Published in final edited form as:

J Stat Theory Pract. 2015 April ; 9(2): 266–287. doi:10.1080/15598608.2013.878888.

Pattern-mixture-type Estimation and Testing of Neuroblastoma Treatment Regimes

Xinyu Tang and

College of Medicine, University of Arkansas for Medical Sciences, Little Rock, AR
xtang@uams.edu

Abdus S. Wahed

Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA wahed@pitt.edu

Abstract

Sequentially randomized designs are commonly used in biomedical research, particularly in clinical trials, to assess and compare the effects of different treatment regimes. In such designs, eligible patients are first randomized to one of the initial therapies, then patients with some intermediate response (e.g. without progressive diseases) are randomized to one of the maintenance therapies. The goal is to evaluate dynamic treatment regimes consisting of an initial therapy, the intermediate response, and a maintenance therapy. In this article, we demonstrate the use of pattern-mixture model (commonly used for analyzing missing data) for estimating the effects of treatment regimes based on familiar survival analysis techniques such as Nelson-Aalen and parametric models. Moreover, we demonstrate how to use estimates from pattern-mixture models to test for the differences across treatment regimes in a weighted log-rank setting. We investigate the properties of the proposed estimators and test in a Monte Carlo simulation study. Finally we demonstrate the methods using the long-term survival data from the high risk neuroblastoma study.

Keywords

dynamic treatment regime; high risk neuroblastoma study; parametric model; pattern-mixture models; proportional hazard model; sequentially randomized design

1 Introduction

Pattern-mixture models (Little 1993) are commonly used in longitudinal data analysis to account for informative dropouts. These models stratify the data according to dropout patterns and formulate a model for each stratum. The final parameter estimate is a weighted average of the stratum-specific estimates. The basic idea of pattern-mixture models is that conditional on the pattern, the missing mechanism is ignorable within a stratum, and thus information from the complete cases can be borrowed to predict the incomplete cases. In

what follows, we show how a pattern-mixture-type estimation and test procedures can be adopted to estimate and test dynamic treatment regimes.

1.1 Treatment Regimes in Children's Neuroblastoma

Sequentially randomized designs are becoming common in biomedical research, particularly in clinical trials. These trials are usually designed to evaluate and compare the effects of different treatment regimes resulting from medical decision making at different stages of therapy. In a sequentially randomized clinical trial, eligible patients are first randomly assigned to receive one of the initial treatments. Patients meeting some criteria (e.g. no progressive diseases) are then randomized to receive one of the maintenance treatments. Such multistage treatment assignment results in dynamic treatment regimes consisting of an initial treatment, the intermediate response, and a second stage treatment. For example, in a neuroblastoma study conducted by the Children's Cancer Group, following the initial cycle of chemotherapy, eligible children with high risk neuroblastoma were randomly assigned to receive either a combination of myeloablative chemotherapy, total-body irradiation, and transplantation of autologous bone marrow purged of cancer cells (ABMT), or chemotherapy alone. Children without progressive diseases were subsequently randomized to receive either 13-cis-retinoic acid (cis-RA) or no further therapy. Thus, four possible treatment regimes could be constructed for the neuroblastoma study: (i) treat with chemotherapy followed by cis-RA if no progressive diseases present (CCR); (ii) treat with chemotherapy followed by no further therapy if no progressive diseases present (CNR); (iii) treat with ABMT followed by cis-RA if no progressive diseases present (ACR); (iv) treat with ABMT followed by no further therapy if no progressive diseases present (ANR). This formulation of dynamic treatment regimes from multistage treatment assignment allows simultaneous assessment of both first and second stage treatments, which has been the focus of such clinical trials.

1.2 Existing Methods, their limitations, and the goal of this study

The last decade has seen considerable advancement in the development of statistical methods for estimating the effects of dynamic treatment regimes (Murphy, van der Laan and Robins 2001; Murphy 2003; Murphy and Bingham 2009; Lunceford, Davidian and Tsiatis 2002; Wahed and Tsiatis 2004; Wahed and Tsiatis 2006; Hernan et al. 2006; Guo and Tsiatis 2005; Lokhnygina and Helterbrand 2007). Murphy et al. (2001) developed marginal models for the mean response for a given dynamic treatment regime. In a follow-up article Murphy (2003) they provided a methodology for constructing the optimal dynamic treatment regime. Lunceford et al. (2002) proposed consistent survival and mean restricted survival estimators for treatment regimes in a two-stage randomization design. Wahed and Tsiatis (2004, 2006) introduced the most efficient estimator utilizing additional information from auxiliary variables. Guo and Tsiatis (2005) derived a weighted risk set estimator (WRSE) on the basis of counting process theory (Fleming and Harrington 1991) using a time-varying measurement for the intermediate response. Hernan et al. (2006) described a simple method to compare dynamic treatment regimes via inverse probability weighting. Murphy and Bingham (2009) used screening experiments to identify potential treatment components and screen out insignificant ones for developing dynamic treatment regimes.

Lokhnygina and Helderbrand (2007) extended the Cox regression method to the estimation of log hazards of dynamic treatment regimes in a two-stage randomization design and applied a robust score test to compare two treatment regimes sharing the same second-stage treatment. Feng and Wahed (2008) presented a modified supremum weighted log-rank test to test the equality of two dynamic treatment regimes. Both tests are applicable only to the comparisons of treatment regimes that share different induction treatments so that the groups being compared are statistically independent. In this article, we use traditional survival analysis and missing data techniques (e.g. Nelson-Aalen (NA) estimators, parametric models, pattern mixture models) to estimate and compare the effects of treatment regimes. Contrary to the existing methods (Lokhnygina and Helderbrand 2007; Feng and Wahed 2008) the proposed test procedure does not require the comparison groups to be independent and hence allows comparison between any pair of regimes.

The article is organized as follows. In Section 2, we show the use of pattern-mixture model in combination with standard survival methods for estimating the survival quantities and for comparing different treatment regimes based on the overall survival. In Section 3, a Monte Carlo simulation study is carried out to examine the performance of the proposed method. In Section 4, we analyze the high risk neuroblastoma dataset to compare overall survival for different neuroblastoma treatment regimes. We wrap up the article with a discussion in Section 5.

2 Methods

The basic structure of a two-stage randomization design where only two treatment options are available is depicted in Figure 1, which reflects the design of the high risk neuroblastoma study described in brief in Section 1 (details are given in Section 4). In a two-stage randomization design, patients are randomly assigned to J initial treatments, namely, $A_1, A_2, \dots, A_{J-1}, A_J$. Then patients with a response (e.g. without progressive diseases) are re-randomized to K maintenance treatments, namely, $B_1, B_2, \dots, B_{K-1}, B_K$, upon consent to further randomization. A dynamic treatment regime in a two-stage treatment setting consists of an initial treatment, the decision rule for choosing the second-stage treatment, and a second-stage treatment. For example, the treatment regime $A_j B_k$ is defined as follows.

Definition 2.1

Regime $A_j B_k$: Treat with A_j (initial treatment), if respond and consent to further treatment, treat with B_k (second-stage treatment).

The regime $A_j B_k$ defined above is a dynamic treatment regime since the assignment of second-stage treatments depends on the intermediate outcome (response). We will denote the above regime by $A_j B_k$. The goal is to estimate survival quantities (e.g. hazards, survival probabilities) under various regimes and compare their effects. The observed data from a two-stage randomization design can be expressed as

$$\{X_{ji}, R_i, R_i Z_{ki}, U_i, \Delta_i\},$$

where X_{ji} is the indicator for the j th initial treatment, $X_{ji} = 1$ if the i th patient was assigned to A_j as initial treatment, and $X_{ji} = 0$ if otherwise; R_i is the indicator for response and consent, $R_i = 1$ if the i th patient responded to the initial treatment and consented to further randomization, and $R_i = 0$ if otherwise; $Z_{ki} = 1$ if the i th patient was assigned to maintenance treatment B_k , and $Z_{ki} = 0$ if otherwise (note that Z_{ki} is defined only when $R_i = 1$); and U_i denotes the observed death ($\delta_i = 1$) or censoring time ($\delta_i = 0$). In other words, when $\delta_i = 1$, $U_i = T_i$, the potential survival time, and when $\delta_i = 0$, $U_i = C_i$, the potential censoring time. We assume independent censoring. Patients who were censored without a response were treated as nonresponders [since a response was not observed before being censored; see Lunceford et al. (2002)].

