TESTS FOR RANDOM SIGNS CENSORING IN COMPETING RISKS

by

Shannon M. Woolley

B.S. in Mathematics, Villanova University, 2012

Submitted to the Graduate Faculty of

the Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2017

UNIVERSITY OF PITTSBURGH

GRADUATE SCHOOL OF PUBLIC HEALTH

This dissertation was presented

by

Shannon M. Woolley

It was defended on

April 10, 2017

and approved by

Jeanine M. Buchanich, MEd, PhD
Research Associate Professor, Department of Biostatistics
Graduate School of Public Health, University of Pittsburgh

Chung-Chou (Joyce) H. Chang, PhD
Professor, Department of Medicine and Biostatistics
School of Medicine, University of Pittsburgh

Manisha Jhamb, MD, MPH
Assistant Professor, Department of Medicine
School of Medicine, University of Pittsburgh

Dissertation Advisor/ Co-advisor:

Abdus S. Wahed, PhD
Professor, Department of Biostatistics
Graduate School of Public Health, University of Pittsburgh,

Jonathan G. Yabes, PhD
Assistant Professor, Department of Medicine and Biostatistics
School of Medicine, University of Pittsburgh

TESTS FOR RANDOM SIGNS CENSORING IN COMPETING RISKS

Shannon M. Woolley, PhD

University of Pittsburgh, 2017

ABSTRACT

In the setting of competing risks, the marginal survival functions of the latent failure times are nonidentifiable without making further assumptions about the joint distribution, the majority of which are untestable. One exception is the random signs censoring assumption which assumes the main event time is independent of the indicator that the main event preceded the competing event. Few methods exist to formally test this assumption, and none consider a stratified test, which detects whether random signs censoring is met within subgroups of a categorical covariate. We develop a nonparametric stratified test for random signs censoring that is easy to implement. In addition, it is often of interest to model the effects of several covariates in relation to the cause of interest. Thus, as an extension of the stratified test, we also propose a test for conditional random signs censoring, which allows for the random signs censoring assumption to be met after adjusting for categorical and/or continuous covariates.

Through Monte Carlo simulations, we show our proposed test statistics have empirical levels close to the nominal level and maintain adequate power even with relatively small sample sizes and random right censoring. Compared to the standard test, both of our proposed tests have nearly equivalent power under random signs censoring and are superior in situations of stratified or conditional random signs censoring, where the standard test fails to detect random signs censoring within subgroups or after adjusting for covariates, respectively. Their ease of implementation and utility are illustrated through an application to liver transplant data from the United Network for Organ Sharing.

Public Health Significance: Clinicians must make decisions affecting patients lives using the information available to them. Relying on research results based on models that use unverifiable assumptions can lead to inaccurate conclusions. The methods proposed here offer a solution to allow for more accurate modeling of marginal survival functions with competing risk data. Through use of these new methods, patient outcomes can be improved over time.

Keywords: Survival analysis, competing risks, random signs censoring, marginal survival function, nonparametric test, cumulative incidence function.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0 INTRODUCTION

Survival analysis is generally defined as a set of methods for analyzing data where the outcome variable is the time until the occurrence of an event of interest. The event of interest can be death, occurrence of a disease, transplant, etc. The time-to-event or survival time can be measured in days, weeks, years, etc. Typically, survival data are not fully observed, but rather are censored. If a patient does not experience the event of interest during the duration of the study, then the subject is "censored" at their last follow-up time, i.e. the last time they were known to be event-free. Censoring can complicate analyses, but standard methods exist for dealing with censored data [Cheng et al., 1998]. Censored observations can be thought of as a form of missing data, and like missing data, they can be non-informative or informative. Non-informative censoring occurs when censoring is independent of the event time, such as when a study participant drops out for non-study related reasons. Censoring can also be dependent or informative, as in the setting of competing risks, where the occurrence of a competing event censors the main event informatively.

A competing risk can be defined as an event whose occurrence either precludes the occurrence of another event under examination or fundamentally alters the probability of occurrence of this other event Gooley et al. [1999]. A classic example of competing risks is found in cancer studies, where disease relapse and death in remission are considered competing events. The occurrence of either event, relapse or death, would prevent the other from happening. A cancer patient could die in remission from a cause entirely unrelated to cancer, such as a motor vehicle accident, and is therefore no longer at risk of relapse. Another example can be found in the analysis of transplantation data. Consider patients in need of a liver transplants who are placed on the waiting list. The transplantation community has attempted to develop an algorithm to prioritize patients for liver allocation by identifying risk

factors associated with pre-transplant mortality. However, while on the list patients could experience competing events, such as receiving a transplant or removal from the list due to deteriorating health, preventing researchers from analyzing the true underlying mortality process.

Competing events complicate research when interest lies in estimating the marginal survival function. The marginal survival function is defined as the probability of a specific event occurring if all other causes of failure were suppressed. However, it is well known that without making further assumptions on the dependence structure between the potential failure times, these quantities are non-identifiable and non-estimable [Tsiatis, 1975].

Conventional methods for survival analysis, such as the Kaplan-Meier estimator or Cox proportional hazards regression, essentially ignore the presence of competing risks by assuming independence between the competing events. This assumption is often not reasonable and can lead to biased results. In particular, the Kaplan-Meier method overestimates the marginal survival function in the presence of positive dependence [Klein et al., 2001]. Overestimation occurs because subjects who fail from a competing risk and thus are no longer at risk for the main event are treated the same as independent censored observations who are still at risk of failing from the main event [Gooley et al., 1999]. Hence, in situations with dependent risks, alternative methods are necessary.

One option is the random signs censoring model introduced by Cooke [1993, 1996], which assumes that the main event time distribution is independent from the indicator that main event has occurred. Random signs censoring is desirable because if the assumption is satisfied, it can be shown that the marginal survival of the main event of interest is identifiable [Lindqvist and Skogsrud, 2008]. It is also verifiable using only the observed data, graphically and via formal testing methods. Cooke [1993] showed that a joint distribution of two competing event times will satisfy the random signs censoring assumption if and only if the normalized subsurvival curve of the main event stochastically dominates that of the competing event. Thus, the random signs censoring assumption can be verified by graphing the normalized subsurvival functions and seeing if they demonstrate the stochastic ordering. This dominance relation is also used in the more recently developed formal testing methods [Dewan et al., 2004, Dauxois et al., 2014]. However, the current tests available are limited in

that they do not allow for the possibility of covariate effects. They have not considered testing if a sample follows stratified random signs censoring, which allows for the possibility that random signs censoring may only be met within one or some covariate groups and not the overall sample. They also cannot test for conditional random signs censoring, which allows for random signs censoring to be met after conditioning on multiple covariates, categorical and/or continuous.

Thus, the aim of this dissertation is to develop two new tests for random signs censoring: stratified and conditional. Chapter 2 focuses on the development of a test for the stratified random signs censoring assumption, while Chapter 3 proposes a test for the conditional random signs censoring assumption. Final conclusions and areas for future research are discussed in Chapter 4.

2.0 TEST FOR STRATIFIED RANDOM SIGNS CENSORING

2.1 INTRODUCTION

Based on Organ Procurement and Transplantation Network (OPTN) data as of May 20, 2016, there are currently over 15,000 adults and children in the United States who have been medically approved for liver transplants and are waiting for donated livers to become available, with more added to the list each day. Around 6000 transplants are performed every year, yet more than 1500 candidates die each year while still on the waiting list. In 2015, a record high of 7127 liver transplants were performed; however, waitlist mortality remains a concern as 1420 candidates on the list died waiting and an additional 1473 were removed as a result of being too sick to undergo transplant.

As part of an ongoing effort to improve liver allocation and decrease deaths on the waiting list, the United Network for Organ Sharing (UNOS), a nonprofit charitable organization, developed a new system for prioritizing candidates waiting for liver transplants based on statistical formulas that predict who needs a liver transplant most urgently. Implemented in February 2002, the model for end-stage liver disease (MELD) and its pediatric counterpart, pediatric end-stage liver disease (PELD), give patients a continuous severity score ranging from less ill to gravely ill based on routine lab test results [Sharing, 2008]. Higher transplant priority is given to patients with higher scores.

Research has shown that MELD and PELD accurately predict most liver patients short-term risk of death without a transplant; however, it has been argued that PELD is less accurate in its ability to predict pre-transplant mortality than MELD [Barshes et al., 2006, Olthoff et al., 2004]. PELD has been criticized for underestimating the severity of illness in pediatric patients leading to increased waiting time and increased patient morbidity.

The PELD score currently uses a Cox proportional hazards model to predict pre-transplant mortality, which treats transplants as non-informative censored events instead of competing events. This is problematic because non-informative censored events are considered independent from the event of interest (pre-transplant death in this case); whereas competing events fundamentally alter the probability of the occurrence of the main event [Gooley et al., 1999]. Because those who receive a transplant tend to be sicker, there is a positive correlation with the underlying mortality process. Therefore, it is not reasonable to assume independence between pre-transplant death and receiving a transplant. In cases of dependence such as this, it is necessary to consider models that do not assume independent competing risks.

Extensive research has been done on analyzing competing risks data. Modern literature has focused on the crude incidence approach, which includes analyses of the cause-specific hazards and cumulative incidence functions. The Cox proportional hazards model [Cox, 1972] has been one of the most commonly used approaches for modeling cause-specific hazards for all causes [Prentice et al., 1978]. In more recent years, an emphasis has been placed on modeling covariate effects on the cumulative incidence function directly [Zhang et al., 2008]. One of the first approaches, a Cox-type proportional hazards model for the subdistribution hazard function, was proposed by Fine and Gray [1999]. Another approach using pseudo-values was developed by Klein and Andersen [2005], while Scheike et al. [2008] suggested an alternative using a direct binomial regression. These, among many other proposed methods, may be useful for modeling the cumulative incidence function, but they do not allow estimation of the marginal survival function.

Fortunately, Cooke [1993]'s Random Signs Censoring assumption, operating under the latent failure time approach, offers an alternative solution. Random signs censoring assumes that the main event time distribution is independent from the indicator of whether the main event would occur or not (in relation to other competing events). Cooke showed that this assumption is verifiable from the observed data. Under this assumption, the marginal distribution of the main event is identifiable. While random signs censoring was introduced more than two decades ago, there was no established test for it until very recently. Dewan et al. [2004] introduced a test, but it did not consider additional independent censoring. Dauxois et al. [2014] introduced a test that accounts for right censoring, but it does not allow

for the possibility of random signs censoring occurring within covariate groups. However, past research has shown that survival may vary among different groups of people, such as males versus females [Cox, 1972]. For this reason, it is necessary to examine censoring and survival patterns within covariate groups. In these situations, the overall sample may not satisfy the random signs censoring assumption, but it is possible that the random signs censoring assumption is met when stratified by a covariate. This is the basic idea of stratified random signs censoring. The stratified random signs censoring assumption is slightly more relaxed than the random signs censoring assumption in that it checks for random signs censoring overall and within a specified covariate. Hence, the aim of this paper is to introduce a test for the random signs censoring assumption that will also allow us to test for stratified random signs censoring, based on a categorical covariate. To the best of our knowledge, our work is the first to propose a test for stratified random signs censoring on a categorical covariate.

This chapter is organized as follows. Section 2.2 introduces some basic notation and reviews the identifiable functions within the competing risks framework. Sections 2.3 and 2.4 discuss the properties and assumptions of specific classes of models relevant to this work. Section 2.3 reviews some of the classical models used in competing risks, while Section 2.4 presents the Random Signs Censoring models, stratified and unstratified. Sections 2.5 and 2.6 are devoted to the development and asymptotic theory of the unstratified and stratified random signs censoring tests. Section 2.7 studies the finite-sample properties of our proposed test statistic through numerical simulations, and an application to the liver transplant data is given in Section 2.8. Concluding remarks and ideas for future work are given in Section 2.9.

2.2 COMPETING RISKS FRAMEWORK

2.2.1 Notation

Suppose an individual is subject to two failure types. Without loss of generality, let T_1 denote the failure time of the event of interest corresponding to the first failure type, and let

T_2 denote the failure time of the competing event corresponding to the second failure type. Later we will introduce the additional possibility of random right censoring, i.e. independent censoring, but for now we discuss the case in which there is no additional censoring.

In the competing risks setting we can only observe the first failure time, $T = min(T_1, T_2)$, and the indicator of the failure type, $\delta = j,\ j = 1\ or\ 2$ indicating the main event or competing event, respectively. The overall survival function is given by $S(t) = pr(T > t)$.

2.2.2 Identifiable Functions

Unfortunately, due to the problem of non-identifiability, the marginal survival functions of the random variables T_1 and T_2, defined as $pr(T_1 > t)$ and $pr(T_2 > t)$ respectively, are not estimable. However, the probability that an event of type j occurs before time t, or the cumulative incidence function for the j^{th} cause-specific event, is estimable. The cumulative incidence functions, also known as the cumulative subdistribution functions, for main and competing events are given by:

$$F_j(t) = pr(T \leq t, \delta = j),\ j = 1, 2;\ t > 0.$$

The sum of these subdistributions is equal to the cumulative distribution function of T, $F(t) = pr(T \leq t) = F_1(t) + F_2(t) = 1 - S(t)$. Similar to how the cumulative distribution function is the sum of the cumulative incidence functions, the overall survival function can be written as a sum of subsurvival functions, such that $S(t) = S_1(t) + S_2(t)$, where $S_j(t)$ is the subsurvival function for the j^{th} event:

$$S_j(t) = pr(T > t, \delta = j) = pr(T_j > t, T_j < T_{j'}, j' \neq j),\ j = 1, 2; j' = 1, 2.$$

Together, the subsurvival functions $S_1(t)$ and $S_2(t)$ make up the overall survival and form a subsurvival pair, the formal definition of which is given below.

Definition 1. *Functions $S_1(t)$ and $S_2(t)$ form a continuous subsurvival pair if:*

1. *$S_1(t)$ and $S_2(t)$ are non-negative non-increasing continuous real functions on $[0, \infty)$ with $S_1(0) \leq 1$ and $S_2(0) \leq 1$,*
2. *$\lim_{t\to\infty} S_1(t) = \lim_{t\to\infty} S_2(t) = 0$, and*

3. $S_1(0) + S_2(0) = 1$.

Note that $S_j(0) = pr(\delta = j), j = 1, 2$. Later we will use the notation $pr(\delta = 1) = \gamma$ and $pr(\delta = 2) = (1 - \gamma)$ for the probability of having an event of type 1 and type 2, respectively. Any function of the data can be written in terms of the subsurvival functions $S_1(t)$ and $S_2(t)$ because they contain all of the information which can be extracted from observing T. Moreover, under certain additional assumptions on the dependence structure between T_1 and T_2, the subsurvival functions can be used to define a unique set of marginals [Balakrishnan, 1995]. (Details of specific models and their corresponding assumptions are discussed in Sections 2.3 and 2.4.)

Conditioning the subsurvival functions on the occurrence of the corresponding event type, we can obtain the normalized (or conditional) subsurvival functions:

$$S_j^*(t) = pr(T > t \mid \delta = j) = pr(T_j > t \mid T_j < T_{j'}, j' \neq j) = \frac{S_j(t)}{S_j(0)}, \ j = 1, 2; j' = 1, 2,$$

which are also estimable. The functions $S_1^*(.)$ and $S_2^*(.)$ portray particular behaviors under different models in the competing risks setting, ultimately giving insight into the dependence structure between T_1 and T_2. For this reason, the properties of $S_1^*(.)$ and $S_2^*(.)$ under the various models are discussed in the following sections and later are used to develop the test statistic.

