall survival. When analyzing such data to optimize treatments, it is common to use disease-free survival (DFS) to compare treatments at each stage. This approach treats disease recurrence and death equivalently. Considering that patients may survive for a long time after disease recurrence, the DFS approach appears to over-penalize disease recurrences, but under-penalize treatment-related mortalities. Therefore, it can give misleading results for treatment sequence optimization. Statistical methods for dynamic treatment regime optimization are used in this article to overcome this problem. These methods utilize all the longitudinal data collected during the multi-stage process of disease recurrences and treatments, to optimize treatments with the goal of prolonging patients' overall survival.

Multistage treatments for recurrent diseases are inevitably dynamic. That is, the choice of next treatment depends on a patient's responses to previous therapies. Dynamic treatment regimes (DTRs) have long been routinely used in clinics, although not necessarily under the same nomenclature. DTRs are also called adaptive treatment strategies [1, 2]. Research on these topics includes experimental designs to compare treatment sequences [3, 4, 5, 6, 7], estimation of the effects of specified treatment sequences [8, 9, 10, 11], and construction of optimal treatment sequences using randomized or observational data [12, 13]. This article falls into the last category, dealing with optimization of DTRs with overall survival as the primary outcome. The fact that the disease can recur and the treatment course will be altered as a consequence makes this a challenging task.

One may think of applying the Cox proportional hazards model [14] or the accelerated failure time (AFT) model [15] by defining disease recurrences and salvage treatments as time-dependent covariates. However, these approaches cannot be used to compare treatment sequences. This is because the inclusion of intermediate outcomes (disease recurrences) as predictors masks and distorts the effects of the initial treatments on survival [16, P.199]. Most of the work on modeling disease recurrences focuses on either their conditional intensity [17] or their marginal rate [18]. While these are important aspects of disease in medical research, they present difficulties when disease recurrences are subject to dependent censoring by death. Some methods have been proposed to solve these problems [19, 20, 21, 22, 23, 24, 25]. However, these methods cannot be used to construct optimal treatment sequences.

The literature on the optimization of DTRs has been growing. Murphy [12] proposed using backward induction to minimize a regret function. Robins [13] proposed an optimal double-regime structural nested mean model. Moodie, Richardson, and Stephens [26] provided a good description of these complicated optimizations methods for DTRs. Henderson, Ansell, and Alshibani [27] enhanced Murphy's method of regret functions by incorporating model building, checking, and comparison into the estimation process. Chakraborty, Murphy, and Strecher [28] considered inference for non-regular parameters in optimal DTRs. Goldberg and Kosorok [29] extended the above optimization methods to censored data and provided their large-sample properties. More recent developments and their applications include Rosthøj, Keiding, and Schmiegelow [30], Zhang et al [31], Zhao et al [32], Chaffee and van der Laan [33], and Wang et al [34]. Bayesian methods on DTRs are still relatively rare, but Zajonc [35] has made a significant contribution in this area.

Two types of methods, called “Q-Learning” [36, 37] and “A-Learning” [12, 13, 38], are commonly used for the optimization of DTRs. A Q-function is a reward at a certain stage in the multiple decision-making points, and the Q-learning method aims to model and maximize such rewards at the end of the study. The A-learning method directly models the advantages of optimal decisions vs. non-optimal ones, and leave the reward functions unspecified. This may help gain some robustness against reward model mis-specification. But the A-learning method needs to specify treatment selection models, and small mis-specifications in such selection models can accumulate over treatment stages and thus cause severe bias and convergence problems [39]. The Q-learning method avoids the specification of such models, so does not suffer from these problems. Our approach, described in the next section, is a modification of the Q-learning method.

Wahed and Thall [40] reported their analysis of a study for 210 patients with acute myeloid leukemia (AML). For these patients, their front-line treatments (denoted as A_1 to A_4) are randomized, but their salvage treatments (B_1 or B_2), administered after disease resistance or relapse, are chosen by the treating physicians on an individual basis. Wahed and Thall [40] estimated and compared the expected overall survival times under 16 different treatment strategies that utilize the information of patient response to the initial treatment. They used two methods: the inverse probability of treatment weighted (IPTW) method and a parametric approach.

This article provides a different approach for dealing with a more challenging task, that is to consider all the available longitudinal information of each patient, and identify the optimal treatment strategy for each individual from an infinite number of possibilities. The results are further individualized optimal DTRs. During this process, all the categorical and continuous variables longitudinally collected from patients can be used in the treatment decision-making. The decision-rules are obtained by statistical modeling. The methods of Wahed and Thall [40] are appropriate for the estimation of discrete treatment strategies, and are limited to comparisons among a finite number of pre-specified treatment strategies. Our proposed method can be used to construct and propose new treatment strategies, i.e., to discover *new* effective treatment sequences. This is a different task than the evaluation of pre-specified treatment sequences. Most commonly used decision-tree approaches involving Markov chain models cannot be used to identify unknown treatment sequences. They can only be used to make comparisons between specified regimes. However, the optimal DTR may not be among them. In contrast, by including the appropriate variables in the proposed models, our method can identify and optimize DTRs for each individual. The identified optimal DTR may be complicated, such as, for example, “start with A_1 , if time to disease recurrence is less than 6 months, use B_1 for salvage; otherwise use B_2 for salvage”. The optimal DTRs may further depend on other individual characteristics. The decision on which individual characteristics to use is made by model selection, and the cut-off values for variables used in treatment decision-making (such as the 6 months in the above example) are obtained by solving statistical models with estimated parameters.

We describe our approach, and give the estimation method and its properties in Section 2. In Section 3, we apply the proposed method to the AML study of 210 patients. Results from

simulation studies are presented in Section 4. We close with some discussion in Section 5 and provide mathematical details in the Appendix.

2. Optimization of Dynamic Treatment Regimes

We will describe the data and notations in Subsection 2.1, then define our estimation goal in 2.2, state the necessary assumptions and models in 2.3, and provide estimation procedures in Subsection 2.4.

2.1. Data and Notation

To simplify the notation and description, hereafter we assume that the total number of treatment stages is $K = 2$, and that there are only two treatment options for each stage. The methods presented below can be extended to situations with $K > 2$ stages and more than two treatment options for each stage. Practical values for K are usually between 2 and 5 for cancer studies, which provided the motivation for this work.

Denote the stage 1 treatment by $A \in \{0, 1\}$. The time point for this stage 1 treatment is the baseline. We define T as the time from baseline to death. Denote by Z_1 the baseline variables (observed prior to A). For a patient who experiences disease recurrence and receives a salvage treatment, denote the stage 2 treatment by $B \in \{0, 1\}$. Let R be the time from initial treatment A to disease recurrence, and S be the time from salvage treatment B to death. Assuming that the time from disease recurrence to salvage treatment B is sufficiently short and thus ignorable, we have $T = R + S$. Let Z_2 be the covariates observed just prior to the second stage treatment B , and $\bar{Z}_2 = (Z_1', A, Z_2', R)'$. Note that some patients may die without disease recurrence, so for them we observe Z_1, A and T , but not R, Z_2, B , or S . Denote η as the indicator for disease recurrence before death.

2.2. Definition of Optimal Dynamic Treatment Regimes

We use a backward induction method [12] to identify the optimal treatment sequences, using overall survival time as the criterion of optimality. For the initial treatments A , a direct comparison between the effects of $A = 0$ and $A = 1$ on survival is difficult because the survival time is also affected by subsequent salvage treatments after disease recurrences. Ignoring such salvage treatments can lead to bias. To compare $A = 0$ with $A = 1$, a sensible approach is to use the survival times patients would experience if they had received their respective optimal subsequent treatment sequences after their disease recurrences. Under this consideration, the optimal treatment sequences are defined and identified in a backward fashion. That is to say, in the first step, we identify the optimal regimes for the second treatment stage. In the second step, we identify the optimal initial treatment by maximizing the overall survival time while incorporating results from the first step, i.e., the predicted optimal residual survival times from the second stage (details described below). The combination of these two steps gives the optimal treatment sequence.

