Determining When One Treatment is Different From Another Based On A Censored Data Regression Model

John P. Klein and Mei-Jie Zhang
Division of Biostatistics
Medical College of Wisconsin
8701 Watertown Plank Road
Milwaukee, WI 53226

Abstract

Often when comparing the survival rates of individuals given either of two treatments the analysis stops with a test of the hypothesis of no treatment difference and perhaps a plot of the two survival functions. The hypothesis test is usually a comparison of the two survival curves over the entire observational period. An alternative approach to this problem is to provide an investigator with a confidence set for the set of times at which the survival rates of the two treatments are the same. We discuss how such confidence sets can be constructed when the proportional hazards or additive regression model is used to adjust the comparison of interest for other factors which may influence survival. These approaches are illustrated on retrospective data gathered to compare the survival rates of allogeneic and autologous bone marrow transplants for acute leukemia.

1 Introduction

A common problem arising in biomedical applications is the comparison of the survival functions or hazard rates of two treatments. Most standard statistical tests are based on comparing the survival curves or equivalently the hazard functions over a given time period. The time period considered is typically the period from initiation of the treatment to some point in time where observation of the patients ceases. This comparison may be made by the log rank test (cf. Andersen et al. 1993), for example, when there are no other covariates that may influence survival. When there are other covariates that may affect outcome in addition to the treatments under consideration, testing of treatment effects is carried out by some type of regression technique. These tests may be based on any number of parametric or semiparametric models, but most common are tests based on either the Cox (1972) proportional hazards model or on Aalen's (1989, 1993) addative hazards model.

The results of these analyses tell the investigator whether the two treatments have the same survival rates or not. When the results of the test indicate that the survival curves are different the natural question posed by most clinicians is "At what times are these two treatments different?" The answer to this question is crucial to a patient and physician in deciding which of the two treatments to use. It is of special importance when one treatment has higher early survival but lower long term survival. This question is of particular interest in applications like bone marrow transplantation where, when comparing disease free survival rates, one procedure may have a higher early mortality rate due to treatment toxicity than the other treatment but among survivors of this early period the relapse rate is lower.

In this note we present methods for constructing a confidence set for the times at which the two treatments have the same survival function based on Aalen's additive hazards model. Confidence sets for the times at which one treatment has a survival probability at least as high as the other treatment are also presented. The confidence sets are found by inverting a test that compares the survival rates for the two treatments at fixed points in time. The set of all times for which this test accepts the hypothesis of no treatment difference provides the desired confidence set. Note that the confidence set is based on a comparison of the survival rates or cumulative hazard rates at fixed points in time as opposed to the usual tests which compare survival for the entire curve.

The random sets, A_α that we construct by this technique are in fact conservative $(1-\alpha)\times 100\%$ confidence sets for the set, Θ, of all times at which the two survival functions are the same. To see this consider the probability that Θ is a subset of A_α. Let t be an element of Θ. For such a t the subset of the sample space for which this t will also be in the set A_α has probability $(1-\alpha)$ by our method of construction. This will be true for any t in Θ, however different t's yield different subsets of the sample space. The coverage probability is the probability of the union of theses different subsets as indexed by t in Θ. Since each subset has probability exactly $(1-\alpha)$ our coverage probability must be at least $(1-\alpha)$.

In the next section we review results from Klein and Zhang (1997) for comparing two treatments when an adjustment for other covariates is needed using the Cox (1972) proportional hazards model. We review both the case where there is no interaction between these other covariates and the main treatment comparison and the case where there is an interaction between the main treatment effect and some of the covariates. In this section the confidence sets are based on a stratified Cox regression model.

In section 3 we show how Aalen's additive model can be used to generate these confidence sets. Here the sets are based on fitting the full additive model to the data and inverting a pointwise test that the regression function for treatment is equal to zero.

In Section 4 we present an example of these confidence sets using data from The International Bone Marrow Transplant Registry and The Autologous Blood And Marrow Registry. The primary comparison of interest is between the leukemia free survival rates of autologous and allogeneic bone marrow transplants for acute leukemia patients. Autologous transplants, where a patient's own marrow is used to re grow their immune system, are typically less toxic then allogeneic transplants where the marrow from an HLA matched sibling is used. Patients do not experience graft-versus-host disease which is a leading contributor to death in

the first several months after transplant. It is well known, however, that graft-versus-host disease has some protective effect against the reoccurrence of the leukemia, so allogeneic patients who survive the initial period tend to have lower leukemia relapse rates, off setting their higher early treatment related mortality. For a patient there is thus a trade off between early high mortality with allogeneic transplants and lower reoccurrence rates. To help in the decision between these two competing treatment modalities a confidence set for the times at which the survival probabilities of the two treatments are the same is of interest. Also, since autologous transplants are easier to perform as no donor is needed, a confidence set for those times where the survival probability for a autologous transplant patient is not smaller than the corresponding survival probability for an allogeneic transplant patient is also of interest.