Our goal is to find the survival and hazard estimators for the treatment regimes $A_j B_k$, and also provide contrasts for comparing regimes $A_j B_k$, $j = 1, \dots, J$ and $k = 1, \dots, K$ based on the overall survival. First note that the triplet of intermediate response status along with the first and second stage treatments cluster patients into $J(K + 1)$ patterns. J of these patterns correspond to J groups of patients who were treated with A_j but did not respond. Let us denote these patterns by P_{j0} , $j = 1, 2, \dots, J$. The remaining JK patterns correspond to the groups of patients actually receiving the treatment sequence A_j followed by B_k upon response. We will refer to these JK patterns by P_{jk} , $j = 1, 2, \dots, J; k = 1, 2, \dots, K$. Let us define $\lambda_{jk}(t)$ and $S_{jk}(t)$ to be the hazard and survival functions respectively, corresponding to the jk th pattern ($j = 1, \dots, J; k = 0, 1, \dots, K$). Note that for given j and k , in only two of the $J(K + 1)$ patterns, namely, P_{j0} (treated with A_j and did not respond) and P_{jk} (treated with A_j , responded, and then treated with B_k), patients are treated consistent with the regime $A_j B_k$. Therefore, the hazard function $\lambda_{A_j B_k}(t)$ of the treatment regime $A_j B_k$ can be expressed as the weighted average of the two pattern-specific hazard functions $\lambda_{j0}(t)$ and $\lambda_{jk}(t)$ as follows:

$$\lambda_{A_j B_k}(t) = \{1 - \phi_{jk}(t)\} \lambda_{j0}(t) + \phi_{jk}(t) \lambda_{jk}(t), \quad j, k = 1, 2, \quad (2.1)$$

Where

$$\phi_{jk}(t) = \frac{\pi_j S_{jk}(t)}{\pi_j S_{jk}(t) + (1 - \pi_j) S_{j0}(t)}, \quad j, k = 1, 2 \quad (2.2)$$

is the conditional probability of following pattern P_{jk} and surviving beyond time t , given that the patient survived beyond time t under regime $A_j B_k$, and π_j is the probability of response for patients receiving A_j as the initial treatment, i.e., $\pi_j = P(R_i = 1 | X_{ji} = 1)$. See Appendix for a derivation leading to expression (2.1). Similarly, the survival function $S_{A_j B_k}(t)$ of the treatment regime $A_j B_k$ can be expressed as the weighted average of the two pattern-specific survival functions S_{j0} and $S_{jk}(t)$ as follows:

$$S_{A_j B_k}(t) = (1 - \pi_j) S_{j0}(t) + \pi_j S_{jk}(t), \quad j, k = 1, 2. \quad (2.3)$$

This representation of the survival function as a weighted average of pattern-specific survivals is often referred to as the g-computation formula or g-formula (Robins 1986, 1987). From this point forward, we will assume $J = K = 2$, realizing that all methods

described extends to general J and K . From (2.1), the hazards under four regimes and A_1B_1 , A_1B_2 , A_2B_1 and A_2B_2 can be written as

$$\lambda_{A_1B_1}(t) = \frac{\pi_1 S_{11}(t) \lambda_{11}(t) + (1 - \pi_1) S_{10}(t) \lambda_{10}(t)}{\pi_1 S_{11}(t) + (1 - \pi_1) S_{10}(t)}, \quad (2.4)$$

$$\lambda_{A_1B_2}(t) = \frac{\pi_1 S_{12}(t) \lambda_{12}(t) + (1 - \pi_1) S_{10}(t) \lambda_{10}(t)}{\pi_1 S_{12}(t) + (1 - \pi_1) S_{10}(t)}, \quad (2.5)$$

$$\lambda_{A_2B_1}(t) = \frac{\pi_2 S_{21}(t) \lambda_{21}(t) + (1 - \pi_2) S_{20}(t) \lambda_{20}(t)}{\pi_2 S_{21}(t) + (1 - \pi_2) S_{20}(t)}, \quad (2.6)$$

and

$$\lambda_{A_2B_2}(t) = \frac{\pi_2 S_{22}(t) \lambda_{22}(t) + (1 - \pi_2) S_{20}(t) \lambda_{20}(t)}{\pi_2 S_{22}(t) + (1 - \pi_2) S_{20}(t)}, \quad (2.7)$$

Corresponding survival functions of four treatment regimes are

$$S_{A_1B_1}(t) = \pi_1 S_{11}(t) + (1 - \pi_1) S_{10}(t), \quad (2.8)$$

$$S_{A_1B_2}(t) = \pi_1 S_{12}(t) + (1 - \pi_1) S_{10}(t), \quad (2.9)$$

$$S_{A_2B_1}(t) = \pi_2 S_{21}(t) + (1 - \pi_2) S_{20}(t), \quad (2.10)$$

and

$$S_{A_2B_2}(t) = \pi_2 S_{22}(t) + (1 - \pi_2) S_{20}(t). \quad (2.11)$$

Thus if one can estimate the probability of response, and the pattern-specific survival and hazard functions $\lambda_{jk}(t)$ and $S_{jk}(t)$, then the above formulae can be used to estimate the hazard and survival functions for different regimes. The probability of response for patients assigned to A_j as initial treatment can be estimated by

$$\hat{\pi}_j = \frac{\sum_{i=1}^n X_{ji} R_i}{\sum_{i=1}^n X_{ji}}.$$

Both the nonparametric methods such as NA estimator, and parametric methods can be used to obtain the estimates of six pattern-specific survival functions [$S_{10}(t)$, $S_{20}(t)$, $S_{11}(t)$, $S_{12}(t)$, $S_{21}(t)$, $S_{22}(t)$] and six pattern-specific hazard functions [$\lambda_{10}(t)$, $\lambda_{20}(t)$, $\lambda_{11}(t)$, $\lambda_{12}(t)$, $\lambda_{21}(t)$, $\lambda_{22}(t)$] from patients following specific pattern (see Sections 2.1-2.2).

Survival functions of four treatment regimes can then be easily obtained through

$$\begin{aligned} \hat{S}_{A_1 B_1}(t) &= \hat{\pi}_1 \hat{S}_{11}(t) \\ &+ (1 - \hat{\pi}_1) \hat{S}_{10}(t), \hat{S}_{A_1 B_2}(t) \\ &= \hat{\pi}_1 \hat{S}_{12}(t) + (1 - \hat{\pi}_1) \hat{S}_{10}(t), \hat{S}_{A_2 B_1}(t) \\ &= \hat{\pi}_2 \hat{S}_{21}(t) \\ &+ (1 - \hat{\pi}_2) \hat{S}_{20}(t), \text{ and } \hat{S}_{A_2 B_2}(t) = \hat{\pi}_2 \hat{S}_{22}(t) + (1 - \hat{\pi}_2) \hat{S}_{20}(t). \end{aligned}$$

Similarly, one can obtain the estimates for the hazard functions from equations (2.4) - (2.7).

2.1 Nonparametric Methods

Based on the counting process notations described in Fleming and Harrington (1991), the cumulative hazard rate for pattern P_{jk} , $j = 1, 2; k = 0, 1, 2$, can be estimated by the NA estimator

$$\hat{\Lambda}_{jk}(t) = \int_0^t \frac{dN_{jk}(u)}{Y_{jk}(u)},$$

where $N_{jk}(u)$ is the event process for patients following pattern P_{jk} at time u ,

$$N_{jk}(u) = \sum_{i=1}^n N_{jki}(u) = \sum_{i=1}^n I[U_i \leq u, \Delta_i = 1, X_{ji} = 1, (1 - R_i) + R_i Z_{ki} = 1],$$

and $Y_{jk}(u)$ is the number of individuals at risk at time u under pattern P_{jk} , i.e.,

$$Y_{jk}(u) = \sum_{i=1}^n I[U_i \geq u, X_{ji} = 1, (1 - R_i) + R_i Z_{ki} = 1].$$

Then the corresponding hazard and survival can be estimated by $\hat{\lambda}_{jk}(t) = d\hat{\Lambda}_{jk}(t) / dt$ and $\hat{S}_{jk}(t) = \exp\{-\hat{\Lambda}_{jk}(t)\}$ respectively. In practice, the estimated hazard $\hat{\lambda}_{jk}(t)$ can be obtained as the numerical derivative of the kernel-smoothed cumulative hazard estimates (Klein and Moeschberger 2003).

2.2 Parametric Methods

A variety of distributions can be chosen to model the pattern-specific survival times. Weibull distribution, log-normal distribution, and log-logistic distribution are the most common parametric models used for this purpose. We focus on the estimation for the Weibull distribution here for demonstration. Under the Weibull distributional assumption, the hazard and survival functions for the pattern P_{jk} is given by

$$S_{jk}(t) = \exp(-\lambda_{jk} t^{\alpha_{jk}}),$$

and

$$\lambda_{jk}(t) = \lambda_{jk} \alpha t^{\alpha_{jk}-1},$$

where, in many instances, λ_{jk} and α_{jk} are reparametrized as $\lambda_{jk} = \exp(-\mu_{jk}/\sigma_{jk})$ and $\alpha_{jk} = 1/\sigma_{jk}$.