Start with the second stage. Only those patients with disease recurrence (denoted by $\eta = 1$) are included in this stage. Consider a subject with covariate history $\bar{Z}_2$. The choices for the second stage treatments are $B = 0, 1$. At this time point, a treatment regime is dynamic

meaning that the decision for B depends on Z_2 , which includes baseline covariate Z_1 , initial treatment A , and response to treatment A . Denote a dynamic decision rule for the second stage by $g_2(Z_2)$. We use superscripts to denote potential survival times under specified dynamic regimes. With the dynamic regime $g_2(Z_2)$, the potential remaining survival time after the second stage treatment is denoted by $S^{g_2(Z_2)}$. Then consider the two potential regimes $g_2(Z_2) = 1$ and $g_2(Z_2) = 0$, and the two corresponding potential survival times, S^1 and S^0 , respectively. We define

$$\gamma_2^*(\bar{Z}_2) = E[\log\{S^1\} - \log\{S^0\} | \eta=1, \bar{Z}_2] \quad (1)$$

as the difference between mean residual lifetimes in log scale between these two competing options. This means that the optimal treatment regime for the second stage is $g_2(\bar{Z}_2) = 1$ if $\gamma_2^*(\bar{Z}_2) > 0$, and $g_2(\bar{Z}_2) = 0$, otherwise. We emphasize that this result is conditional on all the information observed prior to the second stage treatment, including the first stage treatment A and baseline covariates Z_1 . That is to say, patients with different initial treatments may have different optimal salvage treatments. Keeping in mind that Z_2 includes A and Z_1 as components, we describe the second stage treatment optimization rule as below,

$$g_2^{\text{opt}}(\bar{Z}_2) = \begin{cases} 1 & \text{if } \gamma_2^*(\bar{Z}_2) > 0, \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

Now we consider the stage 1 (initial) treatment optimization. Suppose we have a patient with baseline covariate Z_1 . The choices for the initial treatment are $A = 0, 1$. Denote a dynamic decision rule for the first stage by $g_1(Z_1)$, or simply g_1 for notational simplicity. As described earlier, the comparison for stage 1 treatments should be based on a hypothetical scenario in which each patient with disease recurrence receives his/her optimal treatment for stage 2, namely $g_2^{\text{opt}}(\bar{Z}_2)$. Under this scenario, denote the potential overall survival time after the initial treatment by $T^{(g_1, g_2^{\text{opt}}(\bar{Z}_2|g_1))}$. Here, we use the notation $(Z_2|g_1)$ to emphasize that the optimal second stage treatment depends on the initial treatment g_1 . Since $T = R + S$, we have that

$$T^{(g_1, g_2^{\text{opt}}(\bar{Z}_2|g_1))} = R^{g_1} + S^{g_2^{\text{opt}}(\bar{Z}_2|g_1)}. \quad (3)$$

Certainly for patients without disease recurrence, the second stage treatment is irrelevant. With this understanding, we still use the above simplified notation $T^{(g_1, g_2^{\text{opt}}(\bar{Z}_2|g_1))}$ to denote their potential overall survival times. Then we define that,

$$\gamma_1^*(Z_1) = E \left[\log\{T^{(g_1=1, g_2^{\text{opt}}(\bar{Z}_2|g_1=1))}\} - \log\{T^{(g_1=0, g_2^{\text{opt}}(\bar{Z}_2|g_1=0))}\} \middle| Z_1 \right] \quad (4)$$

The above expectation is taken with respect to the distributions of Z_2 and T , conditional on Z_1 . By this definition, the optimal initial treatment regimes are as follows,

$$g_1^{\text{opt}}(Z_1) = \begin{cases} 1 & \text{if } \gamma_1^*(Z_1) > 0, \\ 0 & \text{otherwise.} \end{cases} \quad (5)$$

Then the optimal personalized dynamic regime is: use $g_1^{\text{opt}}(Z_1)$ as the initial treatment; if the disease recurs, use $g_2^{\text{opt}}(\bar{Z}_2)$ as the salvage treatment.

2.3. Identification and AFT models

To identify the optimal dynamic treatment regimes, we make the following assumptions.

- A.1** Independence between subjects.
- A.2** No unmeasured confounders—that is, treatment assignments are independent of future outcomes given observed history. In longitudinal settings this assumption is usually referred to as “sequential ignorability” [41].
- A.3** Consistency: for any patient who has followed a specific treatment regime, then the potential outcome for this patient under this specific treatment regime should be equal to the observed outcome of this patient [42].

To operationalize the optimization, we assume some parametric models for $\gamma_2^*(\bar{Z}_2)$ and $\gamma_1^*(\bar{Z}_1)$ in (1) and (4), respectively. Note that under A.1 – A.3, we can rewrite equation (1) as

$$\gamma_2^*(\bar{Z}_2) = E\{\log(S) | B=1, \eta=1, \bar{Z}_2\} - E\{\log(S) | B=0, \eta=1, \bar{Z}_2\}. \quad (6)$$

Then, we assume the following AFT model for estimating the stage 2 (salvage) treatment effect,

$$\log(S) = \bar{Z}_2' \beta_{21} + B \bar{Z}_2' \beta_{22} + \varepsilon_2, \quad (7)$$

where the distribution of ε_2 is unspecified except that it has mean zero, and ε_2 is independent and identically distributed (i.i.d.) across subjects. Hereafter, denote $\beta_2 = (\beta_{21}', \beta_{22}')'$. Note that in the above equation, the main effect for B is absorbed into $B \bar{Z}_2' \beta_{22}$ by letting $\bar{Z}_2$ have an intercept. Also, note that B is a scalar, so there are no matrix conformity problems here. The inclusion of an interaction term between salvage treatment B and covariates $\bar{Z}_2$ allows us to identify the optimal salvage treatment for each individual. The AFT model (7) implies that the function $\gamma_2^*(\bar{Z}_2)$ in (1) is equal to

$$\gamma_2(\bar{Z}_2) = \bar{Z}_2' \beta_{22}. \quad (8)$$

For a patient who did not receive his/her optimal second stage treatment, i.e., $B \neq g_2^{\text{opt}}(\bar{Z}_2)$, using equations (6) and (8), we can predict as below his/her potential survival time under the optimal second stage treatment, which is denoted as $S^{g_2^{\text{opt}}(\bar{Z}_2|g_1)}$ in (3). Since this prediction depends on β_2 , we denote it as $H_2(\beta_2)$.

$$\log(S) + |\gamma_2(\bar{Z}_2)| I\{B \neq g_2^{\text{opt}}(\bar{Z}_2)\} \triangleq H_2(\beta_2). \quad (9)$$

We also apply the above equation to patients who actually received their optimal second stage treatments. For them, the potential survival times are equal to the observed ones, i.e., $S^{g_2^{\text{opt}}(\bar{Z}_2|g_1)} = \exp\{H_2(\beta_2)\} = S$. This satisfies the consistency assumption A.3. This $H_2(\beta_2)$ is a modification from standard Q-learning method, which substitutes the actual treatment B in model (7) by $g_2^{\text{opt}}(\bar{Z}_2)$ to obtain $H_2^Q(\beta_2) = \bar{Z}_2' \beta_{21} + g_2^{\text{opt}}(\bar{Z}_2) \bar{Z}_2' \beta_{22}$. This $H_2^Q(\beta_2)$ is a deterministic function of predictors Z_2 , and loses the randomness in the observed outcome S . It does not satisfy the consistency assumption A.3 because $\exp\{H_2^Q(\beta_2)\}$ may not be equal to S when a patient actually received his/her optimal second stage treatment. Regarding this point, our method offers an advantage. Moreover, by retaining the randomness of S in the estimation of $S^{g_2^{\text{opt}}(\bar{Z}_2|g_1)}$, our method should be more robust against mis-specification of model (7).