2.3 COMMONLY USED MODELS

This section provides a basic summary of some specific models commonly used, or rather misused, in the modeling of competing risks. In particular, we review the Independent Competing Risks Model (2.3.1) and the Delay Time Model (2.3.2). When estimating the marginal survival function of latent failure times, both of these models make certain assumptions in order to allow for the identifiability of these quantities.

2.3.1 Independent Competing Risks

This model assumes that the main event time T_1 is independent of the competing event time T_2. Assuming independence between the competing events allows one to uniquely determine the marginal survival functions of T_1 and T_2 from the joint survival function of (T, δ), such that:

$$S(t) = S(t_1, t_2) = pr(T_1 > t_1, T_2 > t_2) = pr(T_1 > t_1)pr(T_2 > t_2).$$

Many conventional methods for survival analysis, such as the Kaplan-Meier estimator and Cox proportional hazards model, follow this assumption by treating T_2 as non-informative censoring. Though widely used, this is a strong, untestable hypothesis.

In a standard survival setting, these methods assume independence between the event of interest and censoring, but this independence assumption is reasonable because subjects who are censored at a specific time point should still be representative of those still at risk.

In the setting of competing risks, these methods treat both censoring times and competing event times as independent from the main event time T_1. However, competing events are in clear violation of this independence assumption as by definition they alter one's probability of experiencing the main event. Consequently, conventional methods like the Kaplan-Meier estimator produce bias results.

For instance, in the liver transplant data example, candidates who receive transplants are typically sicker and at a higher risk of pre-transplant death than other candidates on the waiting list. Yet, the Kaplan-Meier estimator ignores this positive correlation between the competing events and treats them as independent. Thus, those candidates who receive transplants are considered the same as those who are censored and still at risk of failing from the main event. As a result, the Kaplan-Meier estimator inflates the probability of pre-transplant survival, thereby underestimating the risk of pre-transplant mortality.

In addition, there is no general result on the behavior of the normalized subsurvival functions $S_1^*(t)$ and $S_2^*(t)$ as they vary based on the given distributions to T_1 and T_2. If, however, T_1 and T_2 have exponential distributions, Cooke [1993] established that $S_1^*(t) = S_2^*(t)$ for all $t > 0$. In this case, it would technically be possible to verify that they are independent, but verification is limited to this extremely restrictive situation.

2.3.2 Delay Time

Another model whose assumptions allow us to identify marginal distributions is the delay time model, also known as the conditional independence model, introduced by Christer and Waller [1984]. The delay time model is typically used in reliability and maintenance due the nature of its applicability. The model defines failure of a component as a two-stage process, where the first stage represents the time when a system crosses some threshold and the second stage represents the remaining time before failure Wang et al. [2011].

We could also apply this model to a clinical setting. For instance, consider the liver transplant data example again. Hypothetically, say there exists a known health status that when detected indicates that a patient needs to have a liver transplant. There should be some time after this initial defect during which the transplantation surgery should be performed to prevent failure. However, if the transplant is not performed, the person could continue to worsen, resulting in death. The time lapse between the health status indicator and death is referred to as the delay time, hence the name of the model. In this model T_1 and T_2 are dependent sharing a common quantity U but independent when conditioned on U. Specifically, the model assumes:

$$T_1 = U + V$$
$$T_2 = U + W$$

where U, V, and W are mutually independent random variables. In this case, the random variable U represents the degradation time of person until a defect arises, i.e. the health status indicator, and the remaining time before failure from death or transplant (event 1 or 2) is represented by V or W, respectively. In reliability settings, event 1 may be functional failure of the machine; whereas, event 2 may not be a failure, but rather preventative maintenance.

Unfortunately, the likelihood for this model is very complicated, making it difficult to assess the goodness-of-fit of the model Baker and Wang [1993]. Some extensions of the model have been proposed but are computationally intensive and still lack diagnostic plots Baker and Wang [1991], Wang et al. [2011]. In the case when U, V, and W are exponentially

distributed, Hokstadt and Jensen [1998] proved the probability of the main event occurring is constant over time resulting in equal normalized survival functions $S_1^*(t) = S_2^*(t)$ for all $t > 0$. Thus, similar to the independent risks with exponential distributions scenario discussed in 2.3.1, we have a case that is verifiable but is extremely restrictive.

2.4 RANDOM SIGNS CENSORING MODELS

The problem with using the models discussed in Sections 2.3.1 and 2.3.2 is that the assumptions cannot be verified, and as a result, the models cannot be validated. In cases of dependent events, such as in the setting of competing risks, these models yield biased results. In this section, we formally define and discuss the properties of the random signs censoring model and the conditional random signs censoring model. Like the previous models, these models also introduce assumptions on the dependence structure between T_1 and T_2, but unlike the previous models, these assumptions are verifiable.

2.4.1 Random Signs Censoring

2.4.1.1 Concept The random signs censoring model was first introduced by Cooke [1993] and is perhaps the simplest dependent competing risk model. As defined in Section 2.1, random signs censoring assumes that the main event time T_1 is independent of the event indicator δ or equivalently, T_1 is independent of the sign of $(T_2 - T_1)$, hence the name of the model. Because of this independence, under the random signs censoring assumption:

$$S_1^*(t) = pr(T > t|\delta = 1) = \frac{pr(T_1 > t, \delta = 1)}{pr(\delta = 1)} = \frac{pr(T_1 > t)pr(\delta = 1)}{pr(\delta = 1)} = pr(T_1 > t).$$

This somewhat surprising result states that under the random signs censoring assumption, the survival function containing only the observed occurrences of T_1 is the same as the marginal survival function of T_1. This simplification under the random signs censoring assumption points to one of its advantages. Marginal survival estimates can be calculated by removing those subjects who experience failure from the competing event. The subjects

who fail from the competing event are not treated as censored observations but rather are excluded from the analysis entirely. After obtaining the subset, standard survival methods can be used to carry out the rest of the analyses and produce consistent, unbiased results. The random signs censoring model allows us to accurately estimate the marginal survival because unlike the other models that make assumptions on the dependency structure between competing events, the random signs censoring assumption is verifiable.

2.4.1.2 Verification

Verifying the random signs censoring assumption stems from a theorem established by Cooke [1993], which we restate below:

Theorem 1. *Survival pairs* (T_1, T_2) *follow random signs censoring if and only if* $S_1^*(t) > S_2^*(t)$ *for all* $t > 0$.

This theorem implies that the random signs censoring assumption T_1 independent of δ is satisfied if and only if the condition $S_1^*(t) > S_2^*(t)$ for all $t > 0$ is also satisfied. This result is of special interest because it provides the foundation for testing the random signs censoring assumption as the normalized subsurvival functions are estimable. If the normalized subsurvival curve of the main event dominates that of the competing event, then a random signs censoring model may be a good fit for the data. Dauxois et al. [2014] used this result to develop a test statistic for testing random signs censoring.

2.4.2 Stratified Random Signs Censoring

2.4.2.1 Concept

In this paper we consider and develop a test statistic to test the stratified random signs censoring model. Consider the categorical covariate Z with possible values $k = 1, .., K$ where K is the total number of strata in Z. The stratified random signs censoring model assumes that the main event failure time and event indicator within the k^{th} stratum, T_{1k} and δ_{1k} respectively, are independent for at least one k. Based on this assumption, we obtain the following result:

$$S_{1k}^*(t) = pr(T_k > t \mid \delta_k = 1) = \frac{pr(T_{1k} > t, \delta_k = 1)}{pr(\delta_k = 1)} = \frac{pr(T_{1k} > t)pr(\delta_k = 1)}{pr(\delta_k = 1)} = pr(T_{1k} > t)$$

for at least one k. This result states that stratified by the categorical covariate Z, the normalized subsurvival function of T_1 is equivalent to the marginal survival function T_1 for at least one stratum k. This result is an extension of the random signs censoring assumption because it accommodates the case where random signs censoring is satisfied within some subgroups, although the assumption is violated overall. Clinical trials and research studies can find this result particularly useful when trying find which subgroups to analyze and compare.

2.4.2.2 Verification In addition, the stratified random signs censoring assumption is just as easy to verify as the random signs censoring assumption. Stratifying by Z, Cooke's random signs censoring theorem can be updated to illustrate the stratified random signs censoring assumption as follows:

Theorem 2. *Let S_{1k} and S_{2k} be a subsurvival pair, and let Z be a categorical covariate with K strata. Then the following are equivalent:*

1. *There exists a survival pair (T_1, T_2) such that T_2 is a stratified random signs censoring of T_1 and:*
$$S^*_{jk}(t) = \frac{S_{jk}(t)}{S_{jk}(0)}, \text{ for } j = 1, 2, \ k = 1, ..., K, \text{ and all } t \geq 0.$$
2. $$S^*_{1k}(t) > S^*_{2k}(t) \text{ for at least one } k \text{ and all } t > 0.$$

This theorem implies that the stratified random signs censoring assumption T_{1k} independent of δ_k for at least one stratum k is satisfied if and only if the condition $S^*_{1k}(t) > S^*_{2k}(t)$ for at least one k and all $t > 0$ is also satisfied. The proof follows directly from that of Cooke [1993]'s *Theorem 2* by conditioning quantities on each of the K strata.

Similar to the random signs censoring result, this stratified random signs censoring result is of special interest because the normalized subsurvival curves stratified by Z are estimable. Hence, if competing event data exhibit this dominance relationship within at least one subgroup, a stratified random signs censoring model may be appropriate and advantageous to accurately model the data.

2.5 DEVELOPMENT OF THE STRATIFIED TEST

As mentioned, we are interested in testing whether or not the stratified random signs censoring assumption holds. Based on the stratified extension of Cooke's theorem (2), testing the stratified random signs censoring assumption that T_{1k} and δ_k are independent for at least one of the K strata is equivalent to testing whether $S^*_{1k}(t)$ dominates $S^*_{2k}(t)$ for at least one k. Thus, if the stratified random signs censoring assumption does not hold, T_{1k} and δ_k are dependent for all k, which is equivalent to testing whether $S^*_{1k}(t)$ is equivalent to $S^*_{2k}(t)$ for all k. Using this relationship, we obtain the null hypothesis:

$$H_0 : S^*_{1k}(t) = S^*_{2k}(t), \text{ for all } k \text{ and all } t > 0 \, ,$$

against the stratified random signs censoring alternative hypothesis:

$$H_1 : S^*_{1k}(t) > S^*_{2k}(t), \text{ for at least one } k \text{ and all } t > 0 \, ,$$

where the normalized subsurvival functions are stratified by the categorical covariate Z.

Dauxois et al. [2014] developed a test for the random signs censoring assumption with null hypothesis:

$$H^*_0 : S^*_1(t) = S^*_2(t), \text{ for all } t > 0$$

and alternative hypothesis:

$$H^*_1 : S^*_1(t) > S^*_2(t), \text{ for all } t > 0.$$

Using equivalent representations of the normalized subsurvival functions, they formulated a test statistic whose numerator is an estimate of the following quantity:

$$\psi^* = \int_0^\tau \{\gamma F_2(t) - (1-\gamma)F_1(t)\} \, dt.$$

To test our hypothesis, we can also develop the test statistic using equivalent representations of the normalized subsurvival functions, but we will use the functions in terms of multiple strata, as in H_0 and H_1, instead of for a single stratum.

Under H_0, we find the functions $S_{1k}(.)$ and $S_k(.)$ are proportional from the following equivalences fulfilled for all $k = 1, ..., K$ and for all $t > 0$:

$$\begin{aligned}
& S^*_{1k}(t) = S^*_{2k}(t) \\
\iff & \frac{S_{1k}(t)}{S_{1k}(0)} = \frac{S_{2k}(t)}{S_{2k}(0)} \\
\iff & (1-\gamma_k)S_{1k}(t) = \gamma_k S_{2k}(t) \\
\iff & S_{1k}(t) = \gamma_k S_k(t) \\
\iff & \gamma_k F_{2k}(t) - (1-\gamma_k)F_{1k}(t) = 0,
\end{aligned}$$

where $\gamma_k = pr(\delta_k = 1)$. Similarly, under H_1 the following equivalent properties are fulfilled for at least one k and for all $t > 0$:

$$S^*_{1k}(t) > S^*_{2k}(t) \tag{2.1}$$

$$\iff \gamma_k F_{2k}(t) - (1-\gamma_k)F_{1k}(t) > 0. \tag{2.2}$$

Using equivalence (2.2), a measure in favor of the alternative is given by summing across all time-points such that:

$$\psi_k = \int_0^{\tau_k} \{\gamma_k F_{2k}(t) - (1-\gamma_k)F_{1k}(t)\}\, dt,$$

where τ_k is the right endpoint of the support of F_k. Letting $p_k = n_k/n$ denote the proportion of the total sample within stratum k, we propose the following weighted average:

$$\psi = \sum_{k=1}^{K} p_k \int_0^{\tau_k} \{\gamma_k F_{2k}(t) - (1-\gamma_k)F_{1k}(t)\}\, dt,$$

which is null under H_0 and positive under H_1.

To allow for the possibility of independent right censoring, let C denote the censoring random variable, independent from the random variable T, with survival distribution $H(.)$. Now, one observes either the failure time or censoring time, so let us define $X = \min(T, C)$ and $\varepsilon = \delta I(T \leq C)$ where $I(.)$ is an indicator function. Therefore, for $i = 1, ..., n$ individuals we observe $(X_i, \varepsilon_i, Z_i)$, where at time X_i the indicator ε_i is equal to 0 when a censoring time has been observed, to 1 when a main event has occurred, and to 2 when a competing event

has occurred. Again, let Z_i represent the categorical covariate, indicating which of the K strata the i^{th} individual is from. For all $t > 0$, we define the following counting processes:

$$N_{jk}(t) = \sum_{i=1}^{n} I(X_i \leq t,\ \varepsilon_i = j,\ Z_i = k), \quad j = 1,\ 2,\ k = 1, ..., K,$$

where $N_{jk}(t)$ is the number of subjects failing from event type j in stratum k during the interval $[0, t]$. Thus, the number of subjects in stratum k failing from either event type 1 or 2 can be denoted as:

$$N_k(.) = \sum_{j=1}^{2} N_{jk}(.).$$

The number of individuals at risk in stratum k is defined by the process:

$$Y_k(t) = \sum_{i=1}^{n} I(X_i \geq t,\ Z_i = k)$$

which counts the number of subjects who have not experienced any event or are uncensored at time t. The Kaplan-Meier estimator of the survival function $S(.)$ of T for stratum k is given by:

$$\hat{S}_k(t) = \prod_{i:x_i \leq t} \left(1 - \frac{\Delta N_k(x_i)}{Y_k(x_i)}\right),$$

where x_i are the ordered event times associated with the sample and $\Delta N_k(x_i) = N_k(t) - N_k(t^-)$. The Aalen-Johansen estimators of the CIFs for the k^{th} stratum are then given by:

$$\hat{F}_{jk}(t) = \int_0^t \hat{S}_k(u^-) \frac{dN_{jk}(u)}{Y_k(u)}, \quad \text{for } j = 1,\ 2, \quad k = 1, ..., K.$$

In order to construct consistent estimates in the presence of censoring, we apply an inverse probability of censoring weight [Robins and Rotnitzky, 1992] denoted by:

$$W_c(t) = \frac{1}{H(t)},$$

where $H(t) = pr(C > t)$ is the censoring survival distribution. Thus, for large n, γ_k can be consistently estimated by:

$$\hat{\gamma}_k = \frac{\int_0^{\tau_k} W_c(t) dN_{1k}(t)}{\int_0^{\tau_k} W_c(t) dN_k(t)}.$$

Using these quantities, an estimate of ψ becomes:

$$\hat{\psi} = \sum_{k=1}^{K} p_k \int_0^{\tau_k} \left\{ \hat{\gamma}_k \hat{F}_{2k}(t) - (1 - \hat{\gamma}_k) \hat{F}_{1k}(t) \right\} dt.$$

Hence, the test statistic for the stratified test takes the form $\sqrt{n}\hat{\psi}/\hat{\sigma_0}$, where $\hat{\sigma_0}$ is a consistent estimator of σ_0, the standard deviation of ψ under H_0, whose details are discussed in Section 2.6.