2 Confidence Set Based On Cox's Proportional Hazards Model

Often there are other risk factors that need to be adjusted for prior to making the main comparison between the two treatments in many experiments. In this section we construct the confidence set based on the proportional hazards model which has become one of the most commenly used model in the analysis of failure time observations.

2.1 Adjustment For Covariates Not Confounded With Outcome

Let $\boldsymbol{Z} = (Z_1, \cdots, Z_p)$ be a vector of fixed time covariates that influence survival. In this section we assume that there is no significant interaction between the comparison of interest (treatment) and any of these covariates. Here we fit a proportional hazards model for the explanatory covariates stratifying on the treatment of interest. That is we fit the model

$$\lambda(t|\boldsymbol{Z}, \text{Treatment}) = \begin{cases} \lambda_{10}(t)\exp\{\beta^T\boldsymbol{Z}\}, & \text{for treatment 1,} \\ \lambda_{20}(t)\exp\{\beta^T\boldsymbol{Z}\}, & \text{for treatment 2.} \end{cases} \tag{2.1}$$

Let $\hat{\beta}$ and $I(\hat{\beta})$ be the partial maximum likelihood estimator and the observed information for this model. An estimator of the baseline cumulative hazard rate for treatment j, $j = 1, 2$ is given by Breslow's (1975) estimator

$$\hat{\Lambda}_{j0}(t) = \int_0^t \frac{dN_j(u)}{S_j^{(0)}(\hat{\beta}, u)}, \quad \text{where} \tag{2.2}$$

$$S_j^{(0)}(\hat{\beta}, u) = \sum_{i=1}^{n} Y_{ij}(u)\exp\{\beta^T\boldsymbol{Z}_i\} \tag{2.3}$$

with $Y_{ij}(u)$ the indicator of whether the ith individual is at risk at time u and is in the jth group.

For an individual with a covariate vector $\boldsymbol{Z_0}$, the two treatments will have the same survival rate at time t_0 if $\Lambda(t|\boldsymbol{Z_0}, \text{Treatment 1}) = \Lambda(t|\boldsymbol{Z_0}, \text{Treatment 2})$, which from (2.1) is equivalent to having $\Lambda_{10}(t_0) = \Lambda_{20}(t_0)$ or $\Delta(t_0) = \Lambda_{20}(t_0) - \Lambda_{10}(t_0) = 0$. Note that this comparison is independent of the value of $\boldsymbol{Z}_0$. The test statistic for this hypothesis is

$$\hat{\Delta}(t_0) = \hat{\Lambda}_{20}(t_0) - \hat{\Lambda}_{10}(t_0). \tag{2.4}$$

Using standard counting process techniques the large sample variance of this statistic can be shown to be

$$\begin{aligned} Var[\hat{\Delta}(t_0)] &= \sum_{j=1}^{2} \int_0^{t_0} \frac{dN_j(u)}{[S_j^{(0)}(\hat{\beta}, u)]^2} + \boldsymbol{W}^T(\hat{\beta}, t_0)[I(\hat{\beta})]^{-1}\boldsymbol{W}(\hat{\beta}, t_0), \quad \text{where} \qquad (2.5) \\ \boldsymbol{W}^T(\hat{\beta}, t_0) &= \int_0^{t_0} \tilde{\boldsymbol{Z}}_{\boldsymbol{2}}(\hat{\beta}, u) d\hat{\Lambda}_{20}(u) - \tilde{\boldsymbol{Z}}_{\boldsymbol{1}}(\hat{\beta}, u) d\hat{\Lambda}_{10}(u), \\ \tilde{\boldsymbol{Z}}_{\boldsymbol{j}}(\hat{\beta}, u) &= \frac{\boldsymbol{S}_j^{(1)}(\hat{\beta}, u)}{S_j^{(0)}(\hat{\beta}, u)}, \quad \text{and} \qquad (2.6) \\ \boldsymbol{S}_j^{(1)}(\hat{\beta}, u) &= \sum_{i=1}^{n} Y_{ij}(u)\boldsymbol{Z}_i \exp\{\hat{\beta}^T \boldsymbol{Z}_i\}. \end{aligned}$$

An α-level test of $H_o : \Delta(t_0) = 0$ versus $H_a : \Delta(t_0) \neq 0$ is accepted when $|\hat{\Delta}(t_0)/\sqrt{Var[\hat{\Delta}(t_0)]}| \leq z_{\alpha/2}$, where z_α is the αth upper quantile of a standard normal random variable. Inverting this test yields a $100 \times (1-\alpha)$ confidence set for the times at which $S_1(t) = S_2(t)$ as

$$\begin{aligned} & \left\{t_0 : -z_{\alpha/2} \leq \hat{\Delta}(t_0)/[Var(\hat{\Delta}(t_0))]^{1/2} \leq z_{\alpha/2}\right\} \\ = & \left\{t_0 : \hat{\Delta}(t_0) - z_{\alpha/2}\sqrt{Var(\hat{\Delta}(t_0))} \leq 0 \leq \hat{\Delta}(t_0) + z_{\alpha/2}\sqrt{Var(\hat{\Delta}(t_0))}\right\} \qquad (2.7) \end{aligned}$$

