Confidence Regions For The Equality Of Two Survival Curves

John P. Klein, Ph.D.
Mei-Jie Zhang, Ph.D.

Technical Report 13

July 1996

Division of Biostatistics
Medical College of Wisconsin
8701 Watertown Plank Road
Milwaukee WI 53226
Phone: (414)456-8280

Confidence Regions For The Equality Of Two Survival Curves

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 region for the set of times at which the survival rates of the two treatments are the same. We discuss how such confidence regions can be constructed in three situations. First, we construct confidence regions when there are no additional factors that need be adjusted for. Second, based on a proportional hazards model, we show how to construct the desired confidence regions adjusted for explanatory covariates that are not confounded with the two treatments. Lastly, we extended these results to allow for explanatory covariates that are confounded with treatment. 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. Tests may be based on any number of parametric or

semi-parametric models, but most common are tests based on the Cox (1972) proportional 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 higher.

In this note we present methods for constructing a confidence region for the times at which the two treatments have the same survival function. Confidence regions for the times at which one treatment has a survival probability at least as high as the other treatment are also presented. The confidence regions 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 region. Note that the confidence region 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.

In the next section we present the results for comparing two treatments when there are no other explanatory covariates that may affect survival. The confidence region is based on a comparison of the Nelson-Aalen estimators (Aalen 1978) of the cumulative hazard rates at each point in time.

In Sections 3 and 4 the problem of constructing a confidence region for the times at which the survival rates are the same for the two treatments is considered for cases where an adjustment for other covariates is needed. In Section 3 we examine the case where there is no interaction between these other covariates and the main treatment comparison. In Section 4 we consider the problem when there is an interaction between the main treatment effect and some of the covariates. In these sections we base the confidence regions on a stratified Cox regression model.

In each section we present an example of these confidence regions 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 region 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 region 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 Regions When There Are No Other Explanatory Covariates

In this section we consider the simple case of comparing two treatments when there are no explanatory covariates to be adjusted for. Suppose we have independent samples of size n_1 and n_2 from each of the two treatments. Let $N_j(t), j = 1, 2$ be the processes that count the number of events that have occurred in sample j at or prior to time t. Let $Y_j(t)$ be the number at risk at time t in the jth sample, $j = 1, 2$. Note in this formulation the data may be right censored and left truncation is allowed (See Andersen et al (1993)). For $j = 1, 2$ let $\lambda_j(t)$ be the hazard rate of the time to the event in the jth group and let $\Lambda_j(t) = \int_0^t \lambda(u)du$ and $S_j(t) = \exp\{-\Lambda_j(t)\}$ be the cumulative hazard rates and survival functions.

To construct a $(1-\alpha) \times 100\%$ confidence region for the set of all times, t_0, for which $S_1(t_0) = S_2(t_0)$, we consider testing the hypothesis $H_0 : S_1(t_0) = S_2(t_0)$ against the alternative hypothesis $H_A : S_1(t_0) \neq S_2(t_0)$. Note that this null hypothesis is testing equality of the two curves at a fixed point in time and is not a test for equality of the two survival curves over the entire time period. This hypothesis about the equality of the two survival functions is equivalent to the hypothesis $H_0 : \Lambda_1(t_0) = \Lambda_2(t_0)$. It is this hypothesis we shall test at each point in time since the asymptotic convergence rates of the estimated cumulative hazard rates tend to be faster than those for the estimated survival functions.

The confidence region is based on the test statistic formed as the difference of the two Nelson-Aalen estimators of the cumulative hazard functions. The test statistic here is an estimator of $\Delta(t_0) = \Lambda_2(t_0) - \Lambda_1(t_0)$ and is given by

$$\hat{\Delta}(t_0) = \int_0^{t_0} \frac{dN_2(u)}{Y_2(u)} - \int_0^{t_0} \frac{dN_1(u)}{Y_1(u)}, \tag{2.1}$$

which has a variance that can be estimated by

$$Var[\hat{\Delta}(t_0)] = \int_0^{t_0} \frac{dN_2(u)}{[Y_2(u)]^2} + \int_0^{t_0} \frac{dN_1(u)}{[Y_1(u)]^2}, \tag{2.2}$$