Now consider the optimization of the stage 1 treatment. As elaborated earlier, to optimize treatment for the first stage, the criterion is to use the potential overall survival time under a situation in which if a patient experience a disease recurrence, then he/she would receive his/her optimal treatment for stage 2. For a patient who received initial treatment A , the potential overall survival time under this situation is denoted by $T^{(g_1=A, g_2^{\text{opt}}(\bar{Z}_2|g_1=A))}$ in (3). It can be predicted by $R + \exp\{H_2(\beta_2)\}$. If a patient does not experience disease recurrence, then the second stage treatment is irrelevant, and we set $T^{(g_1=A, g_2^{\text{opt}}(\bar{Z}_2|g_1=A))}$. Then, in general, we can predict $T^{(g_1=A, g_2^{\text{opt}}(\bar{Z}_2|g_1=A))}$ by

$$\eta[R + \exp\{H_2(\beta_2)\}] + (1 - \eta)T \triangleq \exp\{H_1(\beta_2)\}. \quad (10)$$

In order to identify g_1^{opt} based on definition (4), we impose another AFT model for the random variable $H_1(\beta_2)$ as

$$H_1(\beta_2) = Z_1' \beta_{11} + AZ_1' \beta_{12} + \varepsilon_1, \quad (11)$$

where the distribution of ε_1 is unspecified except that ε_1 has mean zero and is i.i.d. across subjects. Hereafter, denote $\beta_1 = (\beta_{11}', \beta_{12}')'$. By the above AFT model, the function $\gamma_1^*(\bar{Z}_2)$ in (4) is equal to

$$\gamma_1(Z_1) = Z_1' \beta_{12}. \quad (12)$$

The correct decisions for g_1 and g_2 depend on the AFT models in (7) and (11) correctly specifying $\gamma_1(\cdot)$ and $\gamma_2(\cdot)$ for $\gamma_1^*(\cdot)$ and $\gamma_2^*(\cdot)$, respectively, in (4) and (1).

2.4. Parameter Estimation

Parameter estimation needs to account for right-censored survival time data. For this we use the inverse probability of censoring weighted (IPCW) method [43], and assume coarsening at random. This concept was formulated by Heitjan and Rubin [44]. Miloslavsky et al [45] specified it for the situation of recurrent events. In our situation, with both recurrent events and death, the assumption is specified as $\lambda\{t|D(\bar{v})\} = \lambda\{t|D(\bar{t})\} = \lambda\{t|D(\bar{t}), T - t\}$, where $\lambda(\cdot)$ is the hazard function of censoring time C , $D(\bar{v})$ represents all the other data (R, S, T, Z_1, A, Z_2 , and B) up to the end of study v , and $D(\bar{t})$ represents the information about D up to the time t . We assume a Cox proportional hazards model [14] for the hazard function of C_i , as follows,

$$\lambda_i\{t|\bar{D}_i(t)\} = \lambda_0(t) \exp\{W_i'(t)\psi\}, \quad (13)$$

where $W_i(t)$ is a summary statistic from $D_i(\bar{t})$, ψ is an unknown parameter, and $\lambda_0(\cdot)$ is a non-specified hazard function. We maximize the partial likelihood to obtain estimator $\hat{\psi}$ for ψ , and obtain the Breslow [46] estimator $\hat{\Lambda}_0(t)$ for $\Lambda_0(t) = \int_0^t \lambda_0(s) ds$. Then we estimate the probability $G_i\{t|D_i(\bar{t})\} = \Pr\{C_i \leq t|D_i(\bar{t})\}$ by $\hat{G}_i\{t|D_i(\bar{t})\} = \exp[-\int_0^t \exp\{W_i'(s)\hat{\psi}\} d\hat{\Lambda}_0(s)]$.

Following the above preparation, we estimate the parameters in (7) and (11). The IPCW estimation method proceeds as follows. Denote $T_i^* = \min(T_i, C_i)$, $\delta_i = I(T_i \leq C_i)$ and $\omega_i = \delta_i / \hat{G}(T_i^* | D_i(T_i^*))$. Then, for stage 2, we can estimate β_2 in (1) by solving the following equation,

$$U_2(\beta_2) = \sum_{i=1}^n \omega_i \eta_i \left(\frac{\bar{Z}_{2i}}{B \bar{Z}_{2i}} \right) \{\log(S_i) - \bar{Z}_{2i}' \beta_{21} - B \bar{Z}_{2i} \beta_{22}\} = 0. \quad (14)$$

After we obtain the estimator $\hat{\beta}_2$, the estimated rule of optimal stage 2 treatment regimes for subject i is

$$\hat{g}_2^{\text{opt}}(Z_{2i}) = \begin{cases} 1 & \text{if } \bar{Z}_{2i}' \hat{\beta}_{22} > 0, \\ 0 & \text{otherwise.} \end{cases} \quad (15)$$

This estimated rule and the estimator $\hat{\beta}_2$ are plugged in equations (9) and (10) to compute $H_1(\hat{\beta}_2)$. Applying the IPCW estimation techniques, we solve the following estimating equation for stage 1,

$$U_1(\beta_1) = \sum_{i=1}^n \omega_i \left(\frac{\bar{Z}_{1i}}{A \bar{Z}_{1i}} \right) [\log\{H_{1i}(\hat{\beta}_2)\} - Z_{1i}' \beta_{11} - A Z_{1i}' \beta_{12}] = 0. \quad (16)$$

After we obtain the estimator $\hat{\beta}_1$, the estimated rule of optimal stage 1 treatment regimes for subject i is

$$\hat{g}_1^{\text{opt}}(Z_{1i}) = \begin{cases} 1 & \text{if } Z'_{1i} \hat{\beta}_{12} > 0, \\ 0 & \text{otherwise.} \end{cases} \quad (17)$$

The consistency and asymptotic normality of the resulting estimators $\hat{\beta}_1$ and $\hat{\beta}_2$ can be established under some regularity conditions by arguments similar to those in Huang and Ning [47]. See the Appendix for details of the regularity conditions. Note also that the estimated parameters in model (7) for the salvage treatment are plugged into equations (9) and (10) to construct counterfactual times. Due to these plug-ins, the variance formulae for the resulting estimators are not straightforward. We derive them by using the so-called “sandwich formula” [48], and provide them in the Appendix.

In this subsection, we have used a simple method for estimating the parameters for the proposed models. Other more efficient and sophisticated methods are available. However, our focus in this article is not estimation efficiency, but the construction of models on the counterfactual outcomes, as done in the previous subsection. Therefore, we use this simple method of inference based on inverse-probability-weighting.

We assume that all the potential confounders for the effects of treatment B on S are included in Z_2 . This is from assumption A.2 as listed in Subsection 2.3. When model (7) is correctly specified, the coefficient β_{22} estimates the causal effects of B on S , which can vary across different levels of Z_2 . Since the initial treatment indicator A is included in Z_2 , the optimal salvage treatment option depends on the initial treatment. The validity of using regression models, as in (7), to perform causal inference has been demonstrated by many authors, for example, Gelman and Hill [49, Chapter 9]. Tan [50] showed that when model (7) is correctly specified, it is more efficient than other methods, such as using estimating equations weighted by the inverse probability of treatment.

To ensure that the mean survival time is finite, we consider survival time restricted to L , e.g., $L = 10$ years. Here, L is chosen for convenience, but is usually slightly lower than the observed maximum follow-up time in the data set. In this case, for patients with $T > L$ with either $\delta = 0$ or $\delta = 1$, we redefine $T = L$ and $\delta = 1$. This also ensures that the probability of not being censored is always larger than a small positive constant so that its inverse is bounded above, and hence the condition C.3 in the Appendix is satisfied. This technique has been used for the same purpose by many authors, e.g., Robins and Rotnitzky [43], Zhao and Tsiatis [51], in their applications of IPCW estimators. Restriction of survival time to L usually does not affect much the analysis results because L is only slightly smaller than the maximum follow-up time so that only a few observations are slightly truncated, and all the other survival times remain the same.