To find sets of time where we are $(1-\alpha) \times 100\%$ confident that $S_1(t) \leq S_2(t)$ consider testing the hypothesis $H_0 : \Lambda_1(t_0) \geq \Lambda_2(t_0)$ versus $H_A : \Lambda_1(t_0) < \Lambda_2(t_0)$. This is equivalent to testing $H_0 : \Delta(t_0) \leq 0$ versus $H_A : \Delta(t_0) > 0$. The desired confidence set for those points in time where treatment 2 is at least as good as treatment 1 ($\Delta(t_0) \leq 0$) is given by

$$\left\{t_0 : \hat{\Delta}(t_0)/\sqrt{Var[\hat{\Delta}(t_0)]} < z_\alpha\right\} = \left\{t_0 : 0 \geq \hat{\Delta}(t_0) - z_\alpha\sqrt{Var[\hat{\Delta}(t_0)]}\right\}.$$

To illustrate these calculations we consider data from a retrospective study of the effectiveness of bone marrow transplantation for patients with acute myelocytic leukemia (AML). Of interest is the comparison of survival rates between patients given either an autologous (auto) or allogeneic (allo) transplant. The data set consists of data on 1,325 patients reported over a four year period to either the International Bone Marrow Transplant Registry (allo transplants) or the Autologous Blood and Marrow Registry (auto transplants). 381 patients received an autologous transplant and 944 a HLA identical sibling allogeneic transplant. In

addition to type of transplant, on each patient includes remission status (1st or second complete remission), age (dichotomized as ≤ 30 or > 30) and Karnofsky score (dichotomized as < 90 or ≥ 90) at transplant. For patients in second complete remission the duration of the first complete remission (dichotomized as ≤ 1 yr or > 1 yr) is also available.

The confidence set is based on the results of fitting a proportional hazards model, stratified on transplant type, with binary covariates for remission status, age, Karnofsky score and duration of first complete remission. We find that a 95% confidence set for the times where the survival probabilities for the two transplant types are not different, adjusted for this set of covariates, is the set of time points given by

$$C2 = \{t_0 |\ t_0 \in [0, 0.132) \cup [0.151, 1.242) \cup [2.281, 2.418)\} \text{ years.}$$

A 95% confidence set for those times where patients given an auto transplant have a survival probability at least as high as patients given an allo transplant is given by

$$C1 = \{t_0 |\ t_0 \in [0, 0.861) \cup [0.872, 1.179)\}\,.$$

2.2 Adjustment For Covariates Confounded With Outcome

In some instances the comparison of the treatments of interest is complicated by some of the explanatory covariates have differential effects on the survival rates for the two treatments. Suppose that the covariate vector can be partitioned as $\boldsymbol{Z}^T = (\boldsymbol{Z}_1^T, \boldsymbol{Z}_2^T)^T$, where $\boldsymbol{Z}_1$ is a vector of length q_1 of the covariates confounded with treatment and Z_2 is a vector of length q_2 of the covariates not confounded with treatment.

To construct the confidence set where the survival rates are the same for the two treatments a stratified proportional hazards model is used. We fit the model

$$\lambda(t|\boldsymbol{Z}, \text{Treatment}) = \begin{cases} \lambda_{10}(t)\exp\{\gamma_1^T \boldsymbol{Z}_1 + \theta^T \boldsymbol{Z}_2\}, & \text{for treatment 1,} \\ \lambda_{20}(t)\exp\{\gamma_2^T \boldsymbol{Z}_1 + \theta^T \boldsymbol{Z}_2\}, & \text{for treatment 2.} \end{cases} \tag{2.8}$$

Estimates for $\beta = (\theta_1^T, \gamma_1^T, \gamma_2^T)$ are found by fitting a Cox model, stratified on treatment group to the data with an augmented covariate vector $\boldsymbol{Z}^T = (\boldsymbol{Z}_2^T, \boldsymbol{Z}_1^T I[\text{Treatment} = 1], \boldsymbol{Z}_1^T I[\text{Treatment} = 2])$. For a given set of confounding factors, $\boldsymbol{Z}_{10}$, the two treatments will have the same survival rate at time t_0 if

$$\Delta(t_0|\boldsymbol{Z}_{10}) = \Lambda_{20}(t_0)\exp\{\gamma_2^T \boldsymbol{Z}_{10}\} - \Lambda_{10}(t_0)\exp\{\gamma_1^T \boldsymbol{Z}_{10}\} \tag{2.9}$$

is equal to zero. The estimator of $\Delta(t_0|\boldsymbol{Z}_{10})$ given by

$$\hat{\Delta}(t_0|\boldsymbol{Z}_{10}) = \hat{\Lambda}_{20}(t_0)\exp\{\hat{\gamma}_2^T \boldsymbol{Z}_{10}\} - \hat{\Lambda}_{10}(t_0)\exp\{\hat{\gamma}_1^T \boldsymbol{Z}_{10}\}$$

follows from the fitted Cox model with $\Lambda_{j0}()$ estimated using Breslow's estimator (2.2).