The estimates of the parameters μ_{jk} and σ_{jk} can be obtained through maximum likelihood method (Klein and Moeschberger 2003). After fitting the Weibull distribution to the subjects following pattern P_{jk} , the maximum likelihood estimates $\hat{\mu}_{jk}$ and $\hat{\sigma}_{jk}$, their respective estimated variance $var(\hat{\mu}_{jk})$ and $var(\hat{\sigma}_{jk})$, and their estimated covariance $cov(\hat{\mu}_{jk}, \hat{\sigma}_{jk})$ can be obtained using the Fisher's information matrix. Then the maximum likelihood estimators of parameters λ_{jk} and α_{jk} are given by $\hat{\lambda}_{jk} = \exp(-\hat{\mu}_{jk}/\hat{\sigma}_{jk})$ and $\hat{\alpha}_{jk} = 1/\hat{\sigma}_{jk}$. Using the delta method, one can obtain the estimated variance of $\hat{\lambda}_{jk}$, and $\hat{\alpha}_{jk}$, and their estimated covariance as follows:

$$var(\hat{\lambda}_{jk}) = \exp\left(-2\frac{\hat{\mu}_{jk}}{\hat{\sigma}_{jk}}\right) \left[\frac{var(\hat{\mu}_{jk})}{\hat{\sigma}_{jk}^2} - 2\frac{\hat{\mu}_{jk}cov(\hat{\mu}_{jk}, \hat{\sigma}_{jk})}{\hat{\sigma}_{jk}^3} + \frac{\hat{\mu}_{jk}^2 var(\hat{\sigma}_{jk})}{\hat{\sigma}_{jk}^4} \right],$$

$$var(\hat{\alpha}_{jk}) = \frac{var(\hat{\sigma}_{jk})}{\hat{\sigma}_{jk}^2},$$

and

$$cov(\hat{\lambda}_{jk}, \hat{\alpha}_{jk}) = \exp\left(-\frac{\hat{\mu}_{jk}}{\hat{\sigma}_{jk}}\right) \left[\frac{cov(\hat{\mu}_{jk}, \hat{\sigma}_{jk})}{\hat{\sigma}_{jk}^3} - \frac{\hat{\mu}_{jk}var(\hat{\sigma}_{jk})}{\hat{\sigma}_{jk}^4} \right].$$

Therefore, the hazard and survival functions can be estimated as $\hat{\lambda}_{jk}(t) = \hat{\lambda}_{jk} \hat{\alpha}_{jk} t^{\hat{\alpha}_{jk}-1}$ and $\hat{S}_{jk}(t) = \exp(-\hat{\lambda}_{jk} t^{\hat{\alpha}_{jk}})$. Similarly, their estimated standard errors can be obtained using the delta method.

2.3 Treatment Regime Comparisons

Comparisons of treatment regimes in terms of their hazards can be carried out by integrating the weighted difference between two cumulative hazard functions (Fleming and Harrington 1991). For example, the weighted log-rank test statistics for comparing cumulative hazard functions $\Lambda_{A_j B_k}(t)$ and $\Lambda_{A_j' B_{k'}}(t)$ is given by

$$D_{A_j B_k / A_j' B_{k'}} = \int_0^\infty K_{A_j B_k / A_j' B_{k'}}(s) \left\{ d\hat{\Lambda}_{A_j B_k}(t) - d\hat{\Lambda}_{A_j' B_{k'}}(t) \right\}$$

$$= \int_0^\infty K_{A_j B_k / A_j' B_{k'}}(s) \left\{ \hat{\lambda}_{A_j B_k}(s) - \hat{\lambda}_{A_j' B_{k'}}(s) \right\} ds,$$

where the weight of K is chosen according to Feng and Wahed (2008) as follows:

$$K_{A_j B_k / A_{j'} B_{k'}}(s) = \frac{\bar{Y}_{jk}(s) \bar{Y}_{j'k'}(s)}{\bar{Y}_{jk}(s) + \bar{Y}_{j'k'}(s)},$$

where $\bar{Y}_{jk}(s) = \sum_{i=1}^n W_{jki} I(U_i \geq s)$ and W_{jki} is the weight function for the i th patient defined as

$$W_{jki} = X_{ji} \{ (1 - R_i) + X_{ji} R_i Z_{ki} / \pi_{jk} \} / \pi_j,$$

where $\pi_j = P(X_{ji} = 1)$, and $\pi_{jk} = P(Z_{ki} = 1 | X_{ji} = 1, R_i = 1)$. The test statistic $D_{A_j B_k / A_{j'} B_{k'}}$ is asymptotically normally distributed with mean 0 and some variance σ^2 . As mentioned in Section 1.2, although both Lokhnygina and Helderbrand (2007) and Feng and Wahed (2008) proposed methods to test the equality of two treatment regimes, their tests can only be applied to the comparison of treatment regimes that share different induction therapies so that the treatment regimes being compared are statistically independent. Because the treatment regimes that share the same induction therapy are more likely to be correlated, it becomes difficult to obtain the explicit form of the variance estimator of the test statistic. Instead we used a bootstrap resampling method to obtain the estimated variance $\hat{\sigma}^2$. The bootstrap variance estimator $\hat{\sigma}^2$ can be obtained as follows:

Step 1: Generate a set of bootstrap samples from the original sample S using sampling with replacement. Define the set of bootstrap samples as $\{S_i, i = 1, \dots, n_k\}$;

Step 2: Calculate the test statistic based on each bootstrap sample S_i and obtain a set of corresponding test statistics $\{D_i, i = 1, \dots, n_k\}$;

Step 3: Obtain the variance estimator $\hat{\sigma}^2$ by computing the variance of the set of test statistics.

Note that this test procedure does not require the two comparator regimes to have independent data. Although several null hypotheses were tested in the simulation study and data analysis, no adjustment was made for multiple comparisons in this article. P-values less than 0.05 were considered to indicate statistical significance. The multiple comparisons are further discussed in Section 5.

3 Simulation Study

A Monte Carlo simulation study was conducted to evaluate the performance of the two methods described in Section 2: NA estimator and parametric models. A simple two-stage randomization design was chosen for the simulation study. The indicator for the initial treatment $A_1(X_1)$ was sampled from *Bernoulli*(0.5) distribution and the intermediate response (R) was drawn from *Bernoulli*(0.5) to reflect the response rate in the neuroblastoma dataset. For the responders indicator for the maintenance treatment $B_1(Z_1)$ was generated from *Bernoulli*(0.5). For the i^{th} patient, the pattern-specific survival times $T_{jki}^*, j=1, 2, k=0, 1, 2$ are generated from different parametric distributions (as described

below). The survival time for the i^{th} patient can then be expressed in terms of the pattern-specific survivals as follows:

$$T_i = \sum_{j=1}^2 X_{ji} \left[(1 - R_i) T_{j0i}^* + R_i \sum_{k=1}^2 Z_{1k} T_{jki}^* \right].$$

We generated T_{jk}^* , $j=1, 2; k=0, 1, 2$ from various distributions (Table 1):

- Under scenario I, the survival times T_{jk}^* for each pattern were generated from Weibull distributions. The survival distributions for treatment regimes A_1B_1 and A_1B_2 are the same. Treatment regimes A_1B_1 and A_1B_2 share the same induction therapy A_1 .
- Under scenario II, the survival times T_{jk}^* followed log-normal distributions. The survival distributions for treatment regimes A_1B_1 and A_2B_1 are the same. Treatment regimes A_1B_1 and A_2B_1 share the same second-stage therapy B_1 .
- Under scenario III, the survival time T_{jk}^* for each pattern followed a log-logistic distribution. The survival distributions for treatment regimes A_1B_1 and A_2B_2 are the same. Treatment regimes A_1B_1 and A_2B_2 have different induction and second-stage therapies.
- Under scenario IV, the survival times T_{jk}^* , $j = 1, 2; k = 1, 2$ (of responders) followed Weibull distributions, while the survival times T_{jk}^* , $j = 1, 2; k = 0$ (of nonresponders) followed log-normal distributions. The survival distributions for treatment regimes A_1B_2 and A_2B_1 are the same. Treatment regimes A_1B_2 and A_2B_1 have different induction and second-stage therapies.
- Under scenario V, the survival times T_{jk}^* , $j = 1, 2; k = 1, 2$ (of responders) followed log-normal distributions, while the survival times T_{jk}^* , $j = 1, 2; k = 0$ (of nonresponders) followed log-logistic distributions. No pattern is specified in the population. Although the survival rates at 1 and 3 years appear to be the same for treatment regimes A_1B_1 and A_1B_2 (Table 1), the survival curve for the treatment regime A_1B_2 declines much faster than that for the treatment regime A_1B_1 after 3 years.

Censoring time C was drawn from a uniform distribution, $Uniform(0, \tau)$, and τ was selected to achieve a censoring percentage of 30% approximately to reflect the censoring rate in the neuroblastoma dataset. For each scenario, 500 samples of size 400 were generated and a total of six different models were fitted to each sample. These models represented the five mixture models presented in scenarios I to V above plus the NA estimator to estimate the pattern-specific distributions (see Sections 2.1 and 2.2). We used a bootstrap resampling method with $n_k = 50$ repetitions to obtain the variance estimator for the weighted log-rank test statistic.

Tables 2 and 3 present the survival estimate (EST), absolute bias (BIAS) in parentheses, asymptotic standard error (ASE), empirical standard deviation (ESD) and coverage probability (CP) of 95% confidence interval under scenarios I - V at 1 and 3 years

respectively. Because the survival data under scenario I were generated based on the Weibull distribution, the survival estimates obtained using the Weibull distribution performed very well, as one would expect, with the coverage probabilities close to 95%. Some of the survival estimates obtained using other parametric distributions, e.g. log-normal distribution, appeared to be biased, and the coverage probability for one model was as low as 71.6%, which occurred for $S_{A_2B_2}(t)$ at 1-year for the log-normal models. The survival estimates based on the NA estimator appeared to be unbiased with well-maintained coverage probabilities, but the estimated standard errors were relatively larger compared to the ones obtained using the other parametric models. Similar results were observed under scenarios II - V. We observed consistent coverage probabilities when the true distributions were fitted to the data. However, the performance of the estimates obtained from other parametric distributions than the true ones were unpredictable. The NA estimator performed very well with the coverage probabilities being well-maintained under all scenarios. Therefore, from the results we observed under scenarios I - V, it is clear that the NA estimator provides unbiased estimates under all true models, which is expected as NA estimator is unbiased in large samples. Moreover, the choice of distribution when fitting the parametric models affects the survival estimates to some extent. Additional simulations were carried out with response rates varying between 20% and 70% (both equal and unequal for the first two arms). The 1-year estimates were very similar to those at 50% response rate. However, because less patients were available for estimation at 3 years, the NA estimates at 3 years were slightly biased when the response rates were relatively smaller (e.g. 40%). When the true distributions were used to fit the parametric models, the estimates appeared to be unbiased under different response rates studied.

