January 2013

A Monte Carlo Approach to Change Point Detection in a Liver Transplant

Alexia Melissa Makris
University of South Florida, amakris@health.usf.edu

A Monte Carlo Approach to Change Point Detection in a Liver Transplant Population

by

Alexia Melissa Athienitis-Makris

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Department of Epidemiology and Biostatistics
College of Public Health
University of South Florida

Major Professor: James Mortimer, Ph.D.
Co-major Professor: Fred Huffer, Ph.D.
Co-major Professor: Meenakshi Devidas, Ph.D.
Angel Alsina, M.D.
Wei Wang, Ph.D.

Date of Approval:
October 1, 2013

Keywords: survival analysis, distance effect, AIC, accelerated failure time

Dedication

This work is dedicated to my loving son Gabriel, my daughter Jennifer, and my husband Mahos who have been supporting me while I work on this dissertation for all these years; and to my parents for their love and support.

Acknowledgments

First of all I would like to acknowledge Dr. Angel Alsina for his unfailing support and mentoring over the years. He has not only taught me various clinical aspects necessary to complete this dissertation but he has also ensured that I persevered even when it seemed that there was no end in sight. Thank you for teaching me how to fight for myself and making me a stronger person. I believe his profound positive impact will continue to benefit my future professional work and I am grateful to have a friend for life.

I would also like to thank Dr. Fred Huffer for answering countless questions at various hours of the day. He has been a mentor and a friend to me since August 2000 when I entered the graduate program at Florida State University and I am pleased and honored to have him as a major professor through this journey.

Dr. Kate Wolfe-Quintero has been a key person in ensuring that I made it through this program and I am grateful to her for helping with my writing skills.

Table of Contents

List of Tables

List of Figures

Abstract

Patient survival post liver transplant (LT) is important to both the patient and the center's accreditation, but over the years physicians have noticed that distant patients struggle with post LT care. I hypothesized that patient's distance from the transplant center had a detrimental effect on post LT survival. I suspected Hepatitis C (HCV) and Hepatocellular Carcinoma (HCC) patients would deteriorate due to their recurrent disease and there is a need for close monitoring post LT. From the current literature it was not clear if patients' distance from a transplant center affects outcomes post LT. Firozvi et al. (Firozvi AA, 2008) reported no difference in outcomes of LT recipients living 3 hours away or less. This study aimed to examine outcomes of LT recipients based on distance from a transplant center. I hypothesized that the effect of distance from a LT center was detrimental after adjusting for HCV and HCC status.

Methods:

This was a retrospective single center study of LT recipients transplanted between 1996 and 2012. 821 LT recipients were identified who qualified for inclusion in the study. Survival analysis was performed using standard methods as well as a newly developed Monte Carlo (MC) approach for change point detection. My new methodology, allowed for detection of both a change point in distance and a time by maximizing the two parameter score function (M_{2p}) over a two dimensional grid of distance and time values. Extensive simulations using both standard distributions and data resembling the LT data structure were used to prove the functionality of the model.

Results:

Five year survival was 0.736 with a standard error of 0.018. Using Cox PH it was demonstrated that patients living beyond 180 miles had a hazard ratio (HR) of 2.68 (p-value<0.004) compared to those within 180 miles from the transplant center. I was able to confirm these results using KM and HCV/HCC adjusted AFT, while HCV and HCC adjusted LR confirmed the distance effect at 180 miles (p=0.0246), one year post LT. The new statistic that has been labeled M_{2p} allows for simultaneous dichotomization of distance in conjunction with the identification of a change point in the hazard function. It performed much better than the previously available statistics in the standard simulations. The best model for the data was found to be $\lambda(t; Z, X) = \lambda_0(t)\exp[\{\beta + \theta I(t \geq \tau)\}I(Z > c) + \eta' X(t)]$ which dichotomizes the distance Z, replacing it by *I(Z>c)*, and then estimates the change point c and τ.

Conclusions:

Distance had a detrimental effect and this effect was observed at 180 miles from the transplant center. Patients living beyond 180 miles from the transplant center had 2.68 times the death rate compared to those living within the 180 mile radius. Recipients with HCV fared the worst with the distance effect being more pronounced (HR of 3.72 vs. 2.68). Extensive simulations using different parameter values in both standard simulations and simulations resembling LT data, proved that these new approaches work for dichotomizing a continuous variable and finding a point beyond which there is an incremental effect from this variable. The recovered values were very close to the true values and p-values were small.

Chapter One

Introduction

1.1 Distance Effect on Post-Transplant Survival

Patient survival post-transplant is important to both the patient and a transplant center's accreditation, especially given the scarcity of donor organs and the increasing demand for them. Physicians and patients work closely to improve survival of transplant patients, so a thorough understanding of factors affecting this is an important key to improving post-transplant survival. Over the years, physicians had noticed that patients who travelled further to visit the transplant center experienced more challenges with post-transplant care. Although there were a few papers on this topic, most used a very short follow up period of one year during which surgical complications come into play and had majority of the patients living closest to the LT center (Firozvi et al., 2008, Axelrod et al., 2008, Kemmer et al., 2011, Zorzi et al., 2012). I decided to revisit this topic with a larger number of patients in a retrospective study of all patients at Tampa General Hospital. I hypothesized that further distance of the patient's residence from the liver transplant center leads to worse survival post LT. This study aimed to examine outcomes of liver transplant recipients based on distance from a transplant center. I examined the distance effect using cubic regression splines to detect nonlinearity, and with Akaike Information Criterion (AIC) and a Monte Carlo (MC) approach used to determine a distance cutoff.

1.2 Literature Review on Outcomes

Since liver transplant was first performed by Dr. Starzl in 1963, it has become widely indicated for patients with fatal acute and chronic end stage liver diseases (ESLD) (Starzl et al., 1982). Over 6000 LTs are performed each year in the U.S., with improved survival noted in the recent years

(Alsina et al., 2009). A donor and recipient disparity in LT currently exists (Merion, 2009; Merion et al., 2005). United Network for Organ Sharing (UNOS) was established in 1987 to ensure equitable access to these organs (Tuttle, Curley, Roh, 1997). It is also important to identify patients who can obtain most survival benefit (Merion et al., 2005). Thus, guidelines for liver allocation have evolved over the years to achieve the best outcome for all patients in need of LT.

Since 2002, UNOS has incorporated the use of Model for End Stage Liver Disease (MELD) to modify liver allocation (Freeman et al., 2002). MELD is a formula based on objective laboratory values, with some exception points such as those given for hepatocellular carcinoma (HCC) to determine patients' risk of death and hence the need for LT. This is an excellent and successful example of how statistical models can be translated into policy.

One of the goals of this dissertation is to investigate whether statistical models can also be used to identify patients at higher risk of death after LT. Freeman et al. (2002) have recognized that post-transplant risk models will be important as they add to the pre-transplant risk models such that scarce organs will not only benefit patients with the highest risk of dying without LT, but also achieve highest chance of survival with the LT.

I was interested in studying the distance effect on outcome after LT. However, most of the geographical literature emphasized allocation disparities before LT. It has been shown that LT candidates who live far from a LT center had less chance of receiving a LT (Tuttle et al., 1997). Distant patients also had higher rate of death and dropout from the list due to worsening of their medical conditions (Zorzi et al., 2012). They were less likely to be called in as a "backup" patient and less likely to be admitted during times of emergency (Firozvi et al., 2008). Given access disparities that arise from distance barriers pre-transplant, I hypothesized that the distance would also affect post LT outcome.

There has been a dearth of literature evaluating the association between residential distance from a LT center and post LT survival. Through literature review, I have identified only two studies from the U.S., both of which yielded unexpected results. Firozvi et al. (2008) showed that patients who lived more than 3 hours away had similar outcomes to those who lived closer for all of the following: listing, transplantation, and 1 year survival after LT. Axelrod et al. (2008) examined the distance effect on outcome via discussion of urban vs. rural settings. Nearly 14% of US population does not live in major urban areas where LT centers are predominately located; patients with long travel distance are those from rural and small town settings. Their study showed that rural and small town residents had lower wait list registration rate and lower transplant rate for liver, compared to urban residents. However, once a patient was wait listed, the wait time among rural residents was not longer, and the post-LT survival was not different (Axelrod et al., 2008).

Due to the donor and recipient disparity in LT that currently exists (Merion, 2009), it is important to maximize the utility of transplantation, whether measured by post LT survival, rejection, access to post-transplant care, quality of life, or other variables. UNOS plans to modify liver allocation in order to make waiting time more uniform across the country for patients with similar disease severity. However, these models do not specifically address post-transplant benefits (Freeman et al., 2002; Washburn, 2008, 2012; Washburn, Pomfret, & Roberts, 2011). One aspect of LT that had not been thoroughly studied was the impact of the patient's distance (residence) from the transplant center on post-transplant survival, although it is already known that distance was a detriment to access and being listed for liver transplantation (Axelrod et al., 2008; Kemmer, Alsina, & Neff, 2011; Park et al., 2012; Park et al., 2011).

Distance or geographical studies in liver transplantation are rare and tightly interwoven with concepts of 1) allocation models, 2) large vs. small centers and their claims of superior ability to care for patients at long distances vs. to serve their community, 3) post-transplant care, 4) insurance carriers and centers of excellence, 5) and overall utility of transplants (Axelrod et al.,

2008; Firozvi et al., 2008; Kemmer et al., 2011; Park et al., 2012; Park et al., 2011). This study was conducted because it remained unclear to us if the patient's distance from the transplant center affected the outcome post liver transplantation.

Five years ago, a study published in *Liver Transplantation*, caught my attention. It concluded that distance does not adversely affect outcomes (Firozvi et al., 2008). The conclusion of this study and the editorial comment (Washburn, 2008) that it received were of much interest, which prompted me to restudy this topic. In the Firozvi study, based on 66 transplant patients, a survival analysis was conducted utilizing a small number of patients at long distances. Lack of statistical power due to the small sample size may have explained the similar survival curves. The study only focused on survival at one year post-transplant, a time period too short to evaluate the multiple difficulties that the patients may encounter in their post-transplant course of care. These include, but are not limited to, Hepatitis C (HCV) and Hepatocellular Carcinoma (HCC) recurrence, compliance, and medical complications of transplantation. In addition, using one specific cutoff point for distance that is predetermined ahead of time, excluded the possibility of discovering a different and more appropriate cutoff that might have been further or closer to the transplant center than what was initially suspected.

Axelrod et al. (2008) found that rural patients had worse survival, while Park et al. (2011) reported that pediatric rural LT patients did not have worse survival. There are some who argued that rural health care, in general, is of lower quality than that provided in urban settings because better prepared or more competent health care providers would be attracted to urban centers where there would be more opportunities and access to cultural and academic functions than can be found in rural settings. Axelrod et al. (2008) looked at 174,630 patients who underwent heart, liver or kidney transplantation with an accrual period of 5 years (1999-2004). Each of these organs was different with respect to follow-up and expected outcomes. Because heart and kidney transplant care can be different from LT care, it is impossible in this study to determine the effect of distance on LT patients' survival. For example, patients who received a kidney

transplant were able to receive follow-up care from their local nephrologist, whereas patients who received a liver transplant had to follow-up with the center that performed the transplant. This made the issue of distance more important in LT recipients as opposed to those receiving a kidney. Also this study examined rural vs. urban residence which was not necessarily the same as the effect of driving distance from the transplant center.

The study by Park et al. (2011) found that rurality did not significantly affect health outcomes after LT in 388 pediatric patients. The authors used urban influence (UI) codes published by the US Department of Agriculture (USDA) to stratify patients as urban or rural depending on county of residence. Again, rural vs. urban does not necessarily reflect distance from the transplant center. This study also differed from mine in that their population consisted of pediatric LT patients. They included patients with UNOS status 1 (fulminant hepatic failure) who received priority on the waiting list because these patients only had hours to live without a transplant. Logistic regression models were used in their papers that disregarded the length of survival. By contrast, I used, Cox PH and AFT models that take into account the actual duration of survival. In addition, other studies used a pre-specified mileage/driving time cutoff whereas my study allowed that cutoff to be data driven. In a later study, Park et al. (2012) examined 3,307 pediatric LT patients and found that rural location had a negative impact on patient health within the first 6 months of LT by increasing the risk for allograft rejection, although patients in rural areas had lower rates of developing post LT lymphoproliferative disorder.

Several studies addressed the effect of distance on pre-transplant survival and waiting time to LT.

Kemmer et al. (2011) examined the time between diagnosis and transplant. While distance did not appear to affect access to transplant, it did not necessarily preclude a distance effect on survival post-transplant. In this paper only 439 patients were studied and the median distance was 36.8 miles with a range of 0.5 to 231 miles. The present study included a larger sample size

of 821 patients living a median distance of 27.8 miles with a range of 0 to 548 miles from the transplant center.

Zorzi et al. (2012) studied 5,673 candidates listed for liver transplant in UNOS Region 4 between 2004 and 2010. The authors established that there was a deterioration in survival on the waiting list for patients living more than 30 miles from the transplant center in patients with Model for End stage Liver Disease (MELD) score over 20 (p-value<0.0001). The 30 mile intervals were pre-specified as 0-30, 30-60 and over 60 with the cutoffs not dictated by the data. Methodology included Kaplan Meier and Cox PH, but survival post-transplant was not studied. It was also specific to those with MELD scores over 20, whereas my study pertained to the entire LT population.

Barritt et al. (2012) assessed the effect of distance on whether or not the patient received a LT. The mean distance to the primary transplant center did not differ between those who were transplanted and those who were not. Age, race, sex, distance to transplant centers, and rural vs. urban residence did not influence the odds of receiving a liver transplant (Barritt, Telloni, Potter, Gerber, & Hayashi, 2012).

It is important to note that distance from a transplant center can lead to selection bias in those receiving a transplant. Liver failure patients living far from a liver transplant center had less chance of receiving transplantation (Benach & Amable, 2004; McCormick et al, 2004). Distant patients had difficulty navigating through the transplant process, even after referral. These distant patients were also less likely to be called in as a “backup” patient (Washburn et al., 2011), which reduced their chances of successful transplantation, therefore reducing survival. The differential probability of transplantation by distance needs to be considered, because it can result in those transplanted patients living at longer distances being in poorer health than those living closer to the transplant center.

My study was different from previous studies in the following ways: First, it utilized a larger number of patients overall (821 vs. 66 in Firozvi et al.) with more at longer distances. Second, it followed liver transplant recipients for up to 5 years post transplantation vs. one year in previous studies (Firozvi et al., 2008). Third, it excluded patients who relocated temporarily to be closer to the transplant center for purposes of receiving a transplant. Hence, this significant bias present in other studies such as that by Firozvi et al., was eliminated. Fourth, adjustment was made for important covariates, such as HCV and HCC which recurred affecting survival.

1.3 Literature Review of the Methods

The Kaplan-Meier (KM) estimator, the most widely used method for estimating survivor functions in biomedicine, is also known as the product-limit estimator because the estimated survival probabilities are computed using a product limit formula (Kleinbaum & Klein, 2012). Researchers were using this method for many years prior to 1958 when Kaplan and Meier showed that it was a nonparametric maximum likelihood estimator, therefore giving it a solid theoretical justification (Allison, 2010). When there are no censored data the KM estimator $\hat{S}(t)$ is just the sample proportion of observations with event time greater than *t*. When censoring is present then for a given time *t* all the event times that are less than or equal to *t* are taken and for each of those event times, the probability of surviving to time t_{j+1}, given that one has survived to time t_j is computed using the following formula:

$$\hat{S}(t) = \prod_{j:\, t_j \le t} \left[1 - \frac{d_j}{n_j}\right]$$

where at each time t_j there are $n_j$individuals at risk of death and d_j is the number of individuals who die at time t_j.

The log-rank test compares two or more survival curves using a null hypothesis of a common survival curve. It is based on the summed observed minus expected score for a given group and its variance estimate with *k-1* degrees of freedom where *k* is the number of groups (Kleinbaum & Klein, 2012). The log-rank statistic can be written as

$$\sum_{j=1}^{r}(d_{1j} - e_{1j})$$

summing over all unique event times in every group. The expected number of events in group 1 at time *j*, $e_{1j} = \frac{n_{1j}d_j}{n_j}$. The Wilcoxon statistic is a weighted sum of the deviations of observed number of deaths from the expected number of deaths given by

$$\sum_{j=1}^{r} n_j(d_{1j} - e_{1j})$$

This test puts more weight on early times as compared to late times since n_j always decreases; therefore it is less sensitive than the log-rank test to differences between groups that occur at later on in time.

The Cox PH model (Cox, 1972) can be written as

$$h_i(t) = \lambda_0(t)\exp\{\beta_1 x_{i1} + \cdots + \beta_k x_{ik}\}$$

This equation illustrates that the hazard for an individual i at time *t* is the product of a linear function of a set of *k* fixed covariates which is then exponentiated and the baseline hazard function $\lambda_0(t)$. Partial likelihood enables estimation of the coefficients β of the proportional hazards model without, having to specify the baseline hazard function, except for the restriction that it must be positive. Assumptions include non-informative censoring and proportional hazards, but the model can be generalized to allow for non-proportional hazards. In my study, Breslow's method was employed for handling ties (Breslow, 1974).

In contrast to the Cox PH, the Accelerated Failure Time (AFT) model, depicts a relationship between the survivor functions of any two individuals. If $S_i(t)$ is the survivor function for individual *i*, then for another individual *j*, the AFT model states that $S_i(t) = S_j(\phi_{ij}t)$ for all *t* where ϕ_{ij} is a constant that is specific to the pair *(i, j)* so what makes one individual different from another is the rate at which they age. The model can be written as

$$\text{Log } T_i = \beta_0 + \beta_1 x_{i1} + \ldots\ldots\ldots + \beta_k x_{ik} + \sigma\varepsilon_i$$

where T_i denotes the event time for the i^{th} individual, $x_{i1}, \ldots, x_{ik}$ are the values of *k* covariates for

that same individual, ε_i is a random disturbance term, and $\beta_0, \ldots, \beta_k$and σ are parameters to be estimated. AFT models have the advantage that they allow distributions of $\boldsymbol{\varepsilon}$ besides the normal distribution but retain assumptions of constant mean and variance, and independence across observations (James, 2005; Keiding, Andersen, & Klein, 1997).

Some typical distributions for the AFT model can be found below. Usually for parametric models, the shape parameter *p* would be held fixed and *λ* re-parameterized in terms of predictor variables and regression parameters. The AFT was used in the past for modeling kidney transplant data (Lambert, Collett, Kimber, & Johnson, 2004) but not LT.

Table 1: Common AFT Distributions

Distribution	f(t)	S(t)	h(t)
Exponential	$\lambda\ \exp(-\lambda t)$	$\exp(-\lambda t)$	λ
Weibull	$\lambda p t^{p-1} \exp(-\lambda t^p)$	$\exp(-\lambda t^p)$	$\lambda p t^{p-1}$
Log-Logistic	$\frac{\lambda p t^{p-1}}{(1+\lambda t^p)^2}$	$\frac{1}{1+\lambda t^p}$	$\frac{\lambda p t^{p-1}}{1+\lambda t^p}$

The logistic regression (LR) model is most appropriate when events can only occur at regular, discrete points in time. With my data, it was used at 1 year post LT with 739 patients eligible for inclusion. Eighty-two patients were excluded from this analysis because they did not have one year's worth of follow-up at the time this study ended (September 25th 2012) although they were alive. I felt it was a useful model for this data since ties arise from grouping continuous data into intervals (Albert & Anderson, 1984; Hilbe, 2009). Allowing P_{it} to be the conditional probability that individual *i* has an event at time *t*, given that the individual has not already had an event, then the logistic regression equation is

$$\log\left(\frac{P_{it}}{1-P_{it}}\right) = \alpha_t + \beta_1 x_{it1} + \cdots + \beta_k x_{itk}$$

The KM and Cox PH were the most frequently used in the literature due to the ease of

interpretation. AFT models required more stringent assumptions about the distribution of the data so they were not used as often. AFT had the advantage of a completely specified hazard and survival function but required an assumption about the underlying distribution. The Cox PH model does not rely on distributional assumptions and the baseline hazard is not necessary for estimation of the hazard ratio. However the distribution of the survival time is unknown and it is less consistent with the theoretical survival function as it is usually a step function. LR ignores valuable information concerning the length of survival post-transplant and provides only a snap shot at a particular point in time (1 year in this study).

Akaike's Information Criterion (AIC) is a measure of the goodness of fit of an estimated statistical model (Bozdogan, 1987). It is an operational way of trading off the complexity of an estimated model against how well the model fits the data. I used this criterion to establish the best possible cutoff point for distance in Cox PH models.

Boisson et al. (2008) studied the survival time as affected by the rate of degradation of quality of life and extended the partial likelihood score statistic to apply to survival time. Goodman et al. (2011) used data driven methods to estimate both the number of change points in the hazard function and where those change points occurred. An alpha spending function was used where $\alpha^*(k) = \frac{\alpha}{2^{k-1}}$ for the overall significance level α to ensure strong evidence for choosing a more complicated model. Nelder and Mead (1965) developed a Wald-type test statistic that uses the Nelder-Mead Simplex optimization algorithm, which is robust but relatively slow. Specifically they tested sequentially for the presence of an additional change point using $H_0: \alpha_{k-1} - \alpha_k = 0$ vs $H_1: \alpha_{k-1} - \alpha_k \neq 0$ where the test statistic was $X_w = \frac{(\hat{\alpha}_{k-1} - \hat{\alpha}_k)^2}{Var\,(\hat{\alpha}_{k-1} - \hat{\alpha}_k)}$ which was distributed as a $\mathcal{X}^2$ with one degree of freedom. Once a null hypothesis could not be rejected then one stops and concludes that there were no more change points. A simulation study successfully demonstrated the strength of this new method and found that power was most affected by sample size. Change points were restricted to larger than the first non-censored survival time and smaller than the last non-censored survival time.

Liang et al (1990) introduced a variant of the Cox PH model and used age as the time variable and τ as the change point. However their method required the time invariant condition that $B(\tau) = \lim_{n\to\infty}\left\{\frac{\partial^2 l}{\partial\theta^2}(0,\beta,\eta,\tau)\right\}^{-1}\left\{\frac{\partial^2 l}{\partial\theta\partial\gamma}(0,\beta,\eta,\tau)\right\}$ was independent of τ. This time invariant condition of $B(\tau)$ usually did not hold when there existed additional covariates *X* and their coefficients $\gamma \neq 0$.

Matthews et al. (1982) presented a likelihood ratio test for detecting a single change point. They used the following hazard function:

$$\lambda(t) = \begin{cases} \lambda_1, & t \leq \tau \\ \rho\lambda_1, & t > \tau \end{cases}$$

Since standard asymptotic likelihood inference on the parameters $\lambda_1, \rho,$ and τ was not possible, they used maximum likelihood estimators and unconditional procedures. The log-likelihood test statistic for the null hypothesis $\tau = 0$, is $\Delta_0 = l(\hat{\lambda}_1, \hat{\rho}, \hat{\tau}) - l(\hat{\lambda}, \hat{\lambda}, 0)$, where $\hat{\lambda}$ was the maximum likelihood estimator of the failure rate in a simple exponential model. Although it would have been naïve to apply asymptotic likelihood ratio theory to conclude that $2\Delta_0$ had a $\mathcal{X}^2$ distribution with 2 degrees of freedom, the percentiles of the $\mathcal{X}^2_{(2)}$ distribution agreed quite well with the simulation results.

Nguyen, Rogers, and Walker (1984) pointed out that the likelihood was unbounded under the alternative hypothesis since a singularity appeared as $\rho \to \infty$ and τ was taken immediately before the largest observation. Matthews et al. (1985) removed the singularity by considering the data as discrete and reformulated the likelihood as a product of probabilities rather than densities. They proposed tests based on maximal score statistics and derived the asymptotic significance levels. Yao (1986) suggested constraining the estimate of τ proposed previously by Matthews, Farewell, and Pyke (1985) so it did not fall in the interval between the two largest observations. Worsley (1988) found that the singularity could be removed if the largest observation was artificially considered to be censored.

Chang, Chen, and Hsiung (1994) proposed an estimator for the change-point $\hat{\tau}$ that was easier to implement and could be considered a nonparametric counterpart of the estimator resulting from the score process proposed by Matthews et al (1985). Henderson (1990) suggested some modified likelihood ratio tests with the most important modification involving a weighted and standardized likelihood ratio value, which leads to a higher power and a smaller mean squared error for τ. Loader (1991) derived large deviation approximations to the significance level of the likelihood ratio test by a random change of time scale for the empirical process.

Akman and Raferty (1986) analyzed a change-point Poisson process and provided point and interval estimates of the change-point. They investigated the small-sample performance of the proposed procedures by means of a Monte-Carlo study. Raftery and Akman (1986) developed a comparable Bayesian approach. A kernel method for the estimation of the change-point of the most rapid change of a continuous hazard function was then proposed by Müller and Wang (1990).

Kleinbaum (1996) found that the assumption of proportional hazards was not always relevant in the whole range of the covariate and the covariate could be dichotomized to satisfy this assumption. Kleinbaum's procedure led to a two-phase Cox model with a change-point according to a threshold that may be fixed or estimated from the data. Other authors considered a non-regular Cox model involving a two-phase regression and time-dependent covariates, with a change-point at an unknown time (Liang, Self, and Liu, 1990; Luo, 1996; Luo, Turnbull, and Clark, 1997).

Pons (2003) studied the asymptotic behavior of the maximum partial likelihood estimator of the parameters in a non-regular Cox model with a change-point according to the unknown threshold of a covariate. In their model, the hazard rate of a survival time T^0 had the form

$\lambda_\theta(\, t \mid Z\,) = \lambda\,(t) \exp\{r_\theta\,(Z(t))\}$ with $r_\theta\big(Z(t)\big) = \alpha^T Z_1(t) + \beta^T Z_2(t) \mathbb{I}_{\{Z_3 \le \zeta\}} + \eta^T Z_2(t)\, \mathbb{I}_{\{Z_3 > \zeta\}}$

where $\theta = (\zeta, \xi^T)^T, \xi = (\alpha^T, \beta^T, \eta^T)^T, \lambda$ was an unknown baseline hazard function and $Z = (Z_1^T, Z_2^T, Z_3)^T$ was a vector of covariates.

Zucker and Lakatos (1990) presented two weighted log rank type statistics designed to have good efficiency over a wide range of lags. One was a maximum efficiency robust statistic, while the second was a simplified version of this statistic. Both of these were substantially more efficient than the conventional log rank statistic. This was an attractive method because it required no modeling assumptions. When there was a lag in the effect of a certain covariate, the proportional hazards assumption was violated causing the log rank test to be inefficient and therefore a weighted version was best. However, choosing weights is a non-trivial issue. Self et al. (1988) suggested an approach for choosing weights when the lag function was equal to some function. Another approach taken by the Physician's Health Committee was to give positive weight only to the portion of the trial during which one felt fairly certain that all or most of the full treatment effect was present, but this approach had serious drawbacks as early adverse effects could be overlooked (Physician's, 1983). Zucker and Lakatos' (1990) approach included all events so there was no risk of proving a treatment to be beneficial when there were early adverse effects; however their approach down-weighed early events even though they did not completely exclude them.

The goal of this study was to search for two types of change points in the hazard function: one with respect to distance *c* and the other with respect to time τ. It was possible that the distance effect was not present initially but became evident later on in the patient follow-up, after a time point τ. In addition I suspected that the distance effect began after a point c, with a potentially increasing effect beyond this point [max(z-c,0)]. Identifying change points in a hazard function was of great importance in survival analysis and I was particularly interested in the Monte Carlo approach used by Liu et al. (2008). It was computationally efficient, avoided technical assumptions that would be difficult to verify, and gave a legitimate p-value for the test for the existence of a change point at an unspecified location, which other methods failed to do.

1.3.1 Background

The hazard function (Liu et al., 2008) could be specified with one change point as

$$\lambda(t; Z, X) = \lambda_0(t) \exp[\{\beta + \theta I(t \leq \tau)\}Z + \eta' X(t)]$$

and for two change points as

$$\lambda(t; Z, X) = \lambda_0(t) \exp[\{\beta + \theta_1 I(t \leq \tau_1) + \theta_2 I(t \leq \tau_2)\}Z + \eta' X(t)]$$

where $\lambda_0(t)$ was an unspecified baseline hazard function, $\tau > 0$ was the change-point parameter, θ was the magnitude of change after the change point, β was the throughout distance effect on the hazard function, Z was distance from LT center which caused the non-constant hazard, η were the coefficients and X(t) were other risk factors that I adjusted for (HCV and HCC). Using this model I determined the presence of change points in the hazard as well as the magnitude of change in distance effect before and after the change-point τ.

Let $T_i = \min(\tilde{T}_i, C_i)$ and $\delta_i = I(\tilde{T}_i \leq C_i)$ where $\tilde{T}_i$ and C_i denote the failure time and censoring time for the i[th] subject. Define $N_i(t) = I(T_i \leq t, \delta_i = 1)$ as the count of failures on the i[th] subject at time t, $Y_i(t) = I(T_i \geq t)$ the at-risk indicator, and $\kappa = \inf\{t; pr(T > t) = 0\}$ the shortest time *t* such that it is not possible for both survival time and censoring time to exceed it. For a trial with *n* subjects, the partial likelihood for the unknown parameters *(θ, β, γ, τ)* based on the observed data is

$$L(\theta, \beta, \gamma, \tau) = \prod_{i=1}^{n} \int_0^{\kappa} \frac{\exp[\{\beta + \theta I(t \leq \tau)\}Z_i + \gamma' X_i(t)]}{\sum_{j=1}^{n} Y_j(t) \exp[\{\beta + \theta I(t \leq \tau)\}Z_j + \gamma' X_j(t)]} dN_i(t)$$

Complications arise with the Likelihood Ratio Test since the likelihood is not a smooth function of the change-point τ and under the null hypothesis of no change-point $H_0: \theta = 0$, the parameter τ disappears from the likelihood (Liang et al. 1990; Matthews et al. 1985). Matthews et al. (1985) proposed a maximal score statistic as an alternative, but the normalized score process indexed by the change-point parameter must converge to an O-U process. An O-U process is a stochastic/random process that is stationary, Gaussian and Markovian. It is a modification of

random walk where the process drifts toward the mean. Since the technical assumptions for the limiting O-U process were difficult to verify in semi-parametric models (Liang et al., 1990), there was a need for improved methodology.

Liu et al (2008) let $\eta = (\beta, \gamma')'$ and $X_i^* = (Z_i, X_i')'$ and defined the non-normalized score statistic as

$$S_\theta(\tilde{\eta};\ \tau) = \frac{\partial\, l(.;\tau)}{\partial\, \theta}\Big|_{\theta=0,\eta=\tilde{\eta}}$$

and the corresponding normalized score statistic as

$$W_\theta(\tilde{\eta};\ \tau) = \frac{S_\theta^2(\tilde{\eta};\ \tau)}{V\ (\tilde{\eta};\ \tau)}$$

where $\tilde{\eta}$ is the restricted maximum partial likelihood estimates under H_0 and $V(\tilde{\eta};\ \tau)$ is the variance estimator of $S_\theta(\tilde{\eta};\ \tau)$ evaluated under the null hypothesis. If the change point is suspected to lie in a region $\mathcal{A}$ then maximizing over this region gives

$$M = \sup_{\tau\in\mathcal{A}} |S_\theta(\tilde{\eta};\ \tau)| \text{ and } M^* = \sup_{\tau\in\mathcal{A}} W_\theta(\tilde{\eta};\ \tau)$$

Liang et al. (1990) proposed something similar to *M**, but they imposed a technical assumption that B(τ) was independent of τ, so when $\mathcal{A}$ was a fixed time interval then the normalized score process converged to an O-U process, where

$B(\tau) = \lim_{n\to\infty}\left\{\frac{\partial^2 l}{\partial\,\theta^2}(0,\beta,\gamma,\tau)\right\}^{-1}\left\{\frac{\partial^2 l}{\partial\,\theta\,\partial\,\gamma}(0,\beta,\gamma,\tau)\right\}$ (Liang et al., 1990)

This time-invariant condition of B(τ) usually did not hold in the presence of additional covariates when the coefficients were not zero (η≠0). As an alternative Liu et al. (2008) used an efficient Monte Carlo method to evaluate the statistical significance for both *M* and *M** which did not require the technical assumption that B(τ) was independent of τ nor the special O-U process covariance structure for the normalized score process. This approach numerically approximated the joint distribution of $\{S_\theta(\tilde{\eta};\ \tau), \tau \in \mathcal{A}\}$ and kept within process correlation intact.

1.3.2 Monte Carlo Method

Under H_0: $\theta = 0$, $dM_i(t) = dN_i(t) - Y_i(t)\exp\{\eta_0' X_i^*(t)\}\, d\Lambda_0(t)$ is a martingale, where η_0 is the true value of η and $\Lambda_0(t) = \int_0^t \lambda_0(u)du$. Let $\mathcal{I}(\theta,\eta;\ \tau)$ be the negative second derivative matrix of the

log partial likelihood at a supposed change-point time τ. By the law of large numbers $n^{-1}\,\mathcal{I}(0,\eta;\,\tau)$ converges to a matrix $\sum(\tau)$ under $H_0\colon \theta = 0$. Using a Taylor series expansion $S_\theta(\tilde{\eta};\,\tau)$ is asymptotically equivalent to $\sum_{i=1}^{n}\tilde{S}_{\theta,i}(0,\eta_0;\,\tau)$ which is a sum of *n* independent random variables with mean zero where $\tilde{S}_{\theta,i}(0,\eta_0;\,\tau) = S_{\theta,i}(0,\eta_0;\,\tau) - \sum_{\theta\eta}(\tau)\sum_{\eta\eta}^{-1}(\tau)\,S_{\eta,i}(0,\eta_0;\,\tau)$ or as a martingale integral

$$\tilde{S}_{\theta,i}(0,\eta_0;\,\tau) = \int_0^k [I(t \le \tau)\{Z_i - \tilde{z}_0(t)\} - \sum_{\theta\eta}(\tau)\sum_{\eta\eta}^{-1}(\tau)\{X_i^*(t) - \bar{x}_0^*(t)\}]dM_i(t) \qquad \text{- (*)}$$

Here $\sum_{\theta\theta}(\tau)$, $\sum_{\eta\eta}(\tau)$ and $\sum_{\theta\eta}(\tau)$ are the components of $\sum(\tau)$ according to the partition of θ and η. Under the null hypothesis $n^{-1/2}S_\theta(\tilde{\eta};\,\tau)$ converges to a zero-mean Gaussian process, the limiting variance is $\tilde{V}(\tau) = \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n}\tilde{S}_{\theta i}^2(0,\,\eta_0;\,\tau)$, and the covariance is

$$\tilde{V}(\tau_1,\tau_2) = \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n}\tilde{S}_{\theta i}(0,\eta_0;\,\tau_1)\,\tilde{S}_{\theta i}(0,\eta_0;\,\tau_2)$$

Replacing the unknown parameters in $\tilde{S}_{\theta,i}(0,\eta_0;\,\tau)$ with consistent estimators under the null (with the exception of τ) $\hat{S}_{\theta,i}(\tau) = \int_0^k [I(t \le \tau)\{Z_i - \tilde{Z}_0(t;\,\tilde{\eta})\} - \mathcal{I}_{\theta\eta}(0,\tilde{\eta};\,\tau)\,\mathcal{I}_{\eta\eta}^{-1}(0,\tilde{\eta};\,\tau)\{X_i^*(t) - \tilde{X}_0^*(t)\}]d\hat{M}_i(t)$

where $d\hat{M}_i(t) = dN_i(t) - Y_i(t)\exp\{\tilde{\eta}'X_i^*\}\,d\tilde{\Lambda}\,(t)$ and $\tilde{\Lambda}\,(t) = \sum_{i=1}^{n}\int_0^t \frac{dN_i(s)}{\sum_{j=1}^{n} Y_j(s)\exp\{\tilde{\eta}'X_j^*(s)\}}$.