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 region for the times at which

$S_1(t) = S_2(t)$ as $\{t_0 : -z_{\alpha/2} \leq \hat{\Delta}(t_0)/\sqrt{Var[\hat{\Delta}(t_0)]} \leq z_{\alpha/2}\}$. Note that the confidence region can also be written as $\{t_0 : \hat{\Delta}(t_0) - z_{\alpha/2}\sqrt{Var[\hat{\Delta}(t_0)]} \leq 0 \leq \hat{\Delta}(t_0) + z_{\alpha/2}\sqrt{Var[\hat{\Delta}(t_0)]}\}$, so that the desired region is the set of all those time points for which a $(1-\alpha)\times 100\%$ pointwise confidence region for $\Delta(t)$ contains 0.

To find regions 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 region is

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

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.

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 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.

Figure 2 is a plot of $\hat{\Delta}(t_0)/\sqrt{Var[\hat{\Delta}(t_0)]}$ against t_0. Here $\Delta(t_0) = \Lambda_{\text{Auto}}(t_0) - \Lambda_{\text{Allo}}(t_0)$. The dotted lines at ± 1.96 are the cutoffs for a 95% confidence region in that all time points for which $\hat{\Delta}(t_0)/\sqrt{Var[\hat{\Delta}(t_0)]}$ lies within that band are included in the region. The 95% confidence region for the points where the survival functions are the same for the two types of transplants is given by the set of all time (in years) in the set

$$C2 = \{t_0 |\ t_0 \in [0, 0.137) \cup [0.143, 0.855) \cup [0.880, 1.102) \cup [1.124, 1.162)\}.$$

The dashed lines at ± 1.645 can be used to find 95% confidence regions for all the times where the survival probability for one treatment is at least as good as for the other treatment. Of interest here is a confidence region for those times where patients given an auto transplant have a survival probability at least as high as patients given an allo transplant. That is the set of times where $S_{\text{Auto}}(t) \geq S_{\text{Allo}}(t)$. A 95%confidence region is found by determining all those times that fall below the line at 1.645. That region is given by the following set

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

This region suggests that in the first year after transplant auto patients do not do any worst then allo patients, but after about one year they have smaller survival probabilities.

3 Adjustment For Covariates Not Confounded With Outcome

In many experiments there are other risk factors that need to be adjusted for prior to making the main comparison between the two treatments. 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 and any of these covariates.

The confidence region, adjusted for these other covariates, is based on the proportional hazards model (Cox (1972)). Here we fit a proportional hazards model for the explanatory covariates stratifying on the treatment of interest. That is we fit the model

$$\lambda(t|\boldsymbol{Z}, \text{Treatment}) = \begin{cases} \lambda_{10}(t)\exp\{\beta^T\boldsymbol{Z}\}, & \text{for treatment 1,} \\ \lambda_{20}(t)\exp\{\beta^T\boldsymbol{Z}\}, & \text{for treatment 2.} \end{cases} \tag{3.1}$$

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

$$\hat{\Lambda}_{j0}(t) = \int_0^t \frac{dN_j(u)}{S_j^{(0)}(\hat{\beta}, u)}, \quad \text{where} \tag{3.2}$$

$$S_j^{(0)}(\hat{\beta}, u) = \sum_{i=1}^{n} Y_{ij}(u)\exp\{\beta^T\boldsymbol{Z}_i\} \tag{3.3}$$

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 (3.1) 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

$$\hat{\Delta}(t_0) = \hat{\Lambda}_{20}(t_0) - \hat{\Lambda}_{10}(t_0). \tag{3.4}$$