An estimator of the asymptotic variance of $\hat{\Delta}(t_0|\boldsymbol{Z}_{10})$ can be shown to be

$$\begin{aligned} Var(\hat{\Delta}(t_0|\boldsymbol{Z}_{10})) &= \sum_{j=1}^{2} \int_0^{t_0} \exp\{2\hat{\gamma}_j^T \boldsymbol{Z}_{10}\} \frac{dN_j(u)}{[S_j^{(0)}(\hat{\beta}, u)]^2} + \\ &+ \left\{\boldsymbol{W}_2(\hat{\beta}, t_0) - \boldsymbol{W}_1(\hat{\beta}, t_0)\right\}^T [\boldsymbol{I}(\hat{\beta})]^{-1} \left\{\boldsymbol{W}_2(\hat{\beta}, t_0) - \boldsymbol{W}_1(\hat{\beta}, t_0)\right\} \end{aligned}$$

Here

$$\boldsymbol{W}_j(\hat{\beta}, t_0) = \exp\{\hat{\gamma}_j^T \boldsymbol{Z}_{10}\} \int_0^{t_0} [\tilde{\boldsymbol{Z}}_j(\hat{\beta}, u) - \boldsymbol{Z}_{(j)}] d\hat{\Lambda}_{j0}(u), \quad j = 1, 2$$

with $\tilde{\boldsymbol{Z}}_j(\hat{\beta}, u)$, defined by (2.6) and $\boldsymbol{Z}_{(1)} = (\boldsymbol{0}^T, \boldsymbol{Z}_{10}^T, \boldsymbol{0}^T)$ and $\boldsymbol{Z}_{(2)} = (\boldsymbol{0}^T, \boldsymbol{0}^T, \boldsymbol{Z}_{10}^T)$.

Since at t_0 an α level test of the equality of the two survival functions for a fixed value of $\boldsymbol{Z}$ is accepted when $\hat{\Delta}(t_0|\boldsymbol{Z}_{10})/[Var(\hat{\Delta}(t_0|\boldsymbol{Z}_{10}))]^{1/2}$ is in the interval $[-z_{\alpha/2}, z_{\alpha/2}]$, a $(1-\alpha)\times 100\%$ confidence set for those times at which the two treatments are not different is given by

$$\left\{t : -z_{\alpha/2} \leq \hat{\Delta}(t_0|\boldsymbol{Z}_{10})/[Var(\hat{\Delta}(t_0|\boldsymbol{Z}_{10}))]^{1/2} \leq z_{\alpha/2}\right\}$$

Similarly a confidence set for those points in time where treatment 2 is at least as good as treatment 1 is given by

$$\left\{t : \hat{\Delta}(t_0|\boldsymbol{Z}_{10})/[Var(\hat{\Delta}(t_0|\boldsymbol{Z}_{10}))]^{1/2} \leq z_{\alpha}\right\}$$

To illustrate this approach we again use the data comparing autologous and allogeneic transplants. Here, based on a standard semi-parametric regression analysis, it appears that age has a differential effect on the two types of transplants. To adjust for this confounding factor a proportional hazards model stratified on type of transplant is fit to the covariates remission status, Karnofsky score (< 90 or ≥ 90), duration of first complete remission (dichotomized as ≤ 1 yr or > 1 yr) and two interaction covariates. The interaction covariates are $Z_{11} = 1$ if age > 30 and allo transplant and $Z_{12} = 1$ if age > 30 and auto transplant. Note that here the estimate of Δ for a patient under age 30 is the difference of the baseline cumulative hazards from the stratified Cox model, while for patients over 30 each of the baseline hazards is multiplied by the factor $\exp[\gamma_j]$ before differencing.

The 95% confidence sets for the times (in years) where the two treatments have the same survival probability are

$$C2_{\leq 30} = \{t_0 | t_0 \in [0, 1.242) \cup [2.349, 2.418)\}$$

for patients age 30 or less and

$$C2_{>30} = \{t_0 | t_0 \in [0, 0.115) \cup [0.118, 0.129) \cup [0.1590, 5.891)\}$$

for patients over age 30. This suggests that for older patients there is no advantage in survival for either type of transplant but for younger patients the two survival rates are different after the first 15 months or so.

A 95% confidence set for those times where patients given an auto transplant have a survival probability at least as high as patients given an allo transplant is given by

$$C1_{\leq 30} = \{t_0 | t_0 \in [0, 0.858) \cup [0.885, 1.162)\}$$

or patients age 30 or less and

$$C1_{>30} = \{t_0 | t_0 \in [0, 5.891)\}$$

for patients over age 30. Note that this suggests that the auto transplant survival rate is at least as good as the allo transplant rate for patients over age 30, but for patients under 30 the survival rate is only as good for a little over a year after transplant.