3. Data Analysis for a Leukemia Study

We apply our method to optimize the two-stage treatments of acute myeloid leukemia (AML) [52]. The process of initial (induction) and salvage treatments is illustrated in Figure 1. A total of 210 patients with AML were equally randomized to receive one of the four therapies as their initial induction treatments, FAI, FAIA, FAIG and FAIGA. FAI is a

combination of fludarabine, cytosine arabinoside (Ara-C) and idarubicin; FAIA is that combination plus all-trans retinoic acid (ATRA); FAIG is the FAI combination plus granulocyte colony stimulating factor (G-CSF); and FAIGA is the FAI combination plus both ATRA and G-CSF. Denote the respective indicators for these four treatments by A_1 , A_2 , A_3 and A_4 . Figure 1 shows all the possible pathways of successive states, transition times and salvage therapy following these induction treatments. Patients may be resistant to their initial treatment. They may also respond to their initial treatments, achieve complete remission (CR), but then subsequently experience disease progress. In the analysis below, the time from initial treatment to experiencing treatment resistance or disease progression is denoted as the disease recurrence time, R , in the previous section. After treatment resistance or disease progression, patients received either high-dose Ara-C (HDAC, denoted by $B = 1$) or other therapies ($B = 0$) as their salvage treatment. These salvage treatments were not randomized. The time from salvage treatment to death is denoted by S , as in the previous section. At the end of the study, the death times of 198 patients were observed, and 12 patients were still alive, and hence their survival times were censored.

Data analysis for such sequential treatment assignment commonly proceeds as follows. First, a multinomial logistic regression model is assumed for the possibility during the initial treatment phase of a patient experiencing death, or treatment resistant, or CR. Then, further models are assumed for the transition times. The problem of this modeling approach is that it cannot directly answer a critical question: Which initial treatment is the best for prolonging the overall patient survival time? Note that the treatment that achieves the highest CR rate may not be the one associated with the best overall survival time because survival is affected by many other factors, including CR duration, and survival time after salvage treatment. A marginal regression of overall survival time on the initial treatments may also be invalid because it ignores the effects of non-randomized salvage treatments. A Cox proportional hazards model (Tang and Wahed [53]) with initial treatments, disease status (treatment resistance, CR, disease progression) and salvage treatments as covariates (some are time-dependent) cannot select the best initial treatment for overall survival either, due to the inclusion of intermediate outcomes (disease status) as predictors in the model.

Wahed and Thall [40] provided two ways to estimate the mean survival time associated with different treatment sequences. This article provides an approach to identify the optimal dynamic treatment regimes for each individual patient. The goal is to utilize all the longitudinal information of each individual to discover the best two-stage treatment strategies.

As described in the previous section, we use backward induction to identify individualized optimal treatment regimes. We start with the last treatment stage, which is the second in the above case for AML patients. To identify the optimal salvage treatments, we need to identify important covariates for Z_2 in model (7). For leukemia patients, the most important variables for prognosis are patient age and cytogenetics. Physicians have classified cytogenetics as poor, intermediate or favorable, but we have only the first two types in this study as a consequence of the enrollment criterion. Other variables include time to resistance (R), time to CR, and time from CR to disease progression. Interactions between the above covariates and the salvage treatment indicators are included in the model. Principles (a)

through (d), as follows, are implemented in the derivation of the the above model. (a) Generally a term is included if its p -value < 0.10 . (b) If an interaction term is included by (a), then both of the marginal terms are included regardless of their p -values. (c) Grouped dummy variables (such as treatment indicators) are all included, as long as one of them satisfies (a). (d) As an exception, even though the p -value > 0.10 for cytogenetics, it is still included because usually it is a very important prognostic variable for AML patients. The parameter estimates are reported in Table 1.

Based the results in Table 1, the optimal salvage treatments are as follows. The optimal salvage treatment for patients who achieved CR and then experienced disease progression is a salvage treatment other than HDAC. For patients who were resistant to their initial therapies, the optimal salvage treatment is HDAC if $R = 0.243$ years = 89 days; otherwise, it is another salvage treatment. Table 2 shows the computation of the potential survival times if each patient had received his/her optimal salvage treatment.

To optimize the initial treatments, we use a model, as shown in (11), and include the initial treatments, cytogenetics, age, and their interactions. After some model selection, the final fitted model is reported in Table 3, and comparisons of the effects of initial treatments on overall survival are presented in Table 4.

With these results, the optimal strategies for the initial treatments are as follows. Among the four treatment options, FAIA seems to be the best for patients with intermediate cytogenetics, and FAI seems to be the best for patients with poor cytogenetics; however, none of these differences is statistically significant.

4. Simulation Studies

We conduct simulation experiments to evaluate the accuracy and precision of our estimators and associated inference. We study two treatment stages and generate data according to the AFT models in Section 2. The experiment is repeated 500 times, with two sample sizes: $n = 200, 400$.

A commonly used natural approach for two stage survival data is to use simple models to generate R and S (times from initial treatment to disease recurrence, and from disease recurrence to death, respectively), and then the overall survival time is $T = R + S$. However, this may result in a complicated model for T and make it difficult to optimize the initial treatment to maximize T . We instead specify an AFT model for S , and an AFT model for the potential overall survival time T under optimized second stage treatment, but leave the model for R unspecified.

Denote $\text{logit}(p) = \log\left(\frac{p}{1-p}\right)$ for any $p \in (0, 1)$. For patient i , the initial treatment is assigned through a logistic model $\text{logit}\{P(A_i = 1)\} = -0.5 + Z_{1i}$, where Z_{1i} is a continuous baseline covariate generated from a Uniform (0,1) distribution. Similarly, the assignment of the salvage treatment is based on a logistic model $\text{logit}\{P(B_i = 1)\} = -0.5 + Z_{2i}$, where Z_{2i} is a binary covariate observed at disease recurrence generated with $P(Z_{2i} = 1) = 0.6$. We first use the AFT model in (11) to generate

$$U_i = \exp(Z_{1i} + 0.5A_i + \varepsilon_{1i}), \quad (18)$$

where ε_{1i} has a Uniform($-0.5, 0.5$) distribution. Then, we designate 75% of patients ($0.75n$ are chosen) to experience disease recurrence. For patients without disease recurrence, we set $T_i = U_i$. For the patients experiencing disease recurrence, we decompose U_i into R_i and S_i , as shown below, to ensure that the model (11) is satisfied. We first generate their disease recurrence time R_i from a Uniform distribution on $(0.1U_i, 0.9U_i)$. Then we set $S_{0i} = U_i - R_i$ and generate their residual survival time after disease recurrence using the AFT model (7) as shown below,

$$S_i = S_{0i} \exp(\beta_0 + 0.5Z_{2i} - 0.5B_i + Z_{2i}B_i), \quad (19)$$

This corresponds to the parameter value $\beta_2 = (0.5, -0.5, 1)$ (intercept omitted) in model (7). For these patients with disease recurrence, the overall survival time $T_i = R_i + S_i$. The above model for S_i indicates that the optimal salvage treatment for patient i is

$$g_2^{\text{opt}}(Z_{2i}) = \begin{cases} 1 & \text{if } Z_{2i} = 1, \\ 0 & \text{if } Z_{2i} = 0. \end{cases} \quad (20)$$

With this optimal salvage treatment, a patient's survival time after disease recurrence is $S_i^{g_2(Z_{2i})} = S_{0i} \exp(\beta_0 + 1)$ if $Z_{2i} = 1$, and $S_i^{g_2(Z_{2i})} = S_{0i} \exp(\beta_0)$ if $Z_{2i} = 0$. The value of the intercept is chosen as $\beta_0 = -0.7085$ to satisfy the constraint $E\{S_i^{g_2(\bar{Z}_{2i})}\} = S_{0i}$, where the expectation is taken with respect to the distributions of Z_{2i} , as defined in equation (4). Then we take the same expectation for the overall survival time under optimized salvage treatments, which is $T_i^{(A_i, g_2(Z_{2i}))} = R_i + S_i^{g_2(Z_{2i})}$ as denoted in (3). Since $E\{S_i^{g_2(\bar{Z}_{2i})}\} = S_{0i}$ and $S_{0i} = U_i - R_i$, we have that,

$$E(T_i^{(A_i, g_2(Z_{2i}))} | Z_{1i}, A_i) = E(U_i) = E\{\exp(Z_{1i} + 0.5A_i + \varepsilon_{1i})\}. \quad (21)$$

This shows that the model in (11) is satisfied with respect to its mean structure, and the parameter $\beta_1 = (1, 0.5)$ with the intercept omitted. It is obvious from equation (21), the optimal initial treatment is $A_i = 1$ for all patients.