Therefore at each time τ, $n^{-1}V\,(\tilde{\eta},\tau) = n^{-1}\sum_{i=1}^{n}\hat{S}_{\theta,i}^2(\tau)$ is a consistent estimate for the asymptotic variance of $n^{-1/2}S_\theta(\tilde{\eta};\,\tau)$ under the null hypothesis $H_0\colon \theta = 0$.

In order to approximate the distribution of the test statistics proposed by Liu et al. (2008), they define

$\hat{S}_\theta(\tilde{\eta};\,\tau) = \sum_{i=1}^{n}\hat{S}_{\theta,i}\,(\tau)G_i$ where $\{G_i, i = 1,, \dots, n\}$ are independent standard normal random variables that act as disturbance factors in the sum and agree that the randomly perturbed variant $\hat{S}_\theta(\tilde{\eta};\,\tau)$, given the observed data, has the same limiting distribution as $S_\theta(\tilde{\eta};\,\tau)$ (proof in (Lin, Wei, & Ying, 1993)). Likewise $W_\theta(\tilde{\eta};\,\tau)$ can be approximated by $\widehat{W}_\theta(\tilde{\eta};\,\tau) = \hat{S}_\theta^2(\,\tilde{\eta};\,\tau)\big/V\,(\tilde{\eta},\,\tau)$.

A large number of randomly disturbed score processes $\hat{S}_\theta(\tilde{\eta};\,\tau)$ and $\widehat{W}_\theta(\tilde{\eta};\,\tau)$ are used to

calculate $\widehat{M} = \sup_{\tau \in \mathcal{A}} |\hat{S}_\theta(\tilde{\eta};\ \tau)|$ and $\widehat{M}^* = \sup_{\tau \in \mathcal{A}} \widehat{W}_\theta(\tilde{\eta};\ \tau)$ respectively, and the empirical quantiles of $\widehat{M}$ and $\widehat{M}^*$.

The Monte Carlo resampling procedure outlined above has several advantages:

1. The components to construct $\hat{S}_{\theta,i}(\tau)$ are only calculated once and reused to find the variance $V\ (\tilde{\eta},\ \tau)$.
2. The score statistic process only jumps at the observed failure times so when searching for a change point over the candidate region $\mathcal{A}$, test statistics are only calculated at the observed death times.

In the event that distance was associated with L change points $\tau_1 < \cdots < \tau_L$ with an effect on the hazard function where L >1, then the hazard function is defined as

$$\lambda(t; Z, X) = \lambda_0(t) \exp[\left\{\beta + \sum_{l=1}^{L} \theta_l I\big(t \in J_l(\boldsymbol{\tau})\big)\right\} Z + \gamma' X(t)]$$

where $J_1(\boldsymbol{\tau}) = (0, \tau_1], J_2(\boldsymbol{\tau}) = (\tau_1, \tau_2], \ldots J_{L-1}(\boldsymbol{\tau}) = (\tau_{L-2}, \tau_{L-1}], J_L(\boldsymbol{\tau}) = (\tau_L, \kappa)$ are intervals subdividing the study period K (which is 5 years in my case). The interval (τ_{L-1}, τ_L) is not needed because the model is already saturated over $(0, \kappa)$ without it.

Under multiple change points where $\boldsymbol{\theta}$ and $\boldsymbol{\tau}$ were L *x1* vectors, the partial likelihood function becomes

$$L\ (\boldsymbol{\theta}, \beta, \eta, \boldsymbol{\tau}) = \prod_{i=1}^{n} \int_0^{\kappa} \frac{\exp[\{\beta + \sum_{l=1}^{L} \theta_l\, I(t \in J_l(\boldsymbol{\tau}))\} Z_i + \gamma' X_i(t)]}{\sum_{j=1}^{n} Y_j(t) \exp[\{\beta + \sum_{l=1}^{L} \theta_l\, I(t \in J_l(\boldsymbol{\tau}))\} Z_j + \gamma'\, X_j(t)]} dN_i(t)$$

The null hypothesis for multiple change points is $H_0^{\#}: \theta_1 = \cdots = \theta_L = 0$. Then the normalized maximal score test statistic for testing $H_0^{\#}$ has the form $M^{\#} = \sup_{\tau_1 \in \mathcal{A}_1, \ldots \tau_L \in \mathcal{A}_L} W_\theta^{\#}(\tilde{\eta};\ \tau_1, .. \tau_L)$ where $\{\mathcal{A}_1, \ldots \mathcal{A}_L\}$ are the proposed regions in which the change points $\{\tau_1, \ldots, \tau_L\}$ lie respectively.

The method used to evaluate statistical significance used numerous randomly disturbed score statistics that were generated over all the potential change points. Those were then used to compute the 100(1-α) quantile of the Monte Carlo resampled maximal score statistics which was used as the critical threshold. The proportion of simulation samples in which the null hypothesis of no effect was rejected at the 0.05 level of significance when the null was false defined the empirical power (Leffondré, Abrahamowicz, & Siemiatycki, 2003; Væth & Skovlund, 2004). Type II error was computed as 1-power. The empirical type I error was the proportion of p-values less than the nominal 0.05 significance level (under a true null) from testing the null hypothesis on each simulated sample (Rempala & Looney, 2006). These values were computed and reported for each of the 4 extensions described next in the methods.

1.4 Statement of the Problems/ Applications and Significance

The methods described above were not sufficient for detecting change points in the hazard while simultaneously dichotomizing distance. This brought forth an opportunity for improvement in the existing methodology of the Monte Carlo approach to change point detection which was first proposed by Liu et al. (2008). This research required a more complicated null hypothesis which simultaneously dichotomized distance at a point c and located the time τ such that the distance effect was observed for times greater than τ. This required maximizing the partial likelihood over a grid of distance and time values. My approach required substantial modifications of standard methods in survival analysis and the use of very recently developed methods that were not yet in common use and described only in journal articles (Liu et al., 2008). The goal was to best establish the mileage point at which survival post-transplant declined and the time point post-transplant at which the hazard changed, while adjusting for the effects of multiple covariates observed at the time of transplant.

Critical barriers to exploring this topic include lack of comprehensive data and methods. The United Network for Organ Sharing (UNOS) database has many limitations that prevented researchers from studying the effects of distance on post LT survival. Elements common to most

forms included primary diagnosis, medical condition at time of transplant, functional status, pre-transplant serology, and transplant procedure type. Pre-transplant risk factors included factors such as portal vein thrombosis and previous abdominal surgery for liver transplants. Elements common to the Transplant Recipient Follow-up (TRF) forms included patient status and cause of death, graft status and cause of graft failure, rejection episodes, and biological/anti-viral and immunosuppressive medications. Patients' place of residence and lab results were not available in the detail that was captured here, especially with regard to distance.

Since the UNOS database does not contain patient address information, this could be the reason behind the lack of published data on the topic of patient distance. The preliminary data consisted of 627 patients who were transplanted at Tampa General Hospital (TGH) between 1996 and 2009. Patients were excluded if they underwent multiple organ transplantation (n=37), had fulminant hepatic failure (n=20), died the day of transplant (n=5), or relocated temporarily to be closer to the transplant center (n=46). An additional 194 patients who were transplanted between 2009 and 2012 were added to the initial 627. None of the previous studies had attempted to use AFT models which, although more restrictive due to the distribution assumption, could prove to be more powerful when the underlying distribution was known or could be reasonably estimated. The proposed method for identifying change points in the hazard function was an extension to recent methods described in journal articles (Liu et al., 2008; Matthews & Farewell, 1982; Zucker & Lakatos, 1990), adapted and specifically designed to dichotomize continuous variables and locate change points in the hazard. The improvements to scientific knowledge offered by this study were a more comprehensive dataset and the extension of newer statistical methodology.

The LT program at TGH began in 1996 and is currently the 9th largest LT program in the country and the 4th busiest transplant center in the nation. Data for this analyses was derived from TGH electronic medical records called Electronic Privacy Information Center (EPIC), transplant databases including Organ Transplant Care Platform (OTTR), Chartview (previous electronic record system at TGH which has been replaced by EPIC), flowcharts, and the United Network for

Organ Sharing database containing TGH transplant data. Those who relocated to temporary housing near the transplant center were excluded from the study. The objective is to study the distance between the patients' permanent residence in relation to the transplant center, as it pertains to difficulty with post LT care, HCV treatments, compliance with tests and other factors. Including patients that came from out of state, relocated near TGH, and left early post-transplant, transferring their care to other centers would invalidate and bias this study.

Data at the time of transplant were extracted from Chartview, EPIC and OTTR then entered into Microsoft Excel for Windows (Microsoft Corporation, Redmond, WA) and statistical analysis conducted using SAS 9.3 (Cary, NC: *SAS* Institute. Inc) and R (R Development Core Team (2010). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org).

This study is invaluable for the LT community, as it pertained to how outcomes could be improved in LT patients. In this study of distance from the LT center, utilized more patients at longer distances, followed for a longer period post LT, and is specific to those patients receiving LT only, since the follow-up process for other organs was different. Distance was calculated using SAS 9.3 using an algorithm based on zipcodes (`Distance_from_home =zipcitydistance(zipcode,TGHzip)`). Methods used to establish a distance cutoff in this study were data driven and therefore more accurate than the vague dichotomization of rural vs. urban used by Park et al. (2012; 2011) and Axelrod et al. (2008). In addition, I studied an adult population, which constitutes the majority of LT patients. Pediatric patients were not transplanted at TGH at the time of this study. Furthermore, the distance portion was unique in that patients relocating temporarily to be near the transplant center for purposes of receiving a transplant were excluded, as well as those receiving multiple organ transplants and those with Fulminant Hepatic Failure (Status 1). Therefore, this significant bias present in other studies was eliminated.

My methods included Kaplan-Meier curves, Cox PH, AFT, and restricted cubic splines. Newer methodologies implemented included the use of AIC/ML to best determine the cutoff point, Monte Carlo to determine change points associated with distance, an additional component for dichotomizing distance and their significance were also implemented.

1.5 Contributions from this Dissertation

This dissertation proposes a more general form of the approach taken previously (Liu et al., 2008) which accommodated dichotomization of a continuous variable, which in this case was distance. The general framework

$$\lambda(t) = \lambda_0(t)\exp\left[\theta' Z(t,\xi) + \eta' X(t)\right]$$

allowed for ξ to include a change point in time τ as well as a distance cutoff *c*. While extension three (described in more detail below) allowed for a distance contribution after the cutoff point *c*, extension four was more flexible allowing for an incremental distance effect after the point *c*. While maximizing the partial likelihood gave us parameter estimates, the Cox PH model could not attach any measure of statistical significance to the necessity of change points τ and *c*. I was interested in knowing if these effects were different from 0 and whether they were actually necessary and significant in the model. The purpose of the MC method was to determine the necessity and significance of τ and *c* in the model.

In this model there was a higher dimension for ξ and θ than in the models of Liu et al (2008), which complicated many aspects of the computation. As examples, I noted that $\mathcal{I}_{\theta\eta}$ was now a matrix and a two dimensional grid search was now required to establish τ and *c*. Furthermore, I no longer had the special structure in $Z_i(t,\xi)$ which allowed evaluation only at observed event times. The approach proposed here required evaluation at all event times in some chosen grid of values, since Z no longer factored out as it did before (Liu et al., 2008).

In addition, the maximum likelihood criterion (via AIC) was applied to repeated Cox PH models while adjusting for multiple covariates, to determine the best dichotomization point for distance. This method was not standard.

Also, this study followed patients up to 5 years post LT, which is 2 years longer than what is used for accreditation of centers and longer than what was previously studied in the literature. I also used a significantly larger population than previous studies.

Change point hazard functions may have implications in health care policy decisions in the future. With an enhanced understanding of the changes in LT recipient population mortality rates one can identify gaps, seek solutions, improve performance, and ultimately, better the public's health.

Table 2: Contributions from this Dissertation

	Existing Approaches	New in my Dissertation
Methods / technical	Use the change point equation proposed by Liu et al (Liu et al., 2008)	$\lambda(t) = \lambda_0(t)\exp\left[\theta' Z(t,\xi) + \eta' X(t)\right]$
		Higher dimension for ξ and θ. $\mathcal{I}_{\theta\eta}$ is now a matrix and a two dimensional grid search is now required.
		Drop β, if distance Z does not have a "throughout" effect, but only an effect after a certain time τ.
		Change $I(t \leq \tau)$ to $I(t \geq \tau)$ which is a statistically equivalent model (if β is retained) but has a more natural interpretation in this database. If β is not retained in the model, then this change means there is only a distance effect for times greater than τ.
		Dichotomize the distance Z, replacing it by *I(Z>c)*, and then extending the approach in Liu *et al* (2008) to estimate the change point c in addition to τ.
		Replace Z by *max(Z-c,0)*, the positive part of Z-c. This gives another way of saying that there is no distance effect for distances less than c, but allows the effect to increase for larger distances.
	Dichotomization criterion	Use ML criterion (via AIC) while adjusting for multiple covariates to determine the optimum cutoff point for distance
Application	Use KM and Cox PH to analyze survival data	Use AFT in addition to the standard methods
	Studied a maximum of 3 years post-transplant	Study 5 years post-transplant, 2 years longer than what is used for accreditation of centers. Use a larger patient population than previous studies.

Chapter Two

Methods

2.1 Data Source

This study was approved by the University of South Florida Institutional Review Board. It involved a retrospective review of 821 adult patients who underwent orthotropic liver transplantation at TGH between January 1st 1996 and September 25th 2012. All transplants were from deceased donor sources. Patients were identified consecutively since the program's inception, and data were complete on all patients. Data were obtained from Electronic Privacy Information Center (EPIC), Organ Transplant Care Platform (OTTR), Chartview and transplant flowcharts. In addition, United Network for Organ Sharing (UNOS) database containing TGH's transplant data was accessed as needed to ensure complete information. Patients were excluded if they underwent multiple organ transplantation, had fulminant hepatic failure (Status 1), died the day of transplant (n=5), or relocated temporarily to be closer to the transplant center (n=42).

This study considered random censoring from only administrative censoring 5 years post LT. Survival time was independent of censoring, satisfying this assumption necessary for the model.

2.2 Etiology Grouping

Etiology was condensed into the following three categories: 1) Nonalcoholic steatohepatitis (NASH) and Alcohol, 2) HCV/HCC, and 3) HBV/ primary biliary cirrhosis (PBC) / primary sclerosing cholangitis (PSC)/ Other. In preliminary analyses, there was a statistically significant difference in survival across these three groups. Log-rank, Wilcoxon, and the likelihood ratio test yielded significant p-values (less than 0.0001). Patients with NASH and alcohol related liver

disease had the best survival, while those with HCV and HCC had the worst survival due to reinfection of the graft by HCV and the recurrence of HCC in the cancer patients.

PBC is a disease of the liver that affects the bile ducts within the liver. Inflammation destroys the bile duct which causes bile to remain in the liver, resulting in injury and damage to the liver cells, causing cirrhosis or scarring of the liver. Cirrhosis leads to scar tissue in the liver so the liver loses its ability to function. Cirrhosis also prevents blood from the intestines from returning to the heart. PSC also affects the bile ducts causing inflammation and subsequent obstruction of bile ducts both at the intrahepatic and extra hepatic level. This inflammation hinders the flow of bile to the gut, causing cirrhosis of the liver, liver failure and liver cancer (Maggs & Chapman, 2008). Since these two diseases were related they were grouped together. Hepatitis B was very rare among this group of patients (n=3) so it was grouped with PSC and PBC along with other rare diseases.

HCV infection is a major risk factor for HCC, and 73% of HCC patients in my dataset also had HCV. There was an incubation period of 20-30 years in most HCV related HCC cases and HCV infection usually resulted in HCC via cirrhosis, although the possibility of direct carcinogenic effects of HCV have also been suggested (Di Bisceglie (1997).

Nonalcoholic steatohepatitis (NASH) is a rare complication of obesity with laboratory and histological features indistinguishable from alcoholic hepatitis (Eriksson, Eriksson, & Bondesson, 1986). Both of these involve a fatty liver. Therefore these 2 diseases were grouped together.

2.3 General Modeling and Methodology

As primary variables of interest, distance from the transplant center adjusting for the presence of HCV and HCC was studied. Distance was calculated from the patient's original home at the time of referral to the transplant center in Tampa, FL using SAS 9.3 using an algorithm based on zip codes (`Distance_from_home =zipcitydistance(zipcode,TGHzip)`).

Survival analysis was performed using Kaplan Meier (KM), Cox Proportional Hazards (Cox PH), and Accelerated Failure Time (AFT) methods. Logistic regression (LR) analysis was used to predict outcome (failed vs. not) 1 year post-transplant. Patients transplanted after September 25, 2011 were excluded from the LR analysis since they did not have a yearlong follow-up before the end of the study on September 25, 2012 (n=739). Continuous descriptive data were presented as mean ± standard deviation, while categorical data were summarized as frequencies.

The Cox PH model was chosen because it is semi parametric with minimal assumptions. Also, with the Cox PH the effect of distance and HCV/HCC could be reported as hazard ratios that are easy to interpret. Patients were censored five years post-transplant. The five year period was more than enough to measure the utility of LT. Long term survival measured beyond five years could be affected by patients' co-morbidities, de novo post LT malignancies, non-adherence, loss to follow-up, suicides, accidents, cardiovascular disease, transfer of patients to other centers whom assume their long term care (loss to follow-up), and other causes. Neither UNOS, nor the Scientific Registry of Transplant Recipients (SRTR), nor the Center for Medicare Services (CMS), holds transplant centers accountable for survival beyond three years. This is the yardstick that was used for maintaining accreditation and comparing centers.

Logistic regression (LR) analysis was the weakest method used, since it disregarded the length of the patient's survival post LT and carried the disadvantage that each time point of interest required a separate model. Also, LR was run on a reduced dataset of 739 patients for one year post LT analysis. Distance models were evaluated with a Cox PH multivariate regression model to adjust survival for HCV and HCC, which were associated with patient mortality. The same conditions applied for AFT. The standard for statistical significance was a p-value less than or equal to 0.05. Analysis was conducted including and excluding patients who died within the first 30 days post LT (n=22) to determine if there were any effects on the model by excluding these patients.

2.3.1 AIC/Maximum Likelihood Approach

In order to choose the best threshold value for dichotomizing distance from transplant center for predicting outcome (failed within 5 years from transplant), the following approach was taken. Akaike's Information Criterion (AIC) is a measure of the goodness of fit of an estimated statistical model (Bozdogan, 1987) and it is an operational way of trading off the complexity of an estimated model against how well the model fit the data. Plots of the AIC from Cox PH models vs. all possible cutoff values c, were computed and plotted on a graph.

$$AIC = 2\mathcal{k} - 2\ln(L)$$

$\mathcal{k}$ = number of parameters in model

L = maximized log-likelihood

Smaller AIC values indicate a better fit. If the number of parameters was constant in this approach, AIC would be equivalent to twice the negative log-likelihood, apart from a constant. I adjusted for multiple covariates using the AIC criterion to detect a change point attributed to distance and ensured that this cutoff remained consistent regardless of the number of covariates included in the model. Since I did adjust for multiple covariates, this approach was different from using the likelihood ratio criterion.

2.3.2 Restricted Cubic Regression Splines

Restricted cubic regression splines are a useful tool in exploratory data analysis since they allow the detection of unknown functional relationships between continuous covariates such as distance and the response variable, survival time in the Cox PH model (Durrleman & Simon, 1989; Heinzl, Kaider, & Zlabinger, 1996). They were employed here to explore the potential non-linear relationship between distance from LT center and the length of survival post LT. The RCS macro (H. Heinzl & Kaider, 2007) was used and modified for use with this data set so the splines would reveal information about the relationship between distance as a continuous variable and survival.

The expression for a restricted (or natural) cubic spline function with k knots, $t_1 < \cdots < t_k$ is given by

$C\ (u) = \ \beta_0 + \ \beta_1 u + \ \sum_{j=1}^{k-2} \theta_j C_j(u)$ where $C_1(u), \ldots C_{k-2}(u)$ are cubic terms,

$$C_j(u) = (u - t_j)_+^3 \ - \ \frac{(u - t_{k-1})^3 + [t_k - \ t_j]}{[t_k - \ t_{k-1}]} + \ \frac{(u - t_k)^3 + [t_{k-1} - \ t_j]}{[t_k - \ t_{k-1}]}, j = 1 \ldots . k - 2.$$

C(u) has continuous first and second derivatives, is linear in u for $u<t_1$ and $u>t_k$ (linear in the tails), and is a linear function with regard to the k parameters $\beta_0, \beta_1, \theta_1, \ldots . \theta_{k-2}$. (Heinzl & Kaider, 1997) Confidence bands *(1-α)* are often helpful in interpretation and those are given by

$$\left[\hat{C}_{low}(u_0), \hat{C}_{upp}(u_0)\right] = \hat{\beta}' U_0 \pm (\eta U_0' V U_0)^{1/2}$$

where $\hat{\beta} = (\hat{\beta}_0, \ \hat{\beta}_1, \hat{\theta}_1, \ldots \hat{\theta}_{k-2})^t$, $U_0 = (1, u_0, C_1(u_0), \ldots C_{k-2}(u_0))^t$, V is the large sample covariance matrix for $\hat{\beta}$, and $\gamma = \ \mathcal{X}^2_{p,1-\alpha}$is the *(1-α)* quantile of the $\mathcal{X}^2$ distribution on *p* degrees of freedom. When *p* is set to the number of covariates this method yields Scheffe-type simultaneous confidence bands, and when *p* is set to 1 it yields standard point wise confidence bands (Hess, 1994). The number of knots was pre-specified. In the literature it was suggested that three to five knots would usually suffice (Durrleman & Simon, 1989; Heinzl et al.,1996; Hess, 1994) although I experimented with a larger numbers of knots too. These knots were placed at quantiles of the observed distribution of *u*, near but not at the extremes and roughly uniform over the quantiles. The result of fitting a cubic spline to the data set was usually insensitive to the location of knots unless they occurred in an extremely non-uniform way over the covariate space (Durrleman & Simon, 1989; Heinzl et al.,1996; Hess, 1994).

Let n=821, then the data could be denoted by (y_i, s_i, z_i) for i=1, …n where *y* was the survival time, *s* was the censoring variable where 1 indicated a death and 0 meant the patient was censored and still alive, and *z* denoted distance from the transplant center. The hazard using Cox PH was $\lambda(t;\ z_i) = \ \lambda_0(t) \exp(\beta z_i)$ and *β* was estimated using partial likelihood requiring no further assumptions about the unknown baseline hazard function $\lambda_0(t)$. The log hazard ratio (LHR) function with respect to Z was

$$\mathrm{LHR(Z)} = \log \frac{\lambda(t;Z)}{\lambda_0(t)} = \ \beta Z.$$

It was assumed that a unit change in distance Z had the same effect on the patients' log hazard ratio across the entire range of Z but to explore the nature of LHR(Z) more flexibility was needed. Hence $\lambda(t;Z) = \lambda_0(t)\exp(f(Z))$ which yielded $LHR(Z) = \log\frac{\lambda(t;Z)}{\lambda_0(t)} = f(Z)$. Since $\lambda_0(t) = \lambda\,(t;Z = 0)$, *f(Z)=0* for *Z=0*. Then $LHR(Z) = \log\frac{\lambda(t;Z)}{\lambda_0(t)} \approx C(Z) - C(0) = \beta_1 Z + \sum_{j=1}^{k-2}\theta_j(C_j(Z) - C_j(0))$.

Since in this scenario it was not meaningful to report hazard ratios relative to a distance of 0 but it was of great interest to report them in relation to a specific mileage point of 180 miles, the LHR function for a given reference value *m* was

$$LHR_m(Z) = \log\frac{h\,(t;Z)}{h(t;m)} = f(Z) - f(m) \approx C(Z) - C(m) = \beta_1(Z-m) + \sum_{j=1}^{k-2}\theta_j(C_j(Z) - C_j(m))$$

Once β_1, $\theta_1 \ldots \theta_{k-2}$ were fitted the reference value *m* could be changed and that only involved a shift in the axis. Confidence intervals were also computed and displayed on the graph (Heinzl & Kaider, 2007).

2.3.3 Estimating Change Points Using Monte Carlo

Given that Firozvi et al. (Firozvi et al., 2008) found no significant distance effect one year post LT, one of the goals was to search for change points in the hazard function that could be attributed to distance in a five year period. I was also interested in the time change point over the five year period. It was possible that the distance effect was not present initially but became evident later on in the patient follow-up (see dichotomizing at c in extension three below). It was also possible that beyond that point, the distance effect was incremental (increasing distance effect in extension four below). Identifying change points in a hazard function is important in survival analysis and I was particularly interested in the Monte Carlo approach used by Liu et al (2008). This method gave a legitimate p-value for the test for the existence of a change point at an unspecified location (known only to belong in a certain set).

The hazard function proposed by Liu et al. (2008) was specified with one change point in the hazard function as

$$\lambda(t; Z, X) = \lambda_0(t) \exp[\{\beta + \theta I(t \leq \tau)\}Z + \gamma' X(t)]$$

where $\lambda_0(t)$ was an unspecified baseline hazard function, $\tau > 0$ was the change-point parameter, θ was the magnitude of change after the change point, β was the throughout distance effect on the hazard function, Z was a covariate (distance from transplant center) which potentially caused the non-constant hazard, η were the coefficients and X(t) were other risk factors that I adjusted for in the model (HCV and HCC). Using this model I determined the presence of change points in the hazard τ, as well as the change point in the distance effect, *c*.

Liu et al. (2008) developed this method to give a legitimate p-value for the test of the existence of a change point at an unspecified location. I was using this method in a very unique way to first dichotomize distance at some value *c*, reducing it to a binary variable. Next I chose a value of time τ such that the distance effect was only observed for times greater than τ. The values of *c* and τ could both be considered change points and the approach in Liu et al. (2008) could be used to simultaneously choose both of them after their methods were suitably generalized.

In the simulation performed by Liu et al. (2008) the authors performed 1000 runs and in each run the thresholds were calculated using 10,000 Monte Carlo resampling samples. Sample size was set to 200-300. It was established that the Monte Carlo resampling-based methods approximated the distribution of the maximal score tests very well. Accuracy of the approximation and power increased with sample size, and instability of the variance was noted when there were few failure events. These authors also established that the maximum normalized score statistic was less sensitive in the detection of a change point than that based on the non-normalized score, under all alternative hypotheses that they considered. The maximum normalized score statistic was also more conservative under the null hypothesis.

2.4 Methodological Development Proposed for this Dissertation

The extension to this methodology was in the form of a more complicated null hypothesis than the one in Liu et al. (2008). As stated above, distance was dichotomized at some value *c*, reducing it

to a binary quantity rather like the treatment assignment in Liu et al. (2008). In addition, a value of time τ was chosen such that the distance effect was only observed for times greater than τ. The values of *c* and τ were both considered change points, and the Monte Carlo approach was used to simultaneously choose both of them.

The general form of the extended model was

$$\lambda(t) = \lambda_0(t)\exp\left[\theta' Z(t,\xi) + \eta' X(t)\right]$$

for parameters θ, η, ξ. This model was linear in θ and η for fixed ξ. The parameter ξ contained the nonlinear parameters τ which represent change points in time and *c* which represented the dichotomization point for distance which was the mileage point beyond which distance had an effect on survival. $Z(t,\xi)$ and *X(t)* were functions of covariates, which could be time-varying covariates. In this particular dataset *X(t)* was not time-varying but the model could be applied in the scenario where *X(t)* included time-varying covariates. Distance in my model was part of $Z\ (t,\xi)$ which made the model in Liu et al. a special case of this extended model in which $\xi = \tau$ and $Z\ (t,\xi) = ZI(t \le \tau)$.

Let $N_i(t) = I(T_i \le t, \delta_i = 1)$ be the count of failures on the ith subject at time t, $Y_i(t) = I(T_i \ge t)$ be the at-risk indicator, and $\kappa = \inf\{t; pr(T > t) = 0\}$. For a trial with *n* subjects, the partial likelihood for the unknown parameters (θ, η, ξ) based on the observed data would be

$$L\ (\theta,\eta,\xi) = \prod_{i=1}^{n} \int_0^{\kappa} \frac{\exp[\theta' Z_i(t,\xi) + \eta' X_i(t)]}{\sum_{j=1}^{n} Y_j(t)\exp[\theta' Z_j(t,\xi) + \eta' X_j(t)]} dN_i(t)$$

The goal was to test the null hypothesis H_0: $\theta = 0$ using the score function

$$S_\theta(\theta,\eta,\xi) = \frac{\partial}{\partial\theta} l(\theta,\eta,\xi)$$

where l represented the logarithm of the partial likelihood. Let $\tilde{\eta}$ be the maximum partial likelihood estimate of η under the null hypothesis. The variance estimator of $S_\theta(0,\tilde{\eta},\xi)$ under H_0 was denoted by $V(\tilde{\eta},\xi)$ (Fleming & Harrington, 2011); a definition is given later.

The score functions are given explicitly by

$$S_\theta(\theta,\eta,\xi) = \sum_{i=1}^{n}\int_0^k \left\{ Z_i(t,\xi) - \frac{\sum_{j=1}^n Z_j(t,\xi)Y_j(t)\exp\{\theta' Z_j(t,\xi)+\eta' X_j(t)\}}{\sum_{j=1}^n Y_j(t)\exp\{\theta' Z_j(t,\xi)+\eta' X_j(t)\}} \right\} dN_i(t)$$

$$S_\eta(\theta,\eta,\xi) = \sum_{i=1}^{n}\int_0^k \left\{ X_i(t) - \frac{\sum_{j=1}^n X_j(t)Y_j(t)\exp\{\theta' Z_j(t,\xi)+\eta' X_j(t)\}}{\sum_{j=1}^n Y_j(t)\exp\{\theta' Z_j(t,\xi)+\eta' X_j(t)\}} \right\} dN_i(t)$$

The corresponding test statistics defined by Liu et al (2008) were

$$M = \sup_{\xi\in\mathcal{A}} |S_\theta(0,\tilde{\eta},\xi)| \text{ and } M^* = \sup_{\xi\in\mathcal{A}} W_\theta(\tilde{\eta},\xi)$$

where $W_\theta(\tilde{\eta},\xi) = \frac{S_\theta^2(0,\tilde{\eta},\xi)}{V(\tilde{\eta},\xi)}$.

In the work of Liu et al (2008) θ is univariate, and consequently $V(\tilde{\eta},\xi)$ is univariate. But in much of the present work θ is bivariate, i.e., $\theta = (\theta_1,\theta_2)$, so that the score $S_\theta(\theta,\eta,\xi)$ is a 2x1 vector which can be written as $S_\theta(\theta,\eta,\xi) = (S_\theta^1(\theta,\eta,\xi), S_\theta^2(\theta,\eta,\xi))$, and the variance estimate $V(\tilde{\eta},\xi)$ is a 2x2 matrix. Thus I have statistics analogous to the M and M* for each of the components of θ, which I define by

$$M_j = \sup_{\xi\in\mathcal{A}} |S_\theta^j(0,\tilde{\eta},\xi)| \text{ and } M_j^* = \sup_{\xi\in\mathcal{A}} W_\theta^j(\tilde{\eta},\xi) \text{ for j=1,2,}$$

where $W_\theta^j(\tilde{\eta},\xi) = \frac{(S_\theta^j(0,\tilde{\eta},\xi))^2}{V_{jj}(\tilde{\eta},\xi)}$. Motivated by standard results on quadratic forms, I also defined a test statistic M_{2p} which combines both components of the score vector as follows

$$M_{2p} = \sup_{\xi\in\mathcal{A}} S_\theta'(0,\tilde{\eta},\xi)[V(\tilde{\eta},\xi)]^{-1} S_\theta(0,\tilde{\eta},\xi).$$

The statistic M_{2p}, which is new to this work, is essentially a 'two parameter' analog of the statistic M*, and in my work this statistic is usually referred to as the 'two parameter score' or identified by the acronym M2P, where '2P' stands for 'two parameter'. In Liu et al (2008) the set $\mathcal{A}$ is one-dimensional so that the Supremum in their definition of M and M* is a one-dimensional maximization. However, in much of the present work $\xi = (\tau, c)$ and the set $\mathcal{A}$ is two-dimensional, which is taken to be a finite grid of time and distance values (τ, c). In computing $M_j, M_j^*, \text{and } M_{2p}$, we maximize over this two-dimensional grid.

$$\hat{S}_{\theta,i}(\xi) = \int_0^k \left[\{Z_i(t,\xi) - \bar{Z}_0(t,\xi,\tilde{\eta})\} - \mathcal{I}_{\theta\eta}(0,\tilde{\eta},\xi)\mathcal{I}_{\eta\eta}^{-1}(0,\tilde{\eta},\xi)\{X_i(t) - \bar{X}_0(t,\tilde{\eta})\}\right] d\hat{M}_i(t)$$

and $d\widehat{M}_i(t) = dN_i(t) - Y_i(t)\exp\{\tilde{\eta}'X_i(t)\}d\widetilde{\Lambda}(t)$ with $\widetilde{\Lambda} = \sum_{i=1}^{n}\int_0^t \frac{dN_i(s)}{\sum_{j=1}^{n} Y_j(s)\exp\{\tilde{\eta}'X_j(s)\}}$.