Table 4 presents the rejection rates of the weighted log-rank tests based on all six methods under scenarios I - V for 500 samples of size 400. A total of four null hypotheses [$H_1: \lambda_{A_1B_1}(t) = \lambda_{A_1B_2}(t)$, $H_2: \lambda_{A_1B_1}(t) = \lambda_{A_2B_1}(t)$, $H_3: \lambda_{A_1B_1}(t) = \lambda_{A_2B_2}(t)$, $H_4: \lambda_{A_1B_2}(t) = \lambda_{A_2B_2}(t)$] were tested. The four null hypotheses were true under scenarios I - IV respectively. The resulted type I errors of the tests were close to 0.05 under scenarios I - IV. In some instances we found that the proposed tests were liberal in the sense that the type I error estimates were mostly around or slightly above the nominal level. This may be due to the use of bootstrap variance estimator. Analytical variance estimates, which are tedious to compute, might improve the type I error rates. We also observed similar results based on all the parametric models under scenario I through scenario V. The rejection rates obtained using the NA estimator were slightly smaller than the ones obtained using the parametric models in most of the cases. Thus, we would conclude that the weighted log-rank test appears to be robust regardless of the methods used to obtain the hazard and survival estimates. Although survival estimates at specific time points varied somewhat across different methods, the weighted log-rank test of difference in survival/hazard curves between two treatment regimes provided very similar conclusions.

4 Analysis of Neuroblastoma Dataset

A high risk neuroblastoma study was conducted by the Children's Cancer Group between 1991 and 1996 with the goal of assessing whether a combination of myeloablative chemotherapy, total-body irradiation, and transplantation of autologous bone marrow purged

of cancer cells (ABMT) improves event-free survival in children with high-risk neuroblastoma compared to intensive non-myeloablative chemotherapy, and to determine whether subsequent treatment with 13-cis-retinoic acid (cis-RA) improves event-free survival further compared to no further therapy (Matthay et al. 1999). A two-stage randomization design was used for the treatment assignment. All patients were first treated with induction chemotherapy. Then 379 eligible patients without progressive diseases participated in the first-stage randomization, with 190 patients assigned to chemotherapy and 189 patients assigned to ABMT. A total of 176 patients either had progressive diseases or declined further randomization, so only 203 patients participated in the second stage randomization, with 102 patients assigned to cis-RA and 101 patients assigned to no further therapy. The initial outcome from the neuroblastoma study was analyzed in Matthay et al. (1999). Matthay et al. (2009) reported the long-term study results. Both articles reported 55 more patients participating in the second randomization, but they were not included in the two-stage randomization setting, because they did not participate in the first randomization. Furthermore, both articles separately compared chemotherapy to ABMT, and cis-RA to no further therapy among those who participated in the second randomization. Such analyses separated the first and second stage treatments and ignored the information from those patients who had progressive diseases or declined further randomization, leading to inefficient use of data.

As described before, there are four possible treatment regimes in this data, namely, CCR, CNR, ACR, and ANR (see Section 1). We first identified the 6 subgroups in the data: patients who were assigned to chemotherapy as initial treatment and did not respond; patients who were assigned to chemotherapy as initial treatment, responded and were assigned to cis-RA as maintenance treatment; patients who were assigned to chemotherapy as initial treatment, responded and were assigned to no further therapy; and three other equivalent subgroups corresponding to those who were assigned to ABMT as initial treatment.

For each subgroup we first computed the survival curve using the NA estimator. Based on these estimates we obtained the survival curves for the four regimes through the formulae (2.8)-(2.11) described in Section 2. The overall survival curves for these four treatment regimes (CCR, CNR, ACR, ANR) are shown in Figure 2(a). The survival curves were close to each other early into the study, but the survival rates for treatment regime ACR appeared to be better than those for other regimes later in the study. We then fitted parametric models to estimate the pattern-specific survival functions. For each subgroup, we fitted Weibull, log-logistic and log-normal distributions and chose the best model with the lowest Akaike information criterion (AIC). Table 5 presents the AICs for a total of 18 models involved. The AIC was minimum when the log-normal (LN) distribution was used as the parametric model for all six subgroups. Figure 2(b) shows the corresponding overall survival curves for the four treatment regimes. The pattern follows that of the curves obtained using the NA estimator. At the late stage of the study, the survival curve for treatment regime ACR seemed to differ somewhat from the other three survival curves (CCR, CNR, ANR), while the overall survival curves for treatment regimes CCR, CNR, and ANR followed each other closely. Figure 3 compares the NA estimator and LN model for each regime separately. The survival curves from NA estimator and LN model were similar to each other, with small

deviations at specific time points. Under all four treatment regimes, the survival estimates from the LN distribution were always slightly larger than the NA estimates during the first five years of the study.

We further compared the treatment regimes using the NA estimator and LN model respectively. A total of six null hypotheses comparing different treatment regimes were tested using the weighted log-rank test (Table 6). For example, we tested if there was a significant difference in the hazards of treatment regimes sharing the same initial treatment of “ABMT.” The null hypothesis can be described as $H_6: \lambda_{ACR}(t) = \lambda_{ANR}(t)$. The tests resulted in p-values of 0.19 and 0.21 based on the NA estimator and LN model respectively, showing that there was no statistically significant difference in the hazards of treatment regimes sharing the same initial treatment of “ABMT.” Thus, for patients assigned to initial treatment of “ABMT”, they would have a similar survival regardless of the maintenance therapy they were assigned to. We also tested if there was a significant difference in the hazards of treatment regimes sharing the same maintenance treatment of “cis-RA.” The tests of the null hypothesis $H_2: \lambda_{CCR}(t) = \lambda_{ACR}(t)$ resulted in p-values of 0.22 and 0.25 for the NA estimator and LN model respectively. Therefore, we would conclude that for patients who were assigned to “cis-RA” as maintenance treatment upon response, there was no significant difference in the overall survival between those who were assigned to “chemotherapy” and those who were assigned to “ABMT” as initial treatment. None of the pairwise comparisons were found to be statistically significant, indicating that there might not be a significant difference among four treatment regimes in the neuroblastoma study.

5 Discussion

In this article, we demonstrated the use of the NA estimator and parametric models in a pattern-mixture-type setting for estimating the effects of treatment regimes from two-stage randomization designs. The methodology adopts the pattern-mixture models from the missing data literature (Molenbergh and Kenward 2007), where the parameter estimates are estimated for each missing data pattern and then overall estimate is calculated as a weighted average of the pattern-specific estimates. We also demonstrated how to compare different regimes in terms of their hazards using the weighted log-rank tests.

The simulation study showed that survival estimates could differ when a model other than the true model was fitted. Thus, the survival estimates were affected somewhat by the choice of distributions when fitting the parametric model. The survival estimates obtained using the NA estimator appear to be unbiased. Thus, the NA estimator can be used with confidence to estimate pattern-specific survival distributions and then combined using pattern-mixture approach to produce regime-specific survival distributions. Parametric models introduce strong assumptions on the distribution of survival patterns and hence are subjected to bias. The performance of the NA estimator and parametric models were also evaluated at other response rates. We observed similar results at various response rates except that the NA estimator provided slightly biased estimates for smaller sample sizes. Note that these methods require that there be enough observations in each pattern to achieve a reasonable fit of the parametric models, or to allow for consistent NA estimators. Therefore, cautions should be exercised in applying these methods when the sample size is small.

Furthermore, the simulation study also showed that the rejection rates obtained using all the methods were similar to each other under all scenarios, demonstrating that the weighted log-rank test was robust regardless of the methods for estimating the survival quantities. Unlike previously proposed test procedures (Lokhnygina and Helderbrand 2007; Feng and Wahed 2008), this test can be used to compare regimes that share the same induction treatment (and thus some of the data are common to both regimes). When we utilized the weighted log-rank tests to compare different treatment regimes in terms of their hazards, six pairs of null hypotheses were tested separately. In this circumstance, how small a p-value is considered significant needs further discussion. As we know, many statistical techniques have been developed for protecting the type I errors in multiple comparisons, e.g. Bonferroni correction, Holm's sequential rejection procedure, etc. Similar techniques could possibly be adopted here. However, we decided to pursue this in a separate research study.

Acknowledgments

We would like to thank COG Neuroblastoma Disease Committee for kind permission to use the neuroblastoma data set, especially Dr. Wendy London for her help during the application process. We thank the Associate Editor and the referees of this article for their insightful comments. Dr. Wahed's research was in part supported by a National Institute of Mental Health Grant P30 MH090333.