We simulate two scenarios for generating censoring times. In the first scenario, the censoring times are independently generated from the exponential distribution $\exp(\tau_1)$, where τ_1 is the parameter to control the censoring percentages for disease recurrence ($C_1\%$) and death ($C_2\%$). Under this setting, the Kaplan-Meier estimator is used to estimate the censoring time distribution. In the second scenario, the censoring times are dependent on the baseline covariate Z_1 through a Cox proportional hazards regression model $\lambda_i(t) = \tau_2 t \exp(-Z_{1i})$, where τ_2 is the parameter to control the degree of censoring. In this scenario, the Cox model is used to estimate the distribution of censoring times, which is plugged into the estimating equations (16) and (14).

Table 1 lists the mean of the estimates, their empirical standard deviation (empirical S.E.), the mean of the estimated asymptotic standard errors (asymptotic S.E.), and the coverage probability (C.P.) of the 95% confidence intervals. The estimates for the intercept terms in β_1 and β_2 are omitted since they were irrelevant for treatment decision-making. The proposed estimators perform well in all the scenarios with various degrees of censoring. The relative biases are small, the empirical and asymptotic SE's match well, and the coverage probabilities of the confidence intervals are close to 95%. More importantly, among all the simulations, the correct optimal dynamic treatment regimes (DTRs) are identified all the time (i.e., 100% of the 500 simulation runs). Certainly, this depends on the magnitude of the regression coefficients, the sample size, and the rates of events of interest and censoring. In the above simulations, the sample sizes are moderate. The regression coefficients, the rates of events and the rates of censoring are all reasonable. Therefore, these simulations indicate that the proposed method is practically useful.

5. Discussion

Dynamic treatment regimes (DTRs) reflect natural process of treating chronic diseases and are routinely used by physicians knowingly or unknowingly. Most of these DTR plans are based on personal or group experience, so they may vary from physician to physician and from hospital to hospital, and hence may not be optimal. Although a systematic optimization of the DTRs used in clinical practice is highly desirable, it poses immense challenge for statisticians due to the complex nature of such dynamic processes. There has been some seminal statistical work on the optimization of DTRs. However, easily manageable statistical methods and corresponding computer software are lacking. In this article, we provide a readily implementable statistical method for the optimization of DTRs for recurrent diseases. The criterion of optimality can be overall survival time, or quality-adjusted survival time (to penalize too many disease recurrences and/or high-grade toxicities), or some other combined end point. AFT models on counterfactual survival times have been developed for this purpose.

Model checking is important to assess goodness-of-fit. Since we use AFT models, the traditional methods for checking these models can be applied [54]. As for model building, when dealing with a large number of covariates, it is important to reduce the dimensions to obtain practical models. Many methods address this critical issue, including subset selection (forward, backward, or stepwise), principal component analysis, non-negative garrote and ridge regression. Recent developments include the least absolute shrinkage and selection operator (LASSO [55]), selection by a smoothly clipped absolute deviation penalty (SCAD [56]), elastic net [57], and the Dantzig selector [58], among others. There are also methods proposed particularly for AFT models [59, 60, 61, 62, 63, 64]. These methods can be used for this type of model building.

It is also important to conduct sensitivity analysis regarding model mis-specification. To do that, we may use our proposed method to identify a few “optimal” DTRs under different assumed models. Then the final outcomes under these DTRs can be evaluated and compared using the methods provided by Huang et al [65] and Murphy et al [11]. These methods do not need to assume AFT models or any other models for the outcomes. Instead, they assume

models on the censoring process and use the inverse probability of censoring weighted method. With these approaches, we can perform sensitivity analyses for the results obtained from the proposed models. Certainly, no methods can be truly assumption-free, but we they might be useful for sensitivity analyses.

For a K -stage study with a large K , although the proposed method should work in theory, many practical problems may arise and an optimal answer may not be easily obtained. One of the problems is that, as the number of stages increases, the subsample size within each specific treatment regime will decrease rapidly. A common approach to avoid such a curse of dimensionality is to use some Markov assumptions. For example, the effects of treatment at each stage k may depend on information from the stage $k - 1$, but not from any further previous stages $< k - 1$. Even with this assumption, methods for studies with an extremely large number of stages are not straightforward and thus need further research. One needs to borrow information across different treatment stages. Moreover, instead of completely relying on large sample theory for inferences, small sample size techniques can be applied, when appropriate, to obtain more accurate results. For example, Strawderman [66] provided higher-order asymptotic approximation using Laplace, saddlepoint, and related methods. It would be helpful to implement these techniques for the proposed modeling framework.

Finally, we note that our approach does not directly specify a model for T in general. It only specifies a model for T in the cases when the second stage treatments are optimized. This is different from the usual ways of modeling survival times to disease recurrence and death. The usual ways are to specify a model for R (time to disease recurrence), and specify another model for S (time from disease recurrence to death). Then the overall survival time is $T = R + S$. While this method is natural, it may result in a complicated model for T , and make the optimization of the first stage treatment A difficult. In contrast, we specify two AFT models in (7) and (11), respectively, for S and the counterfactual T under the optimized second stage treatment. It is straightforward to directly use these models to optimize the salvage and initial treatments. The model for R is completely unspecified. There are many possible models for R such that both AFT models in (7) and (11) will hold. However, leaving the model for R unspecified makes it inconvenient for us to conduct simulation studies, as can be seen in Section 4. Nevertheless, the purpose of a statistical model is not to make simulation conduct convenient, but to be applied to data analysis, statistical inference and decision making. Our proposed approach serves this purpose well. Data sets similar as that in Subsection 2.1 are commonly seen in biomedical research. With such a data set, it is very simple and convenient to estimate the optimal dynamic treatment regime for each individual by following the steps in Section 2. Sections 3 and 4 have demonstrated this procedure by simulated and real data sets.

Acknowledgments

This research was supported by the U.S.A. National Institutes of Health and National Cancer Institute grants U54 CA096300, U01 CA152958, P50 CA100632 and 5PO1 CA055164. We are grateful to our colleague, Peter Thall, for bringing this research topic and the AML data set to us. We appreciate the helpful comments of two anonymous reviewers that have led to significant improvement of this article.

Appendix

To simplify the notations, for equation (7), denote $X_2 = (\bar{Z}_2', B\bar{Z}_2')'$, and $\beta_2 = (\beta_{21}', \beta_{22}')'$. Then we have,

$$\log(S) = X_2' \beta_2 + \varepsilon_2. \quad (1)$$

Denote $X_1 = (Z_1', AZ_1')'$, and $\beta_1 = (\beta_{11}', \beta_{12}')'$. Then we rewrite model (11) as shown below,

$$H_1(\beta_2) = X_1' \beta_1 + \varepsilon_1. \quad (2)$$

Assume that $\beta_{01} = 0$ and $\beta_{02} = 0$ are the true values of β_1 and β_2 , respectively. Assume the following regularity conditions.