$\bar{Z}_0(t,\xi,\tilde{\eta})$ and $\bar{X}_0(t,\tilde{\eta})$ could be found from $\bar{Z}_0(t,\xi,\tilde{\eta}) = \frac{S_Z^{(1)}(\tilde{\eta},\xi)}{S_Z^{(0)}(\tilde{\eta})}$ and $\bar{X}_0(t,\tilde{\eta}) = \frac{S_X^{(1)}(\tilde{\eta})}{S_X^{(0)}(\tilde{\eta})}$

where $S_Z^{(0)}(\tilde{\eta}) = S_X^{(0)}(\tilde{\eta}) = \ n^{-1}\sum_{i=1}^{n} Y_i(t)\exp\{\tilde{\eta}'X_i(t)\}$ and

$S_Z^{(1)}(\tilde{\eta},\xi) = \ n^{-1}\sum_{i=1}^{n} Z_i(t,\xi)Y_i(t)\exp\{\tilde{\eta}'X_i(t)\}$ and similarly for $S_X^{(1)}(\tilde{\eta})$. The information matrix $\mathcal{I}$ was partitioned as $\begin{pmatrix}\mathcal{I}_{\theta\theta} & \mathcal{I}_{\theta\eta}\\ \mathcal{I}_{\eta\theta} & \mathcal{I}_{\eta\eta}\end{pmatrix}$.

At each fixed point ξ, let $V(\tilde{\eta},\xi) = \sum_{i=1}^{n}\hat{S}_{\theta i}(\xi)\hat{S}'_{\theta i}(\xi)$ which reduces to $V(\tilde{\eta},\xi) = \sum_{i=1}^{n}\hat{S}^2_{\theta i}(\xi)$ in the univariate case considered by Liu et al (2008). Then $n^{-1}V(\tilde{\eta},\xi)$ is a consistent estimate for the asymptotic variance of $n^{-\frac{1}{2}}S_\theta(0,\tilde{\eta},\xi)$ under the null hypothesis. The Monte Carlo null distribution of $S_\theta(0,\tilde{\eta},\xi)$ was found by simulating $\hat{S}_\theta(0,\tilde{\eta},\xi) = \sum_{i=1}^{n}\hat{S}_{\theta,i}(\xi)G_i$ where $\{G_i, i = 1,\dots,n\}$ were independent standard normal random variables serving as perturbation factors in the sum.

The values of $\hat{S}_\theta(0,\tilde{\eta},\xi)$ may be used to obtain Monte Carlo approximations to the null distributions of the statistics $M_j, M_j^*, \text{and } M_{2p}$ by using these values in place of $S_\theta(0,\tilde{\eta},\xi)$ in the computation of these statistics. That is I computed Monte Carlo replicates of

$\widehat{M}_j = \ \sup_{\xi\in\mathcal{A}}|\hat{S}_\theta^j(0,\tilde{\eta},\xi)|,$

$\widehat{M}_j^* = \ \sup_{\xi\in\mathcal{A}}\frac{(\hat{S}_\theta^j(0,\tilde{\eta},\xi))^2}{V_{jj}(\tilde{\eta},\xi)}$, and $\widehat{M}_{2p} = \ \sup_{\xi\in\mathcal{A}}\hat{S}'_\theta(0,\tilde{\eta},\xi)[V(\tilde{\eta},\xi)]^{-1}\hat{S}_\theta(0,\tilde{\eta},\xi).$

The empirical 100(1-α)th quartile of the randomly perturbed maximal score statistics provided the critical values of nominal level α for the observed maximal score statistics.

The specific extensions I considered were:

1. Dropped β, to indicate that distance Z (dichotomized at 180 miles) did not have a throughout effect, but only an effect after a certain time τ.

$$\lambda(t;Z,X) = \lambda_0(t)\exp[\{\theta I(t \le \tau)\}Z + \gamma' X(t)]$$

2. Changed $I(t \le \tau)$ to $I(t \ge \tau)$ which was a statistically equivalent model (if β is retained) but had a more natural interpretation in this situation. If β was not retained in the model, then this change meant there was only a distance effect for times greater than τ. In this model Z was dichotomized at 180 miles.

$$\lambda(t;Z,X) = \lambda_0(t)\exp[\{\beta + \theta I(t \ge \tau)\}Z + \gamma' X(t)]$$

3. Dichotomized the distance Z, replacing it by *I(Z>c)*, and then extending the approach in Liu et al. (2008) to estimate the change point c in addition to τ.

$$\lambda(t;Z,X) = \lambda_0(t)\exp[\{\beta + \theta I(t \ge \tau)\}I(Z > c) + \gamma' X(t)]$$

This fit into the general framework

$$\lambda(t) = \lambda_0(t)\exp[\boldsymbol{\theta}' Z(t,\xi) + \eta' X(t)]$$

by taking $\boldsymbol{\theta} = \binom{\beta}{\theta}, \eta = \gamma, \xi = \binom{\tau}{c}, Z(t,\xi) = \begin{pmatrix} I(Z > c) \\ I(t \ge \tau)I(Z > c) \end{pmatrix}, X(t) = X(t).$

4. Alternatively, I replaced Z by *max(Z-c,0)*, the positive part of Z-c. This gave another way of saying that there was no distance effect for distances less than c, but allowed the effect to increase for larger distances.

$$\lambda(t;Z,X) = \lambda_0(t)\exp[\{\beta + \theta I(t \ge \tau)\}(Z - c)_+ + \gamma' X(t)]$$

This also fit into the general framework

$$\lambda(t) = \lambda_0(t)\exp[\boldsymbol{\theta}' Z(t,\xi) + \eta' X(t)]$$

by taking $\boldsymbol{\theta} = \begin{pmatrix}\beta\\ \theta\end{pmatrix}, \eta = \gamma, \xi = \begin{pmatrix}\tau\\ c\end{pmatrix}, Z(t,\xi) = \begin{pmatrix}(Z-c)_+\\ I(t \geq \tau)(Z-c)_+\end{pmatrix}, X(t) = X(t)$. Here $(Z-c)_+ = \max(Z-c, 0)$.

The additional challenge for extensions three and four is that β, θ, τ, c and γ must be estimated simultaneously. There are 2 possible approaches to this and we tried both:

The first approach involved the estimation of β, θ, and γ for fixed values of c and τ. This implied fitting a standard Cox model for each of c and τ, which was efficient, particularly since reasonable initial estimates were used.

Maximizing the likelihoods could be done by standard algorithms, such as simulated annealing, variants of gradient ascent or algorithms like Nelder-Mead which required only repeated function evaluations. The log-likelihood function was optimized using the Nelder-Mead algorithm, a variable metric algorithm (Broyden, Fletcher, Goldfarb and Shanno, 1970), and a simulated annealing algorithm to estimate $\tau, c, \beta, \theta, \gamma_1, \text{and } \gamma_2$.

First I used the simplex search algorithm proposed by Nelder and Mead (1965). This method is very effective for multidimensional unconstrained optimization without derivatives (Price & Coope, 2003). Only function values are used and it is robust but relatively slow. This method performs a sequence of transformations on the working simplex, defined by *n+1* points $x_0, \ldots x_n \in \mathbb{R}^n$ that are considered as the vertices of a working simplex *S*. The goal is to decrease the function values at the vertices. In each iteration of the Nelder–Mead algorithm, the vertex with the worst function value is removed and replaced with another point which has a superior value. The new point is obtained by reflecting, expanding, shrinking or contracting the simplex along the line joining the worst vertex with the centroid of the remaining vertices. If one cannot find an improved point, only the vertex with the best function value is retained, and the simplex is shrunk by moving all other vertices toward this value. This process is terminated when the working simplex *S* becomes

sufficiently small, or when the function values are close enough. It has been proven to work reasonably well for non-differentiable functions (Kolda, Lewis, & Torczon, 2003; Nocedal & Wright, 1999).

A quasi-Newton method which is also known as a variable metric algorithm (Broyden, 1970; Powell, 1976), that uses function values and gradients to build up a picture of the surface to be optimized was also used. It was published simultaneously in 1970 by Broyden, Fletcher, Goldfarb and Shanno (BFGS). Often BFGS may not move away from the initial values of τ and c because the numerically estimated gradient is exactly zero. The log-likelihood is flat as a function of τ and c in the neighborhood of $\tau = 3.66$ and $c = 180$ (see Chapter 3). I anticipated this method to behave poorly because it was designed for differentiable functions and the log-likelihood in this case is discontinuous with respect to τ and c but the estimates it produced were not different from the other 2 methods.

Next a variant of simulated annealing (SANN) was used (Bélisle, 1992). Simulated annealing is a Monte Carlo technique for solving optimization problems. It is a stochastic global optimization method and since it is random, it may often not find anything better than the initial parameters. Only the function value is used, it is relatively slow, and was chosen because it works for non-differentiable functions. This implementation uses the Metropolis function *p=p(x,y,t)* for the acceptance probability of the next candidate point *y*, given the current state *x* and the temperature *t*. By default the next candidate point is generated from a Gaussian Markov kernel with scale proportional to the actual temperature *t*. Temperature is decreased according to a logarithmic cooling schedule (Bélisle 1992, p. 890). A disadvantage is that it depends critically on the settings of the control parameters and it is not a general-purpose method but can be very useful in getting to a good value on a very rough surface.

In the frequentist MC approach, the parameter estimates for the covariate effects not associated with the change point were estimated under the null hypothesis that there was no change point.

So after the MC approach had demonstrated the necessity for a change point, these parameters were re-estimated. One potential approach involved simultaneous estimation of all the parameters, including the change point, by maximizing the partial likelihood and incorporating a grid search for maximizing over the change point. Another choice was setting the change point equal to the value which maximized the score statistic and then re-estimating the other parameters by fitting a standard Cox model as in the example in Section 3.4 of Liu et al. (2008). The models included both a change point (in time) and a dichotomization point (for the distance).

2.5 Simulation

I conducted a simulation study using the hazard functions in extensions one to four. There were two types of simulations:

1. Using data that was generated from standard distributions
2. A situation which resembled the actual LT data set.

The purpose of the first simulation was to demonstrate that the methodology worked and that of the second simulation to determine how the methodology worked with the actual LT data structure.

2.5.1 Simulation 1

Distance Z was generated from a random uniform distribution on the interval (0, 1) for extensions three and four and a Bernoulli distribution with a success probability of 0.5 in extensions one and two, to resemble a dichotomous variable. Two additional risk factors were generated using a uniform distribution on (0, 1) and an exponential distribution with mean 1. The change point parameter was set to $\tau = 0.25$ and the distance cutoff to c=50.

For 1000 simulated data sets, critical thresholds based on 1,000 Monte Carlo resampling samples were calculated for each of the four extensions above. The sample size was set to 200 and 300 for approaches one and two and 600 for approaches three and four. A grid was defined over the suspected region where the change point τ and the cutoff c were suspected to lie. The

proposed test procedure was implemented searching for a change point over the grid of a time interval and the distance cutoff range in miles. The simulation results are summarized in tables 31 through 36 where the partial likelihood was maximized over the above stated grid.

The empirical threshold corresponding to a nominal level α was determined by the $100(1-\alpha)^{th}$ quantile of the sample test statistics from 1000 runs of simulations under the null hypothesis. The average Monte Carlo resampling-based threshold was defined as the mean of 1000 thresholds found by the Monte Carlo resampling approach.

The results were tabulated in section 3.9.1 to compare the empirical quantiles of the test statistic based on 1000 simulated data sets with the Monte Carlo resampling thresholds.

2.5.2 Simulation 2

The second simulation effort aimed at resembling the actual situation in the LT data set, with the goal of determining the level of performance that might be expected of my methods in this situation.

HCV, HCC and distance from the original dataset were used. Survival times were generated from the Cox PH model to fit to this data set for each of the hazard functions in extensions three and four (i.e., using the estimated parameter values and a baseline hazard obtained by smoothing Breslow's estimator) with censoring. A piecewise exponential was used to generate the censoring time and the censoring variable was created by determining which of the two was smaller. For simulations under the alternative hypothesis, this testing procedure was repeated for a range of values for β and θ, and the accuracy of estimation for τ and c was recorded.

Since this study involved the estimation of both a change point in time τ and a distance cutoff *c*, the computations were significantly more complex than those done previously (Liu et al., 2008) with an appreciable programming effort required. I evaluated this model over a grid of time and

distance values with the all times considered, not just the event times since distance Z no longer factors out as it did in the simpler case of Liu et al (2008).

Chapter Three

Results and Applications

Four extensions to the MC approach were developed as described in the previous chapter. In this chapter, these models were applied to a specific dataset of liver transplant recipients. In this retrospective single center study of 821 liver transplant recipients at TGH; Hepatitis C (HCV), Hepatocellular Carcinoma (HCC), and patient survival were collected in addition to distance from the transplant center. This methodology allowed us to establish a change point in the hazard function τ as well as a dichotomization point for distance c.

This chapter starts with an introduction of the dataset and the Kaplan Meier curves, Cox Proportional Hazards, Accelerated Failure Time and Logistic Regression models. Data were analyzed both including and excluding early deaths (n=22).

Then I proceeded to fit the model introduced by Liu et al. (2008) followed by the four extensions that I proposed. Lastly, I performed two types of simulations described in section 2.5 above to illustrate the use and application of this methodology.

3.1 Introduction

3.1.1 General Description of the Data

Recipients included 603 males and 218 females with a mean age of 53.5 (standard deviation 9). 22 patients died within the first 30 days post LT. The analyses were conducted with and without these patients. Of the 821 patients, 178 died within five years post LT and 643 (78.32%) were still alive at the five year mark. Hepatitis C (HCV) or Hepatocellular Carcinoma (HCC) were the

most common reason for patient transplant. Table 3 below outlines the patient characteristics including and excluding the 22 early deaths.

Table 3: Liver Transplant Recipient Characteristics

	All Patients	Excluding Early Deaths
	Total (n=821)	n=799
Age (years) mean ± SD	53.48 ± 8.97	53.49 ± 9
Male (%)	603 (73.45%)	587 (73.47%)
Race (Caucasian)	667 (81.24%)	650 (82.70%)
Primary cause of ESLD (%)		
HBV/PBC/PSC/Other	130 (15.83%)	125 (15.64%)
HCV/HCC	470 (57.25%)	459 (57.45%)
NASH and Alcohol	221 (26.92%)	215 (26.91%)

Table 4 describes patient characteristics of 2 groups, those who lived within 180 miles and those who lived beyond this distance. Mean age, gender, race, and transplant etiology were not significantly different between these groups. 570 patients who were transplanted after February 2002 had MELD scores calculated prior to receiving LT, of which 553 patients were from within 180 miles, and 17 patients were from beyond 180 miles.

The mean MELD score was significantly lower for the distant group (p-value=0.008) but the prevalence of HCC (p-value=0.23) and HCV (p-value=0.68) were not. It is suspected that the more severely ill patients with higher MELD scores, who lived further away, did not actually make it to the transplant stage. Due to the fact that there were 251 patients transplanted in the pre-MELD era who did not have MELD scores on file and MELD alone was not a significant predictor of survival using KM (LR p-value=0.1370) and Cox PH (p-value=0.2050), I opted not to adjust for it in the model. However, MELD score at the time of transplant indicates that the distant patients were healthier compared to those living within the 180 mile radius (p-value=0.008).

Table 4: Clinical and Demographic Features of Patients by Distance Category

	Within 180 miles (n=802)	Beyond 180 miles (n=19)	P-value
Mean Age (SD)	53.48 (8.99)	53.6 (8.40)	0.9662
Gender			
Male	590 (73.56%)	13 (68.42%)	0.6157
Female	212 (26.43%)	6 (31.58%)	
Race/Ethnicity			0.9777
Asian	3 (0.37%)	0	
Black	49 (6.11%)	1 (5.26%)	
Caucasian	652 (81.30%)	15 (78.95%)	
Hispanic	64 (7.98%)	1 (5.26%)	
Other	22 (2.74%)	1 (5.26%)	
MELD Score at Transplant (SD)	22.09 (7.18)	17.35 (8.85)	0.008
Etiologies			
Alcohol	120 (14.96%)	5 (26.31%)	0.403
Autoimmune/PBC/PSC	88 (10.97%)	0	
Cryptogenic/NASH	94 (11.72%)	2 (10.53%)	
HBV/Others	42 (5.23%)	0	
HCC	98 (12.22%)	4 (21.05%)	
HCV	285 (35.54%)	7 (36.84%)	
HCV and Alcohol	75 (9.35%)	1 (5.26%)	

Next I looked at the effects of disease etiology, gender and insurance type separately. Disease etiology was grouped into three categories and survival differed by etiology as detected by the KM curves (p-value<0.0001).

3.1.2 Disease Etiology

Differences in survival by disease etiology were significant in the Kaplan Meier curves above according to log-rank, Wilcoxon and likelihood-ratio tests (p-values<0.0001). NASH and alcohol patients had the best survival while those with HCV and HCC fared the worst. This phenomenon was likely due to disease recurrence in these patient populations.

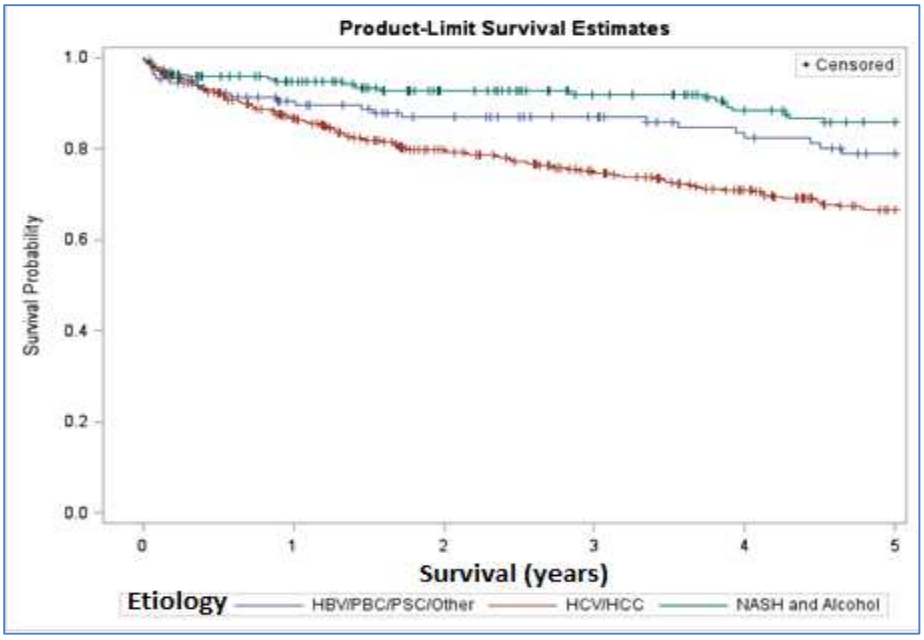


Figure 1: KM by Etiology All Patients

Table 5 below confirms that HCV/HCC were the deadliest with 131 (27.87%) deaths, followed by HBV/PBC/PSC/Other with 23 (17.69%) deaths, and NASH and Alcohol had the best survival with 24 (10.86%) deaths.

Table 5: Disease Etiology for Transplant Patients (All Patients)

Summary of the Number of Censored and Uncensored Values					
Stratum	**Etiology Condensed**	**Total**	**Failed**	**Censored**	**Percent Censored**
1	**HBV/PBC/PSC/Other**	130	23	107	82.31
2	**HCV/HCC**	470	131	339	72.13
3	**NASH and Alcohol**	221	24	197	89.14
Total		821	178	643	78.32

Next, I repeated the KM for 799 patients who survived past 30 days post LT.

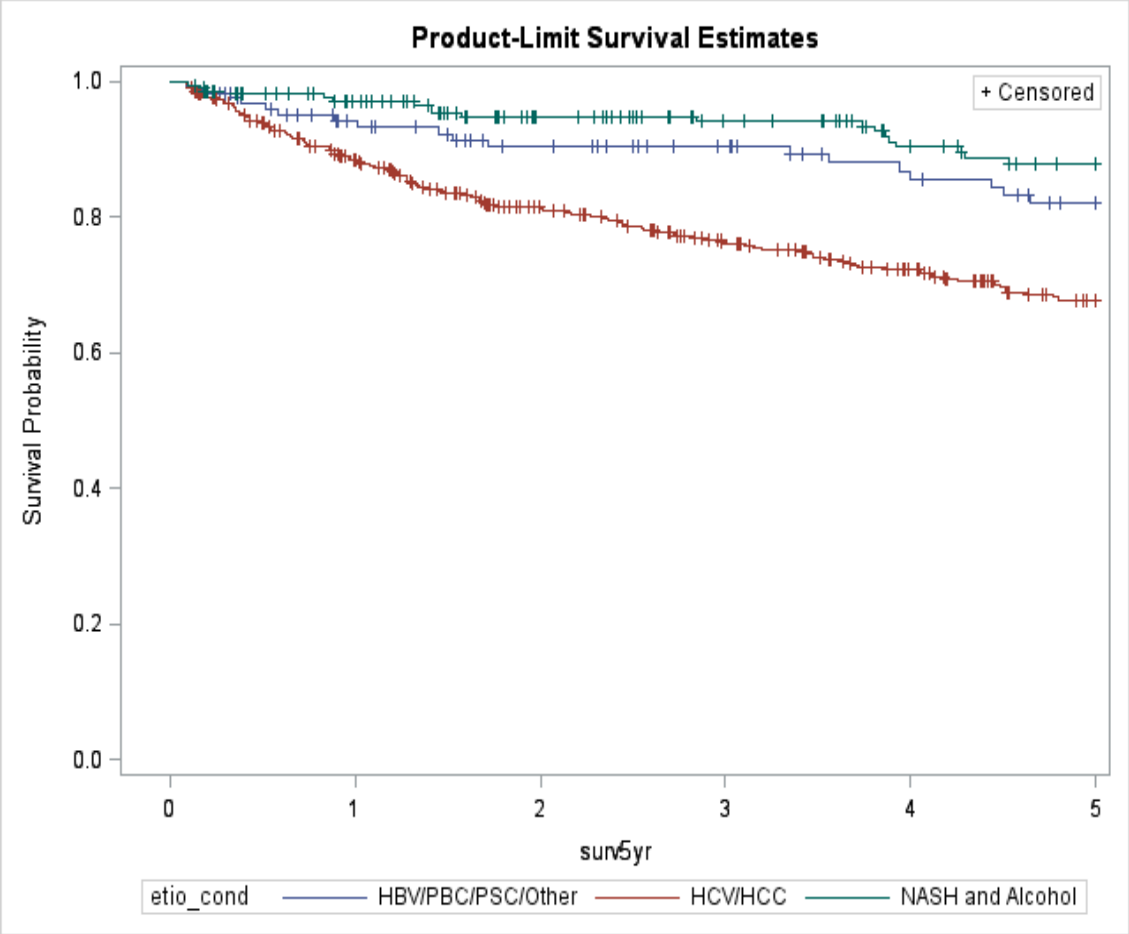


Figure 2: KM by Etiology Excluding Early Deaths

Table 6: Disease Etiology for Transplant Patients Excluding Early Deaths

Summary of the Number of Censored and Uncensored Values					
Stratum	**Etiology Condensed**	**Total**	**Failed**	**Censored**	**Percent Censored**
1	**HBV/PBC/PSC/Other**	125	18	107	85.6
2	**HCV/HCC**	459	122	337	73.42
3	**NASH and Alcohol**	215	19	196	91.16
Total		799	159	640	80.1

The same pattern of survival by disease etiology appeared when early deaths were excluded (n=799), with 122 (26.58%) deaths for patients with HCV/HCC, 18 (14.4%) for patients with HBV/PBC/PSC/Other, and 19 (8.84%) for those with NASH and Alcohol. Differences in survival by disease etiology were significant according to log-rank, Wilcoxon and likelihood-ratio tests (p-values<0.0001). In conclusion, NASH and alcohol patients had the best survival while those with HCV and HCC fared the worst.

3.1.3 Gender Distribution

There were 218 females in the study and 603 males. Although no significant difference in survival by gender was detected using KM (log-rank p-value=0.0935, Wilcoxon p-value=0.2007, likelihood ratio p-value=0.0868), I noticed a separation of the curves after the second year post LT indicating that females had a slightly better survival after the first three years as compared to males.

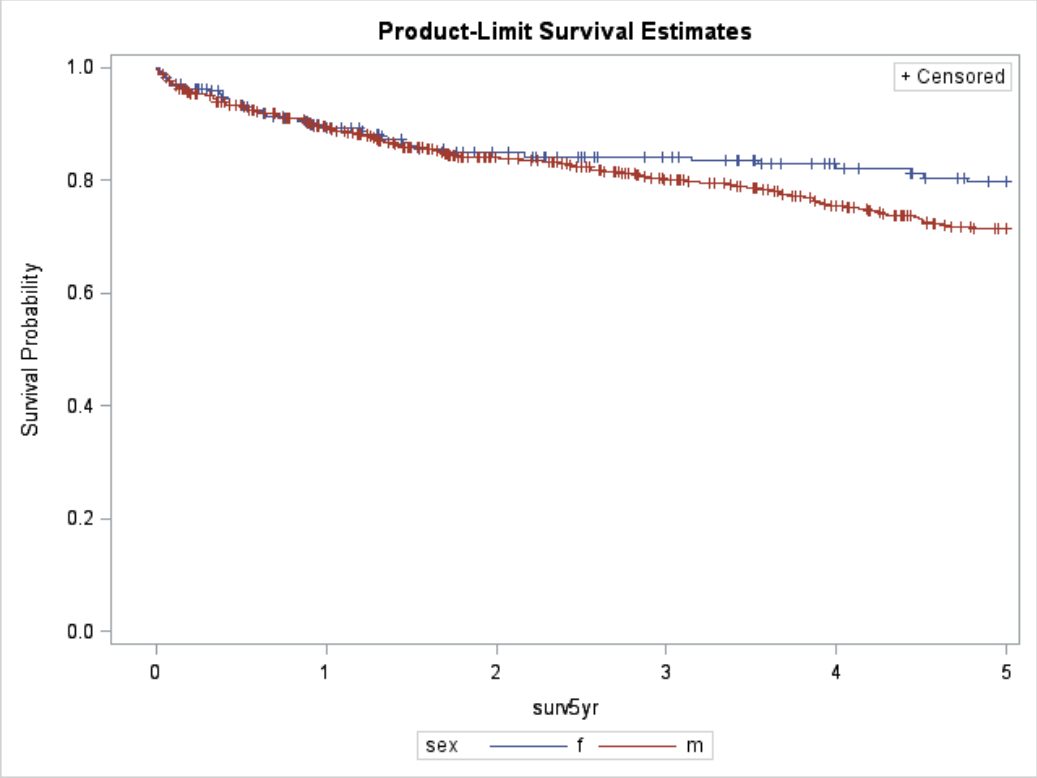


Figure 3: KM by Gender All Patients

On the reduced dataset excluding 22 patients who died in the first month, there were 212 females in the study and 587 males. Although no significant difference in survival by gender was detected using KM (log-rank p-value=0.0775, Wilcoxon p-value=0.1641, likelihood ratio p-value=0.0690), I noticed a separation of the curves after the second year post-transplant indicating that females had a slightly better survival after 3 years as compared to males. P-values were lower in the reduced dataset which excluded 22 early deaths indicating more of a gender difference in survival for that dataset.

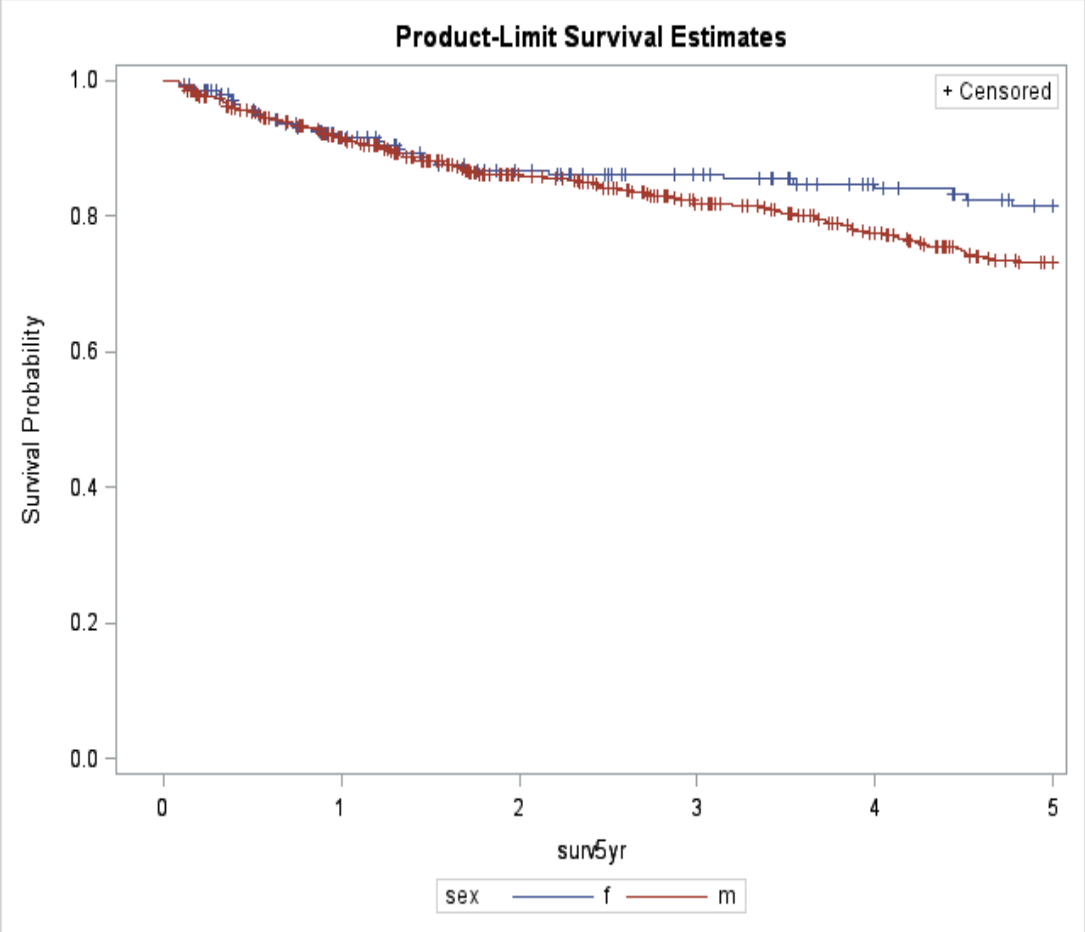


Figure 4: KM by Gender Excluding Early Deaths

3.1.4 Insurance

Insurance was grouped in 4 categories but failed to yield any significant difference in survival as seen by the log-rank (p-value=0.7596), Wilcoxon (p-value=0.6341) and likelihood-ratio (p-value=0.8584) tests. This indicated that the insurance did not impact survival post-transplant even though the literature indicated that there was an effect on access to transplant (Kemmer et al., 2011).

Table 7: Insurance for All Patients

Summary of the Number of Censored and Uncensored Values					
Stratum	**Insurance Grouped**	**Total**	**Failed**	**Censored**	**Percent Censored**
1	**Medicaid**	106	25	81	76.42
2	**Medicare**	179	29	150	83.8
3	**Other**	32	8	24	75
4	**Private**	504	116	388	76.98
Total		821	177	643	78.32

Table 6 above and Figure 5 below indicate that patients on Medicare had the best post LT survival while those with other insurance had the worst, but these differences failed to yield any statistical significance.

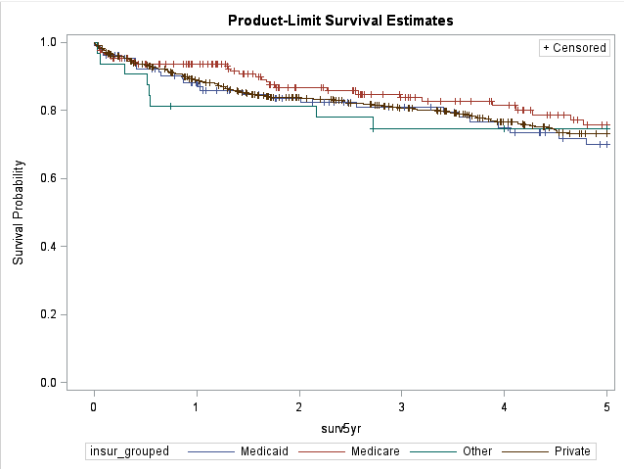


Figure 5: KM by Insurance for All Patients

Table 8: Insurance Excluding Early Deaths

Summary of the Number of Censored and Uncensored Values					
Stratum	**Insurance**	**Total**	**Failed**	**Censored**	**Percent Censored**
1	**Medicaid**	103	22	81	78.64
2	**Medicare**	173	24	149	86.13
3	**Other**	30	6	24	80
4	**Private**	493	107	386	78.3
Total		799	159	640	80.1

The same pattern with Medicare patients having the best survival persisted when patients who died in the first 30 days were excluded but this was not statistically significant. In this subset, patients with private insurance fared the worst which was rather surprising.