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

$$Var[\hat{\Delta}(t_0)] = \sum_{j=1}^{2}\int_0^{t_0} \frac{dN_j(u)}{[S_j^{(0)}(\hat{\beta}, u)]^2} + \boldsymbol{W}^T(\hat{\beta}, t_0)[I(\hat{\beta})]^{-1}\boldsymbol{W}(\hat{\beta}, t_0), \quad \text{where} \tag{3.5}$$

$$\boldsymbol{W}^T(\hat{\beta}, t_0) = \int_0^{t_0} \tilde{\boldsymbol{Z}}_{\boldsymbol{2}}(\hat{\beta}, u)d\hat{\Lambda}_{20}(u) - \tilde{\boldsymbol{Z}}_{\boldsymbol{1}}(\hat{\beta}, u)d\hat{\Lambda}_{10}(u),$$

$$\tilde{\boldsymbol{Z}}_{\boldsymbol{j}}(\hat{\beta}, u) = \frac{\boldsymbol{S}_j^{(1)}(\hat{\beta}, u)}{S_j^{(0)}(\hat{\beta}, u)}, \quad \text{and} \tag{3.6}$$

$$\boldsymbol{S}_j^{(1)}(\hat{\beta}, u) = \sum_{i=1}^{n} Y_{ij}(u)\boldsymbol{Z}_i\exp\{\hat{\beta}^T\boldsymbol{Z}_i\}.$$

Since at t_0 an α level test of the equality of the two survival functions is accepted when $\hat{\Delta}(t_0)/[Var(\hat{\Delta}(t_0))]^{1/2}$ is in the interval $[-z_{\alpha/2}, z_{\alpha/2}]$, a $(1-\alpha)\times 100\%$ confidence region for those times at which the two treatments have the same survival probability is

$$\begin{aligned} & \left\{t_0 : -z_{\alpha/2} \leq \hat{\Delta}(t_0)/[Var(\hat{\Delta}(t_0))]^{1/2} \leq z_{\alpha/2}\right\} \\ = & \left\{t_0 : \hat{\Delta}(t_0) - z_{\alpha/2}\sqrt{Var(\hat{\Delta}(t_0))} \leq 0 \leq \hat{\Delta}(t_0) + z_{\alpha/2}\sqrt{Var(\hat{\Delta}(t_0))}\right\} \end{aligned} \tag{3.7}$$

Similarly a confidence region for those points in time where treatment 2 is at least as good as treatment 1 ($\Delta(t_0) \leq 0$) is given by

$$\begin{aligned} & \left\{t_0 : \hat{\Delta}(t_0)/\sqrt{Var(\hat{\Delta}(t_0))} < z_{\alpha/2}\right\} \\ = & \left\{t_0 : 0 \geq \hat{\Delta}(t_0) - z_{\alpha/2}\sqrt{Var(\hat{\Delta}(t_0))}\right\}. \end{aligned}$$

We shall illustrate these calculations on the data set discussed in Section 2. Additional information, 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 region 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. Figure 3 shows a plot of $\hat{\Delta}(t_0)/\sqrt{Var(\hat{\Delta}(t_0))}$ versus t. Using this Figure we find that a 95% confidence region 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 region 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

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

4 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 region where the survival rates are the same for the two treatments a stratified proportional hazards model is used. We fit the model

$$\lambda(t|\boldsymbol{Z}, \text{Treatment}) = \begin{cases} \lambda_{10}(t)\exp\{\gamma_1^T\boldsymbol{Z}_1 + \theta^T\boldsymbol{Z}_2\}, & \text{for treatment 1,} \\ \lambda_{20}(t)\exp\{\gamma_2^T\boldsymbol{Z}_1 + \theta^T\boldsymbol{Z}_2\}, & \text{for treatment 2.} \end{cases} \tag{4.1}$$

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

$$\Delta(t_0|\boldsymbol{Z}_{10}) = \Lambda_{20}(t_0)\exp\{\gamma_2^T\boldsymbol{Z}_{10}\} - \Lambda_{10}(t_0)\exp\{\gamma_1^T\boldsymbol{Z}_{10}\} \tag{4.2}$$

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 (3.2).