Appendix: Derivation of expression (2.1)

Denote by T_{jk}^* be the potential survival time of a generic patient had he/she followed pattern P_{jk} , $j = 1, 2, \dots, J$, $k = 0, 1, \dots, K$. Note that in practice we could observe only one of the $J(K+1)$ of these potential survival times for each patient, because one patient can belong to only one pattern. Let T_{jk} denote the survival time under the regime $A_j B_k$. By definition, the regime $A_j B_k$ dictates that patient receive A_j as the initial treatment. Therefore, the following derivation will be conditional on receiving A_j as the initial treatment. By definition, the hazard function of the regime $A_j B_k$ is given by

$$\begin{aligned}
 \lambda_{A_j B_k}(t) &= \lim_{\Delta t \rightarrow 0} \frac{P(t < T_{jk} < t + \Delta t)}{P(T_{jk} > t) \Delta t} \\
 &= \lim_{\Delta t \rightarrow 0} \frac{\sum_{r=0}^1 P(t < T_{jk}^* < t + \Delta t | R=r) P(R=r)}{\sum_{r=0}^1 P(T_{jk}^* > t | R=r) P(R=r) \Delta t} \\
 &= \lim_{\Delta t \rightarrow 0} \frac{\sum_{r=0}^1 P(t < T_{jk}^* < t + \Delta t | R=r) \pi_j^r (1 - \pi_j)^{1-r}}{\sum_{r=0}^1 P(T_{jk}^* > t | R=r) \pi_j^r (1 - \pi_j)^{1-r} \Delta t} \\
 &= \lim_{\Delta t \rightarrow 0} \frac{P(t < T_{jk}^* < t + \Delta t) \pi_j + P(t < T_{j0}^* < t + \Delta t) (1 - \pi_j)}{[P(T_{jk}^* > t) \pi_j + P(T_{j0}^* > t) (1 - \pi_j)] \Delta t} \\
 &= \frac{S_{jk}(t) S_{j0}(t)}{\pi_j S_{jk}(t) + (1 - \pi_j) S_{j0}(t)} \\
 &= \lim_{\Delta t \rightarrow 0} \left\{ \frac{P(t < T_{jk}^* < t + \Delta t)}{P(T_{jk}^* > t) \Delta t} \frac{\pi_j}{S_{j0}(t)} + \frac{P(t < T_{j0}^* < t + \Delta t)}{P(T_{j0}^* > t) \Delta t} \frac{1 - \pi_j}{S_{jk}(t)} \right\} \\
 &= \frac{\pi_j S_{jk}(t)}{\pi_j S_{jk}(t) + (1 - \pi_j) S_{j0}(t)} \lim_{\Delta t \rightarrow 0} \frac{P(t < T_{jk}^* < t + \Delta t)}{P(T_{jk}^* > t) \Delta t} \\
 &\quad + \frac{(1 - \pi_j) S_{j0}(t)}{\pi_j S_{jk}(t) + (1 - \pi_j) S_{j0}(t)} \lim_{\Delta t \rightarrow 0} \frac{P(t < T_{j0}^* < t + \Delta t)}{P(T_{j0}^* > t) \Delta t} \\
 &= \phi_{jk}(t) \lambda_{jk}(t) + \{1 - \phi_{jk}(t)\} \lambda_{j0}(t),
 \end{aligned}$$

where $\varphi_{jk}(t)$ is defined in (2.2).

References

- Feng W, Wahed AS. Supremum weighted log-rank test and sample size for comparing two-stage adaptive treatment strategies. *Biometrika*. 2008; 95:695–707.
- Fleming, TR.; Harrington, DP. Counting Processes and Survival Analysis. John Wiley & Sons, Inc.; New York: 1991.
- Guo X, Tsiatis AA. A Weighted risk estimator for survival distributions in two-stage randomization designs with censored survival data. *The International Journal of Biostatistics*. 2005; 1:1–15.
- Hernan MA, Lanoy E, Costagliola D, Robins JM. Comparison of dynamic treatment regimes via inverse probability weighting. *Basic and Clinical Pharmacology and Toxicology*. 2006; 98:237–242. [PubMed: 16611197]
- Klein, JP.; Moeschberger, ML. Survival Analysis Techniques for Censored and Truncated Data. Springer; New York: 2003.
- Little RJA. Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association*. 1993; 88:125–134.
- Lokhnygina Y, Helderbrand JD. Cox regression methods for two-stage randomization designs. *Biometrics*. 2007; 63:422–428. [PubMed: 17425633]
- Lunceford JK, Davidian M, Tsiatis AA. Estimation of survival distributions of treatment policies in two-stage randomization designs in clinical trials. *Biometrics*. 2002; 58:48–57. [PubMed: 11890326]
- Matthay KK, Villablanca JG, Seeger RC, Stram DO, Harris RE, Ramsay NK, Swift P, Shimada H, Black CT, Brodeur GM, Gerbing RB, Reynolds CP. Treatment of high-risk neuroblastoma with intensive chemotherapy, radiotherapy, autologous bone marrow transplantation, and 13-cis-retinoic acid. *The New England Journal of Medicine*. 1999; 341:1165–1173. [PubMed: 10519894]
- Matthay KK, Reynolds CP, Seeger RC, Shimada H, Adkins ES, Haas-Kogan D, Gerbing RB, London WB, Villablanca JG. Long-term results for children with high-risk neuroblastoma treated on a randomized trial of myeloablative therapy followed by 13-cisretinoic acid: A Children's Oncology Group study. *Journal of Clinical Oncology*. 2009; 27:1007–1013. [PubMed: 19171716]
- Molenbergh, G.; Kenward, MG. Missing Data in Clinical Studies. John Wiley & Sons Ltd; England: 2007.
- Murphy SA. Optimal dynamic treatment regimes. *Journal of the Royal Statistical Society: Series B*. 2003; 65:331–355.
- Murphy SA, Bingham D. Screening experiments for developing dynamic treatment regimes. *Journal of the American Statistical Association*. 2009; 104:391–409. [PubMed: 20589222]
- Murphy SA, van der Laan MJ, Robins JM. Marginal mean models for dynamic regimes. *Journal of the American Statistical Association*. 2001; 96:1410–1423. [PubMed: 20019887]
- Robins JM. A new approach to causal inference in mortality studies with sustained exposure periods-application to control of the healthy survivor effect. *Mathematical Modelling*. 1986; 7:1393–1512.
- Robins JM. Addendum to “A new approach to causal inference in mortality studies with sustained exposure periods-application to control of the healthy survivor effect” [Math Model 1986;7:1393-1512]. *Computers & Mathematics with Applications*. 1987; 14:923–945.
- Wahed AS, Tsiatis AA. Optimal estimator for the survival distribution and related quantities for treatment policies in two-stage randomization designs in clinical trials. *Biometrics*. 2004; 60:124–133. [PubMed: 15032782]
- Wahed AS, Tsiatis AA. Semiparametric efficient estimation of survival distributions in two-stage randomization designs in clinical trials with censored data. *Biometrika*. 2006; 93:163–177.

Figure 1.

Diagram for a simple two-stage randomization design where only two treatment options are available at each stage

Figure 2.

Overall survival curves under four treatment regimes in the neuroblastoma study. CCR: “treat with chemotherapy followed by cis-RA if no progressive disease”, CNR: “treat with chemotherapy followed by no further therapy if no progressive disease”, AC R: “treat with ABMT followed by cis-RA if no progressive disease”, ANR: “treat with ABMT followed by no further therapy if no progressive disease”

Figure 3.

Overall survival curves under four treatment regimes in the neuroblastoma study. CCR:

“treat with chemotherapy followed by cis-RA if no progressive disease”, CNR: “treat with chemotherapy followed by no further therapy if no progressive disease”, ACR: “treat with ABMT followed by cis-RA if no progressive disease”, ANR: “treat with ABMT followed by no further therapy if no progressive disease” LN: “Log-normal Model”, NA: “Nelson-Aalen Estimator”

Table 1

Simulation scenarios

Scenario	Response Rate	Censoring Percentage	Responders	Nonresponders	True Survival Rates				
					t(year)	$S_{A_1B_1}(t)$	$S_{A_1B_2}(t)$	$S_{A_2B_1}(t)$	$S_{A_2B_2}(t)$
I	50%	30%	Weibull	Weibull	1	0.76	0.76	0.63	0.56
					3	0.52	0.52	0.37	0.25
II	50%	30%	log-normal	log-normal	1	0.73	0.81	0.73	0.83
					3	0.43	0.55	0.43	0.35
III	50%	30%	log-logistic	log-logistic	1	0.67	0.70	0.58	0.67
					3	0.39	0.44	0.31	0.39
IV	50%	30%	Weibull	log-normal	1	0.83	0.79	0.79	0.75
					3	0.58	0.48	0.48	0.38
V	50%	30%	log-normal	log-logistic	1	0.75	0.75	0.54	0.49
					3	0.61	0.61	0.28	0.26

Table 2

Survival estimate (EST), absolute bias (BIAS) in parentheses, asymptotic standard error (ASE), empirical standard deviation (ESD) and coverage probability (CP) of 95% confidence interval at $t = 1$ year under scenarios I - V.