To construct the test statistic for detecting random signs censoring only (unstratified), the above quantities can be applied to the function ψ^* in a similar manner except that one would ignore the stratification covariate Z when calculating all quantities to obtain:

$$\hat{\psi}^* = \int_0^{\tau} \left\{ \hat{\gamma} \hat{F}_2(t) - (1 - \hat{\gamma}) \hat{F}_1(t) \right\} dt.$$

Letting $\hat{\sigma_0^*}$ denote a consistent estimator of σ_0^*, the standard deviation of ψ^* under H_0, the unstratified test statistic can be written as $\sqrt{n}\hat{\psi}^*/\hat{\sigma_0^*}$. Note that this unstratified test is similar to the one presented by Dauxois et al. [2014] with the weight function equal to one. It is null under H_0 and positive under H_1^* but cannot test H_1.

2.6 ASYMPTOTIC PROPERTIES

In this section, we will discuss the asymptotic properties of our test statistic. The asymptotic distribution of our test statistic is based on the following theorem:

Theorem 3. *Let us suppose that:*

$$\int_0^{\tau_k} W_c(u) dF_k(u) < \infty, \ k = 1, ..., K. \tag{2.3}$$

Then $\sqrt{n}(\hat{\psi} - \psi)$ converges weakly to a mean zero normal random variable Z, with finite variance σ^2. Under H_0 the limiting variance can be expressed in the form of:

$$\sigma_0^2 = \sum_{k=1}^{K} \left(\frac{n_k}{n} \right)^2 \sigma_{0k}^2,$$

where σ_{0k}^2 *is given by:*

$$\begin{aligned}\sigma_{0k}^2 = \gamma_k \int_0^{\tau_k}\int_0^{\tau_k}\int_0^{min(s,t)} W_c(u)dF_{2k}(u)dtds \\ + \frac{\gamma_k}{(1-\gamma_k)^2}\left\{\int_0^{\tau_k} F_{2k}(t)dt\right\}^2 \int_0^{\tau_k} W_c(u)dF_{2k}(u) \\ - \frac{2\gamma_k}{(1-\gamma_k)}\left\{\int_0^{\tau_k} F_{2k}(t)dt\right\}\int_0^{\tau_k}\int_0^{t} W_c(u)dF_{2k}(u)dt.\end{aligned}$$

Proof of Theorem 3. Dauxois and Guilloux [2008] proved the following weak convergence result for a single stratum. Because the strata are independent, this result can also be applied to the within strata quantities such that under assumption (2.3), the following weak convergence holds in $D^3[0,\infty]$:

$$\hat{G}_k = \begin{bmatrix}\hat{G}_{0k}\\ \hat{G}_{1k}\\ \hat{G}_{2k}\end{bmatrix} = \sqrt{n}\begin{bmatrix}\hat{S}_k - S_k\\ \hat{F}_{1k} - F_{1k}\\ \hat{F}_{2k} - F_{2k}\end{bmatrix} \xrightarrow{D} G_k = \begin{bmatrix}G_{0k}\\ G_{1k}\\ G_{2k}\end{bmatrix}, \ as\ n\to\infty, \tag{2.4}$$

where G_k is a mean zero Gaussian process defined by:

$$\begin{aligned}G_{0k}(\cdot) &= S_k(\cdot)U_{0k}(\cdot),\\ G_{jk}(\cdot) &= \int_0^{\cdot}\{F_{jk}(\cdot) - F_{jk}(u)\}\,dU_{0k}(u) + \int_0^{\cdot} S_k(u)dU_{jk}(u),\ j=1,2,\ k=1,...,K,\end{aligned}$$

and U_{1k} and U_{2k} are mean zero Gaussian, square integrable and orthogonal local martingales with covariance function:

$$\langle U_{jk}(s), U_{jk}(t)\rangle = \int_0^{min(s,t)} \frac{W_c(u)dF_{jk}(u)}{S_k^2(u)},\ j=1,2,\ \ and\ \ k=1,...,K$$

and $U_{0k} = -(U_{1k}+U_{2k})$. We can write:

$$\sqrt{n_k}(\hat{\psi}_k - \psi_k) = \sqrt{n_k}\left\{\Psi(\hat{F}_{1k},\hat{F}_{2k}) - \Psi(F_{1k},F_{2k})\right\},$$

where

$$\Psi(F_{1k},F_{2k}) = \int_0^{\tau_k}\{\gamma_k F_{2k}(t) - (1-\gamma_k)F_{1k}(t)\}\,dt$$

and $\gamma_k = F_{1k}(\tau_k)$. The function $\Psi(F_{1k}, F_{2k})$ is Hadamard-differentiable (see e.g. Van Der Vaart and Wellner [1996]); hence, we have the derivative:

$$D_{\Psi}^{F_{1k},F_{2k}}(\alpha_{1k}, \alpha_{2k}) = \int_0^{\tau_k} \left[F_{1k}(\tau_k)\alpha_{2k}(t) + \alpha_{1k}(\tau_k)F_{2k}(t) - \{1 - F_{1k}(\tau_k)\}\, \alpha_{1k}(t) + \alpha_{1k}(\tau_k)F_{1k}(t)\right] dt.$$

The Hadamard differentiability of the function and convergence result (2.4) allow us to apply the functional delta method as described in Theorem 3.9.5 of Van Der Vaart and Wellner [1996] so that we have:

$$\sqrt{n_k}(\hat{\psi}_k - \psi_k) \xrightarrow{D} D_{\Psi}^{F_{1k},F_{2k}}(G_{1k}, G_{2k}) \; as \; n_k \to \infty,$$

where

$$D_{\Psi}^{F_{1k},F_{2k}}(G_{1k}, G_{2k}) = \int_0^{\tau_k} \{\gamma_k G_{2k}(t) - (1 - \gamma_k)G_{1k}(t)\}\, dt - G_{1k}(\tau_k) \int_0^{\tau_k} F_k(t)dt.$$

The limiting distribution of the random variable is Gaussian with mean zero and variance function:

$$\begin{aligned}
\sigma_k^2 &= var\left[\int_0^{\tau_k} \{\gamma_k G_{2k}(t) - (1 - \gamma_k)G_{1k}(t)\}\, dt - G_{1k}(\tau_k) \int_0^{\tau_k} F_k(t)dt\right] \\
&= var\left[\int_0^{\tau_k} \{\gamma_k G_{2k}(t) - (1 - \gamma_k)G_{1k}(t)\}\, dt\right] + var\left\{G_{1k}(\tau_k)\right\} \left\{\int_0^{\tau_k} F_k(t)dt\right\}^2 \\
&\quad - 2\left[\int_0^{\tau_k} \{\gamma_k \langle G_{2k}(t), G_{1k}(\tau_k)\rangle - (1 - \gamma_k)\langle G_{1k}(t), G_{1k}(\tau_k)\rangle\}\, dt\right] \int_0^{\tau_k} F_k(t)dt
\end{aligned}$$

The covariance structure of G_{jk} for $j = 1, 2$ and $k = 1, ..., K$, is given by:

$$\begin{aligned}
\langle G_{ik}(s), G_{jk}(t)\rangle = &\int_0^s \int_0^t \{F_{ik}(s) - F_{ik}(u)\}\, \{F_{jk}(t) - F_{jk}(v)\}\, d\langle U_{0k}(u), U_{0k}(v)\rangle \\
&+ \int_0^s \int_0^t \{F_{ik}(s) - F_{ik}(u)\}\, S_k(v) d\langle U_{0k}(u), U_{jk}(v)\rangle \\
&+ \int_0^s \int_0^t S_k(u)\, \{F_{jk}(t) - F_{jk}(v)\}\, d\langle U_{ik}(u), U_{0k}(v)\rangle \\
&+ \int_0^s \int_0^t S_k(u) S_k(v) d\langle U_{ik}(u), U_{jk}(v)\rangle.
\end{aligned}$$

However, due to the orthogonality of U_{1k} and U_{2k}, we have:

$$\langle U_{0k}(u), U_{0k}(v)\rangle = \langle U_{1k}(u), U_{1k}(v)\rangle + \langle U_{2k}(u), U_{2k}(v)\rangle = \int_0^{min(u,v)} \frac{W_c(w)dF(w)}{S^2(w)},$$
$$\langle U_{0k}(u), U_{jk}(v)\rangle = -\langle U_{jk}(u), U_{jk}(v)\rangle = -\int_0^{min(u,v)} \frac{W_c(w)dF_j(w)}{S^2(w)} (j = 1, 2),$$
$$\text{and } \langle U_{ik}(u), U_{jk}(v)\rangle = \delta_{ijk} \int_0^{min(u,v)} \frac{W_c(w)dF_i(w)}{S^2(w)},$$

where δ_{ijk} is the Kronecker delta. Therefore, we can write the covariance function of G as:

$$\begin{aligned}\langle G_{ik}(s), G_{jk}(t)\rangle = &\int_0^{min(s,t)} \{F_{ik}(s) - F_{ik}(u)\} \{F_{jk}(t) - F_{jk}(u)\} \frac{W_c(u)dF_k(u)}{S_k^2(u)} \\ &- \int_0^{min(s,t)} \{F_{ik}(s) - F_{ik}(u)\} \frac{W_c(u)dF_{jk}(u)}{S_k(u)} \\ &- \int_0^{min(s,t)} \{F_{jk}(t) - F_{jk}(u)\} \frac{W_c(u)dF_{ik}(u)}{S_k(u)} + \delta_{ijk} \int_0^{min(s,t)} W_c(u)dF_{ik}(u).\end{aligned}$$

Furthermore, we can write:

$$\begin{aligned}&var\left[\int_0^{\tau_k} \{\gamma_k G_{2k}(t) - (1-\gamma_k)G_{1k}(t)\} dt\right] \\ &= \int_0^{\tau_k}\int_0^{\tau_k} \left\{\gamma_k^2\langle G_{2k}(t), G_{2k}(s)\rangle - 2\gamma_k(1-\gamma_k)\langle G_{1k}(t), G_{2k}(s)\rangle + (1-\gamma_k)^2\langle G_{1k}(t), G_{1k}(s)\rangle\right\} dsdt.\end{aligned}$$

Hence, we have:

$$\begin{aligned}\sigma_k^2 = &\int_0^{\tau_k}\int_0^{\tau_k} \left\{\gamma_k^2\langle G_{2k}(t), G_{2k}(s)\rangle - 2\gamma_k(1-\gamma_k)\langle G_{1k}(t), G_{2k}(s)\rangle + (1-\gamma_k)^2\langle G_{1k}(t), G_{1k}(s)\rangle\right\} dsdt \\ &+ \langle G_{1k}(\tau_k), G_{1k}(\tau_k)\rangle \left\{\int_0^{\tau_k} F_k(t)dt\right\}^2 \\ &- 2\left[\int_0^{\tau_k} \{\gamma_k\langle G_{2k}(t), G_{1k}(\tau_k)\rangle - (1-\gamma_k)\langle G_{1k}(t), G_{1k}(\tau_k)\rangle\} dt\right] \int_0^{\tau_k} F_k(t)dt.\end{aligned}$$

Under H_0, the equation:

$$F_{1k}(.) = \frac{\gamma_k}{(1-\gamma_k)} F_{2k}(.)$$

is true and can be used to simplify the expression of the variance σ_k^2. After some tedious algebra, the variance expression reduces to:

$$\begin{aligned}\sigma_{0k}^2 = \gamma_k &\int_0^{\tau_k}\int_0^{\tau_k}\int_0^{min(s,t)} W_c(u)dF_{2k}(u)dtds \\ &+ \frac{\gamma_k}{(1-\gamma_k)^2}\left\{\int_0^{\tau_k} F_{2k}(t)dt\right\}^2 \int_0^{\tau_k} W_c(u)dF_{2k}(u) \\ &- \frac{2\gamma_k}{(1-\gamma_k)}\left\{\int_0^{\tau_k} F_{2k}(t)dt\right\}\int_0^{\tau_k}\int_0^{t} W_c(u)dF_{2k}(u)dt.\end{aligned}$$

□

2.7 SIMULATION STUDIES

2.7.1 Data Generation

To evaluate the performance of our test statistic in various scenarios, we conducted a series of Monte Carlo simulations under the null and alternative hypotheses. We were interested in the finite sample properties of our proposed stratified random signs censoring test statistic as well as how it compared to the unstratified random signs censoring test statistic. Thus, we assessed the type I error rate under the null hypothesis and the power under the random signs censoring and stratified random signs censoring alternative hypotheses for both test statistics.

In all simulations, datasets of sample sizes $n = 500,\ 1000\ and\ 1500$ were generated under different censoring, competing event, and covariate group proportions and replicated 1000 times. The non-informative censoring time C was generated independently from an exponential distribution with parameter η which was varied to produce censoring rates of 0%, 10%, 25%, and 50%. The covariate of interest Z is a binary variable in all scenarios generated from a Bernoulli distribution, where the probability of success parameter q indicated the probability of being in group B versus group A. We present the simulation results for $q = 1/2$; though not shown, various values of q were assessed.

Generation of event times vary for the different hypotheses; thus, in addition to the simulation results, the remaining data generation methods are discussed individually in the following subsections.

2.7.2 Simulation under the Null Hypothesis

In order to generate data under the null hypothesis, we used the Delay Time model whose properties are discussed in Section 2.3.2, or for greater detail one can refer to Baker and Wang [1993] or Hokstadt and Jensen [1998]. Recall that in this model, T_1 and T_2 are generated from a sum of exponentially distributed variables U, V, and W, such that:

$$T_1 = U + V$$
$$T_2 = U + W$$

The random variable U was generated from an exponential distribution with rate equal to 1, while the rates of V and W, were set to either 1 or 2 to create different proportions of main event occurrences. We produced proportions equal to about 1/3, 1/2, and 2/3 to represent situations where there are few main events, an equal number of main and competing events, and a large number of main events, respectively.

Table 1 shows the Monte Carlo estimates of the empirical levels under each design with varying the censoring percentages and sample sizes. We considered a nominal level of 5% in all three scenarios.

Under each design, the results for the stratified test are very similar to those of the unstratified test. In addition, there is little variation between the results from each design, indicating that changing the proportion of main events has minimal effect on the estimates. The empirical levels do seem to be closest to the nominal level when the percentage of main events is highest, but they are still close to the nominal levels in the other scenarios. Increases in sample size have little effect, and even sample sizes where $n = 500$ have fairly stable estimates. The largest fluctuations in the estimates are seen by changes in the number of censored observations. At high rates of censoring, such as 50%, the estimates begin to move farther away from nominal level, which is to be expected when the outcomes of the majority of the sample are not observed.

2.7.3 Simulation under Random Signs Censoring

For the random signs censoring alternative hypothesis H_1^*, we generated data that follow the random signs censoring assumption overall, and thereby follow the stratified random signs censoring assumption as well. To simulate this data, we generated T_1 from an exponential distribution with rate equal to 1. We then generated a random variable, ξ, from a uniform distribution from 0 to T_1. We also generated a random variable, π, from a Bernoulli distribution with parameter, p, which sets the probability of having a main event. The competing event time distribution, T_2, was then formed by using the following expression:

$$T_2 = T_1 - \{(2\pi - 1)\xi\}\,.$$

This formulation ensures the dominance relation between T_1 and T_2 necessary for the random signs censoring assumption to be met. Similar to the null scenario, we considered situations with various amounts of main events by letting $p = 0.25$, 0.50, or 0.75. Table 2 compares the power levels of the random signs censoring test statistic to our stratified random signs censoring test statistic under the various H_1^* scenarios.