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 (3.6) 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 region for those times at which the two treatments are not different is given by

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

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

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

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. Figure 4a and 4b show the standardized difference between the cumulative hazard rates for patients under 30 and over 30 years of age, respectively. 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 regions 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 region 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.

5 Discussion

In this note we have presented an approach to finding confidence regions for the times at which the survival functions of two treatments are the same. The confidence regions are based on pointwise comparisons of the cumulative hazard rates estimated from a stratified proportional hazards model. Our use of the cumulative hazard rate as the basic statistic for making these comparisons is motivated by the work of Borgan and Liestl (1990) and Andersen et al (1996) who found the small to moderate sample size performance of inference techniques based on the cumulative hazard rate to be better than that of similar techniques based on the estimated survival function. We could have developed the confidence regions discussed in this paper based on a direct test of the equality of the two survival function estimators. For example, in the no covariate case the confidence region is based on the

difference of the two Kaplan-Meier estimators of the survival function. We find that for most samples we have examined this alternative approach gives quite similar results to those presented here.

Confidence regions can be based on other non- and semi-parametric models. One possible model is to base the regions on an additive regression model (Aalen (1989, 1993)). Here the estimate of the difference in cumulative hazards and its standard error follows directly from the fitted model. An other possible model is to use a proportional hazards model where, rather then stratifying on the treatments, a series of time dependent indicators are used for the treatment effects in subintervals of time.

The intervals we presented are useful for clinicians to instruct their patients on the long and short term effects of their decision on which treatment to follow. They are one result of a long process of analyzing a complex survival experiment. The process includes construction by usual techniques of a final regression model and a careful check of the appropriateness of the final model. The plots and confidence regions serve as an aide to understanding the results of this complex analysis.

6 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. (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.

Borgan, Ø. and Liestl, K. (1990). A Note On Confidence Intervals And Bands For The

Survival Curve Based On Transformations. *Scandinavian Journal Of Statistics*, **17**, 35-41.

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.