Scenario	$A_1 B_1$			$A_1 B_2$			$A_2 B_1$			$A_2 B_2$		
	EST(BIAS)	ASE/ ESD	CP (%)	EST(BIAS)	ASE/ ESD	CP (%)	EST(BIAS)	ASE/ ESD	CP (%)	EST(BIAS)	ASE/ ESD	CP (%)
I	W	0.76(0.029)	8	0.76(0.029)	0.029	95.6	0.64(0.029)	0.033	94.2	0.56(0.036)	0.036	93.0
	EI	0.00(0.00)	8	0.00(0.00)	0.00	6	0.00(0.00)	0.03	2	0.00(0.00)	0.03	0
	LN	0.72(0.03)	1	0.72(0.03)	0.03	81.6	0.60(0.04)	0.034	78.6	0.51(0.05)	0.036	71.6
	LG	0.74(0.01)	0	0.74(0.01)	0.01	94.0	0.62(0.01)	0.035	93.8	0.54(0.02)	0.038	88.8
	W	0.73(0.03)	9	0.73(0.03)	0.03	86.2	0.61(0.02)	0.034	84.0	0.54(0.02)	0.036	90.4
	EI	0.03(0.02)	9	0.03(0.02)	0.02	2	0.03(0.03)	0.03	0	0.03(0.03)	0.03	4
	LN	0.74(0.02)	2	0.74(0.02)	0.02	90.4	0.61(0.03)	0.035	90.2	0.52(0.04)	0.036	77.8
	LG	0.76(0.00)	3	0.76(0.00)	0.00	95.0	0.64(0.00)	0.039	94.8	0.56(0.04)	0.043	94.4
	W	0.75(0.02)	8	0.75(0.02)	0.02	87.8	0.75(0.02)	0.031	92.4	0.82(0.01)	0.024	97.4
	EI	0.02(0.00)	8	0.02(0.00)	0.00	8	0.02(0.01)	0.02	2	0.01(0.01)	0.02	4
II	LN	0.74(0.00)	3	0.74(0.00)	0.00	93.2	0.74(0.00)	0.031	93.4	0.84(0.00)	0.023	94.8
	LG	0.75(0.01)	4	0.75(0.01)	0.01	88.8	0.75(0.01)	0.032	92.8	0.84(0.00)	0.023	93.6
	W	0.75(0.02)	8	0.75(0.02)	0.02	87.8	0.75(0.02)	0.031	92.4	0.82(0.01)	0.024	97.4
	EI	0.02(0.00)	8	0.02(0.00)	0.00	8	0.02(0.01)	0.02	2	0.01(0.01)	0.02	4
	LN	0.74(0.00)	3	0.74(0.00)	0.00	93.2	0.74(0.00)	0.031	93.4	0.84(0.00)	0.023	94.8
	LG	0.75(0.01)	4	0.75(0.01)	0.01	88.8	0.75(0.01)	0.032	92.8	0.84(0.00)	0.023	93.6
	W	0.75(0.02)	8	0.75(0.02)	0.02	87.8	0.75(0.02)	0.031	92.4	0.82(0.01)	0.024	97.4
	EI	0.02(0.00)	8	0.02(0.00)	0.00	8	0.02(0.01)	0.02	2	0.01(0.01)	0.02	4
	LN	0.74(0.00)	3	0.74(0.00)	0.00	93.2	0.74(0.00)	0.031	93.4	0.84(0.00)	0.023	94.8
	LG	0.75(0.01)	4	0.75(0.01)	0.01	88.8	0.75(0.01)	0.032	92.8	0.84(0.00)	0.023	93.6

	$A_1 B_1$			$A_1 B_2$			$A_2 B_1$			$A_2 B_2$		
	$A_1 B_1$			$A_1 B_2$			$A_2 B_1$			$A_2 B_2$		
	EST(BIAS)	ASE/ ESD	CP (%)	EST(BIAS)	ASE/ ESD	CP (%)	EST(BIAS)	ASE/ ESD	CP (%)	EST(BIAS)	ASE/ ESD	CP (%)
W	0.74(0.01)	0.031 /0.03	92.0	0.81(0.00)	0.025 /0.02	94.2	0.74(0.01)	0.031 /0.02	94.2	0.82(0.02)	0.024 /0.02	93.2
EI		1		4				8			2	
LN												
LN	0.74(0.01)	0.032 /0.03	91.4	0.82(0.01)	0.026 /0.02	91.4	0.74(0.01)	0.032 /0.03	92.6	0.84(0.01)	0.023 /0.02	92.6
LG		4		6				1			3	
NA	0.73(0.00)	0.038 /0.04	92.6	0.81(0.00)	0.031 /0.03	95.0	0.73(0.00)	0.038 /0.03	93.2	0.84(0.00)	0.028 /0.02	94.0
		0		0				9			8	
<hr/>												
III	$A_1 B_1$			$A_1 B_2$			$A_2 B_1$			$A_2 B_2$		
	$A_1 B_1$			$A_1 B_2$			$A_2 B_1$			$A_2 B_2$		
	EST(BIAS)	ASE/ ESD	CP (%)	EST(BIAS)	ASE/ ESD	CP (%)	EST(BIAS)	ASE/ ESD	CP (%)	EST(BIAS)	ASE/ ESD	CP (%)
W	0.69(0.02)	0.032 /0.02	89.4	0.73(0.02)	0.030 /0.02	89.8	0.62(0.04)	0.035 /0.03	82.6	0.69(0.02)	0.032 /0.03	87.8
EI		8		8				3			1	
LN	0.66(0.01)	0.032 /0.03	94.6	0.70(0.00)	0.030 /0.03	93.0	0.58(0.01)	0.035 /0.03	94.4	0.66(0.01)	0.033 /0.03	93.0
		3		2				7			6	
LG	0.67(0.00)	0.034 /0.03	95.8	0.70(0.00)	0.031 /0.03	95.4	0.59(0.00)	0.037 /0.03	94.4	0.67(0.00)	0.034 /0.03	94.8
		2		1				8			5	
W	0.67(0.00)	0.032 /0.03	96.4	0.70(0.00)	0.030 /0.03	94.6	0.59(0.01)	0.035 /0.03	93.6	0.67(0.00)	0.033 /0.03	94.6
EI		0		0				5			3	
LN												
LN	0.66(0.01)	0.033 /0.03	94.0	0.70(0.00)	0.031 /0.03	94.2	0.58(0.01)	0.036 /0.03	95.0	0.66(0.01)	0.034 /0.03	93.0
LG		4		3				8			7	
NA	0.67(0.00)	0.039 /0.03	95.2	0.71(0.00)	0.036 /0.03	93.8	0.59(0.00)	0.042 /0.04	92.6	0.67(0.00)	0.039 /0.04	93.4
		8		6				5			0	

		$A_1 B_1$			$A_1 B_2$			$A_2 B_1$			$A_2 B_2$		
		EST(BIAS)	ASE/ ESD	CP (%)	EST(BIAS)	ASE/ ESD	CP (%)	EST(BIAS)	ASE/ ESD	CP (%)	EST(BIAS)	ASE/ ESD	CP (%)
IV	W	0.83(0.00)	0.025 /0.02	95. 8	0.79(0.00)	0.029 /0.02	96. 8	0.79(0.00)	0.029 /0.02	95. 0	0.75(0.00)	0.031 /0.02	96. 6
	EI		3			6			7			8	
	LN	0.83(0.00)	0.026 /0.02	92. 0	0.78(0.01)	0.030 /0.03	91. 8	0.78(0.01)	0.029 /0.03	89. 0	0.73(0.02)	0.032 /0.03	88. 4
			8			3			5			5	
LG	W	0.84(0.01)	0.026 /0.02	92. 4	0.80(0.01)	0.030 /0.03	92. 4	0.80(0.01)	0.030 /0.03	91. 6	0.75(0.01)	0.033 /0.03	93. 2
	EI		6			0			2			4	
	LN	0.83(0.00)	0.025 /0.02	93. 0	0.79(0.00)	0.029 /0.02	94. 6	0.79(0.00)	0.029 /0.03	94. 2	0.75(0.00)	0.031 /0.03	95. 0
			6			8			0			0	
LN	W	0.83(0.00)	0.026 /0.02	91. 6	0.78(0.01)	0.030 /0.03	92. 2	0.78(0.00)	0.030 /0.03	90. 2	0.73(0.02)	0.032 /0.03	91. 4
	EI		8			3			5			5	
	LN	0.83(0.00)	0.031 /0.02	95. 8	0.79(0.00)	0.035 /0.03	95. 2	0.79(0.00)	0.035 /0.03	93. 4	0.75(0.00)	0.038 /0.03	94. 8
	NA		9			3			5			7	