3 Confidence Sets Based On The Additive Hazards Model

3.1 Estimation In The Additive Model

An alternative to the proportional hazards model is the additive hazards model first suggested by Aalen (1980). This model allows for covariate effects which vary over time since the regression coefficients are functions of time as opposed to the Cox model where they are constants. This approach uses a linear model for the conditional hazard rate and estimates regression coefficient functions by a least squares technique.

To define the model suppose we have an individual with covariates $Z_1(t), \cdots, Z_p(t)$. For such an individual the model for the conditional hazard rate is given by

$$\lambda(t|Z_1(t), ..., Z_p(t)) = \begin{cases} \alpha_0(t) + \sum_{K=1}^{p} \alpha_k(t) Z_k(t) & \text{if this individual is at risk at time } t, \\ 0 & \text{otherwise} \end{cases}$$

Here the $\alpha_j(t)$'s, $j = 0, \ldots, p$ are functions of time to be estimated form the data.

Suppose we observe n individuals. Associated with each individual is a p-vector of possibly time dependent covariates, $Z_i(t) = (1, Z_i(t), \cdots, Z_p(t))$. (Here the first element of the covariate vector is 1 to allow for a baseline intensity.) Let $\lambda_i(t)$ denote the intensity at which the event occurs for the ith subject. To write the model in matrix notation let $Y(t)$ be the $n \times (p+1)$ matrix whose ith row is $Z_i(t)$ if individual i is at risk at time t and is a row of zeros if this subject is not at risk at time t. Then the additive regression model is

$$\lambda(t) = Y(t)\alpha(t). \quad (3.1)$$

Here the first element of $\alpha(t)$ is a baseline intensity and the remaining elements are the regression functions which describe the effect of the covariate over time on survival.

The only restriction on covariates which can be used in this model is that they are predictable in the sense that their value is known just prior to time t (cf. Aalen 1978). In the data set to be used here to illustrate these techniques all covariates are known at the time of transplant so this condition is satisfied.

Estimation for the additive model is based on a least squares approach. Direct estimation of $\alpha(t)$ is difficult so we estimate instead the cumulative regression function, $A(t) = (A_1(t), ..., A_p(t))^T$, where

$$A_j(t) = \int_0^t \alpha_j(t)dt, \quad j = 0, 1, ..., p.$$

Let $T_1 < T_2 < ...$ be the ordered observed times at which events occur. Then Aalen (1980, 1989) shows that the least squares estimator of $A(t)$ is given by

$$\hat{A}(t) = sum_{T_K \leq t} X(T_k) I_k, where (3.2)$$

$X(t)$ is a generalized inverse of $Y(t)$, and I_k is the n-vector of whose ith element is 1 if subject i experiences the event at time T_k and is 0 if they don't. The estimator (3.2) is only defined over the range where the matrix $Y(t)$ is of full rank. Let τ be the random point in time where $Y()$ loses its full rank.

Any generalized inverse can be used in computing the estimator (3.1). By analogy to the usual linear models analysis we shall use the generalized inverse suggested by Aalen (1980), Huffer and McKeague (1991), McKeague (1988), namely

$$X(t) = (Y(t)^T Y(t))^{-1} Y(t)^T \quad (3.3)$$

An alternative choice of the generalized inverse is a weighted inverse which leads to the analog of a weighted least squares estimate (See Huffer and McKeague (1991), McKeague (1988)).

The variance matrix of $\hat{A}(t)$ can be estimated consistently by

$$\Sigma(t) = \sum_{T_k} X(T_k) D_k X(T_k)^T \quad (3.4)$$

where D is the diagonal matrix with I_k as the diagonal. One can show (cf. Aalen 1980, Andersen et al (1993)) that $\hat{A}(t)$ converges weakly to a Gaussian process with independent increments under a wide set of regularity conditions. A SAS Macro to perform the calculations need to obtain $\hat{A}(t)$ and $\Sigma(t)$ is described in Howell and Klein (1996).

3.2 Confidence Sets Adjusted For Other Covariates Not Confounded With Treatment

As for the proportional hazards model, to find a confidence set for those times where the two treatments are the same adjusting for a p-variate set of covariates $Z_1, ..., Z_p$ we base the set on a series of pointwise tests of equality of the adjusted cumulative hazard rates for the two treatments. For each individual define the $p+2$ dimensional vector $Z_i(t) = (1, Z_1(t), ..., Z_p(t), W)$, where $W = 1$ if this individual received treatment 2 and 0 otherwise. Using this coding of the covariates we compute $\hat{A}(t)$ and $\Sigma(t)$. Now the difference in cumulative hazard rates between an individual given treatment 2 and an individual given treatment 1 is

$$\begin{aligned}
\Delta(t|Z) &= \Lambda(t|Z, \text{ Treatment 2}) - \Lambda(t|Z, \text{ Treatment 1}) \\
&= \{A_0(t) + \sum_{k=1}^{p} A_k(t) Z_k(t) + A_{p+1}(t)\} - \{A_0(t) + \sum_{k=1}^{p} A_k(t) Z_k(t)\} \\
&= A_{p+1}(t)
\end{aligned}$$

The variance of this estimator is found directly from $\Sigma(t)$.