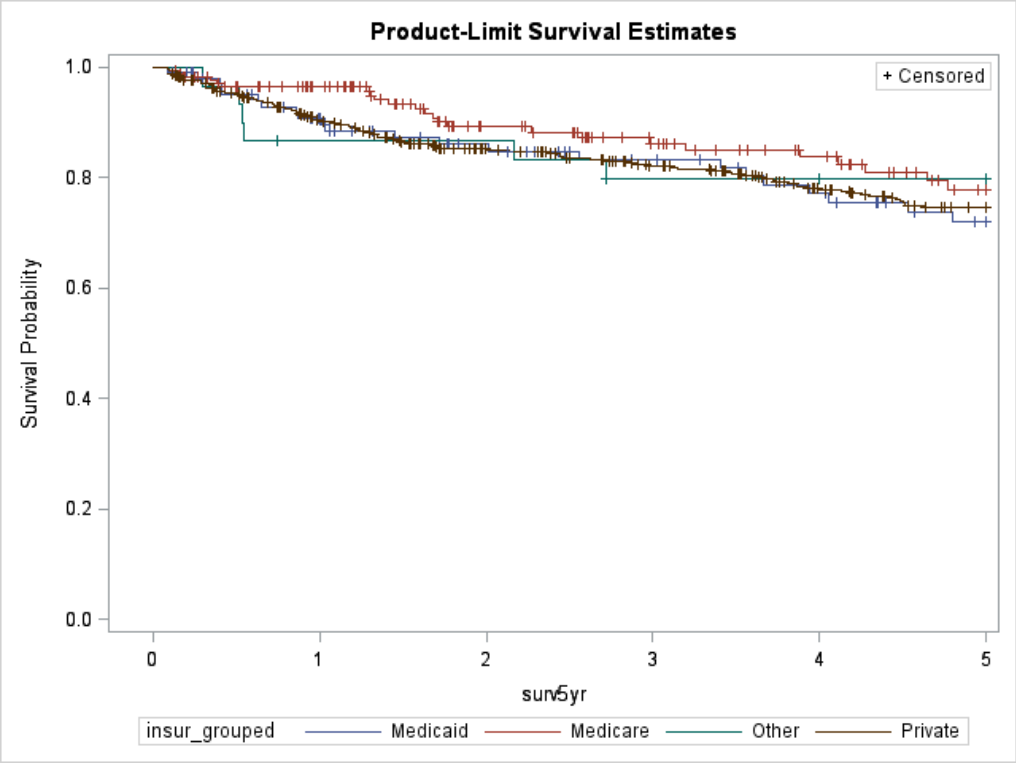


Figure 6: KM by Insurance Excluding Early Deaths

However, difference in survival by insurance failed to yield any significant difference in survival as seen by the log-rank (p-value=0.5442), Wilcoxon (p-value=0.4296) and likelihood ratio (p-value=0.6183) tests but p-values were again lower than those for the KM including early deaths. From this data it appears that there is insufficient evidence that private insurance leads to any survival benefit as one might suspect.

3.2 AIC/ML Approach to the Cox PH Model

I began with the AIC approach to the Cox PH model described in the methods in chapter 2. The eight models displayed in table 9 were fit, all of which pointed to a distance cutoff of 180 miles as the value which minimized the AIC and maximized the partial likelihood function.

The following were also considered as covariates in the model: creatinine, bilirubin, insurance, presence of HCV, HCC, etiology, age, and abuse of alcohol. The data suggested a distinct cutoff value for distance, and that the plots convincingly indicated such a case.

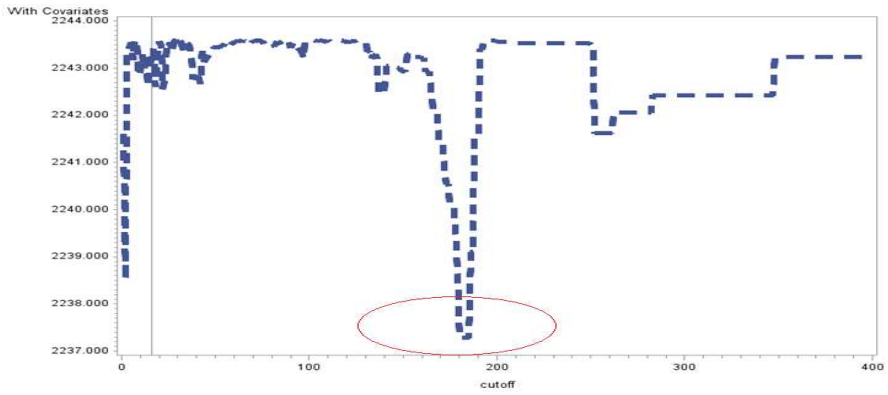
Table 9: AIC Model for the Covariates

	HCC	HCV	Etiology	Alcohol	Creatinine	Bilirubin	Age	Insurance	BMI	AIC
Model 1	√	√								2243.238
Model 2	√	√	√							2244.004
Model 3	√	√	√	√						2244.754
Model 4	√	√	√	√	√					2231.950
Model 5	√	√	√	√	√	√				2218.825
Model 6	√	√	√	√	√	√	√			2220.754
Model 7	√	√	√	√	√	√	√	√		2223.990
Model 8	√	√	√	√	√	√	√	√	√	2225.971

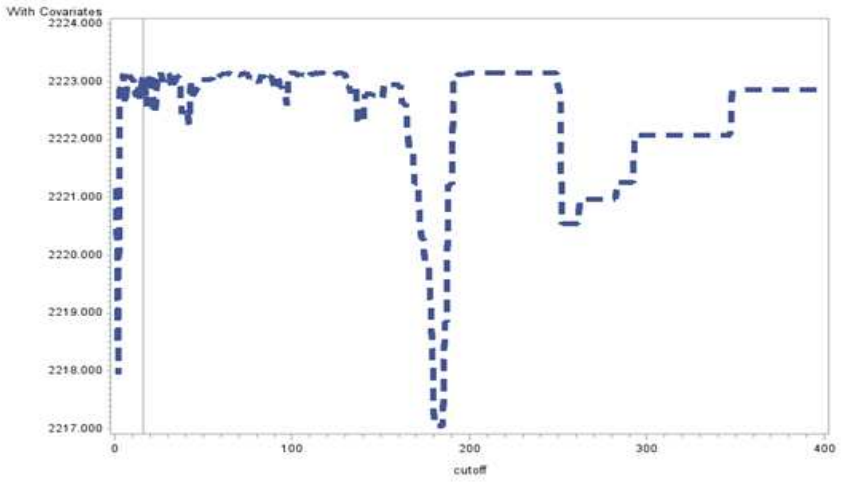
The model adjusting for HCC, HCV, etiology, alcohol, creatinine, and bilirubin had the lowest AIC of 2218.825 indicating that it was the best model for the data. Using this criterion I concluded that 180 miles was the best cutoff point regardless of which covariates I adjusted for. Figure 7 below

shows distance from the transplant center on the horizontal axis and the resulting AIC on the vertical axis. It was evident from the graph that the AIC was minimized at 180 miles.

The algorithm was rerun using the reduced dataset excluding deaths within the first 30 days post-transplant (n=22) and the AIC criterion consistently pointed to 180 days as a critical dichotomization point.



(a) HCV and HCC adjusted only



(b) HCC, HCV, Etiology, Alcohol, CR and Bilirubin adjusted

Figure 7: Using AIC to Estimate Optimal Cutoff

3.3 Restricted Cubic Regression Spline

The restricted cubic regression spline was employed to detect a potential non-linear relationship between distance as a continuous variable and post-transplant survival. Seven knots were placed at 1, 7.36, 21.32, 38.60, 87.88, 174.15, and 267.77 miles to correspond to the quantiles.

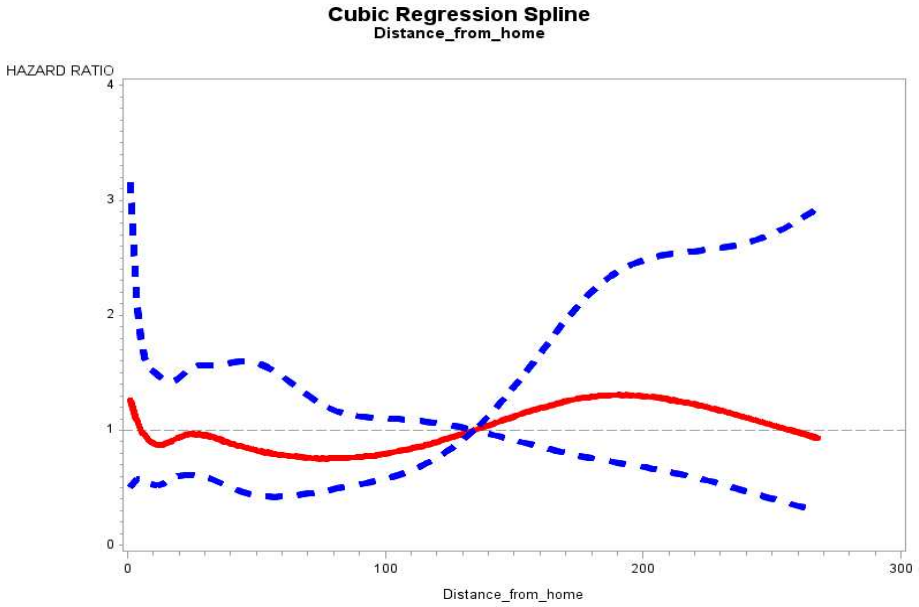


Figure 8: Restricted Cubic Regression Spline with 7 knots

From the curve above there was insufficient evidence of a distance effect as indicated by the Wald (p-value=0.87) for testing the hypothesis that distance had no effect on post LT survival as a continuous variable. This p-value only decreased slightly (0.79) when the restricted cubic regression spline was fit on the reduced dataset which excluded early deaths (n=799). This indicated there was a need to dichotomize distance and establish a cutoff point for distance beyond which there was an impact on survival. This lead to suspicion that there was no throughout effect of distance over the five year period post LT which was the reasoning behind the second modification to the Monte Carlo approach where β was dropped from the model.

The restricted cubic regression spline was also run with 5 knots which made no difference to the significance (p-value=0.9147). Using 9 knots (figure 12) at the following mileage points 2.8, 5.6, 8.9, 16.2, 29.1, 72.8, 114.2, 145.3, and 196 miles there was still no statistically significant effect of distance on survival when distance was specified as a continuous variable.

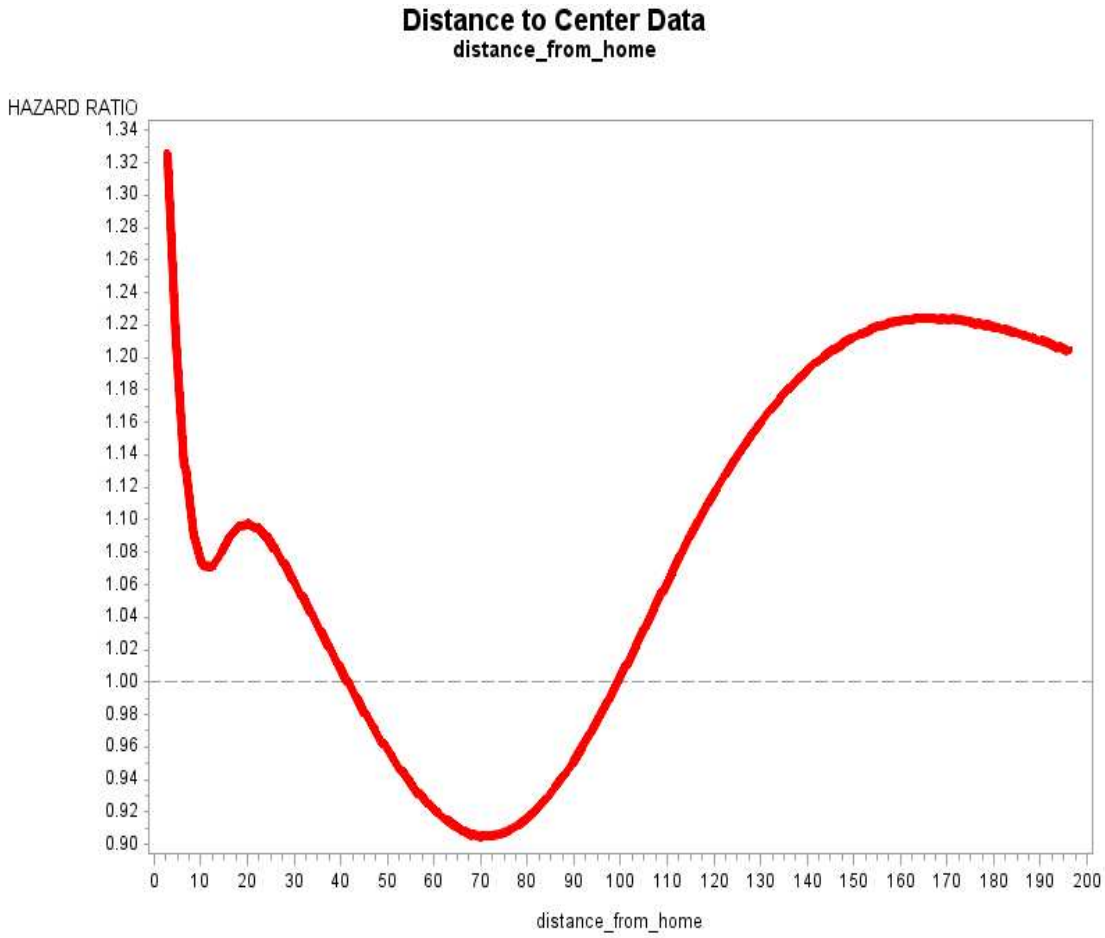


Figure 9: Restricted Cubic Regression Spline with 9 knots

As indicated using the AIC/ML criterion applied to Cox PH in figure 10, 180 miles was the best cutoff point regardless the other covariates I adjusted for. The KM for distance dichotomized at 180 miles is presented in the next section.

3.4 Kaplan Meier

The KM plot for this dichotomization is shown below in figure 10, and p-values for the log-rank (p-value=0.0049), Wilcoxon (p-value=0.0077) and likelihood ratio (p-value=0.0154) indicated that survival was improved for patients who had to travel less than 180 miles to the transplant center as opposed to those who lived more than 180 miles away.

Table 10: Dichotomizing Distance at 180 Miles for All Patients

Summary of the Number of Censored and Uncensored Values					
Stratum	**Distance**	**Total**	**Failed**	**Censored**	**Percent Censored**
1	**within 180**	802	169	633	78.93
2	**beyond 180**	19	9	10	52.63
Total		821	178	643	78.32

From table 10 above I noticed that 21% expired among those living within 180 miles while that increases to 47% for those living over 180 miles from the transplant center. In figure 13 below, it was evident that the survival curve dropped faster for those living beyond 180 miles. It remained steadily below that for patients living within the 180 mile radius indicating impaired survival at longer distances.

Table 11: Dichotomizing Distance at 180 Miles Excluding Early Deaths

Summary of the Number of Censored and Uncensored Values					
Stratum	**Distance**	**Total**	**Failed**	**Censored**	**Percent Censored**
1	**within 180**	780	150	630	80.77
2	**beyond 180**	19	9	10	52.63
Total		799	159	640	80.1

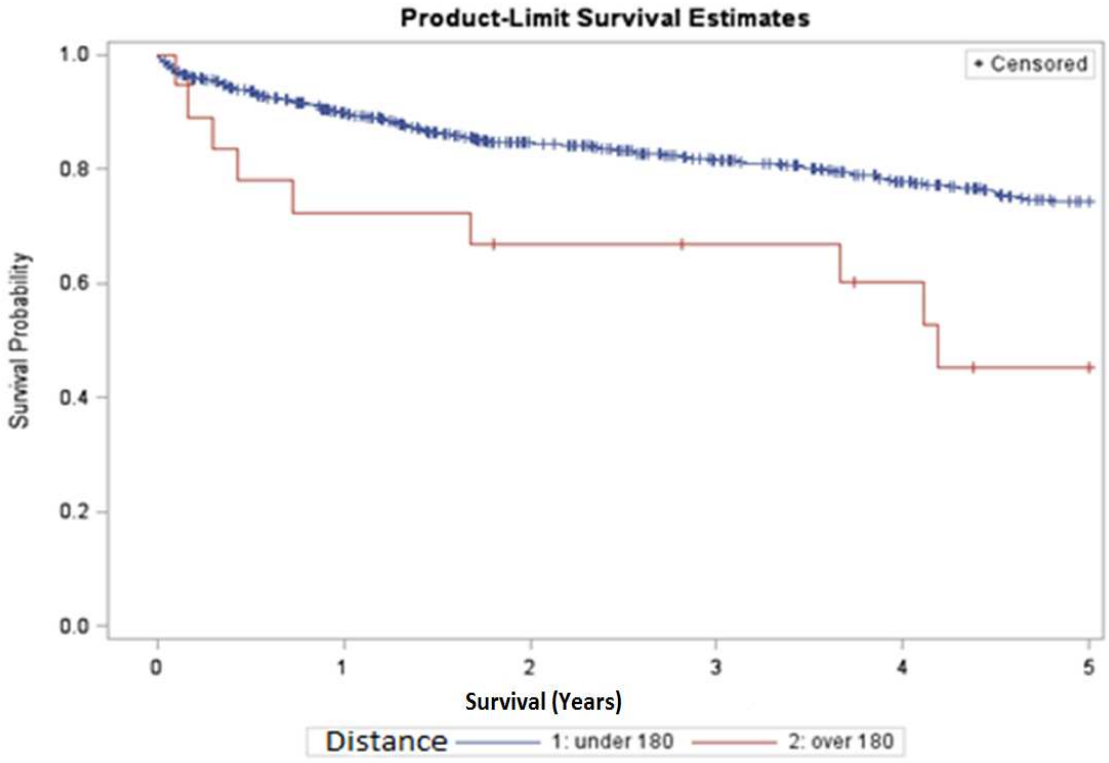


Figure 10: KM at 180 miles

3.5 Cox Proportional Hazards

The Cox PH model below with the 180 mile dichotomization point for distance, adjusted for HCV and HCC, was statistically significant according to likelihood ratio, Score and Wald p-values (all less than 0.0001). The hazard ratio indicated that patients living beyond 180 miles from the transplant center had 2.68 times the death rate compared to those living within 180 miles after adjusting for HCV and HCC (p-value=0.0040). Breslow's estimator was used to handle ties.

The Cox PH model was

$$\lambda_i(t) = \lambda_0(t) \exp\{0.25\ (HCC\ only)_i + 0.80\ (HCV\ only)_i + 0.95\ (both)_i - 0.99\ (under\ 180)\}$$

Table 12: Cox PH for 180 Mile Dichotomization

Analysis of Maximum Likelihood Estimates					
		All Patients (n=821)		Excluding Early Deaths (n=799)	
Parameter	**DF**	**P-value**	**Hazard Ratio**	**P-value**	**Hazard Ratio**
HCC only	1	0.6011	1.279	0.2884	1.660
HCV only	1	<.0001	2.227	<.0001	2.639
Both HCV and HCC	1	<.0001	2.577	<.0001	3.136
Distance beyond 180	1	0.0040	2.681	0.0009	3.145

P-values decreased when patients who died within the first 30 days post-transplant were excluded as noted in table 12 above. All of the 22 patients excluded for early death resided within the 180 mile radius.

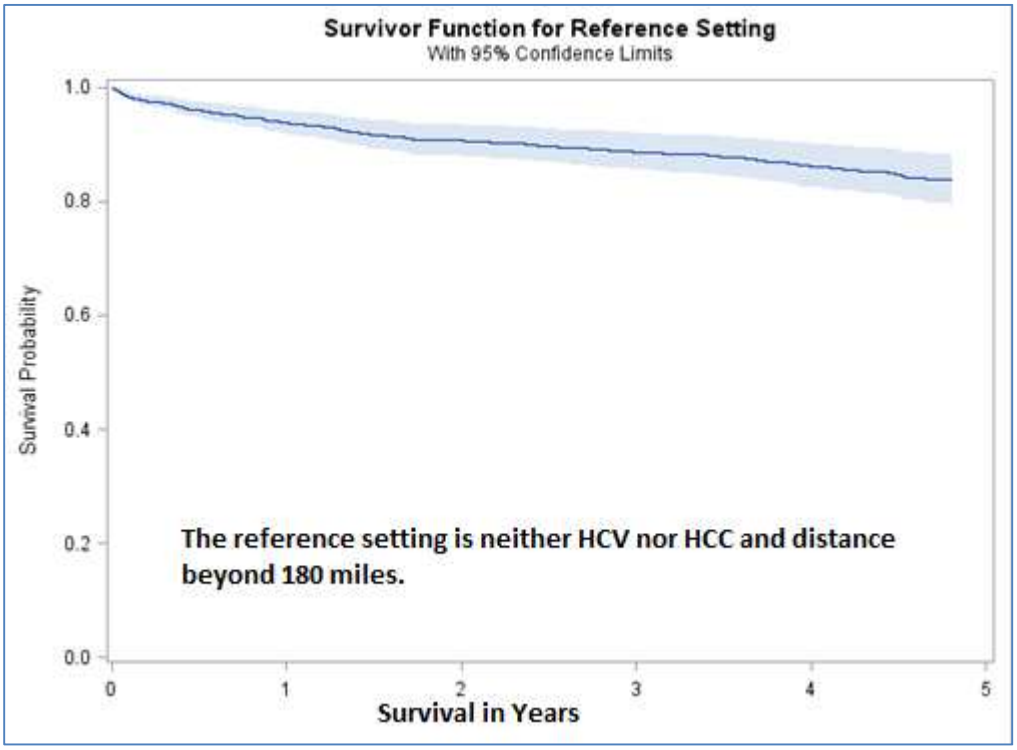


Figure 11: Survivor Function for Cox PH

Separate Cox PH models were fit for patients with HCV and HCC respectively. The effect of distance was more pronounced in patients with HCV (HR of 3.72 for 434 HCV patients vs. 2.5 for total 821 patients). The distance effect was even stronger in patients with HCC (HR of 5.24 for 134 HCC patients).

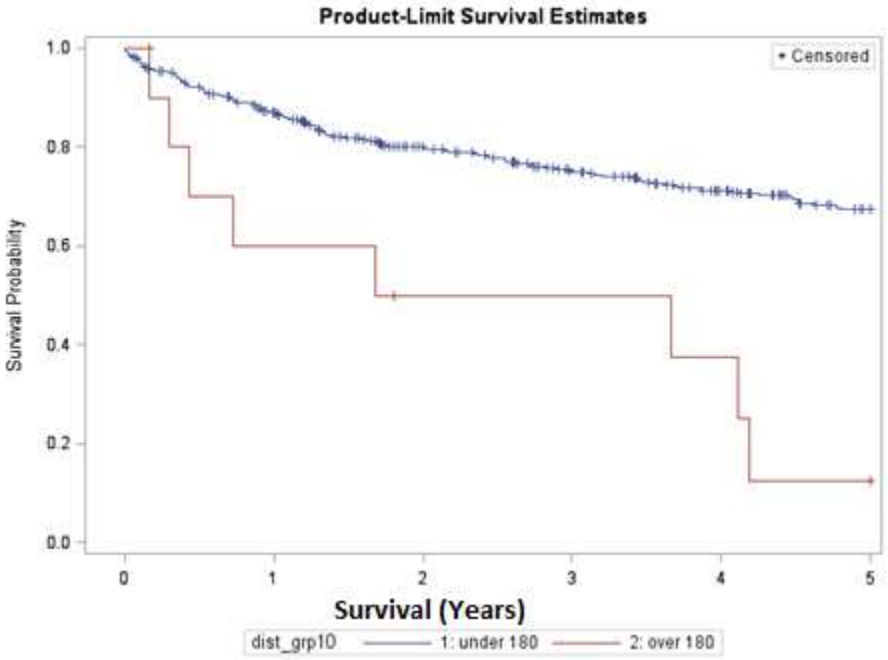


Figure 12: KM Curve for HCV Patients

To assess the fit of this model I first looked at deviance residuals. They were symmetrically distributed around 0 and had a standard deviation of approximately 1.0. Residuals were positive for patients with shorter survival time than expected and negative for patients with longer survival times than expected. Very low or very high values suggested that the patient may be an outlier and therefore in need of attention. Below in figure 13, the residuals were plotted against the covariate distance (continuous variable), and unusual patterns would have suggested features of the data that had not been adequately fitted by the model. This dataset contained censored data so caution was exercised because censoring could produce striking patterns that don't necessarily imply any problem with the model.

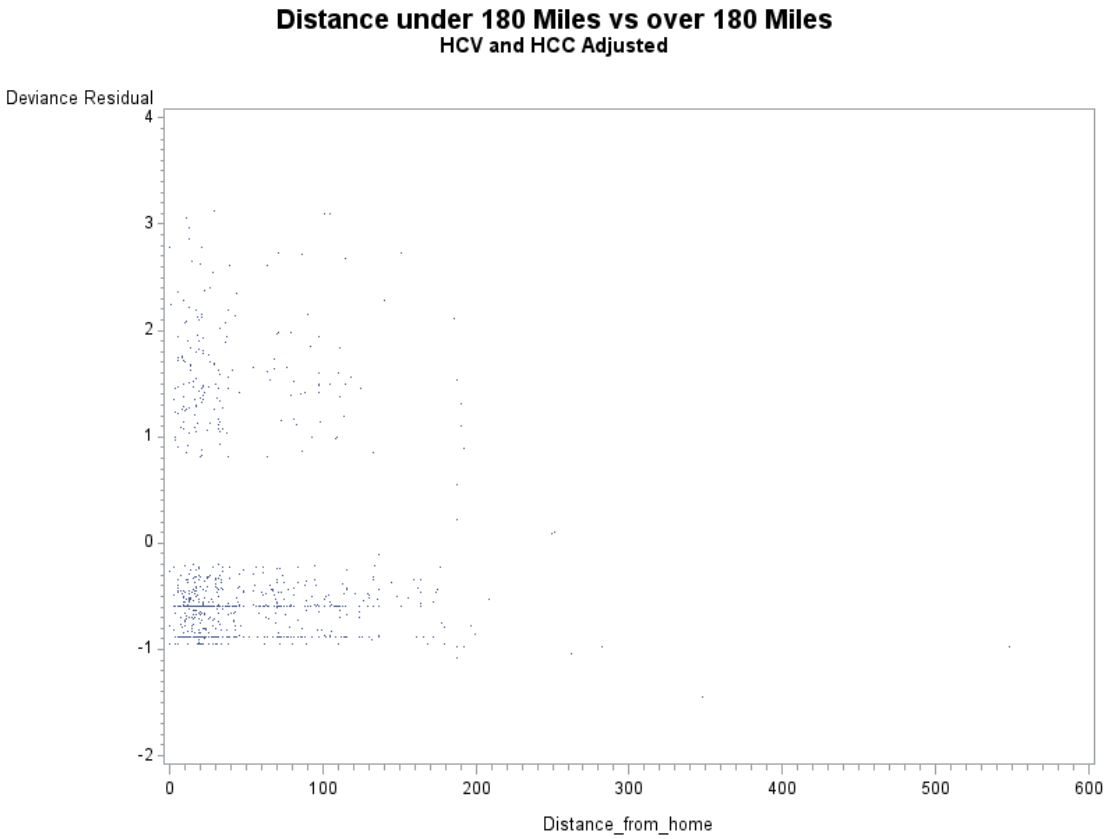


Figure 13: Graph of Deviance Residuals by Distance as a Continuous Variable

A clear disjunction between the two groups of observations was noted. The cluster of points toward the bottom were all censored observations, while the more widely dispersed points in the upper portion of the graph were uncensored observations.

Covariate-wise residuals including Schoenfeld residuals, weighted Schoenfeld residuals and score residuals all had a separate residual for each covariate for each patient. They also sum to approximately zero in the sample. However Schoenfeld residuals were not defined for censored observations. Since distance was a dichotomous variable, the graph for the residuals was not very informative and therefore was omitted.

Influence statistics were computed to detect whether any particular patient would change the results if it were removed from the model (Collett, 2003). Influence on the model as a whole is

measured by the likelihood displacement (LD) statistic which detects how much the log-likelihood multiplied by two, would change if the individual patient was removed from the sample. LD was plotted below in figure 14 against distance, and since all values were relatively small I concluded that there were no influential patients who needed to be revisited.

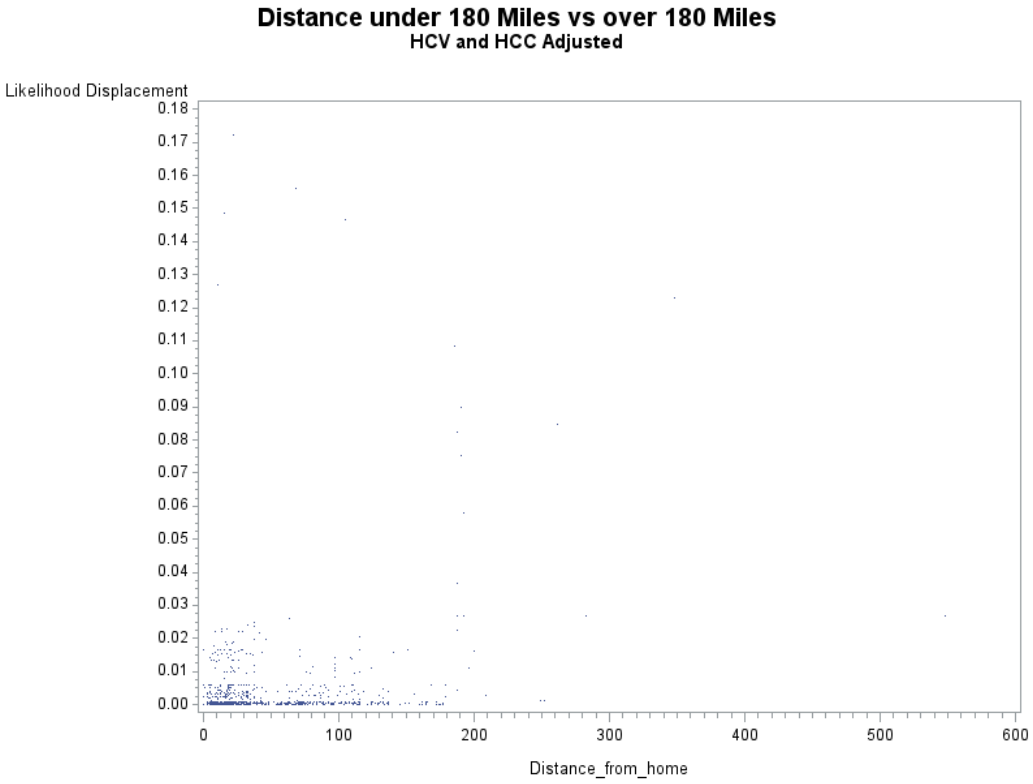


Figure 14: Likelihood Displacement Against Distance to Look for Influential Patients

Another method used to check the proportional hazards (PH) assumption was to look at the cumulative residuals using the Assess statement and ph option in the proc phreg function (Gharibvand, 2008).

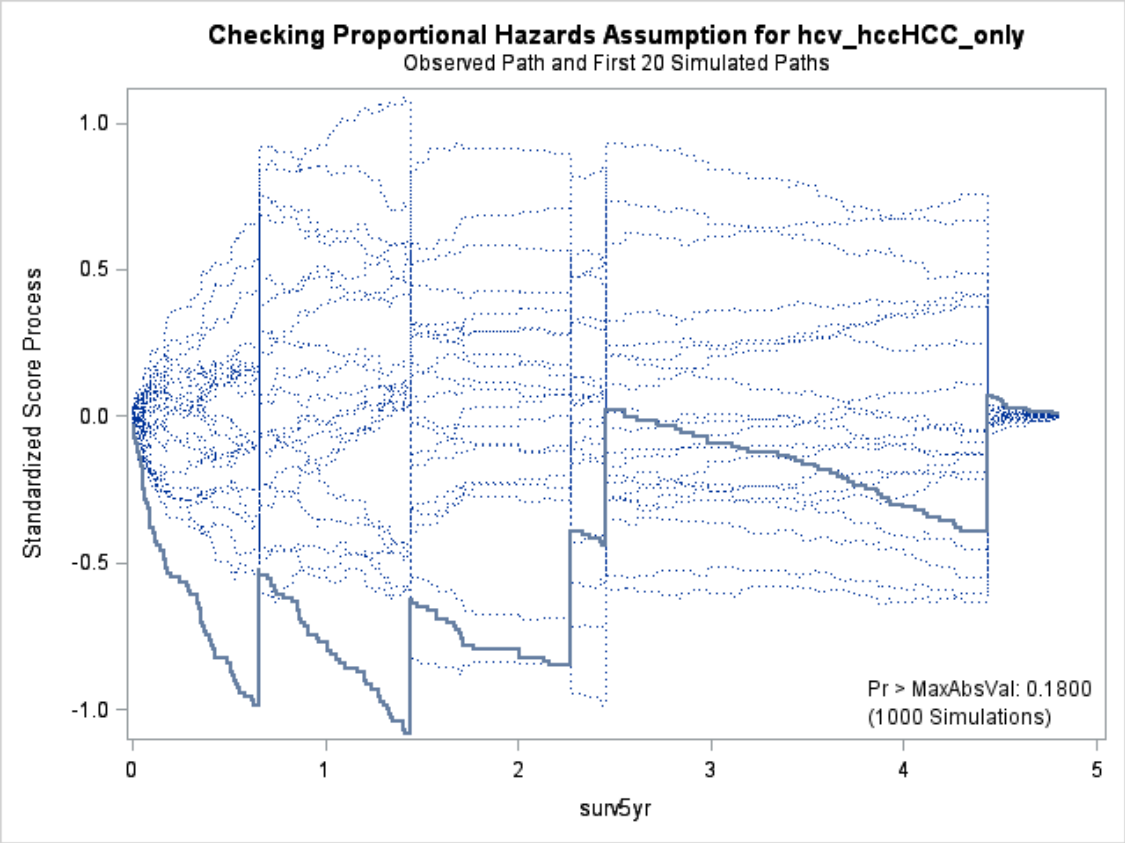


Figure 15: Checking the PH Assumption for Patients with HCC

A plot of the cumulative martingale residuals against the values of the 5 year survival and a p-value of a Kolmogorov-type supremum test based on a sample of 1,000 simulated residual patterns are presented.

The plot in figure 15 displays the observed cumulative martingale residual process for survival together with 20 simulated realizations from the null distribution. It is obvious that the observed process is typical compared to the simulated realizations. Also, some of the 1,000 simulated realizations have an absolute maximum exceeding that of the observed cumulative martingale residual process. Both the graphical and numerical results indicate that a transform is not deemed necessary in the model.