Figure 1. Kaplan-Meier Estimates Of Leukemia Free Survival

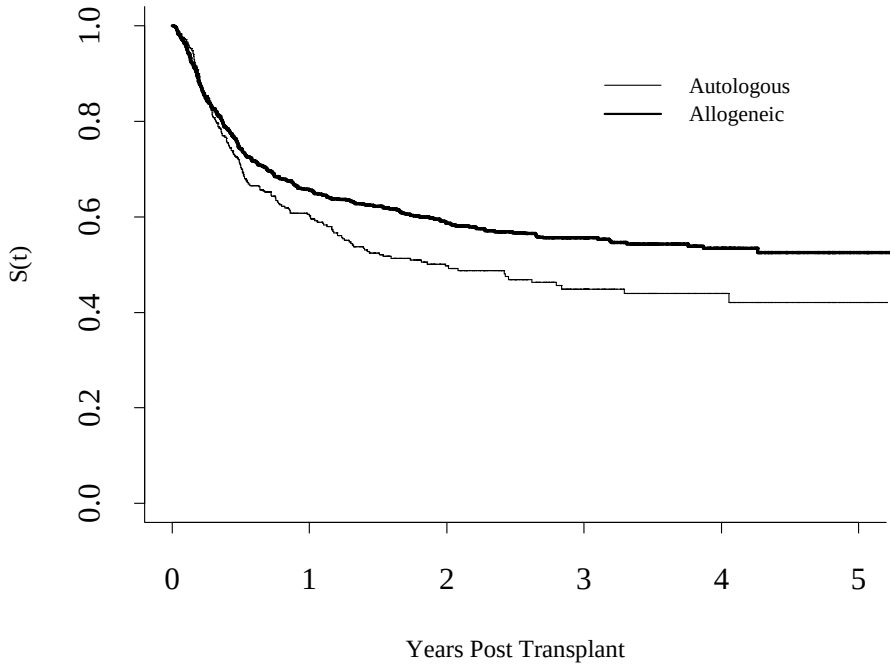


Figure 2. Standardized Unadjusted Difference
In The Cumulative Hazards

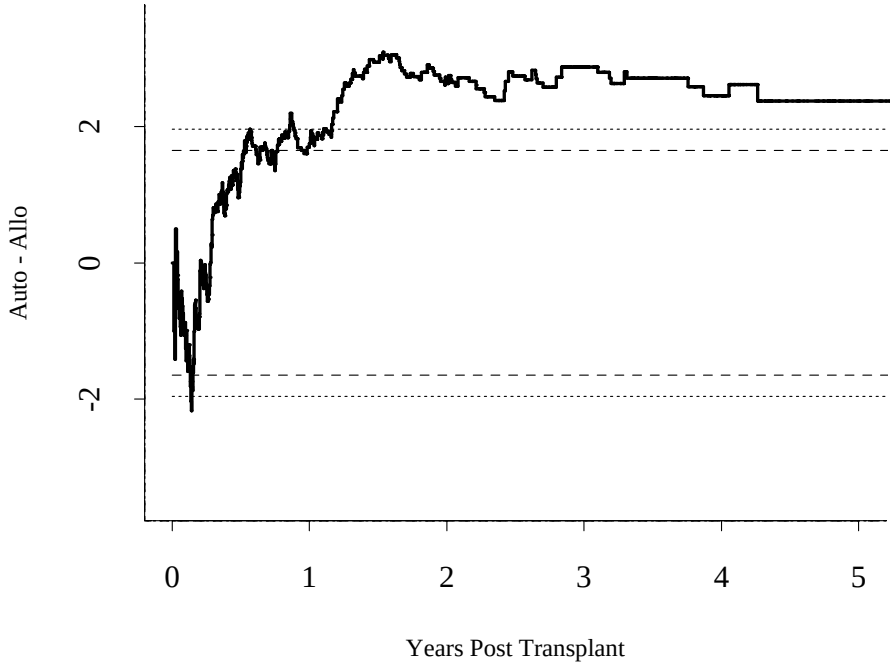


Figure 3. Standardized Difference In The Cumulative Hazards
(No Confounding)