An α-level test of $H_o : \Delta(t_0) = 0$ versus $H_a : \Delta(t_0) \neq 0$ is accepted when $|\hat{\Delta}(t_0)/\sqrt{Var[\hat{\Delta}(t_0)]}| \leq z_{\alpha/2}$, where z_α is the αth upper quantile of a standard normal random

variable. Inverting this test yields a $100 \times (1-\alpha)$ confidence set for the times at which $S_1(t) = S_2(t)$ as

$$\begin{aligned} & \left\{t_0 : -z_{\alpha/2} \leq \hat{\Delta}(t_0)/[Var(\hat{\Delta}(t_0))]^{1/2} \leq z_{\alpha/2}\right\} \\ = & \left\{t_0 : \hat{\Delta}(t_0) - z_{\alpha/2}\sqrt{Var(\hat{\Delta}(t_0))} \leq 0 \leq \hat{\Delta}(t_0) + z_{\alpha/2}\sqrt{Var(\hat{\Delta}(t_0))}\right\} \end{aligned} \tag{3.10}$$

To find sets of time where we are $(1-\alpha) \times 100\%$ confident that $S_1(t) \leq S_2(t)$ consider testing the hypothesis $H_0 : \Lambda_1(t_0) \geq \Lambda_2(t_0)$ versus $H_A : \Lambda_1(t_0) < \Lambda_2(t_0)$. This is equivalent to testing $H_0 : \Delta(t_0) \leq 0$ versus $H_A : \Delta(t_0) > 0$. The desired confidence set for those points in time where treatment 2 is at least as good as treatment 1 ($\Delta(t_0) \leq 0$) is given by

$$\left\{t_0 : \hat{\Delta}(t_0)/\sqrt{Var[\hat{\Delta}(t_0)]} < z_\alpha\right\} = \left\{t_0 : 0 \geq \hat{\Delta}(t_0) - z_\alpha\sqrt{Var[\hat{\Delta}(t_0)]}\right\}.$$

3.3 Confidence sets when the covariates are confounded with treatment

Suppose that the covaraites vector can be partitioned into as a set $\boldsymbol{Z_1}$ of dimension q_1 of covariates confounded with treatment and a set $\boldsymbol{Z_2}$ of dimension q_2 of covariates not confounded with treatment. For this case we define for each individual the $q_2 + 2q_1 + 2$ dimensional vector $\boldsymbol{Z}_i(t) = (1, \boldsymbol{Z}_2, W\boldsymbol{Z}_2, [1-W]\boldsymbol{Z}_2, W)$. Here W is again the indicator of an individual being in treatment group 2. Note that for this covariate vector we have the following cumulative hazard rates for the two treatment groups:
Treatment 2: $A_0(t) + \sum_{k=1}^{q_1} A_k(t)Z_{2k}(t) + \sum_{k=1}^{q_2} A_{1+q_1+k}(t)Z_{1k}(t) + A_{2+q_1+q_2+k}(t)$
Treatment 1: $A_0(t) + \sum_{k=1}^{q_1} A_k(t)Z_{2k}(t) + \sum_{k=1}^{q_2} A_{1+q_1+q_2+k}(t)Z_{1k}(t)$

For an individual with a set $\boldsymbol{Z}_{10}$ of covariates for $\boldsymbol{Z}_1$, the survival functions will be the same at time t if these two cumulative hazard rates are the same. That is if

$$\Delta(t|\boldsymbol{Z}_{10} = A_{2+q_1+q_2+k}(t) \sum_{k=1}^{q_2} [A_{1+q_1+q_2+k}(t) - A_{1+q_1+k}(t)]Z_{1k}(t) = 0.$$

Note that if we let $\boldsymbol{C} = (\boldsymbol{0}, \boldsymbol{Z}_{10}, -\boldsymbol{Z}_{10}, 1)$, where $\boldsymbol{0}$ is the q_1 vector of zero's then $\Delta(t|\boldsymbol{Z}_{10} = \boldsymbol{C}\boldsymbol{A}(t)$ and the variance of $\hat{\Delta}(t|\boldsymbol{Z}_{10} = \boldsymbol{C}\hat{\boldsymbol{A}}(t)$ is estimated by $\boldsymbol{C}\boldsymbol{\Sigma}(t)\boldsymbol{C}^T$.