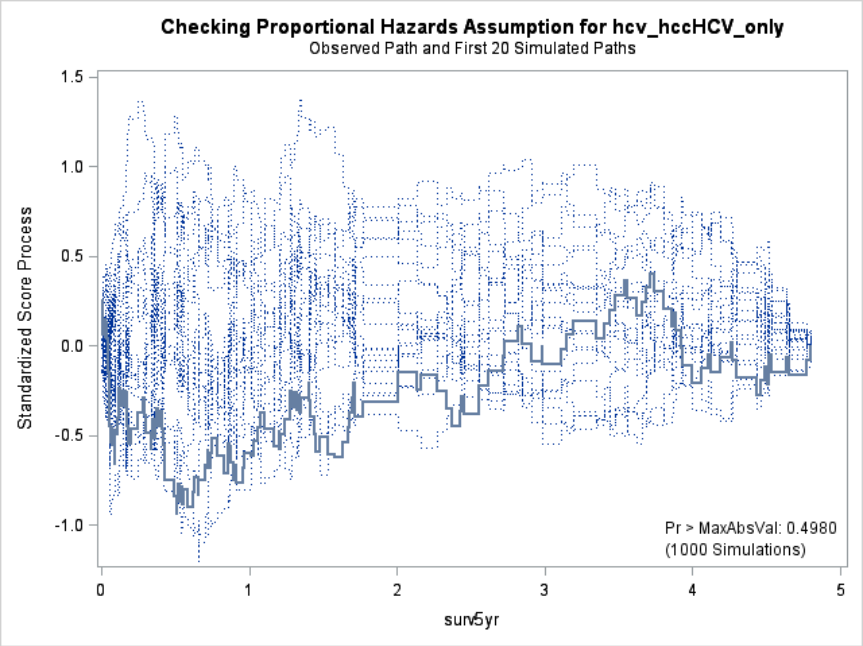


Figure 16: Checking the PH Assumption for Patients with HCV

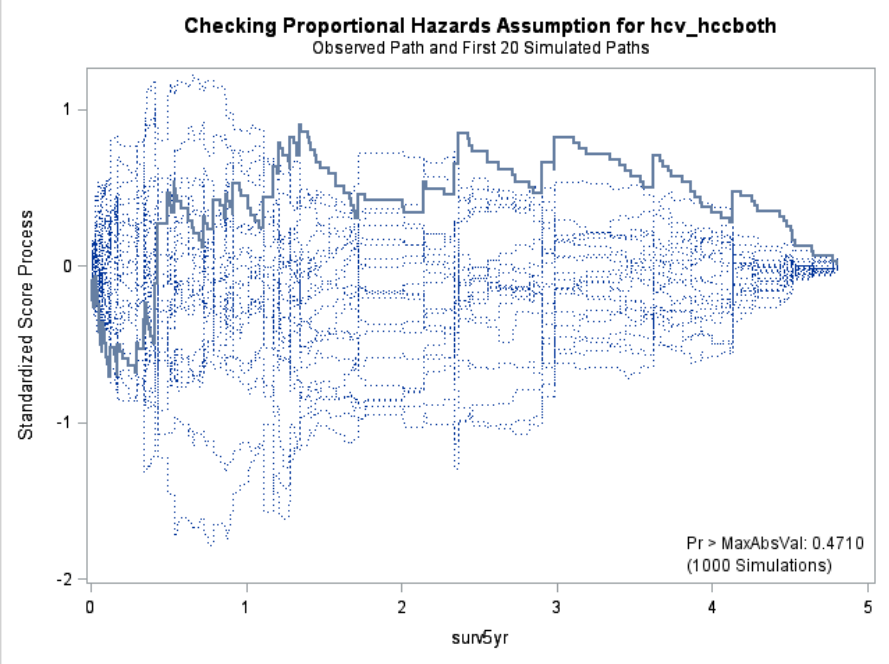


Figure 17: Checking the PH Assumption for Patients with Both HCV and HCC

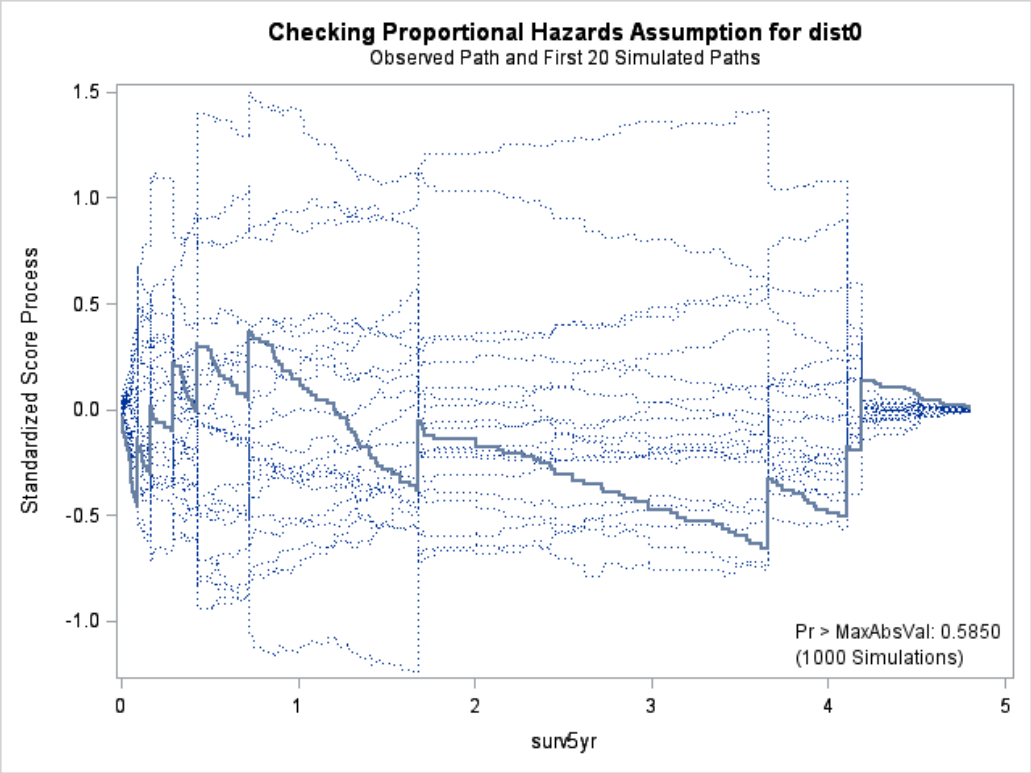


Figure 18: Checking the PH Assumption for Distance Beyond 180 miles

Table 13: Supremum Test for Proportional Hazards Assumption

Supremum Test for Proportional Hazards Assumption				
Variable	**Maximum Absolute Value**	**Replications**	**Seed**	**Pr > MaxAbsVal**
HCC only	1.0828	1000	19	0.18
HCV only	0.9383	1000	19	0.498
Both HCV and HCC	0.8996	1000	19	0.471
Distance Beyond 180	0.656	1000	19	0.585

Based on table 13 and figures 13-18 above, I concluded that there was no serious violation of the proportional hazards (PH) assumption. Since the p-value for all covariates exceeded 0.05, I am

95% confident that there was no relationship between residuals and time. This also confirmed the previous conclusion that the PH assumption was not violated and they were no time-dependent covariates. Therefore, I did not need to do any further stratification.

In the literature, the Cox PH model was the overwhelmingly favored method for regression analysis of survival data. This method required no assumptions about the shape of the distribution of survival times, allowed for time-dependent covariates, could be used with both discrete and continuous-time data, could stratify on categorical control variables and could be extended to non-proportional hazards. However, one cannot test hypotheses about the shape of the hazard function so I resorted to another method known as the accelerated failure time model.

3.6 Accelerated Failure Time Model

An AFT model for the 180 mile distance cutoff was used assuming a Weibull distribution. Other distributions were fit, but Weibull was the best for the LT data. Since the underlying distribution of survival time conditional on the covariates was estimated to be Weibull with σ=0.7, this indicated that the hazard was increasing at a decreasing rate. The survivor function was

$$S_i(t) = \exp\left\{-\left[t_i e^{-2.75+0.37\,HCC\,+1.22\,HCV+1.43\,Both-1.5\,Dist\_under\_180}\right]^{1.53}\right\}$$

where HCV, HCC, Both and Dist_under_180 were all indicator variables that determined whether or not the condition was present.

The log survival became

$$\log T = 2.75 - 0.37\,HCC - 1.22\,HCV - 1.43\,Both + 1.50\,(Distance\ under\ 180) + \sigma\varepsilon$$

Therefore all other covariates held constant, living within 180 miles increased survival by $e^{1.5}$=4.48 units. The 'mean' survival time in patients living within 180 miles was estimated to be 4.48 times longer than that in the group living beyond 180 miles after controlling for HCV and HCC. The hazard function was illustrated in figure 19 below.

Table 14: AFT Model with Weibull Distribution 180 Mile Dichotomization All Patients

Analysis of Maximum Likelihood Parameter Estimates						
				95% Confidence		
Parameter	**DF**	**Estimate**	**Standard Error**	**Limits**		**P-value**
Intercept	1	2.75	0.55	1.67	3.83	<.0001
HCC only	1	-0.37	0.72	-1.77	1.03	0.6060
HCV only	1	-1.22	0.28	-1.77	-0.66	<.0001
Both HCV and HCC	1	-1.43	0.37	-2.15	-0.71	<.0001
Distance under 180	1	1.50	0.53	0.46	2.54	0.0046
Scale	1	1.52	0.11	1.33	1.74	
Weibull Shape	1	0.66	0.05	0.57	0.75	

Table 15: AFT Model with Weibull Distribution 180 Mile Dichotomization Excluding Early Deaths

Analysis of Maximum Likelihood Parameter Estimates						
				95% Confidence		
Parameter	DF	Estimate	Standard Error	Limits		P-value
Intercept	1	2.56	0.44	1.70	3.42	<.0001
HCC only	1	-0.62	0.57	-1.73	0.50	0.2790
HCV only	1	-1.16	0.25	-1.64	-0.68	<.0001
Both HCV and HCC	1	-1.37	0.31	-1.98	-0.76	<.0001
Distance under 180	1	1.36	0.42	0.54	2.18	0.0012
Scale	1	1.19	0.09	1.03	1.37	
Weibull Shape	1	0.84	0.06	0.72	0.97	

By excluding early deaths (n=22) the estimators became more precise with tighter confidence intervals, and the p-values decreased giving a more precise model.

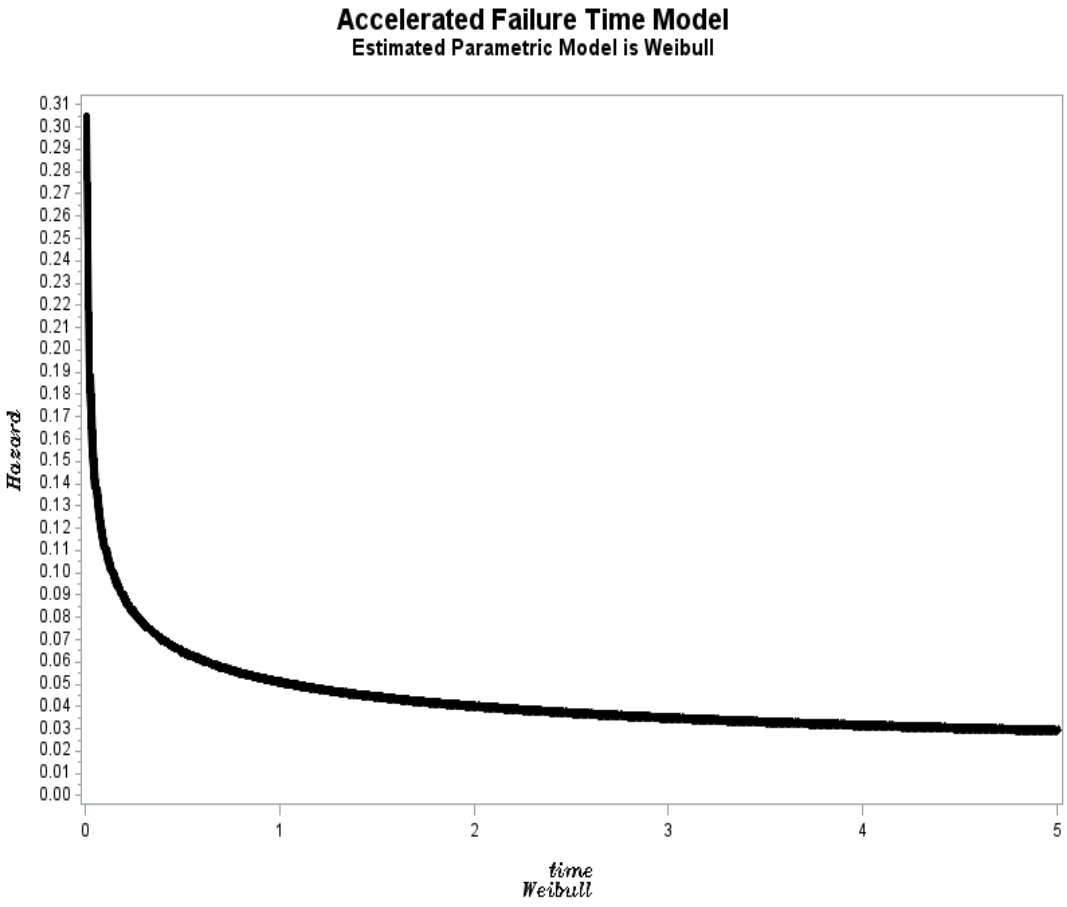


Figure 19: Hazard Function

The Cox-Snell residuals were defined as . These were always positive and when the model was correct they followed an exponential distribution with parameter λ=1. In order to evaluate this graphically I computed the KM estimator of the survivor function, and plotted that against the residuals *e*. The resulting graph was expected to be a straight line which passed through the origin with a slope of 1 (see Figure 20).

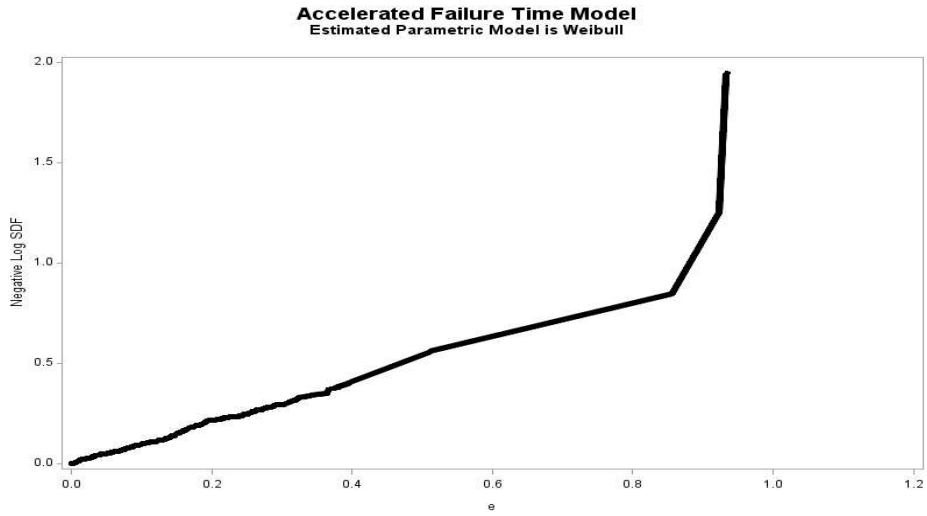


Figure 20: Residual Plot (Weibull Model)

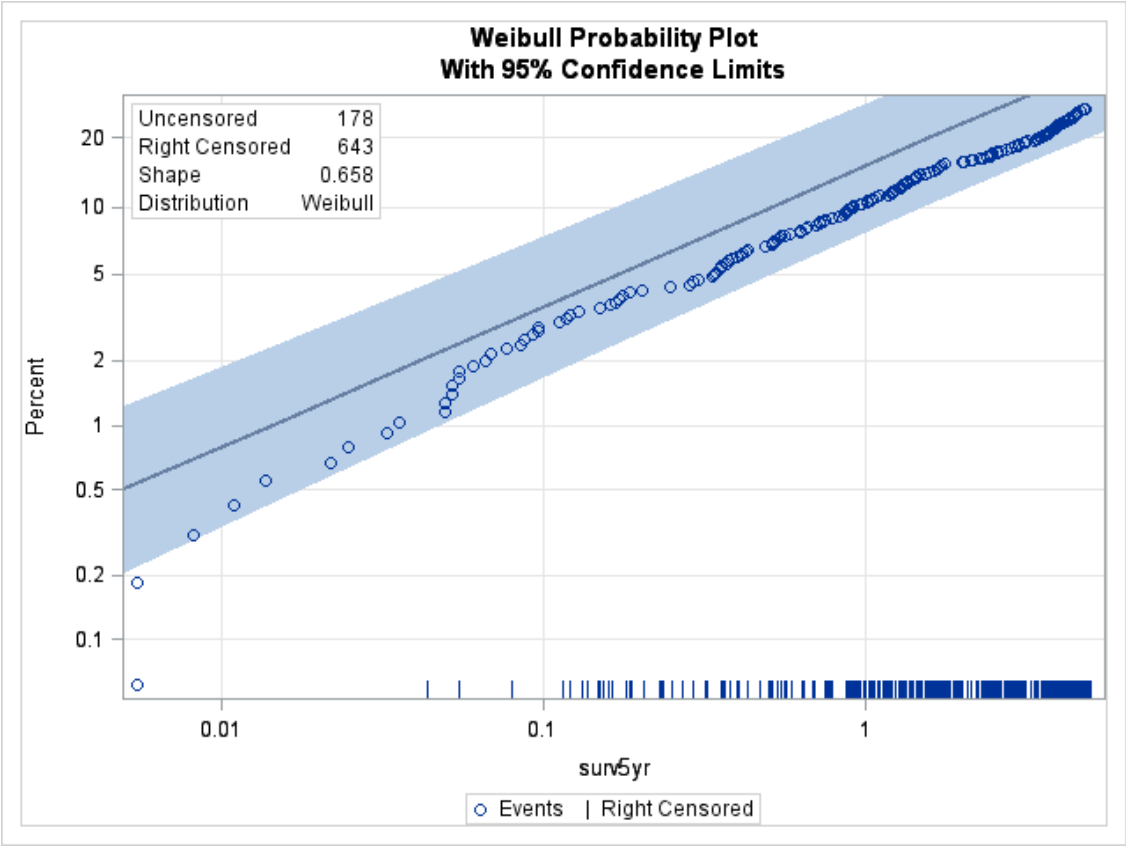


Figure 21: Weibull Probability Plot

The estimated cumulative density function, a line representing the maximum likelihood fit, and point wise parametric confidence bands were plotted in the body of figure 21. The values of right-

censored observations were plotted along the bottom of the graph (Kay & Kinnersley, 2002). All the events fell within the 95% confidence bands indicating a good fit for this model.

Next I fit the same model using the exponential and lognormal distributions and it was evident that the fit was not as good (figure 22 and 23).

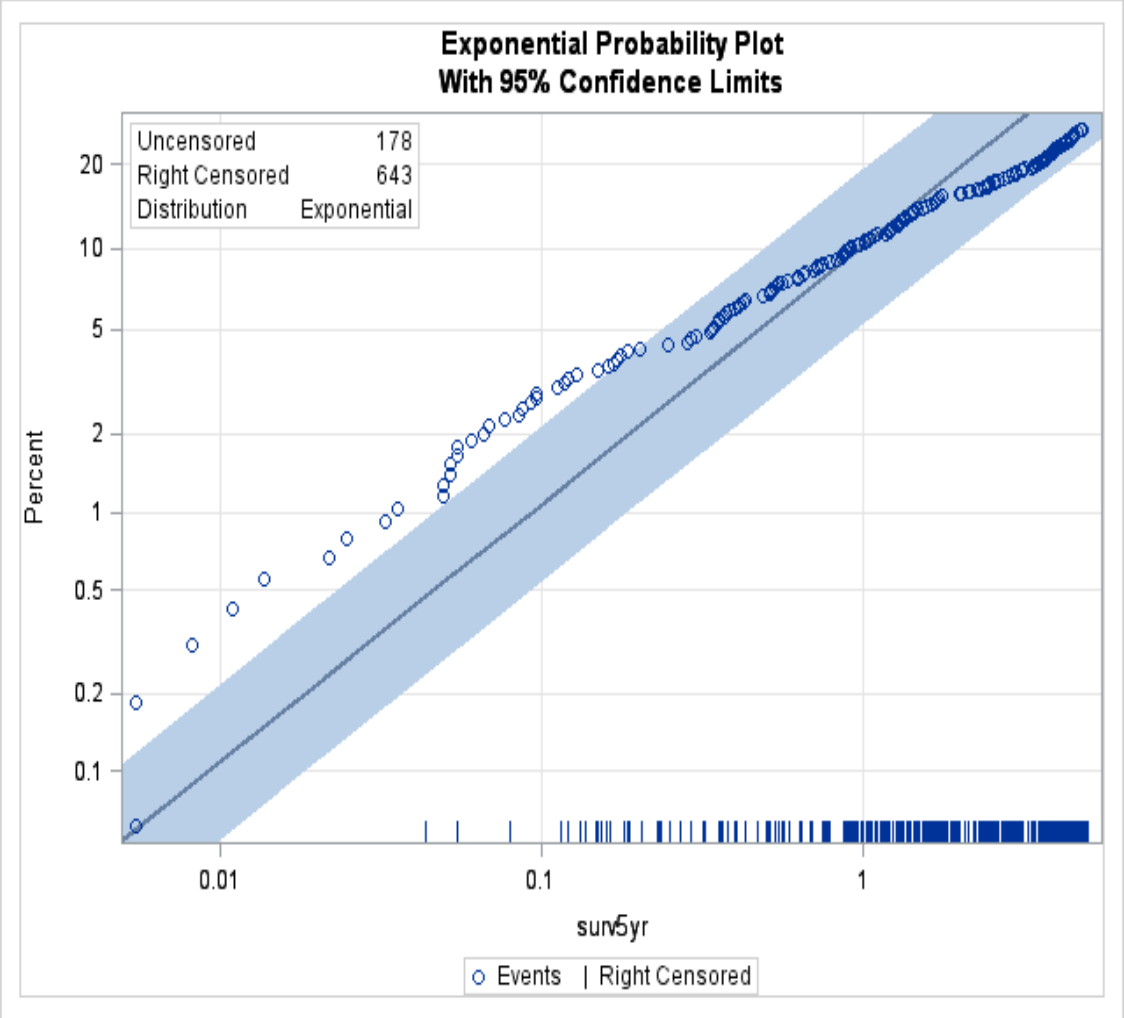


Figure 22: Exponential Probability Plot

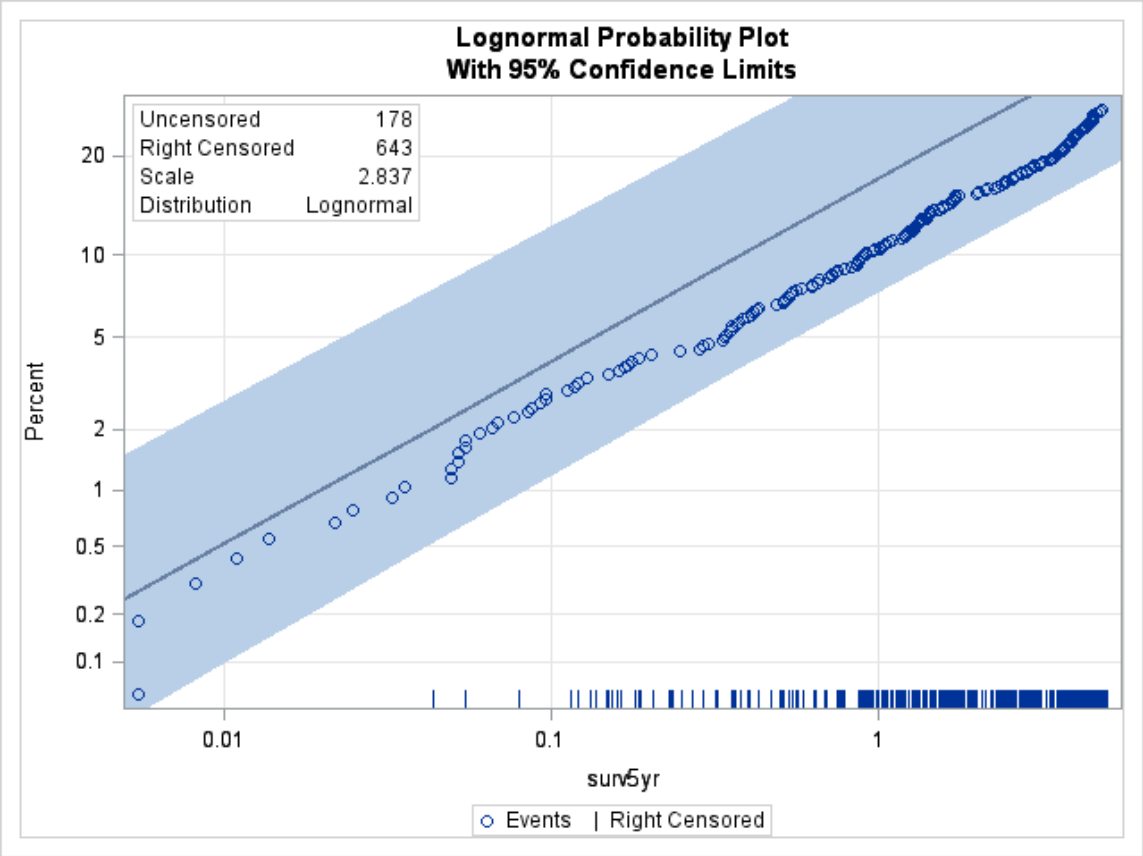


Figure 23: Lognormal Probability Plot

3.7 Logistic Regression

LR model is useful in predicting dichotomous outcomes, in this case, whether the patient is dead or alive at 1 year post-transplant. Patients transplanted after the time period (September 25, 2011) were excluded from the study (n=82) because they had not yet received one year of follow-up. Distance models were computed, adjusting for HCV for the cutoff of 1 year post LT. It was established that at a 180 mile distance cutoff, survival was affected (p-value =0.0246) (Table 15). The odds of death 1 year post LT were 3.4 times higher for patients beyond 180 miles, compared to those within. When the model was rerun excluding early deaths, the odds of death 1 year post LT increased to 4.57 compared to those who lived within the 180 mile radius. However, LR discarded valuable information by ignoring the length of patient survival and by reducing outcomes to a dichotomous variable (dead/alive) at a particular time point. It was also computed

on a reduced dataset of 739 patients compared to the Cox PH and AFT models which were able to utilize all 821 patients in the entire study group.

A total of 739 patients were included in the study and the model was adjusted for the presence of HCV alone. This was done, as opposed to both HCV and HCC, to avoid non-convergence issues. Assuming P_{it} represents the probability that individual *i* had an event at time t, conditional on the fact that an event had not already occurred to that individual. When distance takes on a value of 1 for within 180 miles and 2 for beyond 180 miles the model can be written as

$$\log\left(\frac{P_{it}}{1-P_{it}}\right) = -1.65_t - 0.40\, HCV_{it} - 0.62\, Distance_{it}$$

Table 16: Logistic Regression 180 Mile Dichotomization All Patients

Analysis of Maximum Likelihood Estimates						
Parameter		**Estimate (SE)**	**Odds Ratio**	**95% Wald CI**		**P-value**
Intercept		-1.65 (0.28)				<.0001
Hepatitis C	No	-0.39 (0.13)	0.459	0.275	0.765	0.0028
Distance	1: under 180	-0.61 (0.27)	0.292	0.1	0.854	0.0246

The same model was rerun excluding patients who died in the first 30 days post LT. A total of 722 patients were included in the model. The resulting model was presented below.

$$\log\left(\frac{P_{it}}{1-P_{it}}\right) = -1.85_t - 0.56\, HCV_{it} - 0.76\, Distance_{it}$$

Table 17: Logistic Regression 180 Mile Dichotomization Excluding Early Deaths

Analysis of Maximum Likelihood Estimates						
Parameter		**Estimate (SE)**	**Odds Ratio**	**95% Wald CI**		**P-value**
Intercept		-1.85 (0.29)				<.0001
Hepatitis C	No	-0.56 (0.16)	0.323	0.174	0.600	0.0003
Distance	1: under 180	-0.76 (0.28)	0.219	0.073	0.654	0.0066

I computed LR models at 3 and 5 years post LT at multiple incremental groupings of 30 miles [30, 60, 90,120,150, 180, and so on], of 60 miles, 90 miles, 120 miles, 150 miles, 180 miles, and 240 miles. Models of distance cutoffs vs. the rest of the patients (30 miles vs. the rest; 60 miles vs. the rest, and so on, up to 240 miles) were also run. None of these LR models was able to pinpoint a distance vs. outcome effect, due to the use of a markedly limited dataset from extensive censoring.

3.8 Monte Carlo Approach to Change Point Detection

Initially I used the method developed by Liu et al. (Liu et al., 2008) with distance dichotomized at 180 miles from the transplant center. Patients were dichotomized according to the distance from transplant center (0: within 180 miles, 1: beyond 180 miles) and KM curves were presented in figure 8. KM curves crossed early on and survival for the distant group remained steadily below that of patients living within the 180 mile radius. This indicated that distance may not affect survival immediately after transplant but may begin at a later point in time. It was also possible that the distant patients were in the hospital for longer and the distance effect did not begin until after their hospital discharge. This was a factor that I investigated further using Cox PH models that included change points.

In its general form the hazard function was specified as

$$\lambda(t) = \lambda_0(t)\exp\left[\theta' Z(t,\xi) + \eta' X(t)\right]$$

where Z represented distance, ξ contained the nonlinear parameters τ which was the change point in time and c the dichotomization point for distance, η contained the coefficients β for distance and γ for the presence of Hepatitis C and HCC, and *X(t)* contained indicator variables for Hepatitis C (HCV) and Hepatocellular Carcinoma (HCC).

Let

$$X_1 = \begin{cases} 0 \text{ if the patient did not have HCV} \\ 1 \text{ if the patient had HCV} \end{cases}$$

$$X_2 = \begin{cases} 0 \text{ if the patient did not have HCC} \\ 1 \text{ if the patient had HCC} \end{cases}$$

3.8.1 Original Hazard with One Change Point

I was interested in testing the existence of a change point associated with the distance cutoff that was determined using the AIC (180 miles) over the time period of 0-5 years post liver transplant. When distance Z was already dichotomized at 180 miles the hazard function above could be simplified to

$$\lambda(t; Z, X) = \lambda_0(t) \exp[\{\beta + \theta I(t \leq \tau)\}Z + \eta' X].$$

Using the Monte Carlo approach by Liu et al., distance Z was defined as 0: within 180 miles and 1: beyond 180 miles, while adjusting for HCV (1: yes; 0: no) and HCC (1: yes; 0: no). The p-values were 0.566 and 0.235 and the overall 5 per cent significance level thresholds were 3.57 and 8.04 for M and M*, respectively, based on 10,000 resampling samples. The thresholds were presented by the horizontal lines in the plot. The potential change point locations in terms of achieving the maximal score statistics were at 3.622 and 4.186 years post-transplant for $|S_\theta(\tilde{\eta};\ \tau)|$ and $W_\theta(\tilde{\eta};\ \tau)$ respectively.

Table 18: Change Point Detection Using 180 Mile Dichotomization

	Maximum Test Profile Value	95% Threshold	P-value	Change Point
sup |S|	1.916	3.609	0.562	3.624
sup W	4.823	8.032	0.233	4.187

*Maximization was over a range of τ between 0.25 and 4.75.

After the change-point was found (even though p-values were high), a Cox model was fit with the indicator variable included to represent the change point. Caution was taken in interpreting the model since the estimation ignored the fact that the candidate change point was determined by the test procedure using the same data set. This research was conducted under the assumption that there was only one change point parameter τ but it can be extended easily to incorporate multiple change points in time.

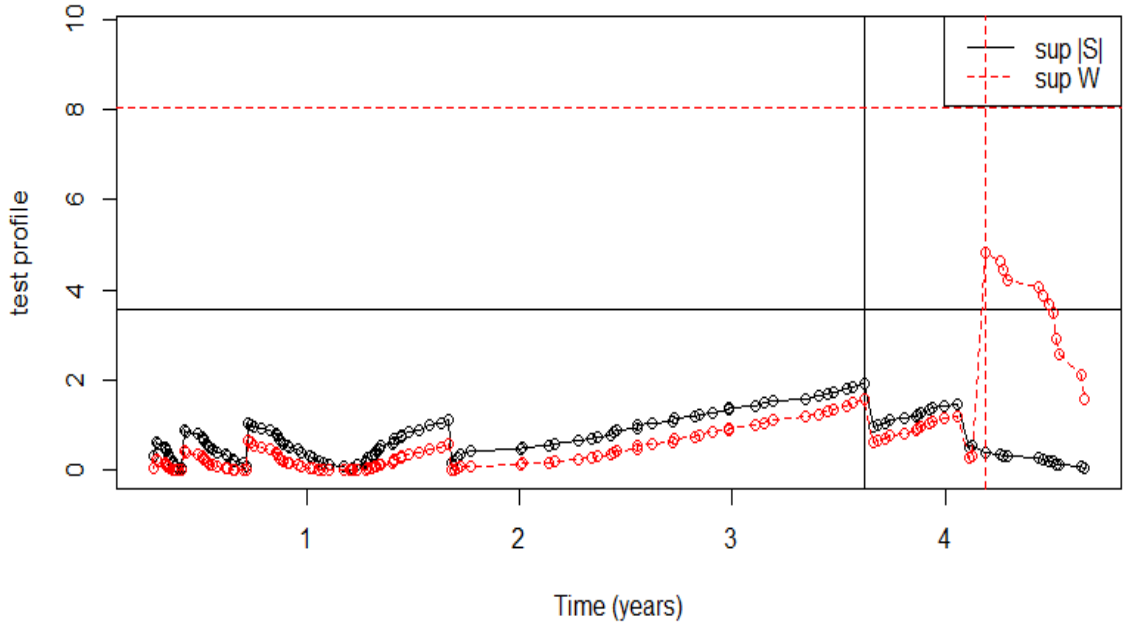


Figure 24: Profiles of Score Test Processes

The resulting Cox PH model with the change point τ set at 4.186 which was indicated by the normalized test statistic yielded the following model:

$$\lambda(t; Z, X) = \lambda_0(t) \exp[\{3.71 - 4.87I(t \leq 4.186)\}Z - 0.70\, HCV + 0.14\, HCC]$$

The reference were patients with both HCV and HCC living within 180 miles with the change point at $\tau = 4.186$. The hazard ratios were reported in table 19 below. Likelihood Ratio, Score and Wald statistics all agreed that this model was statistically significant (p-value<0.0001) and the AIC value was 1919.314.

Table 19: Cox PH Model with $\tau = 4.186$

Analysis of Maximum Likelihood Estimates							
Parameter		**DF**	**Parameter Estimate**	**Standard Error**	**Chi-Square**	**P-value**	**Hazard Ratio**
HCV	n	1	-0.70363	0.16878	17.3802	<.0001	0.495
HCC	n	1	0.14235	0.19212	0.549	0.4587	1.153
β	0	1	3.70805	0.43698	72.0053	<.0001	40.774
θ	0	1	-4.87466	0.44645	119.2172	<.0001	0.008

Next I fit the model with the change point $\tau = 3.622$ which was indicated by the non-normalized score statistic and below are the resulting estimates.