One can note that the power of the stratified test is nearly equal to, if not higher than, that of the random signs censoring test. Similar to the simulation results under the null hypothesis, the proportion of main events has minimal effect, but there is a slight improvement with a higher main event rate. Increases in sample size have little to no gain in power until the rate of censoring is greater than 10%. Noticeable losses in power are not apparent until the percent of censored observations reaches 50%. However, even at the high censoring rates, the power levels are still reasonable.

2.7.4 Simulation under Stratified Random Signs Censoring

We also considered the stratified random signs censoring situation, H_1, where both covariate groups satisfy random signs censoring individually but not as an overall sample. This situation is of particular interest because the former random signs censoring test was not developed to handle such a situation. To create this scenario, we generated T_1 from a Cox proportional hazards model with a baseline exponential distribution and conditional on fixed

categorical covariate Z. By varying the β coefficient corresponding to covariate Z, we were able to change the range of times within the groups. The probability of having a main event p also needs to be conditional on the group; thus, we generated p from a logistic model with covariate Z. The value of the coefficient α in this case altered the probability of having a main event within in each. For the first group, group A, we set the probability of having a main event to 50%, and for group B, we varied the probability of having a main event between very low and high percentages, i.e. about 10% to 90%, for the different scenarios. Similar to the random signs censoring simulation, we generated a random variable, ξ, but here we used a triangle distribution from 0 to T_1 with mode equal to 0. We then generated the random variable, π, from a Bernoulli distribution with parameter p that was produced from the logistic distribution. The competing event time distribution, T_2, was then formed by using the following expression:

$$T_2 = T_1 - \{(2\pi - 1)\xi\}.$$

Generating the data in this manner creates the stratified random signs censoring situation where stratified by the covariate Z, the data follow random signs censoring, but overall (unstratified), they do not.

Table 3 shows the power of our stratified random signs censoring test statistic as well as the probability of the unstratified test statistic detecting random signs censoring under the different H_1 scenarios. The difference between the two tests is quite notable. The stratified test has high power in both scenarios, especially when the rate of censoring is not too high. There are some small decreases in power as censoring reaches 25%, and power noticeably drops when the censoring rate reaches 50%. The power of the test improves with sample size, the largest increases from which are seen at the higher rates of censoring. In addition, the stratified test seems to have slightly better power in the second scenario, where there is a higher percentage of main events in both groups.

On the other hand, the unstratified test has a small chance of detecting a random signs censoring under either scenario. Even with no addition right-censoring and a large sample size of $n = 1500$, there is only a slightly less than 40% chance the unstratified test will reject the null hypothesis. In smaller sample sizes or higher rates or censoring, the probability

of rejecting diminishes to less than 10%. One could conduct the unstratified test on each subgroup individually; however, there would be a loss in power due to the smaller sample sizes of each subgroup and type I error would be inflated due to the multiple comparisons. Hence, in scenarios such as these, there is a clear advantage to using the stratified random signs censoring test over the unstratified test.

2.8 APPLICATION

We applied the random signs censoring tests, stratified and unstratified, to data extracted from the United Network for Organ Sharing liver transplant waiting list. The aim of our analysis was to estimate the marginal survival distribution of death without liver transplantation, i.e. pre-transplant mortality, by treating the competing event, liver transplantation, as a random signs censoring.

The final cohort consisted of 2006 pediatric patients who were on the list February 27, 2002 through June 25, 2010, also referred to as the PELD era. Patients were excluded if they had living donors and a PELD score greater than or equal to 18. The sample consisted of nearly equal representations of the sexes, 52% females ($n = 1046$) and 48% males ($n = 960$). The average PELD score of the sample was about 5. Of the possible event outcomes, 73.1% were transplantations ($n = 1467$), 6.3% were deaths ($n = 146$), and 20.6% were right-censored ($n = 413$). To account for non-informative censoring, we used inverse probability censoring weight product-limit type estimators to estimate the normalized subsurvival curves of the main and competing events, death and transplant respectively.

We first considered the sample as a whole to see if it followed the random signs censoring assumption. In Figure 2.8, the estimated normalized subsurvival curves of competing events death and transplant are plotted for the unstratified sample. The graph leads to the conjecture that the overall sample may follow random signs censoring; however, the test was not statistically significant ($p = 0.216$). The graph fails to illustrate some of the characteristics of the data, such as the proportion of people with main versus competing events.

The numerical test accounts for these differences and concludes the unstratified data do not follow the assumption.

We then stratified the sample based on each patient's serum total bilirubin level, a factor known to be associated with liver health [Freeman et al., 2006, Wiesner et al., 2001, 2003]. The median bilirubin level of the sample, $3.7mg/dL$, was used as a cutpoint, such that those patients with a bilirubin level below $3.7mg/dL$ were considered the "Low Bilirubin" group while those with a level greater than or equal to $3.7mg/dL$ were named "High Bilirubin".

Figure 2.8 shows the estimated normalized subsurvival curves for the stratified sample. Based on the graph, it appears the Low Bilirubin group demonstrates the random signs censoring relationship while the High Bilirubin group does not. In fact, the dominance relationships between the normalized subsurvival curves of death and transplant are nearly opposite for the Low versus High Bilirubin groups, illustrating why treating the sample as a whole is insufficient.

The test for stratified random signs censoring was statistically significant ($p = 0.039$) at the 5% level, confirming the graphical findings. The unstratified test was unable to detect a random signs censoring relationship, likely because it was masked by the lack there of in the High Bilirubin group. One of the advantages to the stratified test is its ability to detect random signs censoring within strata, as demonstrated in this liver transplantation data application.

2.9 DISCUSSION /TABLES AND FIGURES

Competing risks are commonly encountered in statistical analyses, yet there are few existing methods that are appropriate for analyzing these data. In particular, there is a lack of methods to allow one to accurately estimate the marginal survival function of the latent failure times. Using the random signs censoring model is one solution to handle dependent risks that allows identifiability of the marginal distribution. Other assumptions on the dependence structure, such as independence, may allow for identifiability of the marginal survival function, but these assumptions are strong and untestable. We developed a test for

the stratified random signs censoring assumption that provides a test to detect presence of random signs censoring after stratifying by a categorical covariate.

We have also shown through simulation studies that our test is as powerful as the (unstratified) random signs censoring test when testing random signs censoring and is superior in stratified random signs censoring scenarios, where the unstratified test is not applicable. Moreover, the stratified random signs censoring test behaves well even with small sample sizes and fairly high rates of censoring. Though, one limitation of the test is its use with censoring rates close to 50% or higher, as these rates can lead to inflated type I error and low power levels. Another limitation of this test is that it is only applicable to categorical covariates. Indeed, it would be of interest to test for stratified random signs censoring in data containing continuous covariates. In addition, it has currently only been used with a single covariate of interest. In Chapter 3, we will consider a conditional random signs censoring test which can be used with multiple covariates, discrete or continuous. Overall, the stratified random signs censoring test is widely applicable in research settings and easy to implement, making it an ideal model candidate to improve the accuracy of current and future statistical models.

Table 1: Monte Carlo estimates of the type I error level of the stratified and unstratified random signs censoring tests under the null hypothesis, H_0

		N=500		N=1000		N=1500	
$\%T_1$	$\%C$	Stratified	Unstratified	Stratified	Unstratified	Stratified	Unstratified
33%	0%	0.048	0.050	0.053	0.053	0.055	0.055
	10%	0.057	0.059	0.045	0.043	0.043	0.044
	25%	0.064	0.060	0.048	0.049	0.047	0.043
	50%	0.056	0.050	0.066	0.057	0.059	0.053
50%	0%	0.048	0.046	0.049	0.047	0.060	0.059
	10%	0.050	0.058	0.051	0.052	0.041	0.041
	25%	0.048	0.050	0.049	0.053	0.045	0.046
	50%	0.065	0.058	0.060	0.057	0.052	0.051
67%	0%	0.061	0.061	0.054	0.054	0.051	0.050
	10%	0.064	0.065	0.055	0.052	0.053	0.059
	25%	0.065	0.065	0.054	0.051	0.051	0.051
	50%	0.093	0.093	0.073	0.073	0.065	0.065

$\%T_1$ denoting % main events and $\%C$ denoting % censored.

Table 2: Monte Carlo estimates of the power of the tests under random signs censoring, H_1^*

		N=500		N=1000		N=1500	
$\%T_1$	$\%C$	Stratified	Unstratified	Stratified	Unstratified	Stratified	Unstratified
25%	0%	1.00	1.00	1.00	1.00	1.00	1.00
	10%	1.00	1.00	1.00	1.00	1.00	1.00
	25%	0.92	0.90	0.99	0.98	1.00	0.99
	50%	0.45	0.44	0.59	0.57	0.60	0.57
50%	0%	1.00	1.00	1.00	1.00	1.00	1.00
	10%	1.00	1.00	1.00	1.00	1.00	1.00
	25%	0.98	0.97	1.00	0.99	1.00	1.00
	50%	0.67	0.64	0.80	0.77	0.82	0.80
75%	0%	1.00	1.00	1.00	1.00	1.00	1.00
	10%	1.00	1.00	1.00	1.00	1.00	1.00
	25%	0.96	0.96	0.99	0.99	1.00	1.00
	50%	0.75	0.74	0.82	0.80	0.86	0.84

$\%T_1$ denoting % main events and $\%C$ denoting % censored.

Table 3: Monte Carlo estimates of the power of the tests under stratified random signs censoring, H_1, with different proportions of main events in Groups A and B

		N=500		N=1000		N=1500	
$\%T_{1A}, T_{1B}$	$\%C$	Stratified	Unstratified	Stratified	Unstratified	Stratified	Unstratified
50%, 18%	0%	0.99	0.22	1.00	0.31	1.00	0.39
	10%	0.89	0.14	0.99	0.16	1.00	0.21
	25%	0.59	0.08	0.74	0.10	0.87	0.10
	50%	0.20	0.08	0.23	0.08	0.29	0.07
50%, 92%	0%	0.96	0.21	1.00	0.27	1.00	0.38
	10%	0.91	0.18	0.99	0.18	1.00	0.21
	25%	0.77	0.12	0.88	0.10	0.95	0.12
	50%	0.41	0.12	0.45	0.10	0.53	0.09

$\%T_{1A}, T_{1B}$ denoting % main events in Groups A and B respectively, and $\%C$ denoting % censored.

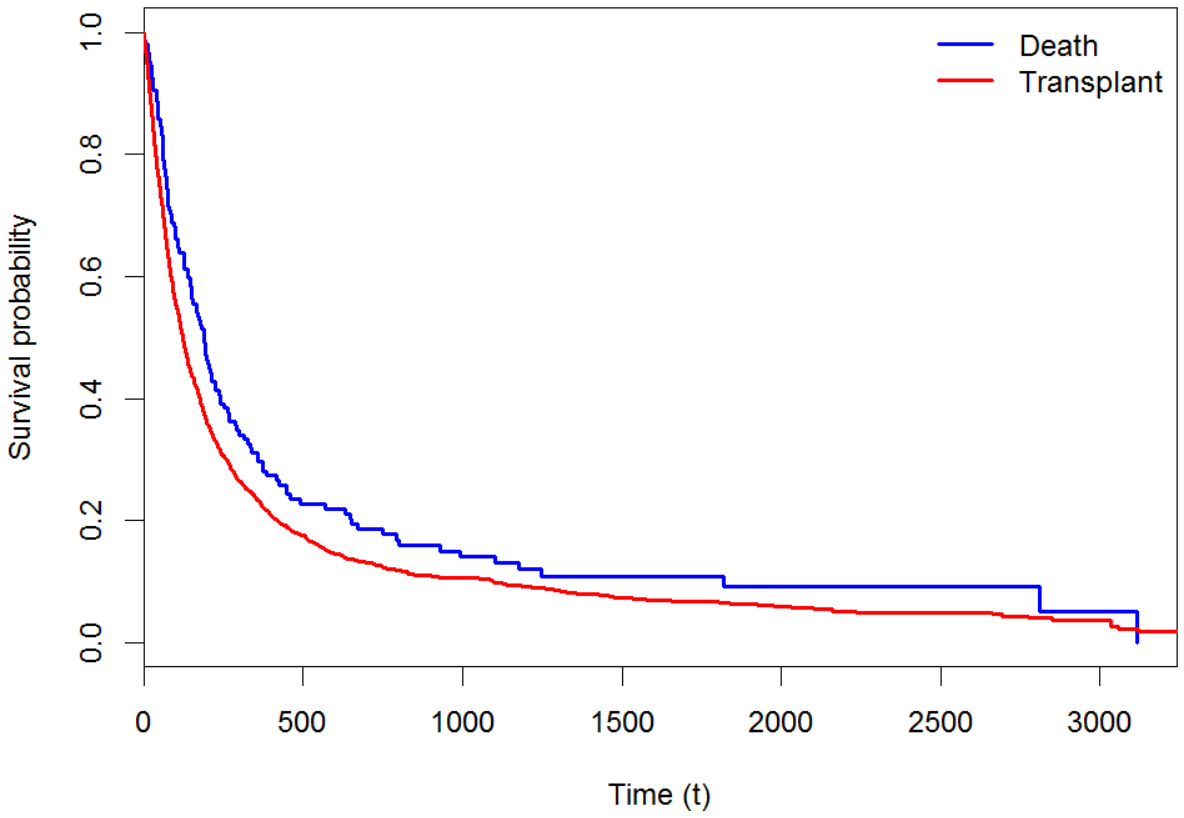

Figure 1: Normalized subsurvival curves of competing events death and transplant, overall

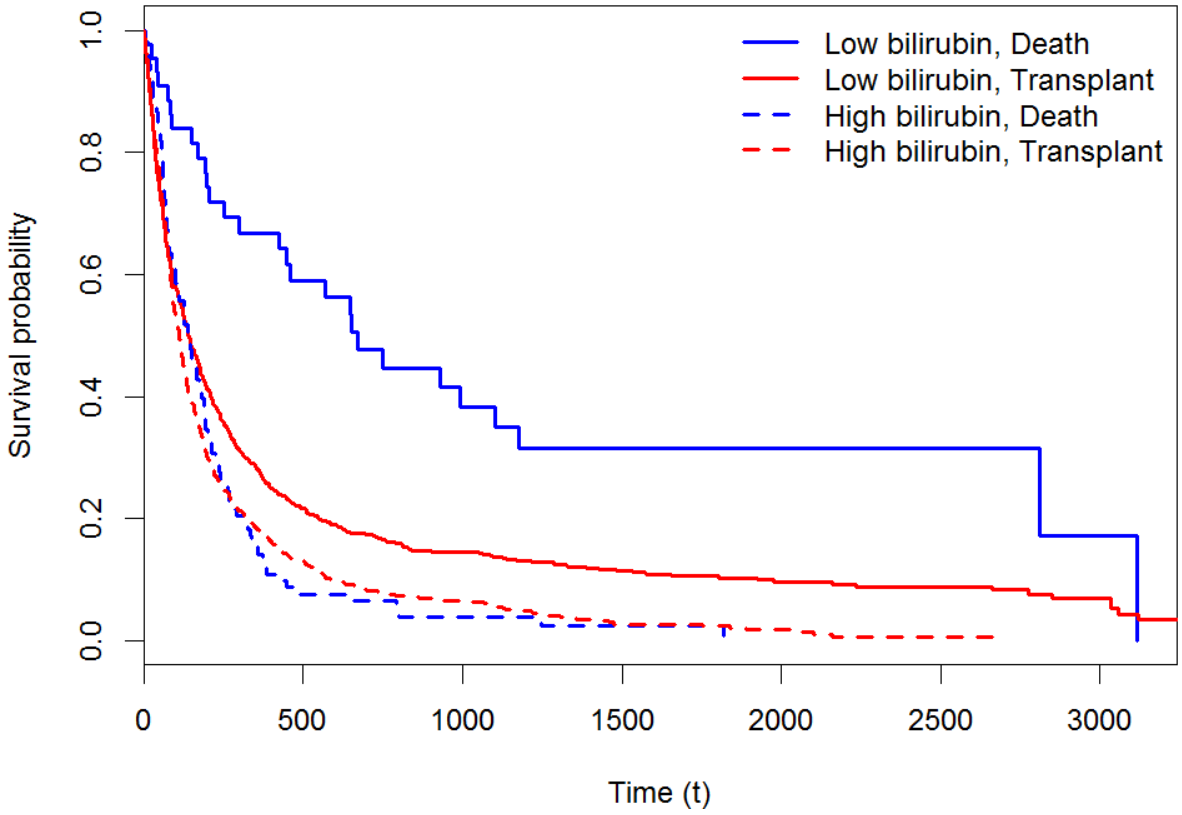

Figure 2: Normalized subsurvival curves of competing events death and transplant stratified by bilirubin level