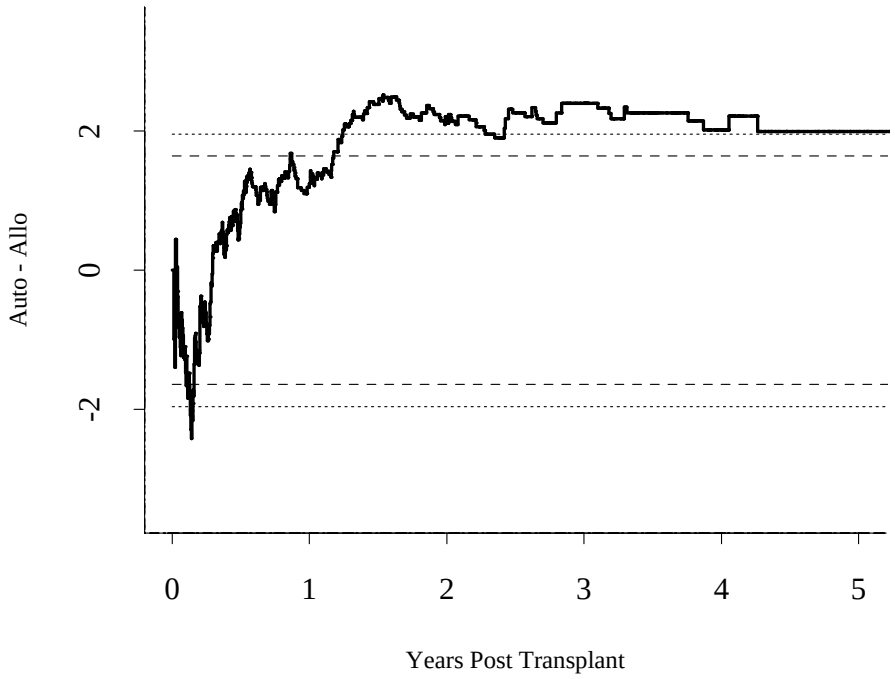
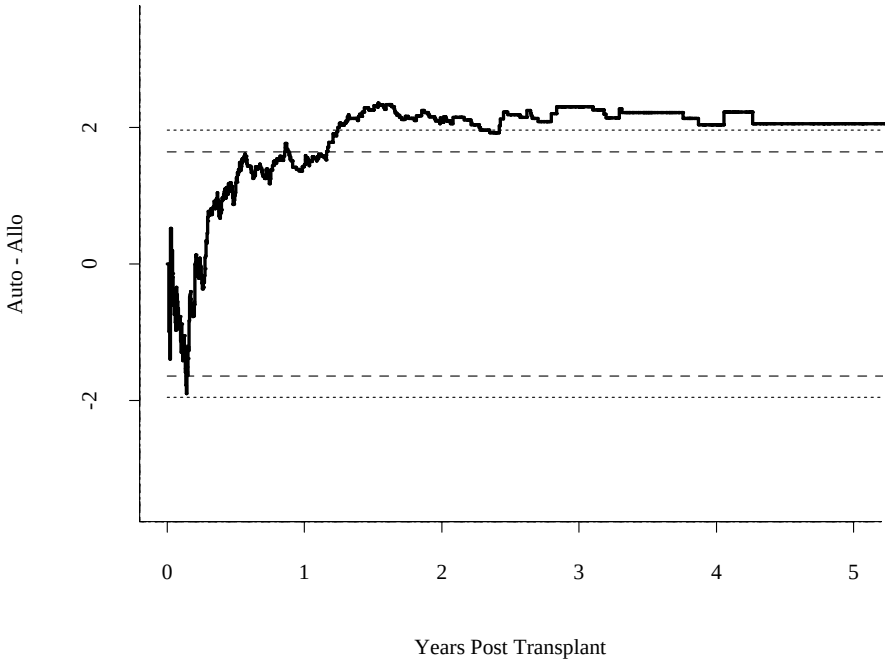
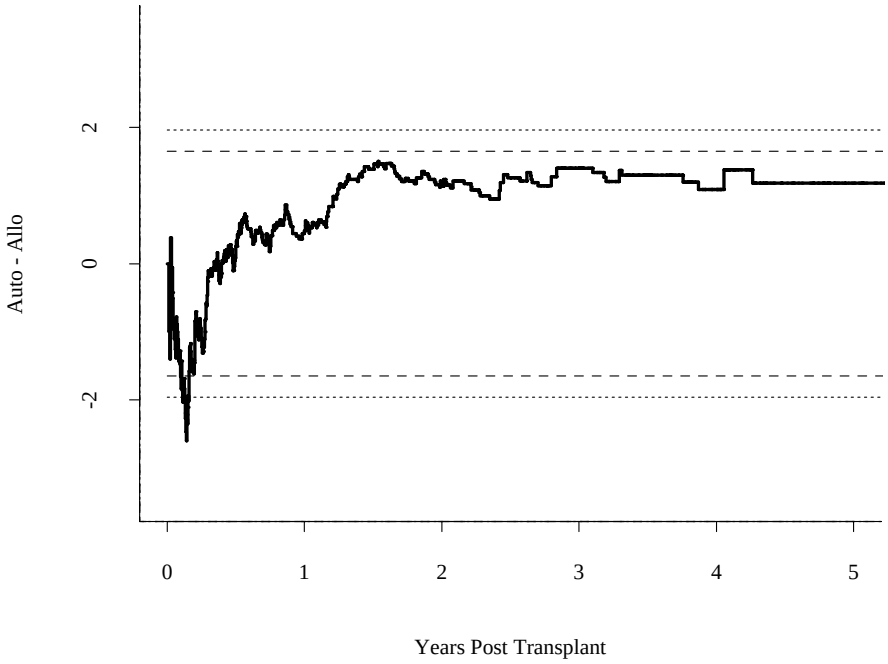


Figure 4. Standardized Difference In The Cumulative Hazards When Age Is Confounded With Type Of Transplant



Age 30 Or Less



Age Over 30