Since at t_0 an α level test of the equality of the two survival functions for a fixed value of $\boldsymbol{Z}$ is accepted when $\hat{\Delta}(t_0|\boldsymbol{Z}_{10})/[Var(\hat{\Delta}(t_0|\boldsymbol{Z}_{10}))]^{1/2}$ is in the interval $[-z_{\alpha/2}, z_{\alpha/2}]$, a $(1-\alpha) \times 100\%$ confidence set for those times at which the two treatments are not different is given by

$$\left\{t : -z_{\alpha/2} \leq \hat{\Delta}(t_0|\boldsymbol{Z}_{10})/[Var(\hat{\Delta}(t_0|\boldsymbol{Z}_{10}))]^{1/2} \leq z_{\alpha/2}\right\}$$

Similarly a confidence set for those points in time where treatment 2 is at least as good as treatment 1 is given by

$$\left\{t : \hat{\Delta}(t_0|\boldsymbol{Z}_{10})/[Var(\hat{\Delta}(t_0|\boldsymbol{Z}_{10}))]^{1/2} \leq z_\alpha\right\}$$

4 Example

To illustrate these calculations we consider data from a retrospective study of the effectiveness of bone marrow transplantation for patients with acute myelocytic leukemia (AML). Of interest is the comparison of survival rates between patients given either an autologous (auto) or allogeneic (allo) transplant. The data set consists of data on 1,325 patients reported over a four year period to either the International Bone Marrow Transplant Registry (allo transplants) or the Autologous Blood and Marrow Registry (auto transplants). 381 patients received an autologous transplant and 944 a HLA identical sibling allogeneic transplant.

The comparison of interest is between the leukemia free survival times (LFS) of the two groups. A patient is considered as an event if they die or their leukemia returns. The event time is the smaller of the time of relapse or death. Figure 1 shows the unadjusted Kaplan-Meier estimators for the two treatment groups. The log rank test of equality of the survival functions in the two treatment groups is rejected with a p-value of 0.0071.

In addition to type of transplant, data on each patient includes remission status (1st or second complete remission), age (dichotomized as ≤ 30 or > 30) and Karnofsky score (dichotomized as < 90 or ≥ 90) at transplant. For patients in second complete remission the duration of the first complete remission is also recorded(dichotomized as ≤ 1 yr or > 1 yr). We wish to determine when the two types of transplants have the same survival rate after adjustment for these fixed explanatory covariates.

We first assume that there is no interaction between these covariates and the type of transplant. For the proportional hazards approach a Cox model is fit, stratified on transplant type, with binary covariates for remission status, age, Karnofsky score and duration of first complete remission. Applying the results in Section 2.2 we find that a 95% confidence set for the times where the survival probabilities for the two transplant types are not different, adjusted for this set of covariates, is the set of time points given by

$$C2 = \{t_0 |\ t_0 \in [0, 0.132) \cup [0.151, 1.242) \cup [2.281, 2.418)\} \text{ years.}$$

A 95% confidence set for those times where patients given an auto transplant have a survival probability at least as high as patients given an allo transplant is given by

$$C1 = \{t_0 |\ t_0 \in [0, 0.861) \cup [0.872, 1.179)\}\,.$$

For the additive model discussed in Section 3.2 we fit the model with covariates for type of transplant and for the four fixed covariates. The 95% confidence set for the times where the survival probabilities for the two transplant types are not different based on this model is given by

$$C2 = \{t_0 |\ t_0 \in [0, 0.137) \cup [0.143, 0.855) \cup [0.880, 1.102) \cup [1.124, 1.1662)\} \text{ years.}$$

The 95% confidence set for those times where patients given an auto transplant have a survival probability at least as high as patients given an allo transplant based on the additive model is given by

$$C1 = \{t_0|\ t_0 \in [0, 0.526) \cup [0.534, 0.537) \cup [0.611, 0.641) \cup [0.688, 0.726) \cup [0.732, 0.768) \cup [0.959, 0.984)\}\ .$$

The sets $C1$ suggest that for a period of time after transplant auto patients do not do any worst then allo patients, but after about this period they have smaller survival probabilities. This time interval is estimated to be a little over a year based on the proportional hazards model and a little under a year based on the additive model.

The above intervals assumed that the fixed covariates were not confounded with treatment. However, Here, based on a standard semi-parametric based on either the proportional hazards or additive hazards model, it appears that age has a differential effect on the two types of transplants.

To adjust for this confounding factor using the proportional hazards approach, a model stratified on type of transplant is fit to the covariates remission status, Karnofsky score, duration of first complete remission and two interaction covariates. The interaction covariates are $Z_{11} = 1$ if age > 30 and allo transplant and $Z_{12} = 1$ if age > 30 and auto transplant.

Using the results in Section 2.2, 95% confidence sets for the times (in years) where the two treatments have the same survival probability are

$$C2_{\leq 30} = \{t_0|t_0 \in [0, 1.242) \cup [2.349, 2.418)\}$$

for patients age 30 or less and

$$C2_{>30} = \{t_0|t_0 \in [0, 0.115) \cup [0.118, 0.129) \cup [0.1590, 5.891)\}$$

for patients over age 30. This suggests that for older patients there is no advantage in survival for either type of transplant but for younger patients the two survival rates are different after the first 15 months or so.