3.0 TEST FOR CONDITIONAL RANDOM SIGNS CENSORING

3.1 INTRODUCTION

In the setting of competing risks, researchers often want to examine and model the effects of several covariates for a specific cause of failure [Zhang et al., 2008]. Therefore, in addition to testing stratified random signs censoring under the competing risks setting, we would like to test another extension of random signs censoring which adjusts for multiple types of covariate effects. Random signs censoring is useful; however it does not take into account covariate effects. We developed a test for stratified random signs censoring in Chapter 2, but it is only applicable with one categorical covariate. In practice we will more often encounter the situation with one or more covariates, some of which continuous. Thus, in order to account for these types of competing risk regression analyses, we need to look at an extension of both the random signs censoring and stratified random signs censoring. Specifically, we are interested in conditional random signs censoring, which allows for random signs censoring to be met after adjusting for covariates. These covariates could be categorical, continuous, or both. After conditioning on potential confounders, we would like to test whether the event indicator is independent of the main event conditional on a set of covariates. In other words, the main event time distribution and event type indicator are only dependent through a common set of covariates. If this assumption is satisfied, the marginal distribution of the latent failure time for the event of interest, T_1, can be modeled with the specified covariates. Modeling the marginal distribution of latent failure time T_1 allows one to assess the relationship between covariate effects and the associated risks specific to the individual event of interest in the absence of the other competing event.

Typical competing risk regression analyses concentrate on the crude incidence approach. In this setting, competing risks are represented as a bivariate random variable, (T, δ), where T is the time to first failure and δ indicates the type of failure observed. The joint distribution of (T, δ) can be completely specified through the cause-specific hazards or through the cumulative incidence functions [Porta Bleda et al., 2007]. These functions represent two distinct quantities, either of which can be modeled given a set of covariates and can appropriately assess risk in the presence of other competing risks [Klein and Andersen, 2005]. Unfortunately, because these quantities are only related to the joint distribution (T, δ) as opposed to that of latent failure times (T_1, T_2), they do not allow estimation of the marginal distribution of T_1. Therefore, they cannot determine if covariate effects are associated with risks involving one or both competing events Dignam et al. [2012]. As a result, one may conclude that a particular covariate is related to improving (or worsening) outcomes for the main event of interest when in reality the opposite is true, and the competing event was masking the effect.

Currently, the only methods available under the latent failure time approach that allow for the assessment of categorical and continuous covariate effects make assumptions on the dependence structure between T_1 and T_2 that are unverifiable [Pintilie, 2006]. Hence, we propose a test for conditional random signs censoring to fill this gap in the literature.

The remainder of this paper is structured as follows. Section 3.2 introduces the observable quantities used in the competing risks regression setting. Section 3.3 describes the details of the conditional random signs censoring assumption, including overall concept and verification. Section 3.4 discusses the development of the proposed test statistic, and Section 3.5 presents the corresponding asymptotic theory. In Section 3.6, the performance of the test is assessed through Monte Carlo simulations. Section 3.7 illustrates the application of the proposed methods using the liver transplant data. This chapter concludes with a final discussion given in Section 3.8.

3.2 COMPETING RISKS REGRESSION QUANTITIES

Compared to standard survival analyses, competing risks analyses are incomplete without investigating competing risks regression models [Kim, 2007]. They are needed to identify risk factors for each competing risk. As mentioned in Section 3.1, competing risks probabilities can be summarized by either the cause-specific (or crude) hazard rate or the cumulative incidence function. Neither makes any assumptions about the relationship between the competing risks, such as independence), and both are directly estimable from the observed data (T, δ) [Klein and Andersen, 2005]. Therefore, we now formally define these quantities and any additional quantities used in the development of the conditional random signs censoring test.

Let $\mathbf{Z}$ be a p-dimensional vector of measured covariates, categorical or continuous. Adjusting for $\mathbf{Z}$, we define the cumulative incidence function for the j^{th} cause:

$$F_j(t|\mathbf{Z}) = pr(T \leq t, \delta = j|\mathbf{Z}), \text{ for } j = 1, 2 \text{ and } t > 0.$$

The sum of these cumulative incidence functions for the main and competing events gives the adjusted cumulative distribution function, $F(t|\mathbf{Z}) = F_1(t|\mathbf{Z}) + F_2(t|\mathbf{Z})$. We introduce the cause-specific hazard rate:

$$\lambda_j(t|\mathbf{Z}) = \lim_{t \to 0} \frac{pr(t \leq t + dt, \delta = j|T \geq t, \mathbf{Z})}{t},$$

and the cumulative cause-specific hazard rate:

$$\Lambda_j(t|\mathbf{Z}) = \int_0^t \lambda_j(u|\mathbf{Z})du.$$

We also consider the overall hazard rate, $\lambda(t|\mathbf{Z}) = \lambda_1(t|\mathbf{Z}) + \lambda_2(t|\mathbf{Z})$, the corresponding cumulative hazard rate $\Lambda(t|\mathbf{Z}) = \int_0^t \lambda(u|\mathbf{Z})du$, and the overall survival function $S(t|\mathbf{Z}) = pr(T > t|\mathbf{Z})$.

Using these functions, the conditional cumulative incidence function of the j^{th} cause is:

$$F_j(t|\mathbf{Z}) = \int_0^t S(u^-|\mathbf{Z})d\Lambda_j(u|\mathbf{Z}),$$

where

$$S(t^{-}|\mathbf{Z}) = \prod_{u<t} \{1 - d\Lambda(u|\mathbf{Z})\}.$$

Lastly, we define the probability of a main event, i.e. event type 1, conditional on covariate vector $\mathbf{Z}$ by $\gamma(\mathbf{Z}) = pr(\delta = 1|\mathbf{Z})$.

3.3 CONDITIONAL RANDOM SIGNS CENSORING

3.3.1 Concept

The conditional random signs censoring model assumes that the main event time and the event type indicator are independent after adjusting for covariate effects. In the stratified random signs censoring setting in Chapter 2, we focused on whether random signs censoring was met within at least one subgroup of a categorical covariate. Conditional random signs censoring extends the stratified setting by allowing one to incorporate both categorical and continuous covariate effects. Instead of verifying whether one or more individual strata of a covariate follows random signs censoring, we are now considering the overall adjusted sample. Hence, we can simultaneously control for multiple covariate effects as opposed to one categorical variable. We now state the conditional random signs censoring assumption as $T_1|\mathbf{Z}$ is independent of $\delta|\mathbf{Z}$. Based on this assumption, we obtain the following result:

$$S_1^*(t|\mathbf{Z}) = pr(T > t|\delta = 1, \mathbf{Z}) = \frac{pr(T_1 > t|\mathbf{Z})pr(\delta = 1|\mathbf{Z})}{pr(\delta = 1|\mathbf{Z})} = pr(T_1 > t|\mathbf{Z}).$$

This result states that after conditioning on the continuous covariate $\mathbf{Z}$, the normalized subsurvival function of T_1 is equivalent to the marginal survival function T_1.

In practical settings, we are often interested in modeling a regression model that takes into account covariate effects. The conditional random signs censoring assumption allows one to incorporate those covariate effects and test whether the adjusted sample satisfies the properties of random signs censoring. If the conditional random signs censoring assumption is met, classic survival regression techniques can be used to model the marginal survival function. Unlike other assumptions used in modeling the marginal survival function, the conditional random signs censoring assumption is verifiable.

3.3.2 Verification

Conditioning on $\mathbf{Z}$, Cooke's Random Sign Censoring theorem can be updated to illustrate the conditional random signs censoring assumption as follows:

Theorem 4. *Let S_1 and S_2 be a subsurvival pair, and let $\mathbf{Z}$ denote a p-dimensional vector, consisting of categorical and/or continuous covariates. Then the following are equivalent:*

1. *There exists a pair (T_1, T_2) of life variables such that $T_2|\mathbf{Z}$ is a conditional random signs censoring of $T_1|\mathbf{Z}$, and:*
$$S_j^*(t|\mathbf{Z}) = \frac{S_j(t|\mathbf{Z})}{S_j(0|\mathbf{Z})}, \forall\, t \geq 0.$$
2. $$S_1^*(t|\mathbf{Z}) > S_2^*(t|\mathbf{Z}),\ \forall\, t > 0.$$

This theorem implies that the conditional random signs censoring assumption $T_1|\mathbf{Z}$ independent of $\delta|\mathbf{Z}$ is satisfied if and only if the condition $S_1^*(t|\mathbf{Z}) > S_2^*(t|\mathbf{Z})$ for all $t > 0$ is also satisfied. This conditional result is a more general extension of both the random signs censoring and stratified random signs censoring assumption in that it allows for estimation of the normalized subsurvival curves after adjusting for categorical and/or continuous covariates.

3.4 DEVELOPMENT OF THE CONDITIONAL TEST

Based on the extension of Cooke's random signs censoring theorem (4), testing the conditional random signs censoring assumption that $T_1|\mathbf{Z}$ and $\delta|\mathbf{Z}$ are independent is equivalent to testing whether $S_1^*(t|\mathbf{Z})$ dominates $S_2^*(t|\mathbf{Z})$ for all $t > 0$. Thus, if the conditional random signs censoring assumption does not hold, $T_1|\mathbf{Z}$ and $\delta|\mathbf{Z}$ are dependent even after conditioning on $\mathbf{Z}$, which is equivalent to testing whether $S_1^*(t|\mathbf{Z})$ is equivalent to $S_2^*(t|\mathbf{Z})$ for any $\mathbf{Z}$. Using this relationship, we obtain the null hypothesis:

$$H_0 : S_1^*(t|\mathbf{Z}) = S_2^*(t|\mathbf{Z}), \forall\, t > 0,$$

against the conditional random signs censoring alternative hypothesis:

$$H_2 : S_1^*(t|\mathbf{Z}) > S_2^*(t|\mathbf{Z}), \forall\, t > 0,$$

where the normalized subsurvival functions are conditional on the covariate vector $\mathbf{Z}$. While the unadjusted sample may not meet the requirements of random signs censoring, this test allows us to check whether the adjusted sample satisfies the conditional random signs censoring assumption.

We can develop the test statistic using equivalent representations of the normalized subsurvival functions, similiar to how we did for the stratified test in Section 2.5. However, now we will use the functions in terms of functions conditional on covariate vector $\mathbf{Z}$ instead of in terms of individual strata.

Under H_0 and noting that $\gamma(\mathbf{Z}) = pr(\delta = 1|\mathbf{Z}) = S_1(0|\mathbf{Z})$, we can show the functions $S_1(.|\mathbf{Z})$ and $S(.|\mathbf{Z})$ are proportional from the following equivalences:

$$\begin{aligned}
& S_1^*(t|\mathbf{Z}) = S_2^*(t|\mathbf{Z}) \\
\iff & \frac{S_1(t|\mathbf{Z})}{S_1(0|\mathbf{Z})} = \frac{S_2(t|\mathbf{Z})}{S_2(0|\mathbf{Z})} \\
\iff & \{1-\gamma(\mathbf{Z})\}\, S_1(t|\mathbf{Z}) = \gamma(\mathbf{Z}) S_2(t|\mathbf{Z}) \\
\iff & S_1(t|\mathbf{Z}) = \gamma(\mathbf{Z}) S(t|\mathbf{Z}) \\
\iff & \gamma(\mathbf{Z}) F_2(t|\mathbf{Z}) - \{1-\gamma(\mathbf{Z})\}\, F_1(t|\mathbf{Z}) = 0,
\end{aligned}$$

which are fulfilled for any $\mathbf{Z}$ and for all $t > 0$. Similarly, under H_2 the following equivalent properties are fulfilled for any $\mathbf{Z}$ and for all $t > 0$:

$$S_1^*(t|\mathbf{Z}) > S_2^*(t|\mathbf{Z}) \tag{3.1}$$

$$\iff \gamma(\mathbf{Z}) F_2(t|\mathbf{Z}) - \{1-\gamma(\mathbf{Z})\}\, F_1(t|\mathbf{Z}) > 0. \tag{3.2}$$

Using equivalence (3.2), a measure in favor of the alternative is given by summing across all time-points such that:

$$\psi(\mathbf{Z}) = \int_0^\tau [\gamma(\mathbf{Z}) F_2(t|\mathbf{Z}) - \{1-\gamma(\mathbf{Z})\}\, F_1(t|\mathbf{Z})] dt,$$

where τ is the right endpoint of the support of $F(t|\mathbf{Z})$. The $\psi(\mathbf{Z})$ function is calculated over the entire sample, but the conditional, or adjusted, values of the functions are plugged into the formula, similar to the idea of conditioning on covariates in a regression model.

As in Section 2.5, we want to allow for the possibility of independent right censoring by letting C denote the censoring random variable, independent from the random variable T. We also let $X = \min(T, C)$ and $\varepsilon = \delta I(T \leq C)$ again. We now observe for $i = 1, ..., n$ individuals $(X_i, \varepsilon_i, \mathbf{Z}_i)$, where X_i denotes the observed time, ε_i indicates the event, and $\mathbf{Z}_i$ gives the vector of covariate values, all specific to the i^{th} individual. The function $\psi(\mathbf{Z})$ can then be estimated by plugging in estimates of its components. Although other consistent estimators can be used, we will estimate $F_j(t|\mathbf{Z})$ using a Cox proportional cause-specific hazards model. That is,

$$\hat{F}_j(t|\mathbf{Z}) = \int_0^t \hat{S}(u^-|\mathbf{Z}) d\hat{\Lambda}_j(u|\mathbf{Z}),\ j = 1, 2,$$

where

$$\hat{S}(t|\mathbf{Z}) = exp\left\{ -\sum_{j=1}^{2} \hat{\Lambda}_{0j}(t) e^{\hat{\beta}_j^T \mathbf{Z}} \right\},$$

and

$$\hat{\Lambda}_j(t|\mathbf{Z}) = \hat{\Lambda}_{0j}(t) e^{\hat{\beta}_j^T \mathbf{Z}},\ j = 1, 2.$$

The vector of regression coefficients β_j can be estimated using a partial likelihood approach. Using counting process notation, the resulting score equation for obtaining $\hat{\beta}_j$ is defined as:

$$\mathbf{U}(\beta_j) = \sum_{i=1}^{n} \int_0^{\tau} \left\{ \mathbf{Z}_i - \frac{\sum_{k=1}^{n} Y_{jk}(t) e^{\beta_j^T \mathbf{Z}_k} \mathbf{Z}_k}{\sum_{k=1}^{n} Y_{jk}(t) e^{\beta_j^T \mathbf{Z}_k}} \right\} dN_j(t),$$

where $Y_{ji}(t) = I(X_i \geq t,\ \varepsilon_i = j)$, $N_j(t) = \sum_{i=1}^{n} I(X_i \leq t, \varepsilon_i = j)$, and τ is a time larger than any observed death time for the j^{th} event. Given $\hat{\beta}_j$, the Breslow [1974] estimates for the baseline cumulative hazard functions are:

$$\hat{\Lambda}_{0j}(t) = \int_0^{\tau} \left\{ \sum_{k=1}^{n} Y_{jk}(t) e^{\hat{\beta}_j^T} \right\}^{-1} dN_j(t),\ j = 1, 2.$$

The probability of having the main event type, $\gamma(\mathbf{Z})$, can also be consistently estimated by $\hat{\gamma}(\mathbf{Z}) = \hat{S}_1(0|\mathbf{Z})$. Similarly, we can estimate $1 - \gamma(\mathbf{Z})$, the probability of having the competing event, by $\hat{S}_2(0|\mathbf{Z}) = 1 - \hat{\gamma}(\mathbf{Z})$. Combining these quantities, an estimate of $\psi(\mathbf{Z})$ becomes:

$$\hat{\psi}(\mathbf{Z}) = \int_0^{\tau} [\hat{\gamma}(\mathbf{Z}) \hat{F}_2(t|\mathbf{Z}) - \{1 - \hat{\gamma}(\mathbf{Z})\} \hat{F}_1(t|\mathbf{Z})] dt.$$

Hence, the test statistic for the stratified test takes the form $\sqrt{n}\hat{\psi}(\mathbf{Z})/\hat{\theta}_0$, where $\hat{\theta}_0$ is a consistent estimator of θ_0, the standard deviation of $\psi(\mathbf{Z})$ under H_0, whose details are discussed in the following section (Section 3.5).