Table 20: Cox PH model with $\tau = 3.622$

Analysis of Maximum Likelihood Estimates							
Parameter		**DF**	**Parameter Estimate**	**Standard Error**	**Chi-Square**	**P-value**	**Hazard Ratio**
HCV	n	1	-0.76966	0.16745	21.1278	<.0001	0.463
HCC	n	1	0.15028	0.19178	0.614	0.4333	1.162
β	0	1	2.99874	0.38993	59.1447	<.0001	20.06
θ	0	1	-4.89206	0.46594	110.2361	<.0001	0.008

AIC was 1922.756 which was larger than 1919.314 from the model with $\tau = 4.186$ indicating worse fit statistically. Likelihood Ratio, Score and Wald statistics were all in agreement that this model was significant (p-value<0.0001).

It was known from the work of Liu et al. (2008) that accuracy of the approximation and power increased with sample size and instability of the variance could be seen when there were few deaths such as the case with this data (78.32% censored). The authors also established that the maximum normalized value of the score statistic was less sensitive in the detection of a change point than that based on the non-normalized score. The maximum normalized value for the test score statistic was also more conservative under the null hypothesis (Liu et al., 2008). Even

though the AIC score was lower for the Cox PH model with the change point established by the normalized score ($\tau = 4.186$), I presented both in this analysis since the non-normalized score (which showed $\tau = 3.622$) was more sensitive to change point detection.

3.8.2 Extension 1: Drop β

I dropped β, to determine if distance Z did not have a throughout effect, but only an effect after a certain time τ.

$$\lambda(t; Z, X) = \lambda_0(t) \exp[\{\theta I(t \leq \tau)\}Z + \eta' X]$$

The p-values were 0.573 and 0.543 and the overall 5 per cent significance level thresholds were 3.387 and 7.592 for M and M*, respectively, based on 100,000 resampling samples. The thresholds were presented by the horizontal lines in the plot. The potential change point locations in terms of achieving the maximal score statistics were at 3.622 years post LT for both $|S_\theta(\tilde{\eta};\ \tau)|$ (non-normalized) and $W_\theta(\tilde{\eta};\ \tau)$ (normalized) respectively.

Table 21: Change Point Detection Without β

	Maximum Test Profile Value	95% Threshold	P-value	Change Point
sup |S|	1.915	3.387	0.573	3.622
sup W	2.624	7.592	0.543	3.622

By dropping beta and increasing the number of resampling samples to 100,000 instead of 10,000, the normalized and non-normalized test statistics agreed at 3.622. A Cox PH model was fit illustrating that distance (dichotomized at 180) was significant (p-value < 0.0001) after 3.622 years post LT. I concluded that the hazard of death for patients living beyond 180 miles of the LT center had 2.023 times the death rate before (and including) 3.622 years compared to those who lived within 180, after adjusting for HCV and HCC. However this model had to be interpreted with caution since there was evidence of a throughout distance effect as indicated in the previous approach.

Table 22: Cox PH Including a Change Point Without β

Analysis of Maximum Likelihood Estimates							
Parameter		**DF**	**Parameter Estimate**	**Standard Error**	**Chi-Square**	**P-value**	**Hazard Ratio**
HCC	n	1	0.2544	0.4709	0.2918	0.5891	1.290
HCV	n	1	0.7794	0.1781	19.1598	<.0001	2.180
Both			0.9280	0.2340	15.7305	<.0001	2.530
Z(t)	0	1	0.7047	0.4172	0.8536	0.0912	2.023

Likelihood Ratio, Score and Wald statistics all agreed that this model was significant (p-value<0.0001). AIC was 1954.936. Since the AIC increased, and β was significant in table 19, I concluded that β should remain in the model and there was evidence of a throughout effect of distance on survival when distance was dichotomized at 180 miles.

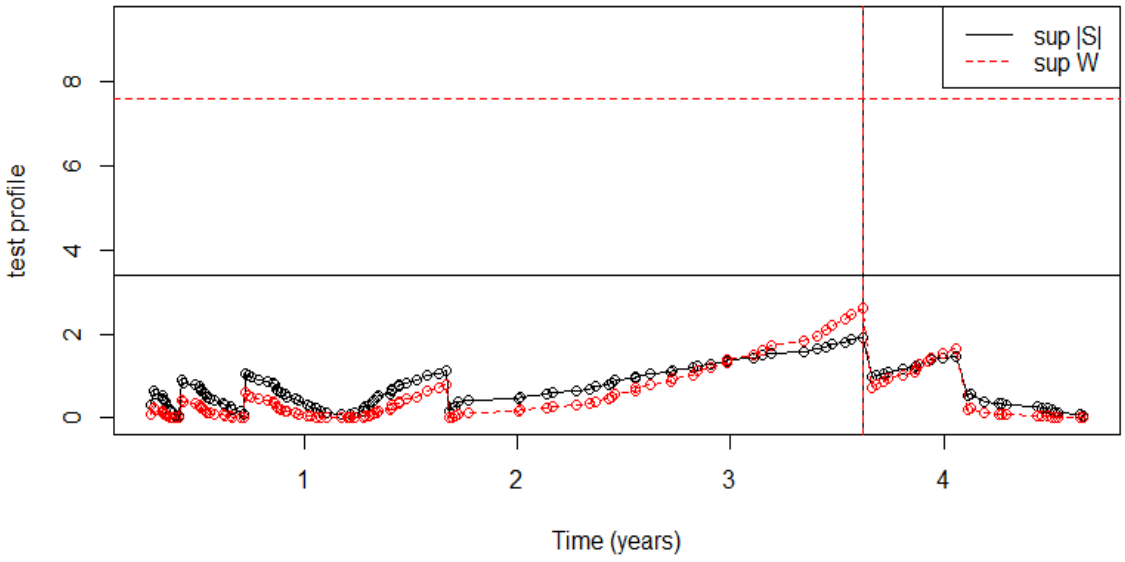


Figure 25: MC Approach With β Dropped

3.8.3 Extension 2: Reversed Inequality Sign

I changed $I(t \leq \tau)$ to $I(t \geq \tau)$ which was a statistically equivalent model but had a more natural interpretation in this dataset. Since β was not retained in the model, this meant there was only a distance effect for times greater than τ and when β was retained, this meant that there was an additional effect due to distance after a certain point in time post-transplant, τ.

$$\lambda(t; Z, X) = \lambda_0(t) \exp[\{\beta + \theta I(t \geq \tau)\}Z + \eta' X]$$

In this model, I adjusted for HCV (1: yes; 0: no) and HCC (1: yes; 0: no) as X(t) before modeling distance Z (1: within 180, 0: beyond 180) since these diseases are known to recur and therefore reduce survival. The p-values were 0.080 and 0.315 and the overall 5 per cent significance level thresholds were 4.855 and 7.491 for M and M*, based on 100,000 resampling samples. The thresholds were presented by the horizontal lines in the plot in figure 25. The potential change point locations in terms of achieving the maximal score statistics were at 0.286 years post LT for $|S_\theta(\tilde{\eta};\ \tau)|$ and $W_\theta(\tilde{\eta};\ \tau)$.

Table 23: Change Point Detection with Inequality reversed

	Maximum Test Profile Value	95% Threshold	P-value	Change Point
sup |S|	4.430	4.855	0.080	0.286
sup W	3.718	7.491	0.315	0.286

The normalized and non-normalized score statistics showed agreement at a change point τ of 0.286 years post-transplant possibly due to random variation in the Monte Carlo. The Cox PH model was fit using this change point and the resulting model is shown in table 24 below. Both β and θ were statistically significant indicating a distance effect throughout the 5 year period (p-value=0.0002 for β) as well as an additional burden of distance beyond 180 miles (p-value<0.0001 for θ) on post-transplant survival.

Table 24: Cox PH with the Inequality Reversed

Analysis of Maximum Likelihood Estimates						
Parameter	**DF**	**Parameter Estimate**	**Standard Error**	**Chi-Square**	**P-value**	**HR**
HCC	1	0.245	0.471	0.271	0.602	1.278
HCV	1	0.782	0.178	19.304	<0.001	2.187
Both	1	0.942	0.234	16.138	<0.001	2.564
Z	1	0.882	0.419	4.436	0.035	2.416

3.8.4 Extension 3: Dichotomizing Distance at *c*

Distance was dichotomized using the model as opposed to a pre-specified value for c=180 miles, replacing Z by *I(Z>c)*, and extending the approach in Liu *et al* (2008) to estimate the change point c in addition to τ. The hazard function used was:

$$\lambda(t; Z, X) = \lambda_0(t) \exp[\{\beta + \theta I(t \geq \tau)\} I(Z > c) + \eta' X]$$

The likelihood function

$$L(\theta, \eta, \xi) = \prod_{i=1}^{n} \int_0^k \frac{\exp\{[\beta + \theta I(t \geq \tau)] I(Z_i > c) + \eta' X_i\}}{\sum_{j=1}^{n} Y_j(t) \exp\{[\beta + \theta I(t \geq \tau)] I(Z_j > c) + \eta' X_j\}} dN_i(t)$$

The log likelihood corresponds to

$$l(\theta, \eta, \xi) = \sum_{i=1}^{n} \int_0^k \left\{ \{[\beta + \theta I(t \geq \tau)] I(Z_i > c) + \eta' X_i\} - \log \sum_{j=1}^{n} Y_j(t) \{[\beta + \theta I(t \geq \tau)] I(Z_j > c) + \eta' X_j\} \right\} dN_i(t)$$

A two-step process was used. First estimates for η were obtained from a Cox PH model with no change point τ, fit with distance pre-dichotomized at 180 miles which seemed to be a reasonable estimate from the previous models. These estimates were $\hat{\gamma}_1 = 0.789$ (for HCV) and $\hat{\gamma}_2 = 0.162$ (for HCC). The values were used as initial values for obtaining estimates for all six parameters simultaneously, which were $\tau, c, \beta, \theta, \gamma_1, \text{and } \gamma_2$.

The log-likelihood function was optimized using the Nelder-Mead, a variable metric algorithm, and a simulated annealing algorithm. The algorithms were used to obtain estimates for $\tau, c, \beta, \theta, \gamma_1, \text{and } \gamma_2$ simultaneously. Convergence was achieved using all three methods and the results were presented in table 25 below.

Table 25: Resulting Estimates for Extension 3

	Initial Values	Nelder-Mead	SANN	BFGS
T	3.66	3.622	4.082	3.474
C	180	179.767	184.691	180.085
Θ	0	1.144	1.455	0.965
B	0.989	0.909	0.960	0.855
γ_1	0.789	0.764	0.827	0.791
γ_2	0.162	0.084	0.171	0.162
Number of function calls		523	10000	50
Gradient		NA	NA	11
Improvement in log-likelihood		1.211	1.300	1.092

The change point in time ranged from 3.474 and 4.082, while the distance cutoff ranged from 179.767 and 184.691. Since these values did not deviate significantly from the initial values, I used different carefully chosen (table 26) as well as random (table 27) starting values and was able to confirm that these estimates maximized the log-likelihood. I found these constituted a global maximum, not just a local one. Although it appeared that the change point in the hazard function τ could not be established with BFGS and SANN, the distance cutoff using SANN was still close to 180 miles, confirming the distance cutoff, and BFGS was not far off either. This led to reinforcing the conclusion that the choice of initial values is critical, and had to be made carefully.

Table 26: Extension 3 with Different Initial Values

(a)

	Initial values	Nelder-Mead	BFGS	SANN
T	3	4.057	NA	NA
C	100	139.591	179.87	180.625
Θ	2	1.142	-0.157	5.057
B	1	0.214	0.987	1.007
γ_1	0.8	0.781	0.789	0.794
γ_2	0.16	0.153	0.162	0.158

(b)

	Initial values	Nelder-Mead	BFGS	SANN
T	4	4.093	3.997	0
C	200	192	200	183.217
Θ	1	4.649	1	0.41
B	0	-1.89	0	1.018
γ_1	0.8	0.876	0.8	0.771
γ_2	0.16	0.16	0.16	0.168

In the scenario in table (a) where the initial values were below what we expect, BFGS and SANN did not converge to anything for distance c. In the scenario in table (b) where the initial values were above what we expected, SANN did not produce an estimate for the change point in time.

Next I used the uniform distribution to generate random initial values with the ranges indicated in table 27 below. The log-likelihood was maximized and the resulting end values were displayed in table 27 below. None of these three algorithms were able to determine a change point in the hazard function when the initial values were generated randomly; they all converged to zero, which was the boundary of the search range.

Table 27: Parameter Estimates for Extension 3 Using Random Initial Values

	Initial Values	Nelder-Mead	BFGS	SANN
T	U (observed death times)	0	0	0
C	(0, 250)	55.833	140.088	183.262
Θ	(-2, 2)	-0.379	0.233	0.619
B	(-2, 2)	0.309	0.076	0.39
γ_1	(-2, 2)	0.78	0.786	0.768
γ_2	(-2, 2)	0.154	0.138	0.195

A grid search was performed over 20 values of τ between 2 and 3.8 and 20 distance values between 120 and 200. The grid was 20x20. The non-normalized score M, normalized score M* and the two parameter score M2P were calculated to account for both the change point and distance cutoff simultaneously, and the values of τ =2.758 and c=195.79 maximized this, although this did not reach statistical significance (p-value=0.118) using the Monte Carlo approach with 1000 repetitions.

Using a finer grid of 100x100 in the same range, the values for τ =2.455 and c=192.727 remained stable and the p-value was 0.146. The maximum value of the score was 7.469 which were also in agreement with the coarse grid (2 parameter score maximized at 7.469). Table 28 summarizes these results. The M_{2p} statistic consistently performed better than M and M*, producing lower p-values. The grid size does not affect the power as seen in section 3.10 below but both grid sizes were used to demonstrate in the table below. Further analysis of a 10x10, 20x20 and 30x30 can be seen in table 38.

Table 28: Score Values for Extension 3

		Results	Max Value		P-value	
Grid	Type of Score	(τ, c)	θ	β	θ	β
Coarse	2 Parameter	(2.758, 195.790)	7.469		0.118	
Coarse	Non-normalized		3.087	5.536	0.279	0.173
Coarse	Normalized		2.411	4.736	0.383	0.123
Fine	2 Parameter	(2.455, 192.727)	7.469		0.146	
Fine	Non-normalized		3.16	5.536	0.256	0.166
Fine	Normalized		2.437	4.736	0.39	0.155

Below in figure 26 is an image plot of values for τ and c. The fine grid (100x100) was used for this plot. The 3D plot in figure 27 was more informative for this approach but both are displayed here for completeness.

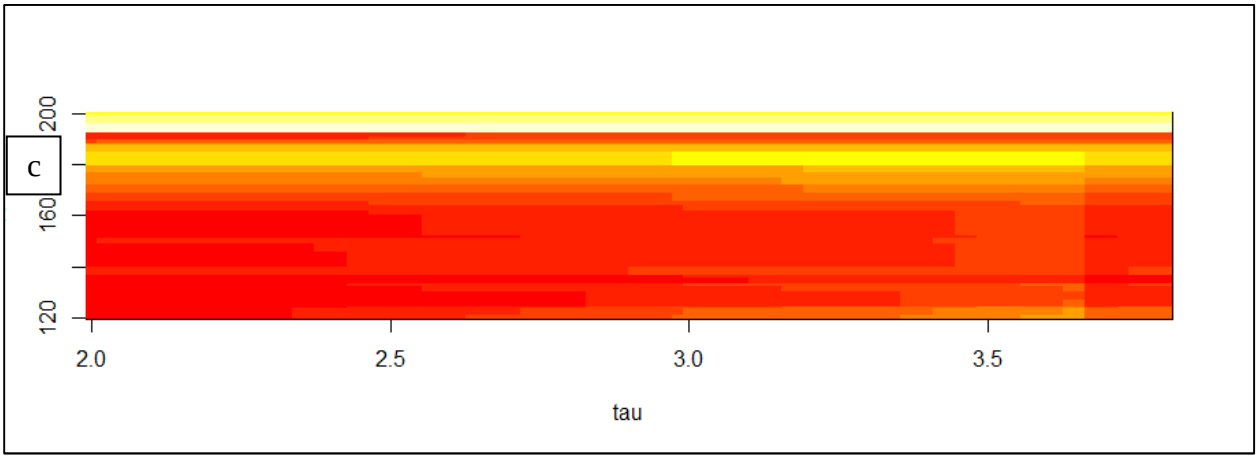


Figure 26: Image Plot of Extension 3

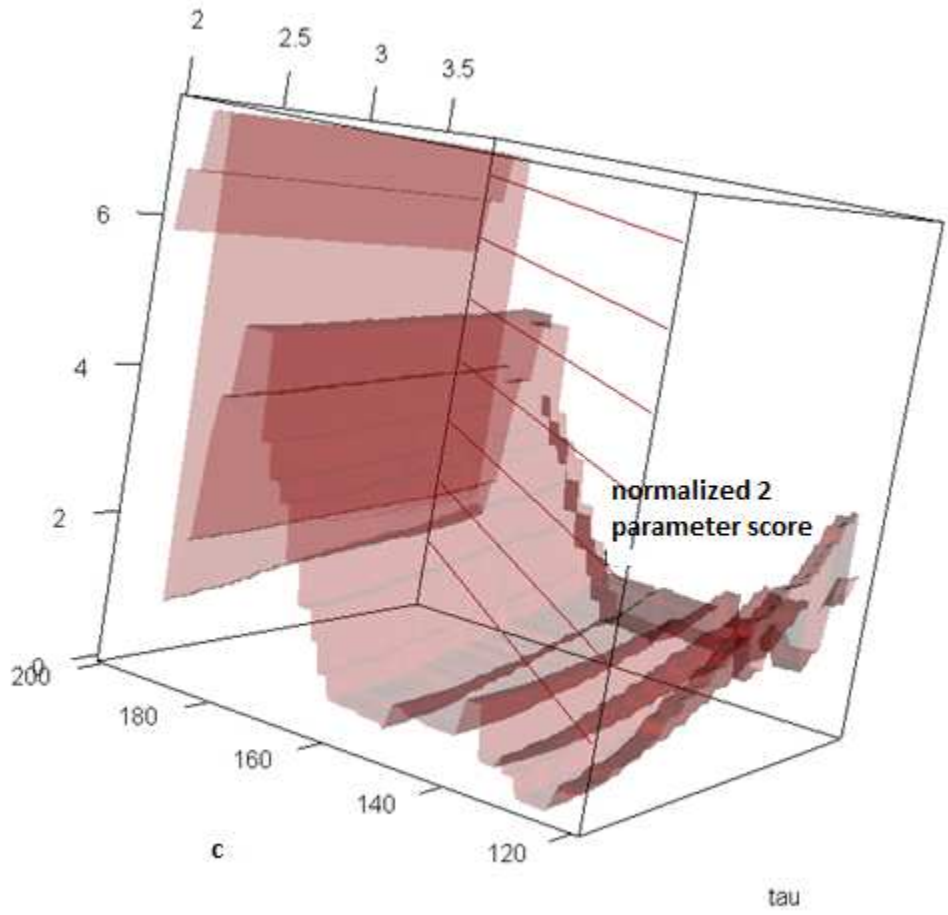


Figure 27: 3D Plot of Extension 3

3.8.5 Extension 4: Increasing Distance Effect

Distance Z was replaced by *max(Z-c,0)*, the positive part of the difference between distance and the cutoff point (Z-c). This meant there was no distance effect for distances less than c, but allowed the effect to increase for larger distances. The hazard function in this case was:

$$\lambda(t; Z, X) = \lambda_0(t) \exp[\{\beta + \theta I(t \geq \tau)\} \max(Z - c, 0) + \eta' X(t)]$$

Using the initial values: $\tau = 3.66$, c=180, $\theta = 0$, $\beta = 0.989$, $\gamma_1 = 0.789$, and $\gamma_2 = 0.162$, the parameter estimates in Table 29 were obtained.

Table 29: Increasing Distance Effect

	Nelder-Mead	SANN	BFGS
T	4.079	1.567	3.649
C	201.209	180.9	179.435
Θ	0.008	0.003	0.006
B	0.844	1.6	0.989
γ_1	0.079	0.824	0.79
γ_2	0.016	0.122	0.162
Number of function calls	149	10000	33
Gradient			5
Improvement in log-likelihood	3.535	251.166	3.483

The distance cutoff was consistently in the region of 180 miles using the variable metric algorithm (BFGS) and simulated annealing (SANN), while the Nelder-Mead approximation was higher at 200 miles. The SANN method yielded the best improvement in the log-likelihood out of all three methods used for optimization. Using different initial values, the results are presented in table 30 below where only the Nelder-Mead approach yielded the first set of initial values.

Table 30: Increasing Distance Effect with Different Initial Values

(a)

	Initial Values	Nelder-Mead
Increasing Distance Effect		
T	4	NA
C	200	219.698
Θ	1	0.569
B	0	-0.336
γ_1	0.8	0.78
γ_2	0.16	0.155
Number of function calls		105

(b)

	Initial Values	Nelder-Mead	BFGS	SANN
T	3	4.109	NA	2.936
C	200	201.14	NA	156.291
Θ	1	0.008	NA	0.002
B	0	0.386	0	0.043
γ_1	0.8	0.802	0.781	0.677
γ_2	0.16	0.014	0.156	0.076
Number of function calls		435	13	10000
Improvement in log-likelihood		12703.45	12699.92	12700.22

(c)

	Initial Values	Nelder-Mead	BFGS	SANN
T	4	4.2	NA	2.844
C	150	194.798	NA	93.307
Θ	1	-0.376	-0.057	0.002
B	0	0.45	0	0.208
γ_1	0.8	0.78	0.78	0.905
γ_2	0.16	0.156	0.156	0.078
Number of function calls		183	18	10000
Improvement in log-likelihood		6307.6	6307.495	6309.171

Nelder-Mead was the most appropriate method for this data, but BFGS and SANN were used for comparison purposes. Different initial values were used and Nelder-Mead approximation produced the most consistent results which was expected.

A grid search was performed over 20 values of τ between 2 and 3.8 and 20 distance values between 120 and 200. The two parameter score was calculated to account for both the change point and distance cutoff simultaneously, and the values of τ =4.237 and c=134.211 maximized this, although this did not achieve statistical significance (p-value=0.118) using the Monte Carlo approach with 1000 repetitions.

Using a finer grid of 100x100 the values in the same range for τ =3.727 and c=192.727 and the p-value was 0.161. The maximum value of the two parameter score was 5.082.

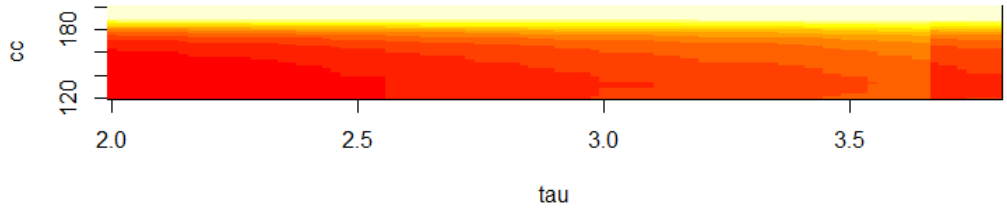
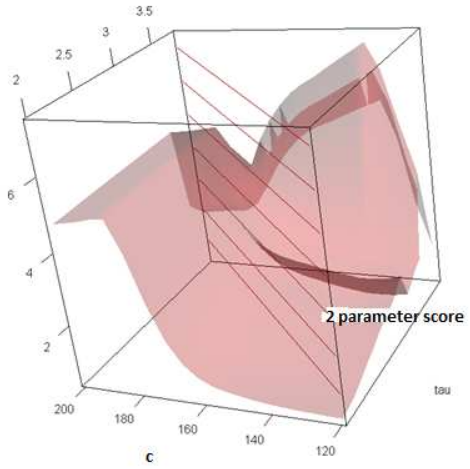
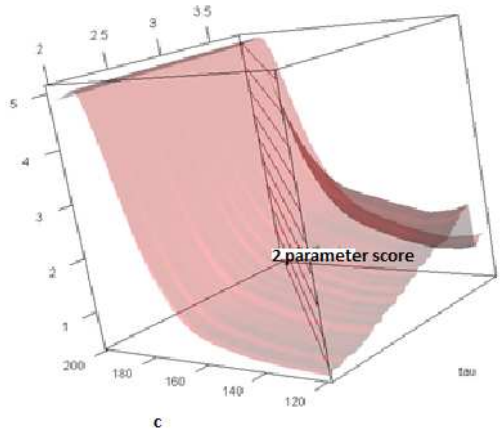


Figure 28: MC Approach of Two Parameter Score for Extension 4



(a) Coarse 20x20 grid



(b) Finer 100x100 grid

Figure 29: 3D Image Plots of Increasing Distance Effect

Table 31: Results for Score Function in Extension 4

		Results	Max Value		P-value	
Grid	Type	(τ, c)	θ	β	θ	β
Coarse	2 Parameter	(4.237, 134.11)	7.344		0.118	
Coarse	Non-normalized		228.677	94.566	0.287	0.792
Coarse	Normalized		4.237	0.906	0.173	0.569
Fine	2 Parameter	(3.727, 192.727)	5.082		0.161	
Fine	Non-normalized		233.101	94.832	0.3	0.789
Fine	Normalized		1.567	0.719	0.368	0.6

3.9 Simulation Results

This section presents simulation results for Extensions 3 and 4 described in Section 2.4. Two types of simulations were run, one using data generated from standard distributions and one with a situation which closely resembled the data set of LT patients. The purpose of the first simulation was to demonstrate that the methodology worked while the second simulation studies its performance in a situation resembling the LT data. The simulations used covariates Z and (X_1, X_2) with distributions described below.

3.9.1 Using Standard Distributions

The covariate X_1 was generated from a uniform distribution with range (0,1), and X_2 was generated from an exponential distribution with mean=1. The sample size was n=600, with 1000 runs of simulations, and the thresholds for 1000 simulated datasets and thresholds based on 1000 Monte Carlo resampling samples. Z was generated from a uniform distribution on (0.100) for extension 3 and (0,10) for extension 4. τ was chosen to be 0.25 and the distance cutoff was chosen to be 50. The grid was 20x20 and τ ranged from 0.15 to 0.30 while c ranged from 20 to 60. The simulation results are summarized below.

For all 3 types of statistics M, M* and M2P, the Monte Carlo resampling-based thresholds matched the empirical values reasonably well (Table 32). More specifically, the empirical

threshold corresponding to a nominal level α was estimated by the 100(1-α)th quantile of the sample test statistics from 1000 data sets simulated under the null hypothesis. The average Monte Carlo resampling-based threshold was defined as the mean of 1000 thresholds found by the Monte Carlo resampling approach. We considered α levels of 0.1, 0.05, and 0.01 to examine the tail approximation by the proposed Monte Carlo approach. The proximity between the empirical threshold and the Monte Carlo resampling-based threshold indicated that the proposed approach well approximated the distribution of the maximal score test statistics.

Table 32: Empirical and Resampling-based Thresholds at the Nominal Level α for Standard Simulations

			Sample		Empirical			Resampling		
Extension	(θ, β ,γ_1, γ_2)		Size		α=0.10	α=0.05	α=0.01	α=0.10	α=0.05	α=0.01
3	(0, 0, 0.78, 0.16)	sup |S|	600	θ	19.2718	20.9852	26.5965	19.0641 (0.6240)	21.3356 (0.7171)	25.7821 (1.0653)
				β	18.8528	21.2806	27.5332	18.9047 (0.6065)	21.3670 (0.7080)	26.0719 (1.0500)
		sup W		θ	6.3836	7.748	11.3765	6.1861 (0.2480)	7.6978 (0.3313)	11.0759 (0.6677)
				β	5.3341	6.829	10.6373	5.3102 (0.2175)	6.7268 (0.2994)	9.9088 (0.6346)
		M2P			9.8826	11.4541	16.395	9.5649 (0.3475)	11.2865 (0.4267)	15.0340 (0.7664)
4	(0, 0, 0.78, 0.16)	sup |S|	600	θ	19.4348	21.6797	26.0303	19.0203 (0.6707)	21.2987 (0.7822)	25.7586 (1.0815)
				β	19.4131	21.638	25.6503	18.8775 (0.6323)	21.3055 (0.7428)	26.0737 (1.1097)
		sup W		θ	6.6696	8.045	11.3061	6.1681 (0.2735)	7.6732 (0.3645)	11.0503 (0.6996)
				β	5.6271	7.1691	10.8428	5.3065 (0.2308)	6.7115 (0.3095)	9.9141 (0.6436)
		M2P			10.4484	12.2452	16.5111	9.5642 (0.3590)	11.2823 (0.4338)	15.0219 (0.7791)

In the 3 situations studied under extension 3 in table 33 below, the median estimated change point τ from 1000 replicated values was 0.2430, 0.2194, and 0.2285 in scenario 1, 2 and 3 respectively, ranging from 0.2194 to 0.2430. The median estimated change point from 1000 values, for the 3 scenarios were 0.2566, 0.2132, and 0.2289. Similarly the mean estimated distance cutoff c was 58.46, 47.39, and 50.34 in scenarios 1, 2 and 3 with distance cutoff values for 1000 replications ranging from 47.39 to 58.46. The median estimated distance cutoff was 60, 49.47, and 49.47. The null was excluded in these estimates given above.

Likewise in table 33, 4 scenarios were studied for extension 4. The mean estimated change points from 1000 values were 0.2197, 0.2195, 0.2155, and 0.2571 for scenarios 4, 5, 6 and 7 respectively. This ranges from 0.2155 to 0.2571 which captures the true value of 0.25. The median estimated change points from 1000 values for the 4 scenarios were 0.2132, 0.2132, 0.2053, and 0.2911. The mean distance of 1000 values in scenarios 4 through 7, were 38.96, 46.61, 46.72, and 55.87 with the mean increasing as the values of β and θ increased. The range was 38.96 to 55.87 which captured the true value of 50 miles. The median estimated cutoffs from 1000 values were 38.95, 51.58, 51.58, and 55.79.

Under the null hypothesis of no change points, the estimators of the change points were spread over the searched interval. When the change points actually existed under the alternatives, the time and distance points yielding the maximal score statistics successfully captured the true change-points of 0.25 years post-transplant and 50 miles as indicated by the median values of their estimators.

Table 33: Standard Simulation Results

				Recovered Value from Simulations				
Extension	Scenario	(β, θ, γ_1, γ_2)	Parameter	True value	Mean	Median	Minimum	Maximum
3	1	(0.02, 0.08,0.78, 0.16)	τ	0.25	0.243	0.2566	0.15	0.3
			c	50	58.46	60	41.05	60
3	2	(0.3, 0.3, 0.78, 0.16)	τ	0.25	0.2194	0.2132	0.15	0.3
			c	50	47.39	49.47	20	60
3	3	(0.6, 0.6, 0.78, 0.16)	τ	0.25	0.2285	0.2289	0.15	0.3
			c	50	50.34	49.47	20	60
4	4	(0, 0.002, 1, 0.5)	τ	0.25	0.2197	0.2132	0.15	0.3
			c	50	38.96	38.95	20	60
4	5	(0.008, 0, 1, 0.5)	τ	0.25	0.2195	0.2132	0.15	0.3
			c	50	46.61	51.58	20	60
4	6	(0.008, 0.002, 1, 0.5)	τ	0.25	0.2155	0.2053	0.15	0.3
			c	50	46.72	51.58	20	60
4	7	(0.15, 0.1,1,0.5)	τ	0.25	0.2571	0.2921	0.15	0.3
			c	50	55.87	55.79	49.47	60

Table 34: Type I Errors and Powers Using Resampling-based Thresholds for Standard Simulations

			Sample	Resampling sup |S|			Resampling sup W			Resampling sup 2 Parameter Score		
Extension	$(\theta, \beta, \gamma_1, \gamma_2)$		Size	α=0.10	α=0.05	α=0.01	α=0.10	α=0.05	α=0.01	α=0.10	α=0.05	α=0.01
3	(0, 0, 0.78, 0.16)	θ	600	0.106	0.042	0.014	0.105	0.055	0.01			
		β		0.1	0.051	0.012	0.105	0.055	0.01	0.113	0.052	0.014
3	(0.3, 0.3, 0.78, 0.16)	θ	600	0.907	0.84	0.653	0.897	0.819	0.632			
		β		0.919	0.862	0.682	0.906	0.846	0.669	0.823	0.752	0.536
3	(0.3, 0, 0.78, 0.16)	θ	600	0.205	0.136	0.039	0.196	0.122	0.033			
		β		0.194	0.122	0.029	0.19	0.119	0.032	0.196	0.124	0.04
3	(0.4, 0.5, 0.78, 0.16)	θ	600	0.782	0.626	0.267	0.712	0.508	0.174			
		β		0.725	0.573	0.232	0.725	0.545	0.182	0.553	0.381	0.165
3	(0.6, 0.8, 0.78, 0.16)	θ	600	0.817	0.67	0.296	0.702	0.504	0.145			
		β		0.843	0.703	0.344	0.81	0.643	0.276	0.664	0.425	0.187
4	(0, 0, 0.78, 0.16)	θ	600	0.108	0.061	0.009	0.117	0.06	0.011			
		β		0.117	0.054	0.009	0.114	0.06	0.011	0.141	0.074	0.013
4	(0.008, 0.002, 1, 0.5)	θ	600	0.566	0.438	0.222	0.551	0.423	0.204			
		β		0.613	0.488	0.264	0.594	0.472	0.259	0.482	0.346	0.159
4	(0.008, 0, 1, 0.5)	θ	600	0.107	0.059	0.01	0.111	0.054	0.01			
		β		0.12	0.056	0.01	0.114	0.06	0.009	0.137	0.073	0.024
4	(0, 0.002, 1, 0.5)	θ	600	0.555	0.435	0.214	0.539	0.418	0.206			
		β		0.606	0.493	0.244	0.588	0.476	0.233	0.466	0.341	0.164

3.9.2 Using Original Data Structure

The purpose of the second simulation was to use the distance, HCV and HCC status from the original data set to generate the survival times using a Cox PH model. The MC model in extensions three and four of this dissertation was used to see how well the parameter estimates could be recovered from this hybrid simulation using the real covariate values.