A 95% confidence set for those times where patients given an auto transplant have a survival probability at least as high as patients given an allo transplant based on the proportional hazards model is given by

$$C1_{\leq 30} = \{t_0|t_0 \in [0, 0.858) \cup [0.885, 1.162)\}$$

for patients age 30 or less and

$$C1_{>30} = \{t_0|t_0 \in [0, 5.891)\}$$

for patients over age 30.

These intervals suggest that for older patients there is little if any advantage in survival for either type of transplant but for younger patients the two survival rates are different after the some period of time. This time is about

For the additive model approach we fit a using a covariate vector with components remission status, Karnofsky score, duration of first complete remission , Z_{11}, $Z_{12} = 1$, and the indicator of type of transplant. Applying the results in Section 3.3 we find that 95%

confidence sets for the times (in years) where the two treatments have the same survival probability based on the additive model are

$$C2_{\leq 30} = \{t_0 | t_0 \in [0, 0.398) \cup [0.622, 0.632) \cup [0.696, 0.721) \cup [0.732, 0.855) \cup [0.872, 1.242) \cup [1.672, 2.837) \cup$$

for patients age 30 or less and

$$C2_{>30} = \{t_0 | t_0 \in [0, 0.066) \cup [0.159, 0.162) \cup [0.165, 0.167) \cup [0.189, 0.195) \cup [0.197, 5.05)\}$$

for patients over age 30. The conclusions are similar to those obtained from the proportional hazards model.

A 95% confidence set for those times where patients given an auto transplant have a survival probability at least as high as patients given an allo transplant based on the additive hazard model is given by

$$C1_{\leq 30} = \{t_0 | t_0 \in [0, 0.356) \cup [2.059, 2.448) \cup [4.260, 5.052)\}$$

for patients age 30 or less and

$$C1_{>30} = \{t_0 | t_0 \in [0, 1.8558) \cup [1.8722, 2.083) \cup [2.215, 2.418) \cup [3.753, 5.052)\}$$

for patients over age 30.

These intervals suggest that for older patients there is little if any advantage in survival for either type of transplant. For younger patients the proportional hazards model suggests that the survival rates are after about two and a half years. Based on the one sided sets constructed by additive hazard model, the inference for younger patients is that they have survival given an auto transplant at least as good as if they were given an allo transplant in the first 3 months after transplant, for a brief period in year two and then again after about 3 and three-fourth years.

5 Acknowledgments

This research was supported by Grant 5 R01 CA54706-05 from the National Cancer Institute for J. P. Klein and M. J. Zhang, and by Grant PO1-CA-40053 from the National Cancer Institute and by Institutional Research Grant 170H from American Cancer Society for M. J. Zhang.

References

Aalen, O. O. (1978). Nonparametric Inference For A Family Of Counting Processes. *Annals Of Statistics*, **6**, 701–726.

Aalen, O. O. (1980). A Model For Non-Parametric Regression Analysis of Counting Processes. *Springer Lecture Notes in Statistics*, **2**, 1–25, Mathematical Statistics and Probability Theory. W. Klonecki A. Kozek and J. Rosinski, editors.

Aalen, O. O. (1989). A Linear Regression Model For The Analysis Of Life Times. *Statistics In Medicine*, **8**, 907–925.

Aalen, O. O. (1993). Further Results On The Non-Parametric Linear Regression Model In Survival Analysis. *Statistics In Medicine*, **12**, 1569–1588.

Andersen, P. K., Bentzon, M. W. and Klein, J. P. (1996). Estimating The Survival Function In The Proportional Hazards Regression Model: A Study Of The Small Sample Size Properties. *Scandinavian Journal Of Statistics*, **23**, 1–12.

Andersen, P. K., Borgan, Ø., Gill, R. D. and Keiding, N. (1993). *Statistical Models Based On Counting Processes*. Springer-Verlag, New York.

Breslow, N. E. (1975). Analysis Of Survival Data Under The Proportional Hazards Model. *International Statistics Review*, **43**, 45–58.

Cox, D. R. (1972). Regression Models And Life-Tables (With Discussion). *Journal Of The Royal Statistical Society B*, **34**, 187–220.

Howell, A. M and Klein, J. P. (1996). A SAS Marco for the Additive Hazards Regression Model *Technical Report Number 18*, Division of Biostatistics, The Medical College of Wisconsin

Huffer, F. W. and McKeague, I.W. (1991). Weighted Least Squares Estimation for Aalen's Additive Risk Model. *JASA*, **86**, 114–129.

Klein, J. P. and Zhang, M.-J. (1997). Confidence Sets For The Equality Of Two Survival Curves *Lifetime Data Analysis*, To Appear.

McKeague, I.W. (1988). Asymptotic Theory For Weighted Least Squares Estimators In Aalen's Additive Risk Model. *Contemp. Math*, **80**, 139–152.