3.5 ASYMPTOTIC PROPERTIES

In this section, we will discuss the asymptotic properties of our test statistic. The asymptotic distribution of our test statistic is based on the following theorem:

Theorem 5. *Let us suppose that:*

$$\int_0^\tau W_c(u)dF(u|\mathbf{Z}) < \infty, \tag{3.3}$$

where $W_c(u)$ is the inverse probability censoring weight. Then $\sqrt{n}\left\{\hat{\psi}(\mathbf{Z}) - \psi(\mathbf{Z})\right\}$ converges weakly to a mean zero normal random variable G, with finite variance θ^2. Under H_0 the limiting variance can be expressed in the form of:

$$\begin{aligned}\theta_0^2 = \gamma(\mathbf{Z}) &\int_0^\tau \int_0^\tau \int_0^{min(s,t)} W_c(u)dF_2(u|\mathbf{Z})dtds \\ &+ \frac{\gamma(\mathbf{Z})}{\{1-\gamma(\mathbf{Z})\}^2}\left\{\int_0^\tau F_2(t|\mathbf{Z})dt\right\}^2 \int_0^\tau W_c(u)dF_2(u|\mathbf{Z}) \\ &- \frac{2\gamma(\mathbf{Z})}{1-\gamma(\mathbf{Z})}\left\{\int_0^\tau F_2(t|\mathbf{Z})dt\right\}\int_0^\tau \int_0^t W_c(u)dF_2(u|\mathbf{Z})dt.\end{aligned}$$

Proof of Theorem 5. The weak convergence of the following quantities was proved for a single stratum [Dauxois et al., 2014] and multiple strata (Section 2.6, Proof 2.6). The result can also be applied to the functions after conditioning on the independent covariate vector $\mathbf{Z}$. Under assumption (3.3), the following weak convergence holds in $D^3[0,\infty]$:

$$\begin{bmatrix}\hat{G}_0(\cdot|\mathbf{Z}) \\ \hat{G}_1(\cdot|\mathbf{Z}) \\ \hat{G}_2(\cdot|\mathbf{Z})\end{bmatrix} = \sqrt{n}\begin{bmatrix}\hat{S}(\cdot|\mathbf{Z}) - S(\cdot|\mathbf{Z}) \\ \hat{F}_1(\cdot|\mathbf{Z}) - F_1(\cdot|\mathbf{Z}) \\ \hat{F}_2(\cdot|\mathbf{Z}) - F_2(\cdot|\mathbf{Z})\end{bmatrix} \xrightarrow{D} \begin{bmatrix}G_0(\cdot|\mathbf{Z}) \\ G_1(\cdot|\mathbf{Z}) \\ G_2(\cdot|\mathbf{Z})\end{bmatrix}, \text{ as } n \to \infty, \tag{3.4}$$

where G is a mean zero Gaussian process defined by:

$$G_0(\cdot|\mathbf{Z}) = S(\cdot|\mathbf{Z})U_0(\cdot|\mathbf{Z}),$$
$$G_j(\cdot|\mathbf{Z}) = \int_0^{\cdot} \{F_j(\cdot|\mathbf{Z}) - F_j(u|\mathbf{Z})\}\, dU_0(u|\mathbf{Z}) + \int_0^{\cdot} S(u|\mathbf{Z}) dU_j(u|\mathbf{Z}),\ j = 1, 2,$$

and U_j's are mean zero Gaussian, square integrable and orthogonal local martingales with covariance function:

$$\langle U_j(s|\mathbf{Z}), U_j(t|\mathbf{Z})\rangle = \int_0^{min(s,t)} \frac{W_c(u) dF_j(u|\mathbf{Z})}{S^2(u|\mathbf{Z})},\ j = 1, 2,$$

and $U_0 = -(U_1 + U_2)$. We can write:

$$\sqrt{n}\left\{\hat{\psi}(\mathbf{Z}) - \psi(\mathbf{Z})\right\} = \sqrt{n}\left\{\Psi(\hat{F}_1, \hat{F}_2) - \Psi(F_1, F_2)\right\},$$

where

$$\Psi(F_1, F_2) = \int_0^{\tau} [\gamma(\mathbf{Z}) F_2(t|\mathbf{Z}) - \{1 - \gamma(\mathbf{Z})\} F_1(t|\mathbf{Z})] dt$$

and $\gamma(\mathbf{Z}) = F_1(\tau|\mathbf{Z})$. The function $\Psi(F_1, F_2)$ is Hadamard-differentiable (see e.g. Van Der Vaart and Wellner [1996]); hence, we have the derivative:

$$D_{\Psi}^{F_1,F_2}(\alpha_1, \alpha_2) = \int_0^{\tau} [F_1(\tau|\mathbf{Z})\alpha_2(t|\mathbf{Z}) + \alpha_1(\tau|\mathbf{Z})F_2(t|\mathbf{Z}) - \{1 - F_1(\tau|\mathbf{Z})\}\alpha_1(t|\mathbf{Z}) + \alpha_1(\tau|\mathbf{Z})F_1(t|\mathbf{Z})]\, dt.$$

The Hadamard differentiability of the function and convergence result (2.4) allow us to apply the functional delta method as described in Theorem 3.9.5 of Van Der Vaart and Wellner [1996] so that we have:

$$\sqrt{n}\left\{\hat{\psi}(\mathbf{Z}) - \psi(\mathbf{Z})\right\} \xrightarrow{D} D_{\Psi}^{F_1,F_2}(G_1, G_2)\ as\ n \to \infty,$$

where

$$D_{\Psi}^{F_1,F_2}(G_1, G_2) = \int_0^{\tau} [\gamma(\mathbf{Z})G_2(t|\mathbf{Z}) - \{1 - \gamma(\mathbf{Z})\} G_1(t|\mathbf{Z})]\, dt - G_1(\tau|\mathbf{Z}) \int_0^{\tau} F(t|\mathbf{Z}) dt.$$

The limiting distribution of the random variable is Gaussian with mean zero and variance function:

$$\begin{aligned}\theta^2 &= var\left(\int_0^\tau \left[\gamma(\mathbf{Z})G_2(t|\mathbf{Z}) - \{1-\gamma(\mathbf{Z})\}\, G_1(t|\mathbf{Z})\right] dt - G_1(\tau|\mathbf{Z})\int_0^\tau F(t|\mathbf{Z})dt\right) \\ &= var\left(\int_0^\tau \left[\gamma(\mathbf{Z})G_2(t|\mathbf{Z}) - \{1-\gamma(\mathbf{Z})\}\, G_1(t|\mathbf{Z})\right] dt\right) + var\left\{G_1(\tau|\mathbf{Z})\right\}\left\{\int_0^\tau F(t|\mathbf{Z})dt\right\}^2 \\ &\quad - 2\left(\int_0^\tau \left[\gamma(\mathbf{Z})\langle G_2(t|\mathbf{Z}), G_1(\tau|\mathbf{Z})\rangle - \{1-\gamma(\mathbf{Z})\}\, \langle G_1(t|\mathbf{Z}), G_1(\tau|\mathbf{Z})\rangle\right] dt\right)\int_0^\tau F(t|\mathbf{Z})dt\end{aligned}$$

The covariance structure of G_j for $j = 1, 2$ is given by:

$$\begin{aligned}\langle G_i(s|\mathbf{Z}), G_j(t|\mathbf{Z})\rangle &= \int_0^s\int_0^t \{F_i(s|\mathbf{Z}) - F_i(u|\mathbf{Z})\}\{F_j(t|\mathbf{Z}) - F_j(v|\mathbf{Z})\}\, d\langle U_0(|\mathbf{Z}u), U_0(v|\mathbf{Z})\rangle \\ &\quad + \int_0^s\int_0^t \{F_i(s|\mathbf{Z}) - F_i(u|\mathbf{Z})\}\, S(v|\mathbf{Z}) d\langle U_0(u|\mathbf{Z}), U_j(v|\mathbf{Z})\rangle \\ &\quad + \int_0^s\int_0^t S(u|\mathbf{Z})\{F_j(t|\mathbf{Z}) - F_j(v|\mathbf{Z})\}\, d\langle U_i(u|\mathbf{Z}), U_0(v|\mathbf{Z})\rangle \\ &\quad + \int_0^s\int_0^t S(u|\mathbf{Z})S(v|\mathbf{Z}) d\langle U_i(u|\mathbf{Z}), U_j(v|\mathbf{Z})\rangle.\end{aligned}$$

However, due to the orthogonality of U_1 and U_2, we have:

$$\begin{aligned}\langle U_0(u|\mathbf{Z}), U_0(v|\mathbf{Z})\rangle &= \langle U_1(u|\mathbf{Z}), U_1(v|\mathbf{Z})\rangle + \langle U_2(u|\mathbf{Z}), U_2(v|\mathbf{Z})\rangle \\ &= \int_0^{min(u,v)} \frac{W_c(w)dF(w|\mathbf{Z})}{S^2(w|\mathbf{Z})}, \\ \langle U_0(u|\mathbf{Z}), U_j(v|\mathbf{Z})\rangle &= -\langle U_j(u|\mathbf{Z}), U_j(v|\mathbf{Z})\rangle = -\int_0^{min(u,v)} \frac{W_c(w)dF_j(w|\mathbf{Z})}{S^2(w|\mathbf{Z})} (j = 1, 2), \\ \text{and } \langle U_i(u|\mathbf{Z}), U_j(v|\mathbf{Z})\rangle &= \delta_{ij}\int_0^{min(u,v)} \frac{W_c(w)dF_i(w|\mathbf{Z})}{S^2(w|\mathbf{Z})},\end{aligned}$$

where δ_{ij} is the Kronecker delta. Therefore, we can write the covariance function of Z as:

$$\begin{aligned}\langle G_i(s|\mathbf{Z}), G_j(t|\mathbf{Z})\rangle &= \int_0^{min(s,t)} \{F_i(s|\mathbf{Z}) - F_i(u|\mathbf{Z})\}\{F_j(t|\mathbf{Z}) - F_j(u|\mathbf{Z})\}\frac{W_c(u)dF_k(u|\mathbf{Z})}{S_k^2(u|\mathbf{Z})} \\ &\quad - \int_0^{min(s,t)} \{F_i(s|\mathbf{Z}) - F_i(u|\mathbf{Z})\}\frac{W_c(u)dF_j(u|\mathbf{Z})}{S_k(u|\mathbf{Z})} \\ &\quad - \int_0^{min(s,t)} \{F_j(t|\mathbf{Z}) - F_j(u|\mathbf{Z})\}\frac{W_c(u)dF_i(u|\mathbf{Z})}{S(u|\mathbf{Z})} + \delta_{ij}\int_0^{min(s,t)} W_c(u)dF_i(u|\mathbf{Z}).\end{aligned}$$

Furthermore, we can write:

$$
\begin{aligned}
&var\left(\int_0^\tau \left[\gamma(\mathbf{Z})G_2(t|\mathbf{Z}) - \{1-\gamma(\mathbf{Z})\}\, G_1(t|\mathbf{Z})\right] dt\right) \\
&= \int_0^\tau \int_0^\tau \gamma(\mathbf{Z})^2 \langle G_2(t|\mathbf{Z}), G_2(s|\mathbf{Z})\rangle - 2\gamma(\mathbf{Z})\,\{1-\gamma(\mathbf{Z})\}\,\langle G_1(t|\mathbf{Z}), G_2(s|\mathbf{Z})\rangle \\
&\quad + \{1-\gamma(\mathbf{Z})\}^2\,\langle G_1(t|\mathbf{Z}), G_1(s|\mathbf{Z})\rangle dsdt.
\end{aligned}
$$

Hence, we have:

$$
\begin{aligned}
\theta^2 = &\int_0^\tau \int_0^\tau \gamma(\mathbf{Z})^2 \langle G_2(t|\mathbf{Z}), G_2(s|\mathbf{Z})\rangle - 2\gamma(\mathbf{Z})\,\{1-\gamma(\mathbf{Z})\}\,\langle G_1(t|\mathbf{Z}), G_2(s|\mathbf{Z})\rangle \\
&+ \{1-\gamma(\mathbf{Z})\}^2\,\langle G_1(t|\mathbf{Z}), G_1(s|\mathbf{Z})\rangle dsdt + var\left\{G_1(\tau|\mathbf{Z})\right\}\left\{\int_0^\tau F(t|\mathbf{Z})dt\right\}^2 \\
&- 2\left(\int_0^\tau \left[\gamma(\mathbf{Z})\langle G_2(t|\mathbf{Z}), G_1(\tau|\mathbf{Z})\rangle - \{1-\gamma(\mathbf{Z})\}\,\langle G_1(t|\mathbf{Z}), G_1(\tau|\mathbf{Z})\rangle\right] dt\right)\int_0^\tau F(t|\mathbf{Z})dt.
\end{aligned}
$$

Under H_0, the equation:

$$
F_1(\cdot|\mathbf{Z}) = \frac{\gamma(\mathbf{Z})}{\{1-\gamma(\mathbf{Z})\}} F_2(\cdot|\mathbf{Z})
$$

is true and can be used to simplify the expression of the variance θ^2. After some tedious algebra, the variance expression reduces to:

$$
\begin{aligned}
\theta_0^2 = &\gamma(\mathbf{Z})\int_0^\tau \int_0^\tau \int_0^{min(s,t)} W_c(u)dF_2(u|\mathbf{Z})dtds \\
&+ \frac{\gamma(\mathbf{Z})}{\{1-\gamma(\mathbf{Z})\}^2}\left\{\int_0^\tau F_2(t|\mathbf{Z})dt\right\}^2 \int_0^\tau W_c(u)dF_2(u|\mathbf{Z}) \\
&- \frac{2\gamma(\mathbf{Z})}{1-\gamma(\mathbf{Z})}\left\{\int_0^\tau F_2(t|\mathbf{Z})dt\right\}\int_0^\tau \int_0^t W_c(u)dF_2(u|\mathbf{Z})dt.
\end{aligned}
$$

□

3.6 SIMULATION STUDIES

We conducted Monte Carlo simulations to evaluate the finite sample properties of our proposed test statistic. We carried out simulations under the null and alternatives hypotheses, creating various scenarios by differing the sample size, proportion of competing events, and rate of random right censoring. We report the estimated type I error for simulations under the null hypothesis and the estimated power for simulations under the alternatives. The results of the proposed conditional random signs censoring test statistic were then compared to the standard (unconditional) random signs censoring statistic.