$A_1 B_1$		$A_1 B_2$			$A_2 B_1$			$A_2 B_2$					
	EST(BIAS)	ASE/ ESD	CP (%)	EST(BIAS)	ASE/ ESD	CP (%)	EST(BIAS)	ASE/ ESD	CP (%)	EST(BIAS)	ASE/ ESD	CP (%)	
V	W	0.77(0.02)	0.026	88. 2	0.77(0.02)	0.026	89. 0	0.58(0.04)	0.035	79. 0	0.54(0.05)	0.036	70. 6
	EI		/0.02	5		/0.02	5		/0.03	0		/0.03	5
	LN	0.75(0.00)	0.027	94. 8	0.75(0.00)	0.027	94. 8	0.55(0.01)	0.035	95. 8	0.50(0.01)	0.036	94. 2
			/0.02	7		/0.02	7		/0.03	3		/0.03	7
LG	W	0.75(0.00)	0.028	94. 6	0.75(0.00)	0.028	95. 2	0.55(0.01)	0.037	96. 4	0.50(0.01)	0.038	94. 4
	EI		/0.02	8		/0.02	8		/0.03	4		/0.03	9
	LN	0.74(0.01)	0.027	94. 4	0.74(0.01)	0.027	94. 4	0.56(0.01)	0.035	96. 4	0.52(0.02)	0.036	90. 2
			/0.02	7		/0.02	7		/0.03	0		/0.03	5
LN	W	0.75(0.00)	0.028	95. 0	0.75(0.00)	0.028	95. 4	0.54(0.00)	0.036	95. 8	0.50(0.00)	0.036	95. 0
	EI		/0.02	8		/0.02	8		/0.03	4		/0.03	8
	LN	0.75(0.00)	0.028	95. 0	0.75(0.00)	0.028	95. 4	0.54(0.00)	0.036	95. 8	0.50(0.00)	0.036	95. 0
			/0.02	8		/0.02	8		/0.03	4		/0.03	8

	$A_1 B_1$			$A_1 B_2$			$A_2 B_1$			$A_2 B_2$		
	EST(	ASE/	CP	EST(	ASE/	CP	EST(	ASE/	CP	EST(	ASE/	CP
	BIAS	ESD	(%)	BIAS	ESD	(%)	BIAS	ESD	(%)	BIAS	ESD	(%)
NA	0.75(	0.031	94.	0.75(	0.031	94.	0.54(	0.042	95.	0.50(	0.043	94.
	0.00)	/0.03	4	0.00)	/0.03	2	0.00)	/0.04	2	0.00)	/0.04	4
		1			1			0			3	

WEI: Weibull distributions, LN: log-normal distributions, LG: log-logistic distributions, WEILN: Weibull distributions for responders and log-normal distributions for nonresponders, LNLG: log-normal distributions for responders and log-logistic distributions for nonresponders, NA: Nelson-Aalen estimators

Table 3

Survival estimate (EST), absolute bias (BIAS) in parentheses, asymptotic standard error (ASE), empirical standard deviation (ESD) and coverage probability (CP) of 95% confidence interval at $t = 3$ year under scenarios I - V.

Scenario	$A_1 B_1$			$A_1 B_2$			$A_2 B_1$			$A_2 B_2$			
	EST (BI AS)	ASE /ES D	C P(%)	EST (BI AS)	ASE /ES D	C P(%)	EST (BI AS)	ASE /ES D	C P(%)	EST (BI AS)	ASE /ES D	C P(%)	
I	W	0.52 (0.0 0)	0.03 7/0.0 36	94 .6	0.52 (0.0 0)	0.03 7/0.0 39	92 .2	0.37 (0.0 0)	0.03 7/0.0 38	94 .0	0.25 (0.0 0)	0.03 4/0.0 36	92 .6
	LN	0.52 (0.0 1)	0.03 7/0.0 36	96 .2	0.51 (0.0 1)	0.03 7/0.0 38	93 .8	0.37 (0.0 0)	0.03 8/0.0 36	95 .6	0.26 (0.0 1)	0.03 5/0.0 31	97 .4
	LG	0.52 (0.0 0)	0.03 7/0.0 36	95 .6	0.52 (0.0 0)	0.03 7/0.0 39	94 .8	0.38 (0.0 1)	0.03 8/0.0 38	95 .4	0.26 (0.0 1)	0.03 5/0.0 33	97 .0
	W	0.53 (0.0 0)	0.03 7/0.0 34	95 .4	0.53 (0.0 0)	0.03 7/0.0 37	94 .8	0.39 (0.0 2)	0.03 7/0.0 37	92 .0	0.27 (0.0 2)	0.03 4/0.0 35	93 .6
	LN	0.51 (0.0 1)	0.03 7/0.0 36	95 .0	0.51 (0.0 1)	0.03 7/0.0 39	93 .6	0.37 (0.0 0)	0.03 7/0.0 35	96 .4	0.26 (0.0 1)	0.03 4/0.0 31	97 .8
	A	0.53 (0.0 0)	0.04 2/0.0 43	93 .8	0.53 (0.0 0)	0.04 2/0.0 45	93 .4	0.38 (0.0 0)	0.04 4/0.0 45	93 .4	0.26 (0.0 1)	0.04 1/0.0 43	93 .4
II	W	0.46 (0.0 4)	0.03 7/0.0 41	80 .0	0.58 (0.0 3)	0.03 5/0.0 35	86 .6	0.46 (0.0 3)	0.03 7/0.0 39	85 .4	0.40 (0.0 5)	0.03 7/0.0 39	72 .6
	LN	0.43 (0.0 0)	0.03 8/0.0 40	93 .4	0.55 (0.0 0)	0.03 6/0.0 35	94 .8	0.43 (0.0 0)	0.03 8/0.0 39	94 .6	0.35 (0.0 0)	0.03 6/0.0 38	93 .8
	LG	0.42 (0.0 0)	0.03 9/0.0 42	93 .6	0.55 (0.0 0)	0.03 7/0.0 35	95 .4	0.42 (0.0 1)	0.03 9/0.0 41	93 .4	0.34 (0.0 1)	0.03 9/0.0 41	91 .4

	$A_1 B_1$			$A_1 B_2$			A^*			$A_2 B_1$			$A_2 B_2$		
	EST (BI AS)	ASE /ES D	C P(%)	EST (BI AS)	ASE /ES D	C P(%)	EST (BI AS)	ASE /ES D	C P(%)	EST (BI AS)	ASE /ES D	C P(%)	EST (BI AS)	ASE /ES D	C P(%)
W	0.44 (0.0 2)	0.03 8/0.0 40	91 .2	0.56 (0.0 1)	0.03 6/0.0 34	94 .4	0.44 (0.0 1)	0.03 8/0.0 38	93 .6	0.44 (0.0 1)	0.03 8/0.0 38	93 .6	0.38 (0.0 3)	0.03 7/0.0 38	86 .8
EI															
LN															
LN	0.42 (0.0 1)	0.03 8/0.0 41	93 .6	0.55 (0.0 0)	0.03 6/0.0 35	95 .2	0.42 (0.0 1)	0.03 8/0.0 39	93 .8	0.42 (0.0 1)	0.03 8/0.0 39	93 .8	0.34 (0.0 1)	0.03 7/0.0 38	93 .0
LG															
N	0.43 (0.0 0)	0.04 4/0.0 45	93 .4	0.56 (0.0 1)	0.04 1/0.0 40	94 .8	0.43 (0.0 0)	0.04 4/0.0 44	94 .8	0.43 (0.0 0)	0.04 4/0.0 44	94 .8	0.36 (0.0 1)	0.04 5/0.0 46	91 .0
A															

	$A_1 B_1$			$A_1 B_2$			$A_2 B_1$			$A_2 B_2$		
	EST (BI AS)	ASE /ES D	C P(%)	EST (BI AS)	ASE /ES D	C P(%)	EST (BI AS)	ASE /ES D	C P(%)	EST (BI AS)	ASE /ES D	C P(%)
III												
W	0.42 (0.0 2)	0.03 8/0.0 40	89 .6	0.46 (0.0 2)	0.03 7/0.0 41	90 .4	0.33 (0.0 2)	0.03 7/0.0 40	89 .0	0.42 (0.0 2)	0.03 8/0.0 42	88 .4
EI												
LN	0.40 (0.0 0)	0.03 8/0.0 38	93 .8	0.44 (0.0 0)	0.03 8/0.0 40	93 .2	0.32 (0.0 1)	0.03 7/0.0 36	95 .8	0.40 (0.0 0)	0.03 8/0.0 41	93 .0
LG												
LN	0.39 (0.0 0)	0.03 6/0.0 39	93 .8	0.44 (0.0 0)	0.03 8/0.0 41	93 .0	0.31 (0.0 0)	0.03 8/0.0 38	94 .6	0.39 (0.0 0)	0.03 9/0.0 41	93 .2
W	0.41 (0.0 2)	0.03 8/0.0 38	92 .8	0.46 (0.0 1)	0.03 7/0.0 39	92 .4	0.33 (0.0 2)	0.03 7/0.0 38	91 .8	0.41 (0.0 2)	0.03 8/0.0 40	91 .0
EI												
LN												
LN	0.39 (0.0 0)	0.03 8/0.0 38	93 .4	0.44 (0.0 1)	0.03 7/0.0 40	92 .2	0.31 (0.0 0)	0.03 7/0.0 36	95 .0	0.39 (0.0 0)	0.03 8/0.0 40	92 .6
LG												
N	0.40 (0.0 0)	0.04 4/0.0 43	95 .2	0.45 (0.0 0)	0.04 3/0.0 46	93 .0	0.31 (0.0 1)	0.04 3/0.0 43	94 .8	0.40 (0.0 0)	0.04 4/0.0 46	94 .0
A												