- (C.1) X_1 and X_2 are bounded; if there exists a constant vector C_1 such that $\Pr(C_1' X_1 = 0) = 1$, then $C_1 = 0$. A similar condition is assumed for X_2 (to avoid collinearity).
- (C.2) $\Pr(R < T) > 0$.
- (C.3) There exists a constant $\tau_c > 0$ such that $\Pr(C > T) > \tau_c$ (to ensure that the IPCW estimator is valid).
- (C.4) $\Gamma_{\beta_1} = E \left\{ \frac{\delta_i (X_{1i} - \bar{X}_1) X_{1i}'}{G_i(T_i)} \right\}$ and $\Gamma_{\beta_2} = E \left\{ \frac{\eta_i \delta_i (X_{2i} - \bar{X}_2) X_{2i}'}{G_i(T_i)} \right\}$ exist (i.e., expectations $< \infty$) and are nonsingular.
- (C.5) The asymptotic variance-covariance matrices of two estimating equations, Σ_{β_1} and Σ_{β_2} , exist.

If regularity conditions C.1–C.5 are satisfied, $\hat{\beta}_1$ and $\hat{\beta}_2$ are consistent estimators of β_1 and β_2 , respectively. Moreover, $\sqrt{n}(\hat{\beta}_1 - \beta_{01}) \rightarrow MVN \left\{ 0, \Gamma_{\beta_1}^{-1} \Sigma_{\beta_1} (\Gamma_{\beta_1}^{-1})' \right\}$ and $\sqrt{n}(\hat{\beta}_2 - \beta_{02}) \rightarrow MVN \left\{ 0, \Gamma_{\beta_2}^{-1} \Sigma_{\beta_2} (\Gamma_{\beta_2}^{-1})' \right\}$, where MVN is the multivariate normal distribution, and estimators for Σ_{β_1} and Σ_{β_2} are provided below.

Here, we provide the variance-covariance formulae for $\hat{\beta}_1$ and $\hat{\beta}_2$ under the assumptions that the censoring time is independent of the covariates. The arguments can be easily generalized to a setting with covariate-dependent censoring.

Asymptotic distribution of $\hat{\beta}_2$

Suppose $S_T(s) = \Pr(T > s)$, $G(s) = \Pr(C > s)$, and that $\Lambda_c(u)$ is the cumulative hazard function of the censoring times C . Denote

$$\begin{aligned}
M_i^c(s) &= I(T_i^* \leq t, \delta_i=0) - \int_0^s I(T_i^* \geq u) d\Lambda_c(u), \\
D(s) &= E \left\{ \frac{I(T_i \geq s) \eta_i \delta_i X_{2i} (\log S_i - X'_{2i} \beta_{02})}{G(T_i)} \right\}, \\
\Upsilon_i &= \frac{\eta_i \delta_i X_{2i}}{G(T_i)} (\log S_i - X'_{2i} \beta_{02}) + \int_0^\infty \frac{D(s)}{G(s) S_T(s)} dM_i^c(s), \\
\Sigma_{\beta_2} &= E\{\Upsilon_i^{\otimes 2}\}.
\end{aligned}$$

Then we have that $\sqrt{n}(\hat{\beta}_2 - \beta_{02})$ converges weakly to a normal distribution with mean zero and variance-covariance matrix $\Gamma_{\beta_2}^{-1} \Sigma_{\beta_2} (\Gamma_{\beta_2}^{-1})'$, with Γ_{β_2} given in condition (C.4) of the theorem, and the matrix Σ_{β_2} can be estimated by $n^{-1} \sum_{i=1}^n \hat{\Upsilon}_i^{\otimes 2}$, obtained by replacing $G(s)$, $S_T(s)$ and other terms in Υ_i by their respective estimators.

Asymptotic distribution of $\hat{\beta}_1$

Recall that β_{22} is a component of β_2 . Suppose $\Gamma_{\beta_{22}}^{-1}$ is the submatrix of $\Gamma_{\beta_2}^{-1}$ corresponding to the parameter β_{22} , and the true value of β_{22} is β_{022} . Denote

$$\begin{aligned}
C_{\beta_{22}} &= E \left\{ \frac{\eta_i \delta_i X_{1i} (-B_i + \text{sign}(\bar{Z}'_{2i} \beta_{022})) S_i e^{-B_i \bar{Z}'_{2i} \beta_{022} + (\bar{Z}'_{2i} \beta_{022})_+}}{G(T_i) \left(R_i + S_i e^{-B_i \bar{Z}'_{2i} \beta_{022} + (\bar{Z}'_{2i} \beta_{022})_+} \right)} \right\} \\
D_{\beta_{22}}(s) &= E \left\{ \frac{\eta_i \delta_i X_{1i} \log \left(R_i + S_i e^{-B_i \bar{Z}'_{2i} \beta_{022} + (\bar{Z}'_{2i} \beta_{022})_+} \right) I(s \leq T_i)}{G(T_i)} \right\} \\
D_{\beta_1}(s) &= E \left\{ \frac{(1 - \eta_i) \delta_i X_{1i} (\log T_i - X'_{1i} \beta_{01}) I(s \leq T_i)}{G(T_i)} \right\}, \\
\Phi_i &= \frac{\eta_i \delta_i X_{1i}}{G(T_i)} \left\{ \log \left(R_i + S_i e^{-B_i \bar{Z}'_{2i} \beta_{022} + (\bar{Z}'_{2i} \beta_{022})_+} \right) - X'_{1i} \beta_{01} \right\} \\
&\quad + \frac{(1 - \eta_i) \delta_i X_{1i}}{G(T_i)} (\log T_i - X'_{1i} \beta_{01}) + C_{\beta_{22}} \Gamma_{\beta_{22}}^{-1} \Upsilon_i \\
&\quad + \int_0^\infty \left[\frac{\{D_{\beta_{22}}(s) + D_{\beta_1}(s)\} dM_i^c(s)}{G(s) S_T(s)} \right]. \\
\Sigma_{\beta_1} &= E\{\Phi_i^{\otimes 2}\}.
\end{aligned}$$

Then, $\sqrt{n}(\hat{\beta}_1 - \beta_{01})$ converges weakly to a normal distribution with mean zero and variance-covariance matrix $\Gamma_{\beta_1}^{-1} \Sigma_{\beta_1} (\Gamma_{\beta_1}^{-1})'$, where Σ_{β_1} is given in condition (C.4) of the theorem, and the matrix Σ_{β_1} can be estimated by $n^{-1} \sum_{i=1}^n \hat{\Phi}_i^{\otimes 2}$, obtained by replacing $G(s)$, $S_T(s)$ and other terms in Φ_i by their respective estimators.

References

1. Lavori PW, Dawson R. A design for testing clinical strategies: Biased adaptive within-subject randomization. *Journal of the Royal Statistical Society, Series A.* 2000; 163:29–38.