For each scenario, datasets with sample sizes of 500, 1000, and 1500 were generated and replicated 1000 times. Independent non-informative censoring time C was generated using an exponential distribution with rate η. Different values of the parameter η were chosen to produce varying amounts of censoring ranging from 0% to 50%. For simplicity, we considered a single continuous covariate Z and generated it from a standard normal distribution.

Due to inherent differences of the event times under the null and alternative hypotheses, further details regarding data generation are discussed in the following subsections.

3.6.1 Simulation under the Null Hypothesis

To investigate the performance of the conditional random signs censoring test statistic under the null hypothesis, we generated data using the Delay Time model as was done previously for the stratified random signs censoring test statistic. Please refer to Section 2.7.2 for details regarding the data generation process of the event times. By varying event time distribution parameters, we once again produced proportions of main event occurrences equal to about 1/3, 1/2, and 2/3 to create scenarios with a few, an equal, and a large number of main versus competing events, respectively.

The Monte Carlo estimates of the type I error levels under each design with varying sample sizes and censoring rates are shown in Table 4. The empirical levels for both test statistics are very close to the nominal level of 0.05 across all combinations. There are some observed increases as the percentage of main events gets larger. The type I error rate for the

conditional test also increases slightly as the rate of censoring increases. On the contrary, the type I error rate for the overall random signs censoring test tends to decrease as the censoring rate increases. Specifically, for a sample size of $n = 1000$ and 33% main events, the conditional test has a type I error rate of 0.050 with 0% censoring and 0.059 with 25% censoring, whereas the unadjusted test has levels of 0.052 and 0.039, respectively.

3.6.2 Simulation under Random Signs Censoring

The data generation process of the event times under the random signs censoring alternative H_1^* is identical to that which is described in Section 2.7.3. The parameter specifications of $p = 0.25$, 0.50, or 0.75 were also used again, resulting in the same overall designs. Aside from the test statistic, the main difference between the set-up for this simulation and the previous version (presented in Section 2.7.3) is that Z is now a continuous covariate instead of categorical.

Table 5 presents the power levels under the various H_1^* scenarios for the conditional and unconditional random signs censoring test statistics. The power levels of both statistics are high under the random signs censoring alternative for the various scenarios. For larger sample sizes of $n = 1000$ and $n = 1500$, the power levels do not decrease until censoring reaches 50% and even then the loss is minimal. When the sample size only consists of $n = 500$, adequate power is still maintained at 50% censoring. Similar to the results under the null hypothesis, there is little change in the estimates as the percentage of main events change but small increases are noted as the number of main events increase. While both statistics have high power estimates for all combinations of censoring, main events, and sample size, the conditional test actually has higher power than the overall test in every scenario.

3.6.3 Simulation under Conditional Random Signs Censoring

Because the type I error rates were upheld, we explored a variety of scenarios under H_2 to investigate the power of the conditional random signs censoring test statistic. Under this alternative, the data do not satisfy the randoms signs censoring assumption marginally; however, they do follow it conditionally, after adjusting for covariate Z. Similar to the

stratified random signs censoring alternative, the standard (unconditional) test was not developed to handle this type of scenario. To generate the event times, we used an approach to similar that of H_1 (see Section 2.7.4). First, T_1 was generated from a Cox proportional hazards model using a baseline exponential distribution and conditional on fixed continuous covariate Z with regression coefficient β. The probability of having a main event, p, was generated from a logistic model with covariate Z and corresponding regression coefficient α. The random variable π, which determines whether a main or competing event occurred, was then generated from a Bernoulli distribution using parameter p. Thus, both the main event time T_1 and event indicator π are conditional on continuous covariate Z. The competing event time T_2 is also conditional on Z because it was generated using the following expression:

$$T_2 = T_1 - \{(2\pi - 1)\xi\},$$

where ξ is a random variable generated from the triangle distribution from 0 to T_1. Hence, this method of data generation creates a random signs censoring relationship between T_1 and T_2 that is conditional on continuous covariate Z.

We generated data this way under two different parameter designs, A and B, to vary the distribution of main and competing events in relation to the covariate values. Under design A, observations with positive covariate values were more likely to have a main event and observations with negative covariate values were more likely to have a competing event. The opposite was true for design B, where negative covariate values increased the likelihood of having a main event and positive values a competing event. Design A also had a larger difference between the average covariate value for a main event versus competing event. The average covariate values for main and competing events under design B were both closer to 0, creating a slightly weaker correlation between the covariate value and event outcome.

Table 6 compares the power of the conditional random signs censoring test statistic to the standard random signs censoring test statistic under this alternative for the different combinations of design type, sample size, and censoring rate.

The difference in results are quite striking. The conditional test has higher power than the standard test across all scenarios. The more notable difference is seen under design A, where the standard random signs censoring test statistic has little to no power regardless

of sample size. The conditional test, on the other hand, has adequate power for all sample sizes and censoring up to 25%. There is some loss in power with the smaller sample size of $n = 500$ relative to the other sample sizes, especially with additional censoring, but large decreases in power are not seen until 50% censoring.

Under design B, the superiority of the conditional test is still seen, but the overall test shows increased power levels compared to the previous levels under design A. This increase is likely due to the weaker conditional relationship or dependency between the observed covariate values and main event occurrences. The power estimates for the conditional test are more stable than the overall test and still larger for every sample size and censoring rate, particularly for a sample size of $n = 500$. The power estimates of the overall test fluctuate, increasing and decreasing as the rate of censoring gets larger. While the overall test may have improved from design A to B (relative to itself), there is a strong and clear advantage in using the conditional test.

3.7 APPLICATION

Using the same UNOS liver transplant data from Section 2.8, we applied the conditional and unconditional random signs censoring tests. Recall the final cohort consisted of 2006 pediatric patients who were on the list during the PELD era, i.e. February 27, 2002 through June 25, 2010. Exclusions included patients who had living donors or a PELD score greater than or equal to 18. For more details regarding the final cohort, please refer to Section 2.8.

We first analyzed the overall sample, not adjusting for any covariates, and performed the random signs censoring test. Because we are using the sample cohort as before and this overall test does not allow for inclusion of covariates of any kind, the results are the same as those presented in Section 2.8. Note that the standard random signs censoring test was not statistically significant ($p = 0.216$).

We then planned to conduct the conditional random signs censoring adjusting for each patient's serum total bilirubin level, a continuous covariate. When we performed the stratified test, we divided the sample into two groups (Low Bilirubin and High Bilirubin) based on the

median bilirubin level of the sample. For the conditional test, we can use the original continuous version of bilirubin as the conditional test can adjust for categorical and/or continuous covariates. However, because the conditional test uses a Cox proportional hazards model, any covariates used must satisfy the proportional hazards assumption. Hence, we tested whether the assumption holds for bilirubin and found evidence to contradict proportionality ($p = 0.033$). To correct for this proportional hazards violation, we added an interaction with time to the Cox model, which did allow us to satisfy the assumption ($p = 0.451$). After confirming the other model assumptions were met, we were able to implement the test for conditional random signs censoring. Figure 3.7 shows the predicted values, after adjusting for serum total bilirubin level, of the normalized subsurvival curves for the competing events, death and transplant. The test was statistically significant ($p = 0.006$) at the 5% level, indicating that the data follow the conditional random signs censoring assumption after adjusting for bilirubin level. This result also verifies that the marginal survival of pre-transplant mortality can be consistently estimated using only the observed data. For the purposes of this demonstration, we only considered adjusting for bilirubin, but future work need not be limited to a single predictor.

3.8 DISCUSSION /TABLES AND FIGURE

Testing the conditional random signs censoring assumption provides a way to detect whether the random signs censoring assumption is met after adjusting for categorical and/or continuous covariates in a competing risks dataset. Data satisfying the conditional random signs censoring assumption can then be used to consistently estimate the marginal survival, whereas current methods rely on unverifiable assumptions and can lead to biased estimates. Moreover, the conditional random signs censoring assumption is verifiable using only the observed data and makes estimation of the marginal survival extremely easy and straightforward.

In this study, we developed the test statistic for the conditional random signs censoring assumption. We derived its asymptotic properties and established asymptotic normality.

Through simulation studies, we showed that our test statistic maintains type I error rates close to the nominal level and has greater power than the standard method under random signs censoring. We also illustrated its dominance over the standard test in cases of conditional random signs censoring, where the standard test fails to detect the random signs censoring relationship that exists after adjusting for covariates. Finally, we implemented the test statistic to an example using the liver transplant data. A limitation of this test is its use in scenarios of high censoring rates of 50% or higher, which can lead to inflated type I error levels. In addition, like many other survival methods, we assumed censoring to be non-informative. Thus, this test may not be valid if censoring is informative, and future research would be needed to incorporate such censoring.

Table 4: Monte Carlo estimates of the type I error level of the conditional and overall random signs censoring tests under the null hypothesis, H_0

		N=500		N=1000		N=1500	
$\%T_1$	Censoring	Conditional	Overall	Conditional	Overall	Conditional	Overall
33%	0%	0.051	0.053	0.050	0.052	0.058	0.057
	10%	0.055	0.046	0.048	0.044	0.060	0.056
	25%	0.054	0.041	0.059	0.039	0.060	0.046
	50%	0.060	0.048	0.070	0.049	0.054	0.035
50%	0%	0.054	0.060	0.053	0.053	0.058	0.065
	10%	0.053	0.050	0.048	0.045	0.065	0.057
	25%	0.060	0.050	0.057	0.044	0.055	0.037
	50%	0.076	0.053	0.077	0.047	0.058	0.034
67%	0%	0.054	0.060	0.050	0.055	0.064	0.065
	10%	0.055	0.063	0.055	0.050	0.059	0.054
	25%	0.069	0.060	0.049	0.039	0.066	0.045
	50%	0.086	0.059	0.074	0.048	0.082	0.052

Table 5: Monte Carlo estimates of the power of the conditional and overall tests under random signs censoring, H_1^*

		N=500		N=1000		N=1500	
$\%T_1$	Censoring	Conditional	Overall	Conditional	Overall	Conditional	Overall
25%	0%	1.00	1.00	1.00	1.00	1.00	1.00
	10%	1.00	1.00	1.00	1.00	1.00	1.00
	25%	0.97	0.93	1.00	0.98	1.00	1.00
	50%	0.66	0.53	0.82	0.67	0.90	0.77
50%	0%	1.00	1.00	1.00	1.00	1.00	1.00
	10%	1.00	1.00	1.00	1.00	1.00	1.00
	25%	1.00	0.97	1.00	0.99	1.00	1.00
	50%	0.85	0.75	0.91	0.80	0.97	0.90
75%	0%	1.00	1.00	1.00	1.00	1.00	1.00
	10%	1.00	1.00	1.00	1.00	1.00	1.00
	25%	0.98	0.96	1.00	0.99	1.00	1.00
	50%	0.84	0.77	0.94	0.87	0.97	0.91

Table 6: Monte Carlo estimates of the power of the tests under conditional random signs censoring, H_2

		N=500		N=1000		N=1500	
Design	Censoring	Conditional	Overall	Conditional	Overall	Conditional	Overall
A	0%	0.86	0.13	0.99	0.18	1.00	0.20
	10%	0.72	0.02	0.94	0.01	0.99	0.01
	25%	0.50	0.00	0.73	0.00	0.86	0.00
	50%	0.17	0.00	0.26	0.00	0.32	0.00
B	0%	0.93	0.49	0.99	0.69	1.00	0.84
	10%	0.85	0.62	0.98	0.82	1.00	0.93
	25%	0.73	0.63	0.91	0.81	0.97	0.90
	50%	0.53	0.51	0.68	0.65	0.79	0.74

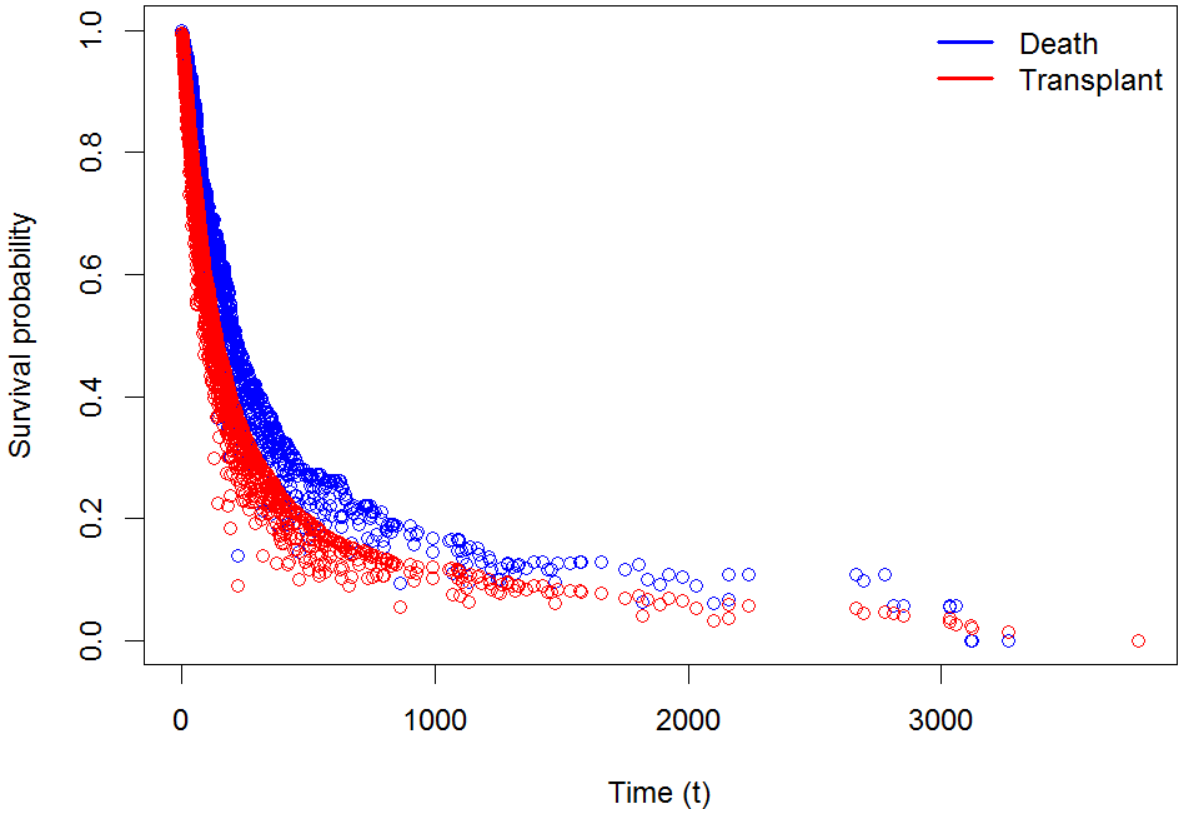

Figure 3: Predicted values of normalized subsurvival of competing events death and transplant conditional on bilirubin level

4.0 SUMMARY

Competing risks are commonly encountered in many biomedical studies where multiple causes of failure are present. Several methods exist for estimating the cause-specific hazards and cumulative incidence functions, but few methods are available for a marginal time-to-event analysis. Estimation of the marginal survival function is desirable so that researchers can assess risks associated with a specific cause of interest as opposed to in-conjunction with other competing events. The majority of the current approaches involve making unverifiable assumptions on the dependency structure between competing events, such as claiming the events are independent. The random signs censoring assumption is an alternative approach that can be verified through the observed data. Formal testing procedures for the random signs censoring assumption have only recently been developed, and they do not allow for covariate effects. It is important to incorporate covariates in order to analyze the risk relationship between them and the event of interest. In this dissertation, we developed two new tests as extensions of the random signs censoring test but that do allow for inclusion of covariate analyses.