The coefficient for HCV was preset to 0.78 and the coefficient for HCC was 0.16 taken from the Cox model estimated under the null hypothesis of no change points. Beta and theta were set to

0.6 and 0.8 respectively for extension three, and 0.08 and 0.02 for extension four. The change point in time τ was set to 0.5 and the distance cutoff was 180. 1000 simulations with 1000 Monte Carlo repetitions were run on a 20x20 grid of time between 0.25 and 0.75 and distance of 150 to 200.

For extension three, the two parameter score function (M_{2P}) was maximized and was statistically significant. In the 4 scenarios in table 35, the mean estimated change point τ, calculated from 1000 estimated values, were recovered at 0.5448, 0.5479, 0.5291, 0.534. These ranged from 0.5291 to 0.5479, depending on the values of the parameters β and θ used, which captured the true value of 0.5. The median estimated change point for the 4 scenarios were 0.5658 and 0.5395 with 2 repetitions of each. The mean distance cutoff c from 1000 estimated values was recovered at 177.2, 176.4, 180.2, and 177.5, which range between 176.4 and 180.2 which are relatively close to the true values of 180. The median estimated change points from 1000 values were 178.99 (three times) and 186.8.

Similarly for extension 4 in table 35 in scenarios 5 through 8, the mean estimated change point τ, calculated from 1000 estimated values, were recovered at 0.5097, 0.4981, 0.4942, and 0.4774, ranging from 0.4774 to 0.5097 which captures the true value of 0.5. The median estimated change points from 1000 values were 0.4868 and 0.4605. The mean value for the distance cutoff c, estimated from 1000 replicated values for scenarios 5 through 8 were 165.1, 152.5 and 161.2 in scenarios 5,6 and 8 where the true value was 180. The median estimated distance cutoff from 1000 values which were generated were 162.9, 163.2 and 159.2. In scenario 7 a true value of 100 miles was used which allowed for more patients with distances above this value and it is noted that the mean recovered value was 94.11 and the median was 92.11 which was closer to the true value of 100 miles. Values for $\gamma_1 \text{and } \gamma_2$ were intentionally varied between 0.78 and 0.16 for one option, and 1 and 0.5 for the second option as seen in table 35 below.

Table 35: Simulations Resembling LT Data

				Recovered Value from Simulations				
Extension	Scenario	$(\beta, \theta, \gamma_1, \gamma_2)$	Parameter	True value	Mean	Median	Minimum	Maximum
3	1	(0.5, 0.7, 0.78, 0.16)	τ	0.5	0.5448	0.5658	0.25	0.75
			c	180	177.2	178.9	150	200
3	2	(0, 0.8, 0.78, 0.16)	τ	0.5	0.5479	0.5658	0.25	0.75
			c	180	176.4	178.9	150	200
3	3	(0.6, 0, 0.78, 0.16)	τ	0.5	0.5291	0.5395	0.25	0.75
			c	180	180.2	186.8	150	200
3	4	(0.6, 0.8, 0.78, 0.16)	τ	0.5	0.534	0.5395	0.25	0.75
			c	180	177.5	178.9	150	200
4	5	(0, 0.02, 1, 0.5)	τ	0.5	0.5097	0.4868	0.25	0.75
			c	180	165.1	162.9	150	185
4	6	(0.008, 0.002,0.78,0.16)	τ	0.5	0.4981	0.4868	0.25	0.75
			c	180	152.5	163.2	100	200
4	7	(0.001, 0.004,1,0.5)	τ	0.5	0.4942	0.4868	0.25	0.75
			c	100	94.11	92.11	50	150
4	8	(0.08, 0.02,0.78,0.16)	τ	0.5	0.4774	0.4605	0.25	0.75
			c	180	161.2	159.2	150	185

The LT data-based simulation produced rather poor results compared to the standard simulation with respect to distance in the fourth extension, probably due to the small number of patients living at longer distances.

For all 3 types of statistics M, M* and M_{2p}, the Monte Carlo resampling-based thresholds did not match the empirical values well (Table 36). M produced the best agreement, M* was rather poor and M2P did not perform well perhaps because of the small number of patients living at longer distances. I considered α level of 0.1, 0.05, and 0.01 to examine the tail approximation by the proposed Monte Carlo approach. The reasonable proximity between the empirical threshold and the Monte Carlo resampling-based threshold indicated that the proposed approach well approximated the distribution of the maximal score test statistics.

Table 36: Empirical and Resampling-based Thresholds at the Nominal Level α for Simulations Resembling the LT Data

			Sample		Empirical			Resampling		
Approach	$(\theta, \beta, \gamma_1, \gamma_2)$		Size		α=0.10	α=0.05	α=0.01	α=0.10	α=0.05	α=0.01
3	(0, 0, 0.78, 0.16)	sup |S|	821	θ	5.319	5.9912	7.4405	4.9572 (0.6752)	5.6746 (0.7627)	7.0773 (0.9430)
				β	5.5346	6.4579	8.2997	5.3186 (0.5623)	6.1400 (0.6482)	7.7493 (0.8381)
		sup W		θ	8.2162	10.945	16.834	5.7283 (0.5024)	7.1647 (0.5579)	10.4095 (0.8227)
				β	6.873	8.485	11.729	5.0819 (0.2782)	6.4699 (0.3354)	9.5737 (0.6331)
		M2P			20.9984	22.933	25.404	8.3956 (0.5691)	10.0184 (0.6250)	13.5376 (0.8878)
4	(0, 0, 0.78, 0.16)	sup |S|	821	θ	327.139	394.15	630.12	307.0252 (147.2931)	361.5709 (174.2347)	468.0873 (226.0252)
				β	338.818	410.84	596.79	334.9368 (136.9262)	397.7352 (162.2611)	519.7167 (212.2840)
		sup W		θ	5.5208	6.6549	8.8947	3.8259 (0.5114)	5.1094 (0.5965)	8.1263 (0.8526)
				β	4.5681	5.332	8.3525	3.3507 (0.2589)	4.5755 (0.3322)	7.5003 (0.6697)
		M2P			13.031	13.7843	14.8645	6.7186 (0.6905)	8.2969 (0.7559)	11.7940 (1.0121)

Under the null hypothesis of no change points, the estimators of the change points were spread over the searched interval. When the change points actually existed under the alternatives, the time and distance points yielding the maximal score statistics successfully captured the true change-points of 0.5 years post-transplant and 180 miles.

Although the methodology was successful in recovering the parameter values used to generate this data, the simulations using standard distributions produced better results than those resembling the LT data. I believe this is due to the small number of patients who lived beyond the 180 mile cutoff (n=19).

Table 37: Type I Errors and Powers Using Resampling-based Thresholds for Simulations Resembling the LT Data

			Sample	Resampling sup |S|			Resampling sup W			Resampling M2P		
Ext	$(\theta, \beta, \gamma_1, \gamma_2)$		Size	α=0.10	α=0.05	α=0.01	α=0.10	α=0.05	α=0.01	α=0.10	α=0.05	α=0.01
3	(0, 0, 0.78, 0.16)	θ	821	0.146	0.068	0.02	0.256	0.146	0.05			
		β		0.116	0.057	0.017	0.201	0.114	0.022	0.609	0.497	0.351
3	(0, 0.6, 0.78, 0.16)	θ	821	0.214	0.113	0.019	0.168	0.065	0.011			
		β		0.255	0.141	0.026	0.224	0.109	0.015	0.4	0.248	0.149
3	(0.7, 0.5, 0.78, 0.16)	θ	821	0.690	0.49	0.145	0.531	0.330	0.068			
		β		0.683	0.515	0.184	0.638	0.439	0.125	0.583	0.359	0.201
3	(0.8, 0, 0.78, 0.16)	θ	821	0.356	0.202	0.029	0.241	0.098	0.007			
		β		0.311	0.154	0.027	0.251	0.116	0.013	0.585	0.434	0.326
3	(0.8, 0.6, 0.78, 0.16)	θ	821	0.817	0.67	0.296	0.702	0.504	0.145			
		β		0.843	0.703	0.344	0.81	0.643	0.276	0.664	0.425	0.187
4	(0, 0, 0.78, 0.16)	θ	821	0.175	0.105	0.029	0.21	0.105	0.017			
		β		0.156	0.093	0.022	0.202	0.085	0.018	0.419	0.305	0.141
4	(0, 0.08, 0.78, 0.16)	θ	821	0.326	0.18	0.018	0.215	0.092	0.008			
		β		1	0.993	0.143	1	0.98	0.074	0.747	0.415	0.046
4	(0.02, 0, 0.78, 0.16)	θ	821	0.53	0.142	0.002	0.125	0.029	0.002			
		β		0.53	0.149	0.002	0.473	0.093	0.001	0.344	0.264	0.126
4	(0.4, 0.6, 0.78, 0.16)	θ	821	0.5	0.340	0.090	0.373	0.198	0.029			
		β		0.507	0.352	0.111	0.466	0.287	0.071	0.508	0.331	0.201
4	(0.001, 0.004,1,0.5)	θ	821	0.433	0.274	0.063	0.357	0.198	0.033			
		β		0.408	0.263	0.072	0.388	0.229	0.047	0.325	0.191	0.090
4	(0.08, 0.02, 0.78, 0.16)	θ	821	0.42	0.249	0.039	0.311	0.15	0.007			
		β		1	0.995	0.224	1	0.992	0.122	0.838	0.526	0.075

In future studies with more patients living beyond the critical cutoff point, this will no longer be an issue and the two types of simulations will produce comparable results.

3.10 Varying Grid Size

Next we explored the grid size to see if there was any impact on the power and we found that there was none.

Table 38: Grid Size

			Resampling sup |S|			Resampling sup W			Resampling sup 2 Parameter Score		
Grid size		Sample Size	α=0.10	α=0.05	α=0.01	α=0.10	α=0.05	α=0.01	α=0.10	α=0.05	α=0.01
10	θ	600	0.9	0.85	0.647	0.888	0.838	0.621			
	β		0.919	0.868	0.705	0.912	0.857	0.685	0.845	0.747	0.538
20	θ	600	0.906	0.835	0.655	0.890	0.822	0.643			
	β		0.923	0.870	0.699	0.916	0.853	0.688	0.819	0.735	0.538
30	θ	600	0.913	0.855	0.667	0.9	0.838	0.655			
	β		0.931	0.883	0.685	0.923	0.868	0.67	0.839	0.75	0.537

$\tau = 0.251, c = 50, \beta = 0.3, \theta = 0.3, \gamma_1 = 1, \gamma_2 = 0.5$

The grid size did not impact the power and this was demonstrated using various grid sizes (10x10, 20x20 and 30x30) while keeping the parameters under the alternative constant as seen in table 37.

Chapter Four

Conclusion

4.1 Summary of Results

My study has demonstrated using extensive methodologies, that those patients who lived outside of the 180 mile radius from TGH had a higher mortality rate during the first 5 years post-transplant. Using the AIC/ML approach applied to repeated Cox PH models, calculated for every possible distance cutoff point (from 1 to 400 miles); it was established that 180 miles was where the AIC was minimized. This corresponded to a driving distance of approximately three hours. It was worth investigating further with other models to confirm the effect of distance on survival post-transplant.

KM curves revealed higher mortality beyond 180 miles at five years post-transplant. Next, Cox PH and AFT models adjusted for HCV and HCC, confirmed a distance effect on survival at 180 miles. LR at one year post-transplant confirmed that 180 miles was also a significant distance cutoff. There are limitations with LR, which discards valuable information by ignoring the length of patient survival and reducing outcomes to a dichotomous variable at a particular time point. Models beyond one year post-transplant were calculated on a reduced dataset due to extensive censoring. Therefore, greater emphasis was put on the findings of the Cox PH and AFT models.

The reason for adjusting for both HCV and HCC was that these diseases cause a higher mortality due to disease recurrence. Patients with HCV and HCC have worse survival post-transplant which was evident in this study. The effect of distance at 180 miles was more pronounced in patients with HCV (HR of 3.72 for 434 HCV patients vs. 2.68 for total 821 patients) and even

stronger in patients with HCC (HR of 5.24 for 134 HCC patients). 98 (12%) patients had both HCV and HCC, as HCV often precedes HCC in liver disease progression.

In this dissertation, I extended current statistical methods for the use of Monte Carlo to establish a p-value for the change point in the hazard function and incorporated both a change point in time and a change point due to a continuous variable, in this case distance. In this dataset of LT recipients, the new model was applied to simultaneously dichotomize distance and establish a point in time where there was a change in the hazard function.

Using the approach proposed by Liu et al. (2008), the non-normalized score function pointed to a change point at 3.622 years post LT but failed to reach statistical significance (p-value=0.566). The normalized score pointed to a change point at 4.186 years post LT but failed to yield statistical significance (p-value=0.235).

My first extension used distance dichotomized at 180 miles and allowed one to drop β, indicating that there was no throughout distance effect. With this model, the distance effect was only present prior to a specific point in time which was established at 3.622 years post-transplant, with a hazard ratio of 0.011, indicating that the hazard of death was lower for those patients living within 180 miles in the first 3.6 years post-transplant. Neither the non-normalized score function nor the normalized score function reached statistical significance with p-values 0.573 and 0.543 respectively.

My second extension enabled reversal of the inequality sign to allow for an additional effect of distance dichotomized at 180 miles, on the hazard beyond a certain time point which was more intuitive for this data. This was found to be 3.663 years post-transplant but failed to reach statistical significance for both the non-normalized and normalized score function with p-values 0.789 and 0.642 respectively.

With the third extension I was able to allow the data to determine the best cutoff for distance while simultaneously detecting the change point in the hazard function. This was done by maximizing the log-likelihood over six parameters $\tau, c, \beta, \theta, \gamma_1, \text{and } \gamma_2$ using the Nelder-Mead, BFGS and SANN optimization methods. The various methods were in close agreement over the distance cutoff at 180 miles and a change point in the hazard between 3.5 and 4 years post-transplant depending on the optimization method used. Using the Monte Carlo approach with a fine grid of 100x100, the change point was established at 2.455 years post LT and the distance cutoff was at 192.727 miles (p-value 0.146).

The fourth approach allowed for an incremental addition to the hazard beyond a mileage cutoff point found at around 180 miles. This implied that for every additional mile beyond 180, there was an incremental impact on survival. The change point in the hazard function was between 3.6 and 4 years post-transplant. Various different initial values were used in the optimization for extensions three and four but the distance effect remained stable and close to 180 miles, confirming what the previous models found. Using the Monte Carlo approach with a fine grid of 100x100, the change point was established at 3.727 years post LT and the distance cutoff was at 192.727 miles (p-value=0.161).

The models studied were consistent in determining that the distance cutoff in survival was 180 miles, although some failed to reach statistical significance. As explained in the results section, I explored various other distance cutoffs within each method and none reached the significance level of the 180 mile cutoff. Extensions three and four in the MC approach provided a novel method to determine the distance cutoff.

The best model for this data was the third approach

$$\lambda(t; Z, X) = \lambda_0(t) \exp[\{\beta + \theta I(t \geq \tau)\} I(Z > c) + \eta' X(t)]$$

which dichotomized distance Z at a point c and then estimated the change points c and τ. This model pointed to a distance effect beyond 192.727 miles with a change point in the hazard due to this, after 2.455 years post-transplant.

Extensive simulations using both standard simulation techniques and a hybrid simulation resembling the LT data, proved that these new approaches work for dichotomizing a continuous variable and finding a point beyond which there was an incremental effect from this variable. Various values of β and θ were used and the median of the recovered values for τ and c, were very close to the true values.

4.2 This Study in Context

Studies in the past have been significantly smaller and conducted survival analysis using a small sample size of 66 LT patients with very few of these at longer distances (Firozvi et al., 2008). That study also calculated survival at one year post-transplant, a time period too short to evaluate the multiple difficulties that the patients may encounter in their follow-up course. These difficulties include, but are not limited to rejections, infections, HCV or HCC recurrence, medication compliance, and other complications of LT.

My study was more comprehensive since I had 821 patients, over the span of 16 years, at longer distance, and with a follow-up of up to 5 years after LT. I excluded patients who had relocated temporarily to be closer to the LT center for purposes of receiving a transplant so this significant bias present in other studies was eliminated. These patients had a transplant center or specialist caring for them near their original residence so TGH was not solely responsible for their follow-up care, and their inclusion would have biased the study. I decided to focus on patients within the state of Florida, where driving distances can be as long as 7 hours for those patients driving from the panhandle. Those with acute liver failure were also excluded, as they had a high early mortality and a different set of circumstances regarding access to transplant centers and compliance. This methodology was not described in other studies (Firozvi et al., 2008). Also

excluded were those with early deaths during their LT hospitalization, as they were not impacted by distance, travel time, or post-transplant care. Furthermore, adjustment was made for important covariates, such as for the presence of HCV and HCC which are diseases most likely to recur. I made a conscious effort to eliminate bias, which I believe has strengthened this study.

Using this unique and extensive methodology, my study was able to detect a distance effect on post-transplant survival beyond 180 miles which other studies were previously unable to find. The MC approach provided additional information that there was a change point in the hazard with an additional distance effect after 3.6 years post-transplant, a factor that had not been studied before in this patient population. The Nelder-Mead approximation, a variable metric algorithm (BFGS), and a simulated annealing algorithm (SANN) were used to simultaneously estimate $\tau, c, \beta, \theta, \gamma_1, \text{and } \gamma_2$. This was necessary since the log-likelihood was discontinuous at τ and c, and therefore not differentiable.

4.3 Contribution

This is the first study to show that the patient's distance from the transplant center can affect survival after the LT. My study is more comprehensive, because it included a large patient pool over the span of 16 years, and the use of multiple statistical models and approximation methods to confirm the findings. I utilized AIC, Kaplan Meier, Cox PH, Accelerated Failure Time models, Logistic Regression and a Monte Carlo approach to change point detection.

The Monte Carlo approach to change point detection and continuous variable dichotomization developed here is applicable to dichotomizing any continuous variable in multiple therapeutic areas. This is a new approach to change point detection that had not been in use before. My fourth approach also allows for an incremental effect beyond a certain cutoff point which is a novel approach.

My study adds to the existing literature and clarifies the role of distance on post-transplant survival for those undergoing LT, a topic somewhat controversial, tightly interwoven to concepts of liver allocation, transplant centers and designations of centers of excellence, post-transplant care, and overall utility of transplants in selected populations.

4.4 Findings

This dissertation is the first of its kind to study the effect of distance five years post LT, since previous papers looked at shorter time periods. It is also the first study to consider a change point in the hazard function due to distance and to demonstrate such a data driven approach to dichotomizing distance using the MC approach.

The findings here indicated that travelling longer distances was detrimental to patient survival which makes a strong argument against fewer but larger transplant centers. Studies in the past conducted survival analysis using a small sample sizes (66 patients) with very few of these at longer distances (Firozvi et al., 2008) and followed for a shorter period of time. This study was more comprehensive in that there were a total of 821 patients over the span of 16 years, at longer distance, and with a follow-up of five years after LT. As the MC approach indicated there is a detrimental effect due to distance beyond 180 miles, and this effect is stronger after 3.6 years post-transplant. Other studies would have missed this since they did not follow patients that long. I recognize that this data would carry more weight if it had a more even distribution of distant patients but the geography of Florida is such that most of the population is concentrated in the larger cities. Given that Tampa Bay Area is considered a large metropolitan region with many nearby suburban and rural areas, 180 miles does approximate to 3 hours of driving. For other LT centers that are located in urban areas, this may not be the case.

Minimizing the AIC as previously described, lead to a distance cutoff of 180 miles. Using Cox PH and AFT, after adjusting for HCV and HCC, I found that patients living beyond 180 miles from the LT center had worse survival in a time period of five years post LT. LR at one year post LT also

revealed that 180 miles was a significant distance cutoff. HCV and HCC impacted survival post LT and were clearly noticeable in this study. Patients living beyond 180 miles had significantly worse survival. This is crucial information because the medical community can monitor this subgroup more aggressively to improve survival in the future. In the case of patients with HCV, the effect of distance on survival post LT appeared to be worse.

There was clearly a prejudicious effect of distance on survival which was not previously confirmed on post LT patients. When dichotomizing distance at 180 miles, those living further had a significant disadvantage in survival (Log-Rank test statistic (p-value=0.0049), Wilcoxon (p-value=0.0077) and the Likelihood Ratio Statistic (p-values=0.0154). Logistic Regression 1 year post LT confirmed that 180 (p-value=0.0292) was significant. This is the first study to demonstrate that the patient's distance from a LT center can affect survival post LT.

The five methods studied (KM, Cox PH, AFT, LR, and MC approach) were consistent in determining that the cutoff in survival was 180 miles, all of which were statistically significant. I explored various other distance cutoffs within each method and none reached the significance level of the 180 mile cutoff.

To limit biases, I excluded patients who had relocated temporarily to be closer to the LT center. This would invalidate any study of the effect of distance since these patients get transferred back to their original institutions. Patients with multiple organ transplants, acute liver failure, and death on the same date as surgery were also excluded, as they had a different course of recovery. Patients who died in the first 30 days (n=22) post LT could also be excluded since they did not have a sufficiently long follow-up, and hence may not be affected much by traveling distance. Therefore, I also reran the models excluding patients with early death, and the hazard ratio increased from 2.68 to 3.15 because of the fact that all 22 of these patients lived within 180 miles of the LT center. This implied that even though patients who lived closer had higher early

mortality, including them did not change the finding that distant patients had higher mortality, and excluding them actually further supports my hypothesis.

4.5 Implication of Findings in Public Health

This research is unique because it required substantial modification of standard methods in survival analysis to incorporate dichotomization and the presence of change points in the hazard function due to distance. I extended and improved upon the Monte Carlo method for change point detection introduced by Liu et al (2008) to accommodate both a change point in the hazard function due to distance at a point in time, and the dichotomization of distance (a continuous variable) with a potential incremental effect. This recently developed method was not yet in common use and described only in a journal article in Statistics in Medicine in 2008, yet I extended it further to accommodate this dataset and the hypothesis that I was trying to prove. A more complicated null hypothesis was employed than that of Liu et al (2008), to simultaneously dichotomize distance at a point *c* and locate the time τ such that an enhanced distance effect is observed for times greater than τ using a Monte Carlo (MC) approach. The MC approach for change point detection that was developed in my dissertation is valuable in multiple scenarios since it is applicable for dichotomizing other biomarkers to predict survival or time to recurrence/graft failure in several other therapeutic areas apart from liver transplants.

By showing that patients at a longer distance exhibit poor survival, one can suggest that patients should be transplanted at the nearest transplant center. Distance also plays a role in how organs could be distributed. Indicating that results were worse at long distances, policy makers could argue that allocation could be improved with multiple smaller LT centers as opposed to fewer larger ones. It is also beneficial for health insurance companies to recognize that travelling a longer distance hinders survival so patients should be encouraged to seek treatment at the nearest LT center regardless of insurance-hospital contracts. From a patient perspective, knowing that distance is a major obstacle to longer survival post-transplant could prompt them to move closer to the center.

To conclude, by using a larger pool of patients and more extensive methodology, I was able to give more strength to this study compared to previous ones found in the literature. Methodology that involved application of AIC to Cox PH and MC to establish the presence of change points in the hazard had very recently been developed and has been extended further here. The MC method that I developed was specially adapted to fit the needs of this dataset, specifically to simultaneously estimate the dichotomization point for distance and the change point in the hazard function, which makes this study unique. My new methodology is valuable in that it is the first of its kind to simultaneously dichotomize a continuous variable, in this case distance, and establish a change point in the hazard function attributed to this variable. Furthermore in the fourth extension, the model allows for an incremental effect of distance beyond a certain cutoff point which was previously not possible to model and obtain the statistical significance for this term. Extensive simulation studies (in chapter 3) illustrate the value of this method. This methodology has numerous applications in a countless therapeutic areas involving a vast array of possible biomarkers.

4.6 Credibility

Bias was minimized by ensuring that both study groups had similar mean age, gender, and race distribution. However, I found that mean MELD scores were significantly higher for patients who lived closer, even though they had lower mortality post LT. High MELD scores (≥21) have been an independent predictor of increased mortality (Heuman et al., 2004). This finding was of interest because it implied that even though patients who lived closer had markers of increased mortality prior to LT, they were not more likely to die after the LT. This further supports that distance effect on mortality was very prominent in this study. However, the prevalence of HCC and HCV did not differ significantly between the two groups. Again, even though distant patients had lower MELD scores indicating a better survival prognosis, they still had increased mortality after LT. The models were adjusted for HCV and HCC in order to eliminate their impact on post LT survival. I demonstrated that these covariates did affect survival post LT. The effect of distance was more

pronounced in patients with HCV (HR of 3.72 for 434 HCV patients vs. 2.5 for total 821 patients). The distance effect was even stronger in patients with HCC (5.24 for 134 HCC patients).

Distance had a deleterious effect on patient survival post LT. This was shown using multiple different statistical modeling techniques with the MC approach which I developed, being new. Due to poor survival, allocation of resources may be needed for this distant population, especially those with HCV and HCC. This study provides the data to support this need. Such resources could include but not limited to, post LT HCV therapy, comorbidities, satellite clinics, mobile units and in general availability of specialists in distant areas deserve further study. This research has indicated that patients living beyond 180 miles from the LT center needed to be followed closely, or timely follow-up at least attempted, so that their medical care is improved. Ideally, one would want strategic partners caring for those patients far away that also have an interest in their progress. If this care does not exist, then their health may suffer.

4.7 Limitations

This study was retrospective, single center and regional bias existed. A larger prospective study would be beneficial in the future, where multiple additional covariates could be collected and adjusted for and plans for this are currently underway. Power is the most important limitation in this study. The small fraction of 19 patients living beyond 180 miles makes it impossible to find statistically significant differences in some of the more complex models. Work status information was incomplete in this retrospective data and therefore could not be studied. The geography of Florida differs to other regions and the lack of mass transit in the state may have caused a different distance effect as compared to other centers.

In this study patients who had relocated temporarily to be closer to the LT center were excluded, since their postoperative travel for follow-up would be altered. The center of residential zip code area was used to approximate the patient's distance to TGH's LT center. I recognize that zip code areas are established by the U.S. Postal Services and could change over the years to

expedite mail delivery. This conceptual area tends to be larger in rural settings. Thus, for the 19 patients who lived over 180 miles from the TGH LT center, I have found that the center of their zip code areas was not a large mileage discrepancy from their physical addresses as zip codes in this area are relatively small. TGH's LT center is located in downtown Tampa, FL, and since Tampa is a part of a large metropolitan region, patients who lived within 180 miles should not have a large zip code area. With this in mind I do not believe that using zip code instead of the full address has compromised my estimate of distance and I believe this study to be accurate.

I did not examine potential difficulties that patients at long distances encountered, such as compliance with appointments and medications. Difficulties that the patients may encounter in their follow-up course include, but are not limited to rejections, infections, HCV or HCC recurrence, medication compliance, and other complications following LT. Even though my study attempted to minimize bias by adjusting for HCV and HCC, as well as making appropriate exclusions, there could be other confounding variables that could affect survival post LT, such as co-morbidities, BMI, perioperative parameters, insurance coverage, socioeconomic status, level of education, or availability of nearby specialists. These factors have been debated regarding disparities in allocation and distribution which affects access and mortality prior to LT, but it is possible that these variables could also affect outcome post LT. In the literature review in chapter one, I listed examples of these disparities and discussed them.

4.8 Using Results to Solve Problems

It was hypothesized that longer distance between the patient's residence and the LT center results in worse outcome post LT. Since this hypothesis was proven to hold true, LT centers should make an effort to provide additional resources, such as finding temporary housing closer to the LT center, scheduling follow-up visits and tests on the same day, offering blood tests in local laboratory centers, fostering more frequent communication, recommending nearby specialists, or even providing satellite clinics.

Possible explanation for poor survival in the distant group, which were not studied here but will be considered in future research, included the inconvenience of travelling a longer distance for follow-up, the expense incurred by the patient when having to travel so far, transportation issues since mass transit and railroad systems are practically non-existent in Florida, fewer follow-up visits as it is suspected that those living further away were more likely to cancel, and the lack of primary providers or specialists in the patient's area.

It is evident that patients living more than 180 miles from the liver transplant center may need to be followed closely. If post LT coordinators are aware of the distance effect they can follow those patients more aggressively to ensure that they receive the necessary follow-up care to maintain a healthy graft.

Health insurance companies should be aware that distance hinders survival, so patients can be encouraged to seek treatment at the nearest LT center. This is not always done and due to insurance contracts some patients are forced to drive long distances to receive LT and follow-up treatment.

4.9 Future Research

There is a fundamental gap in understanding which socioeconomic factors affect long term survival of LT. LT is the only therapy currently available for patients with end-stage liver disease. Approximately 4,000 liver transplants are done yearly, but waiting lists are lengthening and a shortage of organs has caused an increase in deaths among patients waiting for LT. After LT, the survival rate appears to be lower for African-Americans than for Caucasians. Continued existence of this gap poses an increasingly difficult problem in the LT community as allocation of organs with limited supply becomes difficult to establish. The long term goal is to better understand how socioeconomic factors affect access to LT and survival post LT. I plan to design a Socioeconomic Predictive Model to predict survival post LT based on socioeconomic and clinical factors with both a retrospective and a prospective component.

I intend to further refine this study, by adding more patients to the dataset, hopefully having a greater distribution of patients outside of 180 miles, adjusting for more confounding variables, obtaining multicenter and multiregional data (discussions with UNC are underway), and by acquiring the necessary software to enable using the entire physical residential address as opposed to just the zip code.

By acquiring a larger sample size, I can randomly divide the data into two groups and carry out the testing procedure and estimation separately on the two independent subsets. Given that I had 821 patients with 19 living beyond 180 miles this was not possible here.

Another topic of interest to me is the association between the economic recession and the distant patients' outcomes, socioeconomic variables interacting with distance and other factors not studied here such as compliance, difficulty maintaining appointments, difficulty maintaining HCV treatment regimens, and lack of specialty providers.

Simulation studies would aid in establishing the properties of using the AIC criterion applied to repeated Cox PH models to dichotomize a continuous variable. More covariates can be adjusted for than what was available here.

The research group at TGH and I, are currently working together to study the effects of BMI on complications measured by the Clavien-Dino classification scale post LT. The hypothesis is that higher BMI will lead to more severe post LT complications, shorter graft survival and patient survival. In addition we have designed a prospective study to assess the quality of life of HCC patients undergoing LT.

4.10 Conclusion

My study has shown that patients who lived farther than 180 miles from a LT center had a higher

mortality after LT than those who lived closer. Therefore, LT centers should consider this information during initial evaluation of patients and provide additional resources to assist these patients, especially those with HCV. This research has indicated that patients living more than 180 miles from the transplant center may need to be followed closely, so that their medical care is improved. Ideally, one would want strategic partners caring for those patients far away that also have an interest in their progress. If this care does not exist, outcomes will be poor.

It has also inspired a variety of difficult questions: Should patients be transplanted at their nearest transplant center? Should distance play a role in how organs are distributed? Should insurance companies send patients to distant transplant centers of excellent quality when the outcome, in certain populations, is poor? Should aggressive centers transplant difficult patients from far away? Should the trend be for fewer, larger centers? My study provides a partial answer to some of these questions.

References

Albert, A., & Anderson, J. A. (1984). On the existence of maximum likelihood estimates in logistic regression models. *Biometrika, 71*(1), 1-10. doi: 10.1093/biomet/71.1.1

Allison, P. D. (2010). *Survival analysis using SAS: A practical guide*: Sas Inst.

Alsina AE, Antúnez I, González H, Bowers V, Leone J, Potter T (2009). Outcome of Liver Transplantation in a Hispanic Population: 100 Liver Transplants in Puerto Ricans. *Puerto Rico Health Sciences Journal, 28*(4).

Axelrod, D. A., Guidinger, M. K., Finlayson, S., Schaubel, D. E., Goodman, D. C., Chobanian, M., & Merion, R. M. (2008). Rates of solid-organ wait-listing, transplantation, and survival among residents of rural and urban areas. *JAMA: The Journal of the American Medical Association, 299*(2), 202-207.

Barritt, A. S. t., Telloni, S. A., Potter, C. W., Gerber, D. A., & Hayashi, P. H. (2012). Local access to subspecialty care influences the chance of receiving a liver transplant. *Liver Transpl*. doi: 10.1002/lt.23588

Bélisle, C. J. (1992). Convergence theorems for a class of simulated annealing algorithms on Rd. *Journal of Applied Probability*, 885-895.

Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika, 52*(3), 345-370.

Breslow, N. (1974). Covariance analysis of censored survival data. *Biometrics*, 89-99.

Broyden, C. G. (1970). The convergence of a class of double-rank minimization algorithms 2. The new algorithm. *IMA Journal of Applied Mathematics, 6*(3), 222-231.

Chang, I. S., Chen, C. H., & Hsiung, C. A. (1994). Estimation in change-point hazard rate models with random censorship. *Lecture Notes-Monograph Series*, 78-92.

Collett, D. (2003). *Modelling survival data in medical research* (Vol. 57): Chapman & Hall/CRC.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 187-220.

Di Bisceglie, A. M. (1997). Hepatitis C and hepatocellular carcinoma. *Hepatology, 26*(S3), 34S-38S. doi: 10.1002/hep.510260706

Durrleman, S., & Simon, R. (1989). Flexible regression models with cubic splines. *Statistics in Medicine, 8*(5), 551-561. doi: 10.1002/sim.4780080504

Eriksson, S., Eriksson, K.-F., & Bondesson, L. (1986). Nonalcoholic Steatohepatitis in Obesity: A Reversible Condition. *Acta Medica Scandinavica, 220*(1), 83-88. doi: 10.1111/j.0954-6820.1986.tb02733.x

Firozvi AA, L. C., Hayashi PH. (2008). Greater travel time to a liver transplant center does not adversely affect clinical outcomes. *Liver Transplantation, 14*(1), 18-24.

Fleming, T. R., & Harrington, D. P. (2011). *Counting Processes and Survival Analysis*. New York: Wiley.

Freeman, R. B., Wiesner, R. H., Harper, A., McDiarmid, S. V., Lake, J., Edwards, E., Teperman, L. (2002). The new liver allocation system: Moving toward evidence-based transplantation policy. *Liver Transplantation, 8*(9), 851-858. doi: 10.1053/jlts.2002.35927

Gharibvand L, J. D., Liao S. (2008). Evaluation of a Hospice Care Referral Program Using Cox Proportional Hazards Model. *SAS Proceedings*.

Heinzl, H., & Kaider, A. (1997). Gaining more flexibility in Cox proportional hazards regression models with cubic spline functions. *Computer Methods and Programs in Biomedicine, 54*(3), 201-208. doi: http://dx.doi.org/10.1016/S0169-2607(97)00043-6

Heinzl, H., & Kaider, A. (2007). Manual for the SAS-Macro RCS.