2. Dawid AP, Didelez V. Identifying the consequences of dynamic treatment strategies: A decision-theoretic overview. *Statistics Surveys*. 2010; 4:184–231.
3. Murphy SA. An experimental design for the development of adaptive treatment strategies. *Statistics in Medicine*. 2005; 24:1455–1481. [PubMed: 15586395]
4. Murphy SA, Bingham D. Screening experiments for developing dynamic treatment regimes. *Journal of the American Statistical Association*. 2009; 104:391–408. [PubMed: 20589222]
5. Thall PF. Adaptive therapy for androgen-independent prostate cancer: A randomized selection trial of four regimens. *Journal of the National Cancer Institute*. 2007; 99:1613–1622. [PubMed: 17971530]
6. Thall PF. Evaluating multiple treatment courses in clinical trials. *Statistics in Medicine*. 2000; 19:1011–1028. [PubMed: 10790677]
7. Zhao Y, Kosorok M, Zeng D. Reinforcement learning design for cancer clinical trials. *Statistics in Medicine*. 2009; 28:3294–3315. [PubMed: 19750510]
8. Robins J. The control of confounding by intermediate variables. *Statistics in Medicine*. 1989; 8:679–701. [PubMed: 2749074]
9. 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*. 1993:24–33.
10. Robins JM. Causal inference from complex longitudinal data. In: Berkane, M., editor. *Latent Variable Modeling and Applications to Causality*. New York: Springer-Verlag; 1997. p. 69–117.
11. Murphy SA, van der Laan MJ, Robins JM. Group CPPR. Marginal mean models for dynamic regimes. *Journal of the American Statistical Association*. 2001; 96:1410–1423. [PubMed: 20019887]
12. Murphy SA. Optimal dynamic treatment regimes (with discussion). *Journal of the Royal Statistical Society, Series B*. 2003; 65:331–366.
13. Robins JM. Optimal structural nested models for optimal sequential decisions. *Lecture Notes in Statistics*. 2004; 179:189–326.
14. Cox DR. Regression models and life tables (with discussion). *Journal of the Royal Statistical Society, Series B*. 1972; 34:187–220.
15. Cox, DR.; Oakes, D. *Analysis of Survival Data*. London, New York: Chapman and Hall; 1984.
16. Kalbfleisch, JD.; Prentice, RL. *The Statistical Analysis of Failure Time Data*. Wiley; 2002.
17. Andersen RK, Gill RD. Cox's regression model for counting processes: a large sample study. *Annals of Statistics*. 1982; 80:1100–1120.
18. Pepe MS, Cai J. Some graphical displays and marginal regression analyses for recurrent failure times and time dependent covariates. *Journal of the American Statistical Association*. 1993; 88:811–820.
19. Ghosh D, Lin DY. Nonparametric analysis of recurrent events and death. *Biometrics*. 2000; 56:554–562. [PubMed: 10877316]
20. Ghosh D, Lin DY. Marginal regression models for recurrent and terminal events. *Statistica Sinica*. 2002; 12:663–688.
21. Ghosh D, Lin DY. Semiparametric analysis of recurrent events data in the presence of dependent censoring. *Biometrics*. 2003; 59:877–885. [PubMed: 14969466]
22. Peng L, Fine JP. Rank estimation of accelerated lifetime models with dependent censoring. *Journal of the American Statistical Association*. 2006; 101:1085–1093.
23. Peng L, Fine JP. Regression modeling of semicompeting risks data. *Biometrics*. 2007; 63:96–108. [PubMed: 17447934]
24. Liu L, Wolfe RA, Huang X. Shared frailty models for recurrent events and a terminal event. *Biometrics*. 2004; 60:747–756. [PubMed: 15339298]
25. Huang X, Liu L. A joint frailty model for survival and gap times between recurrent events. *Biometrics*. 2007; 63:389–397. [PubMed: 17688491]
26. Moodie EM, Richardson TS, Stephens DA. Demystifying optimal dynamic treatment regimes. *Biometrics*. 2007; 63:447–455. [PubMed: 17688497]

27. Henderson R, Ansell P, Alshibani D. Regret-regression for optimal dynamic treatment regimes. *Biometrics*. 2010; 66:1192–1201. [PubMed: 20002404]
28. Chakraborty B, Murphy S, Strecher V. Inference for non-regular parameters in optimal dynamic treatment regimes. *Statistical Methods in Medical Research*. 2010; 19:317–343. [PubMed: 19608604]
29. Goldberg Y, Kosorok MR. Q-learning with censored data. *Annals of Statistics*. 2012; 40:529–560. [PubMed: 22754029]
30. Rosthøj S, Keiding N, Schmiegelow K. Estimation of dynamic treatment strategies for maintenance therapy of children with acute lymphoblastic leukaemia: an application of history-adjusted marginal structural models. *Statistics in Medicine*. 2012; 31:470–488. [PubMed: 22086750]
31. Zhang B, Tsiatis A, Laber E, Davidian M. A robust method for estimating optimal treatment regimes. *Biometrics*. 2012; 68:101–1018. [PubMed: 21950348]
32. Zhao Y, Zeng D, Socinski M, Kosorok M. Reinforcement learning strategies for clinical trials in non small cell lung cancer. *Biometrics*. 2011; 67:1422–1433. [PubMed: 21385164]
33. Chaffee, P.; van der Laan, M. Targeted maximum likelihood estimation for dynamic treatment regimes in sequential randomized controlled trials; U.C. Berkeley Division of Biostatistics Working Paper Series; 2011. <http://biostats.bepress.com/ucbbiostat/paper277>
34. Wang L, Rotnitzky A, Lin X, Millikan R, Thall P. Evaluation of viable dynamic treatment regimes in a sequentially randomized trial of advanced prostate cancer. *Journal of the American Statistical Association*. 2012; 498:493–508. [PubMed: 22956855]
35. Zajonc T. Bayesian inference for dynamic treatment regimes: Mobility, equity, and efficiency in student tracking. *Journal of the American Statistical Association*. 2012; 107:80–92.
36. Watkins CJCH, Dayan P. Q-learning. *Machine Learning*. 1992; 8:279–292.
37. Nahum-Shani I, Qian M, Almirall D, Pelham WE, Gnagy B, Fabiano GA, Waxmonsky JG, Yu J, Murphy SA. Q-learning: A data analysis method for constructing adaptive interventions. *Psychological Methods*. 2010; 17:478–494. [PubMed: 23025434]
38. Blatt D, Murphy SA, Zhu J. A-learning for approximate planning. 2004 <http://dept.stat.lsa.umich.edu/samurphy/papers/Alearning2004.pdf>.
39. Rosthøj S, Fullwood C, Henderson R, Stewart S. Estimation of optimal dynamic anticoagulation regimes from observational data: a regret-based approach. *Statistics in Medicine*. 2006; 25:4197–4215. [PubMed: 16981226]
40. Wahed AS, Thall PF. Evaluating joint effects of induction salvage treatment regimes on overall survival in acute leukaemia. *Journal of the Royal Statistical Society: Series C*. 2013; 62:67–83.
41. Robins JM. Robust estimation in sequentially ignorable missing data and causal inference models. 1999 Proceedings of the American Statistical Association Section on Bayesian Statistical Science. 2000:6–10.
42. Robins J. A new approach to causal inference in mortality studies with a sustained exposure period application to control of the healthy worker survivor effect. *Mathematical Modelling*. 1986; 7:9–12.
43. Robins, JM.; Rotnitzky, A. Recovery of information and adjustment for dependent censoring using surrogate markers. Boston, MA: Birkhuser; 1992.
44. Heitjan D, Rubin D. Ignorability and coarse data. *The Annals of Statistics*. 1991; 19:2244–2253.
45. Miloslavsky M, Keles S, ver der Laan MJ, 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.
46. 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.
47. Huang X, Ning J. Analysis of multi-stage treatments for recurrent diseases. *Statistics in Medicine*. 2012; 31:2805–2821. [PubMed: 22826139]
48. White H. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Halbert White Econometrica*. 1980; 48:817–838.