In Chapter 2, we proposed a test for the stratified random signs censoring assumption, which tests whether the random signs censoring assumption is met within at least one group of a categorical covariate. Even if the overall sample does not satisfy the random signs censoring assumption, it is possible that a specified subgroup may. If one were to try to apply the unstratified test to each subgroup individually, not only would it be tedious and time consuming, but it would inflate the type I error rate. The stratified test offers an omnibus test that can check all of the subgroups of a categorical covariate at once. We showed analytically that our test statistic has an asymptotically normal distribution. Simulation studies showed that even with random right censoring rates up to about 50% and a relatively small sample

size of $n = 200$, the proposed test statistic maintains a type I error rate close to the nominal level. An example using the liver transplant data exhibited the utility of the stratified test.

In Chapter 3, we further extended the random signs censoring assumption to allow for adjustment of categorical and/or continuous covariates and developed a test for conditional random signs censoring. This test is a promising development in the competing risks field, allowing researchers the potential to assess multiple covariate effects on the marginal survival function for an event of interest. We established the asymptotic properties of our test statistic and derived an estimator for the variance. We were able to show that this test statistic also has an asymptotically normal distribution. Lastly, we evaluated its finite sample properties through simulation and demonstrated its use through an application to the liver transplant data.

Future work in this area of competing risks can consider possible extensions of the stratified and/or conditional random signs censoring tests. It would be interesting to explore other potential model choices (aside from the Cox model) to estimate the cumulative incidence functions for the conditional random signs censoring test. Due to the model-based nature of the test, improvement could be seen from choosing a more accurate underlying model to estimate the cause-specific hazard functions. In addition, both the stratified and conditional tests are currently limited to fixed covariates, but it could be useful to be able to include time-dependent covariates.

APPENDIX

PROOF OF CONDITIONAL RANDOM SIGNS CENSORING THEOREM

Proof of Theorem 4. Let ξ denote a random variable such that, $\xi = T_1 - T_2$. If $T_2|\mathbf{Z}$ is a conditional random signs censoring of $T_1|\mathbf{Z}$, then by definition, $T_1|\mathbf{Z}$ is independent of the sign of $(T_1 - T_2)|\mathbf{Z}$, i.e. $T_1 \perp sign(\xi)|\mathbf{Z}$.

Using this notation, we first prove part $1 \rightarrow 2$:

Since $T_1 \perp sign(\xi)|\mathbf{Z}$,

$$
\begin{aligned}
S_1(t|\mathbf{Z}) &= pr\{T_1 > t, T_1 < T_2|\mathbf{Z}\}, \text{ by definition} \\
&= pr\{T_1 > t, \xi < 0|\mathbf{Z}\}, \text{ since } \xi = T_1 - T_2 \\
&= pr\{T_1 > t|\xi < 0, \mathbf{Z}\}pr\{\xi < 0|\mathbf{Z}\}, \text{ by Bayes' Theroem} \\
&= pr\{T_1 > t|\xi < 0, \mathbf{Z}\}S_1(0|\mathbf{Z}), \text{since } S_1(0|\mathbf{Z}) = pr\{\xi < 0|\mathbf{Z}\} \\
&= pr\{T_1 > t|\xi > 0, \mathbf{Z}\}S_1(0|\mathbf{Z}), \text{ because } T_1 \perp sign(\xi)|\mathbf{Z}, \text{ and} \\
S_2(t|\mathbf{Z}) &= pr\{T_2 > t, T_2 < T_1|\mathbf{Z}\}, \text{ by definition} \\
&= pr\{T_1 - \xi > t, \xi > 0|\mathbf{Z}\}, \text{ since } \xi = T_1 - T_2 \\
&= pr\{T_1 - \xi > t|\xi > 0, \mathbf{Z}\}pr\{\xi > 0|\mathbf{Z}\}, \text{ by Bayes' Theroem} \\
&= pr\{T_1 - \xi > t|\xi > 0, \mathbf{Z}\}S_2(0|\mathbf{Z}), \text{ since } S_2(0|\mathbf{Z}) = pr\{\xi > 0|\mathbf{Z}\} \\
&= pr\{T_1 > t + \xi|\xi > 0, \mathbf{Z}\}S_2(0|\mathbf{Z}).
\end{aligned}
$$

Thus, for $t > 0$,

$$\begin{aligned} & pr\{T_1 > t|\xi > 0, \mathbf{Z}\} > pr\{T_1 > t + \xi|\xi > 0, \mathbf{Z}\} \\ \iff & \frac{S_1(t|\mathbf{Z})}{S_1(0|\mathbf{Z})} > \frac{S_2(t|\mathbf{Z})}{S_2(0|\mathbf{Z})} \\ \iff & S_1^*(t|\mathbf{Z}) > S_2^*(t|\mathbf{Z}). \end{aligned}$$

Next, we prove part $2 \to 1$:

Let T_1 and T_2 be random variables with normalized subsurvival functions conditional on covariate vector $\mathbf{Z}$, $S_j^*(t|\mathbf{Z})$ for $j = 1, 2$. Then $S_1^*(t|\mathbf{Z}) > S_2^*(t|\mathbf{Z})\ \forall\ \mathbf{Z}$ and $\forall\ t > 0$, and $S_1^{*-1}(t|\mathbf{Z})$ and $S_2^{*-1}(t|\mathbf{Z})$ exist. Choose a random variable δ, such that $\delta \perp T_1|\mathbf{Z}$ with

$$pr\{\delta = 1|\mathbf{Z}\} = S_1(0|\mathbf{Z}) \text{ and } pr\{\delta = 2|\mathbf{Z}\} = S_2(0|\mathbf{Z}),$$

and put

$$\xi(\mathbf{Z}) = I\{\delta = 2|\mathbf{Z}\}\left(T_1(\mathbf{Z}) - S_2^{*-1}S_1^*(T_1|\mathbf{Z})\right) - (I\{\delta = 1|\mathbf{Z}\}),\ S_1^*(t|\mathbf{Z}) > S_2^*(t|\mathbf{Z}).$$

Hence we have $\{\xi = -1|\mathbf{Z}\} = \{\delta = 1|\mathbf{Z}\}$ and $\{\xi > 0|\mathbf{Z}\} = \{\delta = 2|\mathbf{Z}\}$.

Therefore, $T_1 \perp sign(\xi)|\mathbf{Z}$ and

$$\begin{aligned} pr\{T_2 > t, T_2 < T_1|\mathbf{Z}\} &= pr\{T_1 - \xi > t, \xi > 0|\mathbf{Z}\},\ \text{since } \xi = T_1 - T_2 \\ &= pr\{T_1 - \xi > t|\xi > 0, \mathbf{Z}\}pr\{\xi > 0|\mathbf{Z}\},\ \text{by Bayes' Theroem} \\ &= pr\{T_1 - \xi > t|\xi > 0, \mathbf{Z}\}S_2(0|\mathbf{Z}),\ \text{since } S_2(0|\mathbf{Z}) = pr\{\xi > 0|\mathbf{Z}\} \\ &= pr\{S_2^{*-1}S_1^*(T_1) > t|\mathbf{Z}\}S_2(0|\mathbf{Z}),\ \text{when } \xi > 0,\ \xi(\mathbf{Z}) = T_1(\mathbf{Z}) - S_2^{*-1}S_1^*(T_1|\mathbf{Z}) \\ &= pr\{T_1 > S_1^{*-1}S_2^*(t)|\mathbf{Z}\}S_2(0|\mathbf{Z}),\ \text{after rearranging} \\ &= \left(S_1^*S_1^{*-1}S_2^*\right)(t|\mathbf{Z})S_2(0|\mathbf{Z}),\ \text{by definition} \\ &= S_2^*(t|\mathbf{Z})S_2(0|\mathbf{Z}),\ \text{since } S_1^*S_1^{*-1} \text{ cancel} \\ &= S_2(t|\mathbf{Z}). \end{aligned}$$

Lastly,

$$\begin{aligned} pr\{T_1 > t, T_1 < T_2|\mathbf{Z}\} &= pr\{T_1 > t, \xi < 0|\mathbf{Z}\}, \text{ since } \xi = T_1 - T_2 \\ &= pr\{T_1 > t|\xi < 0, \mathbf{Z}\}pr\{\xi < 0|\mathbf{Z}\}, \text{ by Bayes' Theroem} \\ &= S_1^*(t|\mathbf{Z})S_1(0|\mathbf{Z}), \text{ by definition} \\ &= S_1(t|\mathbf{Z}). \end{aligned}$$

□

BIBLIOGRAPHY

R. D. Baker and W. Wang. Developing and testing the delay-time model. *The Journal of the Operational Research Society*, 44(4):361–374, 1993.

R.D. Baker and W. Wang. Estimating the delay-time distribution of faults in repairable machinery from failure data. *IMA Journal of Management Mathematics*, 3(4):259–281, 1991.

Narayanaswamy Balakrishnan. *Recent advances in life-testing and reliability*. CRC press, 1995.

Neal R Barshes, Timothy C Lee, Ian W Udell, Christine A O'Mahoney, Saul J Karpen, Beth A Carter, and John A Goss. The pediatric endstage liver disease (peld) model as a predictor of survival benefit and posttransplant survival in pediatric liver transplant recipients. *Liver transplantation*, 12(3):475–480, 2006.

Norman Breslow. Covariance analysis of censored survival data. *Biometrics*, pages 89–99, 1974.

SC Cheng, Jason P Fine, and LJ Wei. Prediction of cumulative incidence function under the proportional hazards model. *Biometrics*, pages 219–228, 1998.

A. H. Christer and W. M. Waller. Delay time models of industrial inspection maintenance problems. *The Journal of the Operational Research Society*, 35(5):401–406, 1984.

Roger M Cooke. The total time on test statistic and age-dependent censoring. *Statistics and probability letters*, 18(4):307–312, 1993.

Roger M Cooke. The design of reliability data bases, part i: review of standard design concepts. *Reliability Engineering and System Safety*, 51(2):137–146, 1996.

David R Cox. Regression models and life tables (with discussion). *Journal of the Royal Statistical Society*, 34:187–220, 1972.

Jean-Yves Dauxois and Agathe Guilloux. Nonparametric inference under competing risks and selection-biased sampling. *Journal of Multivariate Analysis*, 99(4):589–605, 2008.

Jean-Yves Dauxois, Sarah Jomhoori, and Fatemeh Yousefzadeh. Testing an" exponential delay time model"against a" random sign censoring model"in reliability. *Journal de la Socit Franaise de Statistique*, 155(3):104–119, 2014.

Isha Dewan, JV Deshpande, and SB Kulathinal. On testing dependence between time to failure and cause of failure via conditional probabilities. *Scandinavian journal of statistics*, 31(1):79–91, 2004.

James J Dignam, Qiang Zhang, and Masha Kocherginsky. The use and interpretation of competing risks regression models. *Clinical Cancer Research*, 18(8):2301–2308, 2012.

Jason P. Fine and Robert J. Gray. A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*, 94(446):496–509, 1999.

Richard B Freeman, Robert G Gish, Ann Harper, Gary L Davis, John Vierling, Leslie Lieblein, Goran Klintmalm, Jamie Blazek, Robert Hunter, and Jeffrey Punch. Model for endstage liver disease (meld) exception guidelines: Results and recommendations from the meld exception study group and conference (message) for the approval of patients who need liver transplantation with diseases not considered by the standard meld formula. *Liver transplantation*, 12(S3), 2006.

Ted A Gooley, Wendy Leisenring, John Crowley, and Barry E Storer. Estimation of failure probabilities in the presence of competing risks: new representations of old estimators. *Statistics in medicine*, 18(6):695–706, 1999.

P Hokstadt and U Jensen. Predicting the failure rate for components that go through a degradation state. *Safety and Reliability, S. Lydersen, GK Hansen and HA Sandtorv (eds), Balkema, Rotterdam*, pages 389–396, 1998.

Haesook T Kim. Cumulative incidence in competing risks data and competing risks regression analysis. *Clinical Cancer Research*, 13(2):559–565, 2007.

John P Klein and Per Kragh Andersen. Regression modeling of competing risks data based on pseudovalues of the cumulative incidence function. *Biometrics*, 61(1):223–229, 2005.

JP Klein, JD Rizzo, MJ Zhang, and N Keiding. Statistical methods for the analysis and presentation of the results of bone marrow transplants. part i: unadjusted analysis. *Bone marrow transplantation*, 28(10):909, 2001.

Bo Henry Lindqvist and Guro Skogsrud. Modeling of dependent competing risks by first passage times of wiener processes. *IIE Transactions*, 41(1):72–80, 2008.

Kim M Olthoff, Robert S Brown, Francis L Delmonico, Richard B Freeman, Sue V McDiarmid, Robert M Merion, J Michael Millis, John P Roberts, Abraham Shaked, and Russell H Wiesner. Summary report of a national conference: evolving concepts in liver allocation in the meld and peld era. *Liver Transplantation*, 10(S10), 2004.

Melania Pintilie. *Competing risks: a practical perspective*, volume 58. John Wiley and Sons, 2006.

Nria Porta Bleda, Guadalupe Gmez Melis, Calle Rosingana, M Luz, and Nria Malats i Riera. Competing risks methods. 2007.

RL Prentice, JD Kalbfleisch, AV Peterson, N Flournoy, VT Farewell, and NE Breslow. The analysis of failure times in the presence of competing risks. *Biometrics*, 34(4):541–554, 1978.

James M Robins and Andrea Rotnitzky. Recovery of information and adjustment for dependent censoring using surrogate markers. In *AIDS Epidemiology*, pages 297–331. Springer, 1992.

Thomas H Scheike, Mei-Jie Zhang, and Thomas A Gerds. Predicting cumulative incidence probability by direct binomial regression. *Biometrika*, 95(1):205–220, 2008.

United Network for Organ Sharing. Questions and answers for transplant candidates about meld and peld. *Talking about transplantation*, 2008.

Anastasios Tsiatis. A nonidentifiability aspect of the problem of competing risks. *Proceedings of the National Academy of Sciences*, 72(1):20–22, 1975.

Aad W Van Der Vaart and Jon A Wellner. Weak convergence. In *Weak Convergence and Empirical Processes*, pages 16–28. Springer, New York, 1996.

Ling Wang, Haijun Hu, Yuqiao Wang, Wei Wu, and Pengfei He. The availability model and parameters estimation method for the delay time model with imperfect maintenance at inspection. *Applied Mathematical Modelling*, 35(6):2855–2863, 2011.

Russell Wiesner, Erick Edwards, Richard Freeman, Ann Harper, Ray Kim, Patrick Kamath, Walter Kremers, John Lake, Todd Howard, and Robert M Merion. Model for end-stage liver disease (meld) and allocation of donor livers. *Gastroenterology*, 124(1):91–96, 2003.

Russell H Wiesner, Sue V McDiarmid, Patrick S Kamath, Eric B Edwards, Michael Malinchoc, Walter K Kremers, Ruud AF Krom, and W Kim. Meld and peld: application of survival models to liver allocation. *Liver transplantation*, 7(7):567–580, 2001.

Mei-Jie Zhang, Xu Zhang, and Thomas H Scheike. Modeling cumulative incidence function for competing risks data. *Expert review of clinical pharmacology*, 1(3):391, 2008.