Heinzl, H., Kaider, A., & Zlabinger, G. (1996). Assessing Interactions of Binary Time-Dependent Covariates with Time in Cox Proportional Hazards Regression Models Using Cubic Spline Functions. *Statistics in Medicine, 15*(23), 2589-2601.

Henderson, R. (1990). A problem with the likelihood ratio test for a change-point hazard rate model. *Biometrika, 77*(4), 835-843.

Hess, K. R. (1994). Assessing time-by-covariate interactions in proportional hazards regression models using cubic spline functions. *Statistics in Medicine, 13*(10), 1045-1062. doi: 10.1002/sim.4780131007

Hilbe, J. (2009). *Logistic regression models*: CRC Press.

James, I. (2005). Accelerated Failure-time Models *Encyclopedia of Biostatistics*: John Wiley & Sons, Ltd.

Kay, R., & Kinnersley, N. (2002). On the use of the accelerated failure time model as an alternative to the proportional hazards model in the treatment of time to event data: a case study in influenza. *Drug information journal, 36*(3), 571-579.

Keiding, N., Andersen, P. K., & Klein, J. P. (1997). The role of frailty models and accelerated failure time models in describing heterogeneity due to omitted covariates. *Statistics in Medicine, 16*(2), 215-224.

Kemmer, N., Alsina, A., & Neff, G. W. (2011). Orthotopic liver transplantation in a multiethnic population: role of spatial accessibility. *Transplant Proc, 43*(10), 3780-3782. doi: 10.1016/j.transproceed.2011.09.041

Kleinbaum, D., & Klein, M. (2012). Kaplan-Meier Survival Curves and the Log-Rank Test *Survival Analysis* (pp. 55-96): Springer New York.

Kolda, T. G., Lewis, R. M., & Torczon, V. (2003). Optimization by direct search: New perspectives on some classical and modern methods. *SIAM review, 45*(3), 385-482.

Lambert, P., Collett, D., Kimber, A., & Johnson, R. (2004). Parametric accelerated failure time models with random effects and an application to kidney transplant survival. *Statistics in Medicine, 23*(20), 3177-3192. doi: 10.1002/sim.1876

Leffondré, K., Abrahamowicz, M., & Siemiatycki, J. (2003). Evaluation of Cox's model and logistic regression for matched case-control data with time-dependent covariates: a simulation study. *Statistics in Medicine, 22*(24), 3781-3794.

Liang, K. Y., Self, S. G., & Liu, X. (1990). The Cox proportional hazards model with change point: An epidemiologic application. *Biometrics*, 783-793.

Lin, D. Y., Wei, L. J., & Ying, Z. (1993). Checking the Cox model with cumulative sums of martingale-based residuals. *Biometrika, 80*(3), 557-572.

Liu, M., Lu, W., & Shao, Y. (2008). A Monte Carlo approach for change-point detection in the Cox proportional hazards model. *Statistics in Medicine, 27*(19), 3894-3909. doi: 10.1002/sim.3214

Loader, C. R. (1991). Inference for a hazard rate change point. *Biometrika, 78*(4), 749-757.

Luo, X. (1996). The asymptotic distribution of MLE of treatment lag threshold. *Journal of statistical planning and inference, 53*(1), 33-61.

Luo, X., Turnbull, B. W., & Clark, L. C. (1997). Likelihood ratio tests for a changepoint with survival data. *Biometrika, 84*(3), 555-565.

Maggs, J. R., & Chapman, R. W. (2008). An update on primary sclerosing cholangitis. *Current Opinion in Gastroenterology, 24*(3), 377-383 310.1097/MOG.1090b1013e3282f1099e1239.

Matthews, D., Farewell, V., & Pyke, R. (1985). Asymptotic score-statistic processes and tests for constant hazard against a change-point alternative. *The Annals of Statistics, 13*(2), 583-591.

Matthews, D. E., & Farewell, V. T. (1982). On testing for a constant hazard against a change-point alternative. *Biometrics*, 463-468.

Merion, R. M. (2009). 2008 SRTR Report on the State of Transplantation. *American Journal of Transplantation, 9*(4p2), 867-868.

Merion, R. M., Ashby, V. B., Wolfe, R. A., Distant, D. A., Hulbert-Shearon, T. E., Metzger, R. A., Port, F. K. (2005). Deceased-donor characteristics and the survival benefit of kidney transplantation. *JAMA: The Journal of the American Medical Association, 294*(21), 2726-2733.

Müller, H. G., & WANG, J. L. (1990). Nonparametric analysis of changes in hazard rates for censored survival data: An alternative to change-point models. *Biometrika, 77*(2), 305-314.

Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. *The computer journal, 7*(4), 308-313.

Nguyen, H., Rogers, G., & Walker, E. (1984). Estimation in change-point hazard rate models. *Biometrika, 71*(2), 299-304.

Nocedal, J., & Wright, S. J. (1999). Springer series in operations research. Numerical optimization: New York: Springer.

Park, K., Bensen, R., Lu, B., Nanda, P., Esquivel, C., & Cox, K. (2012). Geographical Rural Status and Health Outcomes in Pediatric Liver Transplantation: An Analysis of 6 Years of National United Network of Organ Sharing Data. *The Journal of Pediatrics*.

Park, K. T., Nanda, P., Bensen, R., Strichartz, D., Esquivel, C., & Cox, K. (2011). Effects of rural status on health outcomes in pediatric liver transplantation: A single center analysis of 388 patients. *Pediatric Transplantation, 15*(3), 300-305. doi: 10.1111/j.1399-3046.2010.01452.x

Physician's, H. S. S. C. (1983). Physician's Health Study Protocol. *Brrokline Mass: Harvard Medical School, Department of Medicine*.

Pons, O. (2003). Estimation in a Cox regression model with a change-point according to a threshold in a covariate. *The Annals of Statistics, 31*(2), 442-463.

Powell, M. (1976). Some global convergence properties of a variable metric algorithm for minimization without exact line searches. *Nonlinear programming, 9*, 53-72.

Price, C., & Coope, I. (2003). Frames and grids in unconstrained and linearly constrained optimization: A nonsmooth approach. *SIAM Journal on Optimization, 14*(2), 415-438.

Rempala, G. A., & Looney, S. W. (2006). Asymptotic properties of a two sample randomized test for partially dependent data. *Journal of statistical planning and inference, 136*(1), 68-89.

Self, S., Prentice, R., Iverson, D., Henderson, M., Thompson, D., Byar, D., Goldman, S. (1988). Statistical design of the Women's Health Trial. *Controlled clinical trials, 9*(2), 119-136.

Starzl, T. E., Iwatsuki, S., Van Thiel, D. H., Carlton Gartner, J., Zitelli, B. J., Jeffrey Malatack, J., Thomas Rosenthal, J. (1982). Evolution of liver transplantation. *Hepatology, 2*(5), 614S-636S.

Tuttle, T. M., Curley, S. A., & Roh, M. S. (1997). Repeat hepatic resection as effective treatment for recurrent colorectal liver metastases. *Annals of surgical oncology, 4*(2), 125-130.

Væth, M., & Skovlund, E. (2004). A simple approach to power and sample size calculations in logistic regression and Cox regression models. *Statistics in Medicine, 23*(11), 1781-1792.

Washburn, K. (2008). Geography, transplant centers, and recipients: What can we learn? *Liver Transplantation, 14*(1), 9-10.

Washburn, K. (2012). Maximizing donor potential; evolving OPO metrics and optimizing organ distribution and allocation in the US. *Liver Transplantation*.

Washburn, K., Pomfret, E., & Roberts, J. (2011). Liver allocation and distribution: Possible next steps. *Liver Transplantation, 17*(9), 1005-1012. doi: 10.1002/lt.22349

Worsley, K. (1988). Exact percentage points of the likelihood-ratio test for a change-point hazard-rate model. *Biometrics*, 259-263.

Yao, Y. C. (1986). Maximum likelihood estimation in hazard rate models with a change-point. *Communications in Statistics-Theory and Methods, 15*(8), 2455-2466.

Zorzi, D., Rastellini, C., Freeman, D. H., Elias, G., Duchini, A., & Cicalese, L. (2012). Increase in mortality rate of liver transplant candidates residing in specific geographic areas: analysis of UNOS data. *Am J Transplant, 12*(8), 2188-2197. doi: 10.1111/j.1600-6143.2012.04083.x

Zucker, D. M., & Lakatos, E. (1990). Weighted log rank type statistics for comparing survival curves when there is a time lag in the effectiveness of treatment. *Biometrika, 77*(4), 853-864.

Appendices

Appendix A: Standard Simulation for Extension 3

```
remove(list=ls())

#ginv function was taken from MASS library of Venables and Ripley
ginv<-
  function (X, tol = sqrt(.Machine$double.eps))
  {
    if (length(dim(X)) > 2L || !(is.numeric(X) || is.complex(X)))
      stop("'X' must be a numeric or complex matrix")
    if (!is.matrix(X))
      X <- as.matrix(X)
    Xsvd <- svd(X)
    if (is.complex(X))
      Xsvd$u <- Conj(Xsvd$u)
    Positive <- Xsvd$d > max(tol * Xsvd$d[1L], 0)
    if (all(Positive))
      Xsvd$v %*% (1/Xsvd$d * t(Xsvd$u))
    else if (!any(Positive))
      array(0, dim(X)[2L:1L])
    else Xsvd$v[, Positive, drop = FALSE] %*% ((1/Xsvd$d[Positive]) *
                                        t(Xsvd$u[, Positive, drop = FALSE]))
  }

mc.ext3<-function(TT,cen.ind,XX,dist,tau,cc,nrep,keep=FALSE) {
  n<-nrow(XX)
  np<-ncol(XX)
  ntau<-length(tau)
  ncc<-length(cc)

  mod<-coxph(Surv(TT,cen.ind)~XX,method="breslow")
  gamma<-mod$coeff

  DD<-TT[cen.ind==1]
  ndead<-length(DD)
  d.ind<-(1:n)[cen.ind==1]
  Dpos<-cbind(d.ind,1:ndead)

  pp<-outer(TT,DD,">=")*exp(as.vector(XX%*%gamma))
  pp<-scale(pp,center=FALSE,scale=colSums(pp))

  S.obs<-array(0,dim=c(2,ntau,ncc))

  score<-array(0,dim=c(n,2,ntau,ncc))
  dimnames(score)<-list(NULL,c("theta","beta"),paste("tau",1:ntau,sep=""),
                        paste("cc",1:ncc,sep=""))
  V<-array(0,dim=c(2,ntau,ncc))
  Vinv<-array(0,dim=c(2,2,ntau,ncc))

  da<-c(n,ndead,np+2)
  X.all<-array(0,dim=da)
  for(i in 1:np)X.all[,,i+2]<-XX[,i]

    for(i in 1:ntau)for(j in 1:ncc){

    X.all[,,1]<-outer(dist>cc[j],DD>=tau[i])
```

```
  X.all[,,2]<-(dist>cc[j])

  Xbar0<-colSums(X.all*as.vector(pp))
  X.all<-sweep(X.all,2:3,Xbar0)

  dM<- -pp
  dM[Dpos]<-dM[Dpos]+1
  raw.score<-rowSums(aperm(X.all*as.vector(dM),c(1,3,2)),dims=2)

  S.obs[,i,j]<-colSums(raw.score[,1:2])

  B<-matrix(aperm(X.all*sqrt(as.vector(pp)),c(3,1,2)),da[3],da[1]*da[2])
  info<-B%*%t(B)

  dI<-nrow(info)
  I12<-info[1:2,3:dI]
  I22<-info[3:dI,3:dI]

  score[,,i,j]<-S<-raw.score[,1:2]-t(I12%*%solve(I22,t(raw.score[,3:dI])))


  tmp<-t(S)%*%S
  V[,i,j]<-diag(tmp)
  Vinv[,,i,j]<- if(is.matrix(B<-try(solve(tmp))))B else ginv(tmp)
}

score<-matrix(score,n,)
V<-matrix(V,2,)
Vinv<-array(Vinv,c(2,2,ntau*ncc))

Mstar<-M<-array(0,c(nrep,2))
M2p<-numeric(nrep)

if(keep)S.mc<-array(0,c(nrep,2,ntau,ncc))

S.obs<-matrix(S.obs,2,)
M.obs<-c(max(abs(S.obs[1,])),max(abs(S.obs[2,])))
W.obs<-S.obs^2/V
Mstar.obs<-c(max(W.obs[1,]),max(W.obs[2,]))
W2p.obs<-S.obs[1,]^2*Vinv[1,1,]+
  2*S.obs[1,]*S.obs[2,]*Vinv[1,2,]+S.obs[2,]^2*Vinv[2,2,]
tauvec<-rep(tau,ncc)
ccvec<-rep(cc,rep(ntau,ncc))
i<-which.max(W2p.obs)
M2p.obs<-W2p.obs[i]
tau.max<-tauvec[i]
cc.max<-ccvec[i]


for(i in 1:nrep){
  S<-matrix(rnorm(n)%*%score,2,)
  M[i,]<-c(max(abs(S[1,])),max(abs(S[2,])))
  W<-S^2/V
  Mstar[i,]<-c(max(W[1,]),max(W[2,]))
  M2p[i]<-max(S[1,]^2*Vinv[1,1,]+2*S[1,]*S[2,]*Vinv[1,2,]+S[2,]^2*Vinv[2,2,])
  if(keep)S.mc[i,,,]<-as.vector(S)
```

```
  }


  ans<-list(S.obs=array(S.obs,dim=c(2,ntau,ncc)),
            W.obs=array(W.obs,dim=c(2,ntau,ncc)),
            W2p.obs=array(W2p.obs,dim=c(ntau,ncc)),
            tau.max=tau.max,cc.max=cc.max,
            M.obs=M.obs,Mstar.obs=Mstar.obs,M2p.obs=M2p.obs,
            p.M=pval(M.obs,M),p.Mstar=pval(Mstar.obs,Mstar),
            p.M2p=pval(M2p.obs,M2p),
            M.mc=M,Mstar.mc=Mstar,M2p.mc=M2p)

  if(keep)ans<-c(ans,list(S.obs=array(S.obs,dim=c(2,ntau,ncc)),S.mc=S.mc))

  ans
}

pval<-function(x,xdist){

  rowMeans(t(xdist)>=x)
}

rchpt<-function(r1,r2,t){
  w<-rexp(length(r1))
  ifelse(w<r1*t,w/r1,t+(w-r1*t)/r2)
}

##########################################################################
############ Simulation ##################################################
##########################################################################

library("survival")

n<-600
nrep<-1000

nsim<-1000

g1<-0.78
g2<-0.16

beta<-0.5
theta<-0.4
tau0<-.25
cc0<-50
ngrid=20
tau<-seq(.15,.3,length=ngrid)
cc<-seq(20,60,length=ngrid)


int.tau0<-c(.15,.35)
int.cc0<-c(40,60)

inside<-function(x,int){(int[1]<=x)&(x<=int[2])}

M<-Mstar<-p.M<-p.Mstar<-matrix(0,nsim,2)
```

```
tau.max<-cc.max<-M2p<-p.M2p<-numeric(nsim)

alpha<-c(.10,.05,.01)

nam.alpha<-as.character(alpha)
nam.param<-c("theta","beta")


tau.max<-cc.max<-M2p<-p.M2p<-numeric(nsim)

M<-matrix(0,nsim,2)
colnames(M)<-nam.param
Mstar<-p.M<-p.Mstar<-M

q.M.mc<-array(0,dim=c(nsim,2,length(alpha)))
dimnames(q.M.mc)<-list(NULL,nam.param,nam.alpha)
q.Mstar.mc<-q.M.mc

q.M2p.mc<-matrix(0,nsim,length(alpha))
colnames(q.M2p.mc)<-nam.alpha


num.dead<-num.dead.box<-numeric(nsim)

for(i in 1:nsim){
  cat(i,"\n")
  x1<-runif(n)
  x2<-rexp(n)
  dist<-100*runif(n)
  z<-(dist>cc0)+0
  lambda1<-exp(r1<-g1*x1+g2*x2+beta*z)
  lambda2<-exp(r1+theta*z)
  ss<-10*rchpt(lambda1,lambda2,tau0)
  uu<-4*rexp(n)
  tt<-pmin(ss,uu)
  delta<-(ss<=uu)+0
  xx<-cbind(x1,x2)
  ans<-mc.ext3(tt,delta,xx,dist,tau,cc,nrep)
  M[i,]<-ans$M.obs
  Mstar[i,]<-ans$Mstar.obs
  p.M[i,]<-ans$p.M
  p.Mstar[i,]<-ans$p.Mstar
  tau.max[i]<-ans$tau.max
  cc.max[i]<-ans$cc.max
  M2p[i]<-ans$M2p.obs
  p.M2p[i]<-ans$p.M2p
  q.M.mc[i,,]<-t(apply(ans$M.mc,2,quantile,probs=1-alpha))
  q.Mstar.mc[i,,]<-t(apply(ans$Mstar.mc,2,quantile,probs=1-alpha))
  q.M2p.mc[i,]<-quantile(ans$M2p.mc,probs=1-alpha)
}


Mall<-cbind(M,Mstar,p.M,p.Mstar,tau.max,cc.max,M2p,p.M2p)

print(summary(Mall))
```

```
print(apply(Mall,2,quantile,probs=seq(0,1,by=.1)))

q.M<-t(apply(M,2,quantile,probs=1-alpha))
q.Mstar<-t(apply(Mstar,2,quantile,probs=1-alpha))
q.M2p<-quantile(M2p,probs=1-alpha)
dimnames(q.M)<-dimnames(q.Mstar)<-list(nam.param,nam.alpha)
names(q.M2p)<-nam.alpha

pow<-function(p,a){colMeans(outer(p,a,"<"))}

power.M<-t(apply(p.M,2,pow,alpha))
power.Mstar<-t(apply(p.Mstar,2,pow,alpha))
dimnames(power.M)<-dimnames(power.Mstar)<-list(nam.param,nam.alpha)
power.M2p<-pow(p.M2p,alpha)
names(power.M2p)<-nam.alpha

mean.q.M.mc<-apply(q.M.mc,c(2,3),mean)
sd.q.M.mc<-apply(q.M.mc,c(2,3),sd)
mean.q.Mstar.mc<-apply(q.Mstar.mc,c(2,3),mean)
sd.q.Mstar.mc<-apply(q.Mstar.mc,c(2,3),sd)
mean.q.M2p.mc<-apply(q.M2p.mc,2,mean)
sd.q.M2p.mc<-apply(q.M2p.mc,2,sd)


OF<-mean(num.dead)/n
LF<-mean(num.dead.box)/n


q.M
q.Mstar
q.M2p


mean.q.M.mc
sd.q.M.mc
mean.q.Mstar.mc
sd.q.Mstar.mc
mean.q.M2p.mc
sd.q.M2p.mc


power.M
power.Mstar
power.M2p


OF
LF
```

Appendix B: Standard Simulation for Extension 4

```
remove(list=ls())

#ginv function taken from MASS library of Venables and Ripley
ginv<-
  function (X, tol = sqrt(.Machine$double.eps))
  {
    if (length(dim(X)) > 2L || !(is.numeric(X) || is.complex(X)))
      stop("'X' must be a numeric or complex matrix")
    if (!is.matrix(X))
      X <- as.matrix(X)
    Xsvd <- svd(X)
    if (is.complex(X))
      Xsvd$u <- Conj(Xsvd$u)
    Positive <- Xsvd$d > max(tol * Xsvd$d[1L], 0)
    if (all(Positive))
      Xsvd$v %*% (1/Xsvd$d * t(Xsvd$u))
    else if (!any(Positive))
      array(0, dim(X)[2L:1L])
    else Xsvd$v[, Positive, drop = FALSE] %*% ((1/Xsvd$d[Positive]) *
                                        t(Xsvd$u[, Positive, drop = FALSE]))
  }

mc.ext4<-function(TT,cen.ind,XX,dist,tau,cc,nrep,keep=FALSE) {


  n<-nrow(XX)
  np<-ncol(XX)
  ntau<-length(tau)
  ncc<-length(cc)

  mod<-coxph(Surv(TT,cen.ind)~XX,method="breslow")
  gamma<-mod$coeff

  DD<-TT[cen.ind==1]
  ndead<-length(DD)
  d.ind<-(1:n)[cen.ind==1]
  Dpos<-cbind(d.ind,1:ndead)

  pp<-outer(TT,DD,">=")*exp(as.vector(XX%*%gamma))
  pp<-scale(pp,center=FALSE,scale=colSums(pp))

  S.obs<-array(0,dim=c(2,ntau,ncc))

  score<-array(0,dim=c(n,2,ntau,ncc))
  dimnames(score)<-list(NULL,c("theta","beta"),paste("tau",1:ntau,sep=""),
                        paste("cc",1:ncc,sep=""))
  V<-array(0,dim=c(2,ntau,ncc))
  Vinv<-array(0,dim=c(2,2,ntau,ncc))

  da<-c(n,ndead,np+2)
  X.all<-array(0,dim=da)
  for(i in 1:np)X.all[,,i+2]<-XX[,i]

  for(i in 1:ntau)for(j in 1:ncc){
```

```
  dmc<-pmax(dist-cc[j],0)
  X.all[,,1]<-outer(dmc,DD>=tau[i])
  X.all[,,2]<-dmc

  Xbar0<-colSums(X.all*as.vector(pp))
  X.all<-sweep(X.all,2:3,Xbar0)

  dM<- -pp
  dM[Dpos]<-dM[Dpos]+1
  raw.score<-rowSums(aperm(X.all*as.vector(dM),c(1,3,2)),dims=2)

  S.obs[,i,j]<-colSums(raw.score[,1:2])

  B<-matrix(aperm(X.all*sqrt(as.vector(pp)),c(3,1,2)),da[3],da[1]*da[2])
  info<-B%*%t(B)

  dI<-nrow(info)
  I12<-info[1:2,3:dI]
  I22<-info[3:dI,3:dI]

  score[,,i,j]<-S<-raw.score[,1:2]-t(I12%*%solve(I22,t(raw.score[,3:dI])))


  tmp<-t(S)%*%S
  V[,i,j]<-diag(tmp)
   Vinv[,,i,j]<- if(is.matrix(B<-try(solve(tmp))))B else ginv(tmp)
}

score<-matrix(score,n,)
V<-matrix(V,2,)
Vinv<-array(Vinv,c(2,2,ntau*ncc))

Mstar<-M<-array(0,c(nrep,2))
M2p<-numeric(nrep)

if(keep)S.mc<-array(0,c(nrep,2,ntau,ncc))

S.obs<-matrix(S.obs,2,)
M.obs<-c(max(abs(S.obs[1,])),max(abs(S.obs[2,])))
W.obs<-S.obs^2/V
Mstar.obs<-c(max(W.obs[1,]),max(W.obs[2,]))
W2p.obs<-S.obs[1,]^2*Vinv[1,1,]+
  2*S.obs[1,]*S.obs[2,]*Vinv[1,2,]+S.obs[2,]^2*Vinv[2,2,]
tauvec<-rep(tau,ncc)
ccvec<-rep(cc,rep(ntau,ncc))
i<-which.max(W2p.obs)
M2p.obs<-W2p.obs[i]
tau.max<-tauvec[i]
cc.max<-ccvec[i]

for(i in 1:nrep){
  S<-matrix(rnorm(n)%*%score,2,)
  M[i,]<-c(max(abs(S[1,])),max(abs(S[2,])))
  W<-S^2/V
  Mstar[i,]<-c(max(W[1,]),max(W[2,]))
```

```
    M2p[i]<-max(S[1,]^2*Vinv[1,1,]+2*S[1,]*S[2,]*Vinv[1,2,]+S[2,]^2*Vinv[2,2,])
    if(keep)S.mc[i,,,]<-as.vector(S)
  }

  ans<-list(S.obs=array(S.obs,dim=c(2,ntau,ncc)),
            W.obs=array(W.obs,dim=c(2,ntau,ncc)),
            W2p.obs=array(W2p.obs,dim=c(ntau,ncc)),
            tau.max=tau.max,cc.max=cc.max,
            M.obs=M.obs,Mstar.obs=Mstar.obs,M2p.obs=M2p.obs,
            p.M=pval(M.obs,M),p.Mstar=pval(Mstar.obs,Mstar),
            p.M2p=pval(M2p.obs,M2p),
            M.mc=M,Mstar.mc=Mstar,M2p.mc=M2p)

  if(keep)ans<-c(ans,list(S.obs=array(S.obs,dim=c(2,ntau,ncc)),S.mc=S.mc))

  ans
}

pval<-function(x,xdist){
  rowMeans(t(xdist)>=x)
}

rchpt<-function(r1,r2,t){
  w<-rexp(length(r1))
  ifelse(w<r1*t,w/r1,t+(w-r1*t)/r2)
}

#############################################################################
############ Simulation ######################################################
#############################################################################

library("survival")

n<-600
nrep<-1000
nsim<-1000

g1<-1
g2<-0.5
beta<-0.008
theta<-0.002
tau0<-.25
cc0<-50
ngrid=20
tau<-seq(.15,.3,length=ngrid)
cc<-seq(20,60,length=ngrid)


int.tau0<-c(.15,.35)
int.cc0<-c(40,60)

inside<-function(x,int){(int[1]<=x)&(x<=int[2])}

M<-Mstar<-p.M<-p.Mstar<-matrix(0,nsim,2)
tau.max<-cc.max<-M2p<-p.M2p<-numeric(nsim)
```

```
baseline<-list(u=c(.15,1.75,3.3),r=c(.15,.048,.018,.038))

L.rpwexp<-setup.rcox.with.chpt(lambda1,lambda2,tau0,baseline)

alpha<-c(.10,.05,.01)


nam.alpha<-as.character(alpha)
nam.param<-c("theta","beta")

tau.max<-cc.max<-M2p<-p.M2p<-numeric(nsim)

M<-matrix(0,nsim,2)
colnames(M)<-nam.param
Mstar<-p.M<-p.Mstar<-M

q.M.mc<-array(0,dim=c(nsim,2,length(alpha)))
dimnames(q.M.mc)<-list(NULL,nam.param,nam.alpha)
q.Mstar.mc<-q.M.mc

q.M2p.mc<-matrix(0,nsim,length(alpha))
colnames(q.M2p.mc)<-nam.alpha


num.dead<-num.dead.box<-numeric(nsim)

for(i in 1:nsim){
  cat(i,"\n")
  x1<-runif(n)
  x2<-rexp(n)
  dist<-100*runif(n)
  z<-pmax((dist-cc0)+0,0)
  lambda1<-exp(r1<-g1*x1+g2*x2+beta*z)
  lambda2<-exp(r1+theta*z)
  ss<-10*rchpt(lambda1,lambda2,tau0)
  uu<-4*rexp(n)
  tt<-pmin(ss,uu)
  delta<-(ss<=uu)+0
  xx<-cbind(x1,x2)
  ans<-mc.ext4(tt,delta,xx,dist,tau,cc,nrep)
  M[i,]<-ans$M.obs
  Mstar[i,]<-ans$Mstar.obs
  p.M[i,]<-ans$p.M
  p.Mstar[i,]<-ans$p.Mstar
  tau.max[i]<-ans$tau.max
  cc.max[i]<-ans$cc.max
  M2p[i]<-ans$M2p.obs
  p.M2p[i]<-ans$p.M2p
  q.M.mc[i,,]<-t(apply(ans$M.mc,2,quantile,probs=1-alpha))
  q.Mstar.mc[i,,]<-t(apply(ans$Mstar.mc,2,quantile,probs=1-alpha))
  q.M2p.mc[i,]<-quantile(ans$M2p.mc,probs=1-alpha)
}

Mall<-cbind(M,Mstar,p.M,p.Mstar,tau.max,cc.max,M2p,p.M2p)

print(summary(Mall))
```

```
print(apply(Mall,2,quantile,probs=seq(0,1,by=.1)))

q.M<-t(apply(M,2,quantile,probs=1-alpha))
q.Mstar<-t(apply(Mstar,2,quantile,probs=1-alpha))
q.M2p<-quantile(M2p,probs=1-alpha)
dimnames(q.M)<-dimnames(q.Mstar)<-list(nam.param,nam.alpha)
names(q.M2p)<-nam.alpha

pow<-function(p,a){colMeans(outer(p,a,"<"))}

power.M<-t(apply(p.M,2,pow,alpha))
power.Mstar<-t(apply(p.Mstar,2,pow,alpha))
dimnames(power.M)<-dimnames(power.Mstar)<-list(nam.param,nam.alpha)
power.M2p<-pow(p.M2p,alpha)
names(power.M2p)<-nam.alpha

mean.q.M.mc<-apply(q.M.mc,c(2,3),mean)
sd.q.M.mc<-apply(q.M.mc,c(2,3),sd)
mean.q.Mstar.mc<-apply(q.Mstar.mc,c(2,3),mean)
sd.q.Mstar.mc<-apply(q.Mstar.mc,c(2,3),sd)
mean.q.M2p.mc<-apply(q.M2p.mc,2,mean)
sd.q.M2p.mc<-apply(q.M2p.mc,2,sd)


OF<-mean(num.dead)/n
LF<-mean(num.dead.box)/n


q.M
q.Mstar
q.M2p

mean.q.M.mc
sd.q.M.mc
mean.q.Mstar.mc
sd.q.Mstar.mc
mean.q.M2p.mc
sd.q.M2p.mc

power.M
power.Mstar
power.M2p

OF
LF
```

Appendix C: Numerical Optimization

```
library(survival)

mydata=read.csv("Jan16_R.csv",header=TRUE)
liver=mydata[c(7,8,10,12,13,14,25)]

dim(liver)
names(liver)

liver=liver[order(liver$surv5yr,-liver$censor5),]
liver=liver[rowSums(is.na(liver))==0,]

N=nrow(liver)

cen.t=liver$surv5yr
cen.ind=liver$censor5
dist=liver$Distance_from_home
hcv<-liver$hcv
hcc<-liver$hcc
cc0<-180
zz<-(dist>cc0)+0

mod<-coxph(Surv(cen.t,cen.ind)~zz+hcv+hcc)
summary(mod)

beta0<-mod$coef["zz"]
gamma0<-mod$coeff[c("hcv","hcc")]
tau0<-3.66
theta0<-0

# Specify initial values for parameters.
x0<-c(tau0,cc0,theta0,beta0,gamma0)

XX<-cbind(hcv,hcc)
DD<-cen.t[cen.ind==1]
nd<-length(DD)
d.ind<-(1:N)[cen.ind==1]
Dpos<-cbind(d.ind,1:nd)
RR<-outer(cen.t,DD,">=")+0
nx<-length(x0)

# f computes the log-likelihood for Extension #3.
f<-function(x) {

  tau<-x[1]
  cc<-x[2]
  theta<-x[3]
  beta<-x[4]
  gamma<-x[5:nx]

  AA<- as.vector(XX%*%gamma)+beta*(dist>cc)+theta*outer(dist>cc,DD>=tau)
  sum(AA[Dpos]-log(colSums(RR*exp(AA))))

}
```

```
#########################################################################

parscale<-c(.5,20,.3,.3,.3,.3)
control<-list(fnscale=-1,parscale=parscale)

control.NM<-c(control,list(maxit=5000))
control.BFGS<-c(control,list(ndeps=c(.5,.5,rep(1e-03,4)),maxit=500))
control.SANN<-c(control,list(maxit=10000))

ans.NM<-optim(x0, f , method = "Nelder-Mead", control=control.NM)
ans.BFGS<-optim(x0, f , method = "BFGS" , control=control.BFGS)
ans.SANN<-optim(x0, f , method = "SANN" , control=control.SANN)

ans.NM
ans.BFGS
ans.SANN

# Different parameter estimates.
rbind(x0,ans.NM$par,ans.BFGS$par,ans.SANN$par)

# Improvements in log-likelihood relative to x0.
c(ans.NM$value,ans.BFGS$value,ans.SANN$value)-f(x0)

# Try random initial points.

rt<-range(DD)
nrep<-20
param.est<-array(NA,c(nrep,3,length(x0)))

for(i in 1:nrep){
  xran<-c(runif(1,rt[1],rt[2]),runif(1,0,250),runif(4,-2,2))
  param.est[i,1,]<-optim(xran, f , method = "Nelder-Mead",
                         control=control.NM)$par
  param.est[i,2,]<-optim(xran, f , method = "BFGS",
                         control=control.BFGS)$par
  param.est[i,3,]<-optim(xran, f , method = "SANN",
                         control=control.SANN)$par
}

# Improvements in log-likelihood relative to x0.
apply(param.est,c(1,2),f)-f(x0)

#########################################################################

# My attempt Number 1 with different initial Values
gamma1=c(0.8,0.16)
x1=c(3,100,2,1,gamma1)

ans.NM.1<-optim(x1, f , method = "Nelder-Mead", control=control.NM)
ans.BFGS.1<-optim(x1, f , method = "BFGS" , control=control.BFGS)
ans.SANN.1<-optim(x1, f , method = "SANN" , control=control.SANN)

# Different parameter estimates.
cbind(x1,ans.NM.1$par,ans.BFGS.1$par,ans.SANN.1$par)
```

```
# My attempt Number 2 with different Initial Values
x2=c(4,200,1,0,gamma1)

ans.NM.2<-optim(x2, f , method = "Nelder-Mead", control=control.NM)
ans.BFGS.2<-optim(x2, f , method = "BFGS" , control=control.BFGS)
ans.SANN.2<-optim(x2, f , method = "SANN" , control=control.SANN)

# Different parameter estimates.
                    cbind(x2,ans.NM.2$par,ans.BFGS.2$par,ans.SANN.2$par)
```

About the Author

Alexia holds a Master of Science in Statistics from Florida State University and has undergraduate degrees in both mathematics and statistics from the University of Florida.

She is a member of the American Statistical Association since August 2000, the Global Health Council since June 2010 and the Florida Public Health Association since March 2010. Alexia was the president of the Statistics undergraduate Society at UF, a member of the Golden Key National Honors Society, Pi Mu Epsilon, Mu Sigma Rho, and Phi Kappa Phi.

Alexia was a Duane Meeter Scholar at Florida State University 2000-2001, received a student award in 2011 from the Florida Chapter of the American Statistical Association, was awarded the Public Health Traineeship for 4 consecutive semesters Fall 2009-Spring 2011, and was the recipient of an outstanding oral presentation award in 2011 from the University of South Florida.