49. Gelman, A.; Hill, J. Data analysis using regression and multilevel/hierarchical models. Cambridge University Press; 2007.
50. Tan Z. Comment: Understanding or, ps and dr. *Statistical Science*. 2007; 22:560–568.
51. Zhao H, Tsiatis AA. Efficient estimation of the distribution of quality-adjusted survival time. *Biometrics*. 1997; 55:1101–1107. [PubMed: 11315054]
52. Estey E, Thall P, Pierce S, Cortes J, Beran M, Kantarjian H, Keating M, Andreeff M, Freireich E. Randomized phase ii study of fludarabine + cytosine arabinoside + idarubicin +/- all-trans retinoic acid +/- granulocyte colony-stimulating factor in poor prognosis newly diagnosed acute myeloid leukemia and myelodysplastic syndrome. *Blood*. 1999; 93:2478–2884. [PubMed: 10194425]
53. Tang X, Wahed AS. Comparison of treatment regimes with adjustment for auxiliary variables. *Journal of Applied Statistics*. 2011; 38(12):2925–2938.
54. Patel K, Kay R, Rowell L. Comparing proportional hazards and accelerated failure time models: an application in influenza. *Pharmaceutical Statistics*. 2006; 5:213–224.
55. Tibshirani R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B: Methodological*. 1996; 58:267–288.
56. Fan J, Li R. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*. 2001; 96:1348–1360.
57. Zou H, Hastie T. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society, Series B: Statistical Methodology*. 2005; 67:301–320.
58. Candes E, Tao T. The dantzig selector: statistical estimation when p is much larger than n. *The Annals of Statistics*. 2005; 35(6):2313–2351.
59. Huang J, Ma S, Xie H. Regularized estimation in the accelerated failure time model with high-dimensional covariates. *Biometrics*. 2006; 62:813–820. [PubMed: 16984324]
60. Huang J, Ma S. Variable selection in the accelerated failure time model via the bridge method. *Lifetime Data Analysis*. 2010; 16
61. Johnson BA, Lin DY, Zeng D. Penalized estimating functions and variable selection in semiparametric regression models. *Journal of the American Statistical Association*. 2008; 103:672–680. [PubMed: 20376193]
62. Wang S, Nan B, Zhu DG, Beer J. Doubly penalized buckley-james method for survival data with high-dimensional covariates. *Biometrics*. 2008; 64:132–140. [PubMed: 17680828]
63. Cai T, Huang J, Tian L. Regularized estimation for the accelerated failure time model. *Biometrics*. 2009; 65:394–404. [PubMed: 18573133]
64. Li Y, Dicker L, Zhao S. The dantzig selector for censored linear regression models. *Statistica Sinica*. 2014; 24:251–268. [PubMed: 24478569]
65. Huang X, Cormier J, Pisters P. Estimation of the causal effects on survival of two-stage nonrandomized treatment sequences for recurrent diseases. *Biometrics*. 2006; 62(3):901–909. [PubMed: 16984334]
66. Strawderman R. Higher-order asymptotic approximation: Laplace, saddlepoint, and related methods. *Journal of the American Statistical Association*. 2000; 95:1358–1364.

Figure 1. Possible pathways of successive states, transition times and salvage therapy following induction treatment in acute leukemia patients.

Table 1

Parameter estimates for the AFT model: for the time from salvage treatment to death

	Covariate	Estimate	Lifetime ratio [*]	SE of estimate	P-value
Patients with disease progression	Intercept	-0.046		0.460	0.921
	Time from CR to progression (years)	0.587	1.799	0.188	0.002
	<i>B</i> : HDAC vs. Others	-0.765	0.465	0.364	0.036
Patients with treatment resistance	Intercept	1.21		0.751	0.107
	<i>R</i> : Time to resistance (years)	-4.587		2.812	0.103
	<i>B</i> : HDAC vs. Others	-3.540		0.909	< 0.001
	Interaction: <i>R</i> · <i>B</i>	14.558		4.702	0.002
Both progression and resistance patients	FAIA vs. FAI	0.659	1.933	0.286	0.021
	FAIG vs. FAI	0.328	1.388	0.292	0.262
	FAIGA vs. FAI	0.555	1.742	0.474	0.242
	Cyto: Poor vs. intermediate	-0.519	0.595	0.274	0.060
	Age (years)	-0.021	0.979	0.007	0.001

* Remark: Only interpretable lifetime ratios are presented. The corresponding terms for intercepts and variables involved in interactions are omitted.

Table 2

Potential survival times after optimized salvage treatments

Response to initial treatment	Time to disease progression or resistance	Actual salvage treatment	Optimal salvage treatment	Number of patients	Estimation of second stage potential survival time under optimal salvage treatment
CR and then progression		HDAC	Other	53	$S \exp(0.765)$
		Other	Other	40	S
Resistance	$R \geq 0.243$ years	HDAC	HDAC	4	S
	$R < 0.243$ years	HDAC	Other	23	$S \exp(3.540 - 14.558 R)$
	$R \geq 0.243$ years	Other	HDAC	5	$S \exp(-3.540 + 14.558 R)$
	$R < 0.243$ years	Other	Other	7	S

Table 3

Parameter estimates for the AFT model for overall survival time (with salvage treatments already optimized respectively for each individual)

Covariate	Estimate	Mean Lifetime ratio	SE of estimate	<i>p</i> -value
Intercept	0.342		0.842	0.685
Age (years)	−0.016	0.984	0.0092	0.081
FAIA	0.759	2.137	0.552	0.169
FAIG	0.080	1.083	0.562	0.887
FAIGA	0.122	1.131	0.705	0.861
FAI × Cyto	−0.216	0.805	0.542	0.691
FAIA × Cyto	−1.538	0.214	0.517	0.003
FAIG × Cyto	−0.685	0.504	0.438	0.118
FAIGA × Cyto	−0.523	0.593	0.600	0.383

Note: Cyto = Poor vs. intermediate cytogenetics

Table 4
Comparisons of initial treatments on overall survival (with salvage treatments already optimized respectively for each individual)

Intermediate Cytogenetics				Poor Cytogenetics		
	Difference	SE	p-value	Difference	SE	p-value
FAIA vs FAI	0.759	0.552	0.169	-0.562	0.500	0.262
FAIG vs FAI	0.080	0.562	0.887	-0.389	0.405	0.337
FAIGA vs FAI	0.122	0.705	0.861	-0.184	0.366	0.615
FAIG vs FAIA	-0.680	0.396	0.086	0.173	0.549	0.752
FAIGA vs FAIA	-0.637	0.593	0.283	0.378	0.521	0.468
FAIGA vs FAIG	0.043	0.603	0.943	0.205	0.430	0.634

Table 5

Simulation study: estimation of β_1 and β_2 of the AFT models

N	($C_1\%$, $C_2\%$)	Mean	Empirical S.E.	Asymptotic S.E.	95%C.P.*
Scenario I: $C \sim \exp(\tau_1)$					
$\beta_1 = (1, 0.5)$					
200	(10, 35)	(1.005, .501)	(.137, .073)	(.131, .075)	(.936, .960)
400	(10, 35)	(.996, .500)	(.097, .054)	(.092, .053)	(.940, .941)
200	(20, 50)	(.995, .497)	(.147, .079)	(.144, .083)	(.930, .962)
400	(20, 50)	(.987, .497)	(.105, .059)	(.102, .057)	(.946, .950)
$\beta_2 = (0.5, -0.5, 1)$					
200	(10, 35)	(.505, -.492, .982)	(.196, .211, .286)	(.192, .212, .279)	(.942, .945, .949)
400	(10, 35)	(.496, -.501, .996)	(.137, .145, .192)	(.132, .143, .188)	(.953, .940, .937)
200	(20, 50)	(.502, -.489, .971)	(.215, .230, .312)	(.211, .229, .307)	(.937, .935, .938)
400	(20, 50)	(.491, -.502, .993)	(.146, .159, .212)	(.149, .162, .217)	(.934, .944, .942)
Scenario II: the hazard function of C is $\lambda_C(t) = \tau_2 \exp(-Z_{t1})$					
$\beta_1 = (1, 0.5)$					
200	(10, 35)	(1.003, .501)	(.138, .076)	(.134, .077)	(.941, .947)
400	(10, 35)	(.999, .499)	(.089, .054)	(.094, .054)	(.965, .951)
200	(20, 50)	(1.001, .488)	(.150, .086)	(.151, .085)	(.944, .950)
400	(20, 50)	(.998, .496)	(.100, .060)	(.107, .060)	(.960, .948)
$\beta_2 = (0.5, -0.5, 1)$					
200	(10, 35)	(.498, -.490, .984)	(.193, .220, .292)	(.195, .213, .282)	(.950, .931, .934)
400	(10, 35)	(.494, -.499, .994)	(.136, .149, .199)	(.138, .151, .199)	(.950, .954, .953)
200	(20, 50)	(.480, -.493, .983)	(.212, .243, .313)	(.212, .230, .304)	(.932, .928, .935)
400	(20, 50)	(.488, -.497, .989)	(.149, .159, .217)	(.149, .161, .216)	(.946, .949, .953)

* Remarks: N = sample size, $C_1\%$ = percentage of subjects with time to disease recurrence (R) being censored, $C_2\%$ = percentage of subjects with time to death (T) being censored, S.E. = standard error, 95% C.P. = coverage probability of 95% confidence interval.