		$A_1 B_1$			$A_1 B_2$			$A_2 B_1$			$A_2 B_2$		
		EST (BI AS)	ASE /ES D	C P(%)	EST (BI AS)	ASE /ES D	C P(%)	EST (BI AS)	ASE /ES D	C P(%)	EST (BI AS)	ASE /ES D	C P(%)
IV	W	0.60	0.03	87	0.51	0.03	90	0.51	0.03	88	0.40	0.03	92
	EI	(0.0 3)	5/0.0 35	.0	(0.0 3)	7/0.0 36	.4	(0.0 3)	7/0.0 39	.0	(0.0 2)	6/0.0 35	.0
LN		0.57	0.03	94	0.46	0.03	92	0.46	0.03	91	0.36	0.03	95
		(0.0 1)	6/0.0 36	.6	(0.0 2)	8/0.0 35	.8	(0.0 2)	7/0.0 38	.6	(0.0 2)	6/0.0 31	.2
LG		0.57	0.03	95	0.47	0.03	94	0.47	0.03	92	0.36	0.03	94
		(0.0 1)	7/0.0 37	.0	(0.0 1)	9/0.0 37	.6	(0.0 1)	9/0.0 41	.4	(0.0 2)	8/0.0 34	.8
W		0.58	0.03	94	0.48	0.03	96	0.48	0.03	93	0.38	0.03	96
		(0.0 0)	6/0.0 35	.4	(0.0 0)	7/0.0 36	.4	(0.0 0)	7/0.0 39	.8	(0.0 0)	6/0.0 35	.2
LN		0.56	0.03	93	0.46	0.03	91	0.46	0.03	89	0.35	0.03	93
		(0.0 1)	7/0.0 37	.8	(0.0 2)	8/0.0 36	.8	(0.0 2)	8/0.0 39	.0	(0.0 2)	7/0.0 32	.4
N		0.58	0.04	94	0.48	0.04	96	0.49	0.04	92	0.38	0.04	96
		(0.0 0)	2/0.0 41	.6	(0.0 0)	5/0.0 42	.4	(0.0 1)	4/0.0 45	.8	(0.0 1)	4/0.0 41	.4

		$A_1 B_1$			$A_1 B_2$			$A_2 B_1$			$A_2 B_2$		
		EST (BI AS)	ASE /ES D	C P(%)	EST (BI AS)	ASE /ES D	C P(%)	EST (BI AS)	ASE /ES D	C P(%)	EST (BI AS)	ASE /ES D	C P(%)
V	W	0.62	0.03	92	0.61	0.03	93	0.30	0.03	89	0.28	0.03	89
	EI	(0.0 1)	3/0.0 34	.8	(0.0 0)	3/0.0 34	.8	(0.0 2)	6/0.0 38	.2	(0.0 2)	6/0.0 40	.8
LN		0.62	0.03	94	0.61	0.03	94	0.28	0.03	93	0.27	0.03	93
		(0.0 0)	3/0.0 33	.2	(0.0 0)	3/0.0 33	.2	(0.0 0)	6/0.0 36	.4	(0.0 0)	5/0.0 38	.8
LG		0.61	0.03	94	0.60	0.03	94	0.28	0.03	92	0.26	0.03	92
		(0.0 0)	4/0.0 33	.8	(0.0 1)	3/0.0 33	.4	(0.0 0)	7/0.0 37	.8	(0.0 0)	6/0.0 39	.2
W		0.61	0.03	94	0.60	0.03	94	0.30	0.03	90	0.28	0.03	89
		(0.0 0)	3/0.0 33	.0	(0.0 1)	4/0.0 32	.8	(0.0 2)	6/0.0 36	.2	(0.0 2)	6/0.0 38	.6
LN		0.61	0.03	94	0.60	0.03	94	0.28	0.03	94	0.26	0.03	94
		(0.0 0)	4/0.0 33	.8	(0.0 0)	4/0.0 33	.8	(0.0 0)	5/0.0 35	.4	(0.0 0)	5/0.0 37	.0

$A_1 B_1$			$A_1 B_2$			$A_2 B_1$			$A_2 B_2$		
EST (BI AS)	ASE /ES D	C P(%)	EST (BI AS)	ASE /ES D	C P(%)	EST (BI AS)	ASE /ES D	C P(%)	EST (BI AS)	ASE /ES D	C P(%)
N	0.61	0.03	94	0.61	0.03	94	0.29	0.04	95	0.27	0.04
A	(0.0	6/0.0	.4	(0.0	7/0.0	.4	(0.0	2/0.0	.0	(0.0	1/0.0
	0)	37		0)	37		0)	41		0)	43

WEI: Weibull distributions, LN: log-normal distributions, LG: log-logistic distributions, WEILN: Weibull distributions for responders and log-normal distributions for nonresponders, LNLG: log-normal distributions for responders and log-logistic distributions for nonresponders, NA: Nelson-Aalen estimators

Table 4

The rejection rates of the weighted log-rank tests under scenarios I - V. H_1 : $\lambda_{A_1B_1}(t) = \lambda_{A_1B_2}(t)$, H_2 : $\lambda_{A_1B_1}(t) = \lambda_{A_1B_1}(t)$, H_3 : $\lambda_{A_1B_1}(t) = \lambda_{A_2B_2}(t)$, H_4 : $\lambda_{A_1B_2}(t) = \lambda_{A_2B_1}(t)$

Scenario		H_1	H_2	H_3	H_4
I	WEI	0.062	0.878	1.000	0.848
	LN	0.062	0.876	1.000	0.840
	LG	0.070	0.878	1.000	0.850
	WEILN	0.062	0.882	1.000	0.844
	LNLG	0.062	0.884	1.000	0.848
	NA	0.050	0.214	0.592	0.212

		H_1	H_2	H_3	H_4
II	WEI	0.850	0.062	0.868	0.768
	LN	0.870	0.054	0.666	0.790
	LG	0.856	0.068	0.628	0.780
	WEILN	0.854	0.060	0.788	0.770
	LNLG	0.870	0.062	0.646	0.784
	NA	0.684	0.042	0.132	0.258

		H_1	H_2	H_3	H_4
III	WEI	0.178	0.294	0.072	0.570
	LN	0.196	0.326	0.060	0.638
	LG	0.190	0.318	0.064	0.630
	WEILN	0.176	0.300	0.064	0.594
	LNLG	0.196	0.324	0.062	0.658
	NA	0.136	0.148	0.036	0.298

		H_1	H_2	H_3	H_4
IV	WEI	0.876	0.736	1.000	0.072
	LN	0.860	0.708	1.000	0.066
	LG	0.868	0.720	1.000	0.064
	WEILN	0.874	0.736	1.000	0.074
	LNLG	0.860	0.704	1.000	0.062
	NA	0.672	0.284	0.734	0.028

		H_1	H_2	H_3	H_4
V	WEI	0.980	1.000	1.000	0.970
	LN	0.980	1.000	1.000	0.994
	LG	0.984	1.000	1.000	0.988
	WEILN	0.976	1.000	1.000	0.982
	LNLG	0.984	1.000	1.000	0.988
	NA	0.870	0.609	0.603	0.413

WEI: Weibull distributions, LN: log-normal distributions, LG: log-logistic distributions, WEILN: Weibull distributions for responders and log-normal distributions for nonresponders, LNLG: log-normal distributions for responders and log-logistic distributions for nonresponders, NA: Nelson-Aalen estimators, boldfaced numbers represent estimated type I errors

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 5

AIC values for a total of 18 models fitted to the neuroblastoma dataset

Initial Therapy	Response	Second-stage Therapy	Weibull	log-logistic	log-normal
Chemotherapy	No		1109.55	1192.78	1092.47
ABMT	No		1173.84	1278.44	1154.31
Chemotherapy	Yes	cis-RA	618.84	655.84	608.46
Chemotherapy	Yes	No Further Therapy	619.04	653.88	609.86
ABMT	Yes	cis-RA	415.74	446.31	409.72
ABMT	Yes	No Further Therapy	486.77	521.44	479.38

Table 6

The weighted log-rank test results for the neuroblastoma dataset.

Nelson-Aalen Estimator			
	Test Statistic	Standard Error	P-value
H_1	0.73	1.078	0.50
H_2	1.72	1.390	0.22
H_3	0.27	1.491	0.86
H_4	1.08	1.245	0.39
H_5	-0.31	1.337	0.82
H_6	-1.37	1.038	0.19

Log-normal Model			
	Test Statistic	Standard Error	P-value
H_1	0.06	1.164	0.96
H_2	1.59	1.373	0.25
H_3	0.17	1.409	0.91
H_4	1.55	1.254	0.22
H_5	0.12	1.293	0.93
H_6	-1.34	1.068	0.21

CCR: "treat with chemotherapy followed by cis-RA if no progressive disease", CNR: "treat with chemotherapy followed by no further therapy if no progressive disease", ACR: "treat with ABMT followed by cis-RA if no progressive disease", ANR: "treat with ABMT followed by no further therapy if no progressive disease"

H_1 : $\lambda_{CCR}(t) = \lambda_{CNR}(t)$, H_2 : $\lambda_{CCR}(t) = \lambda_{ACR}(t)$, H_3 : $\lambda_{CCR}(t) = \lambda_{ANR}(t)$,

H_4 : $\lambda_{CNR}(t) = \lambda_{ACR}(t)$, H_5 : $\lambda_{CNR}(t) = \lambda_{ANR}(t)$, H_6 : $\lambda_{ACR}(t) = \lambda_{ANR}(t)$