5-2015

Analytical Comparison of Contrasting Approaches to Estimating Competing Risks Models

Brian Stephen Rickard
University of Arkansas, Fayetteville

Analytical Comparison of Contrasting Approaches to Estimating Competing Risks Models

Analytical Comparison of Contrasting Approaches to Estimating Competing Risks Models

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in Educational Statistics and Research Methods

by

Brian Stephen Rickard
University of Arkansas
Bachelor of Science in Mathematics, 2005
University of Arkansas
Master of Education in Higher Education, 2008

May 2015
University of Arkansas

This dissertation is approved for recommendation to the Graduate Council.

Dr. Wen-Juo Lo
Dissertation Director

Dr. Ronna C. Turner
Committee Member

Dr. Charles E. Stegman
Committee Member

Dr. Brandon C. LeBeau
Committee Member

Abstract

Survival analysis is a commonly used tool in many fields but has seen little use in education research despite a common number of research questions for which it is well suited. Researchers often use logistic regression instead; however, this omits useful information. In research on retention and graduation for example, the timing of the event is an important piece of information omitted when using logistic regression. A simulation study was conducted to evaluate four methods of analyzing competing risks survival data, Cox proportional hazards regression, Weibull regression, Fine and Gray's Method, and Cox proportional hazards regression with frailty. College student retention and graduation is presented as an example. The results indicate that there is no one best model for all simulated scenarios. Instead, it appears the selection of the method of analysis should be based on the characteristics of the data. Both Cox proportional hazards and the Weibull regression are accurate with the base combination (sample size of 500 per group, continuous event time format, no correlation between event times, homogeneous shape parameter for both events for both groups, homogeneous failure rates for both events for both groups, and no frailty) as well as when one parameter is changed from the base combination. In addition, for data where the event time distribution shape does not differ by event, the accuracy of the models is quite similar. However, differences begin to emerge with some combinations of conditions. Cox performs especially poorly with data sets containing both differing event time distribution shapes by event and differing failure rates by group or event while Weibull is least accurate with the combination of homogeneous event time distribution shape, heterogeneous failure rate by group and/or event, and discrete format time. Fine and Gray's method was often ranked last by accuracy, but there are some situations where its accuracy is quite good including retention and graduation data. Cox proportional hazards

regression with frailty performed very similarly to the Cox regression without frailty with no clear benefits.

Acknowledgements

I am immensely grateful for my family, especially my parents, who have been incredibly supportive and encouraging and who instilled in me the value of education and that finding a job you enjoy doing is far more valuable than money. The culmination of my education with a doctoral degree was possible in very large part due to them.

Thank you to my friends for a great many things, but notably for continually ignoring the faux pas of asking when I would finish this dissertation and motivating me to finish.

I am also grateful for Dr. Deborah Korth, Dr. Suzanne McCray, and Dr. Charlotte Lee who not only allowed me to continue my education while in their employ, but actively encouraged it.

I would also like to thank all the professors of my program, whose support, guidance, and wisdom have been invaluable. I would especially like to thank Dr. Lo for encouraging me to choose a topic outside of the norm, and Dr. Denny who taught my first statistics class which forever changed the trajectory of my career.

Dedication

To the memory of Bates "Grandad" Rickard and Daniel "Pops" Hawkins.

Who believed in the importance of education and the glory of God. The lives they led provided some of the greatest lessons I've ever learned.

And to Dr. Mike Compton

My close friend and mentor, whose life touched countless people around the world. "For I have derived much joy and comfort from your love, my brother, because the hearts of the saints have been refreshed through you" (Philemon 1:7 English Standard Version).

Table of Contents

Chapter I: Introduction

Attempts to understand the world around us using basic statistics is far from being a recent trend. Early empires collected data on geographical size, wealth, and population. Navigators, astronomers, and cartographers collected observational data on planets, locations, and directions. As human curiosity and numerical methods increased, the process of analyzing these collected data began to take shape.

Over the course of history, the study of survival has received much attention. One of the early pioneers in the field of statistical analysis, John Graunt, was a relatively uneducated man who was a member of London's Common Council and who became fascinated with mortality data. Graunt poured over seventy years of the Bills of Mortality (lists of the deceased) which were published in England starting in 1592 and gained popularity during the Great Plague in the early seventeenth century as well as christening records from local churches. Using this information he wrote *Natural and Political Observations mentioned in a following index, and made upon the Bills of Mortality With reference to the Government, Religion, Trade, Growth, Ayre, diseases, and the several Changes of the said City,* which summarized these data and arguably birthed modern statistics. In his own words, his aim was "to have reduced several great confused volumes into a few perspicuous Tables, and abridged such Observations as naturally flowed from them, into a few succinct Paragraphs, without any long series of multiloquious Deductions" (Graunt, 1662, pp. 6–7).

Though his methods were rudimentary, his contributions to our understanding of mortality and to the field of statistics were invaluable. Aside from his mortality contributions, which included time-trends for diseases, a refutation of some incorrect beliefs about the spreading of the plague, and incidence of various causes of death, Graunt also introduced the world to statistical sampling and what have since become known as life tables. Using these life tables, which are extensively used in the insurance industry in modern day, Graunt was able to predict survival rates to successive ages and life expectancy for different groups ("John Graunt," 2004).

In scientific research, two of the most basic questions asked are when and why. In biostatistics the when and why are often centered on the event of death. In criminology researchers investigate time to arrest, rehabilitation, and recidivism. In workforce development researchers are interested in the timing and occurrence of job acquisition, promotions, layoffs, and retirement. Demographical studies focus on births, deaths, marriages, and divorces. In engineering researchers investigate time-to-failure of various machines and materials. In education, researchers are interested in the progression of a student's education, generally marked by a graduation of one stage to the next. In many areas of statistics the change over time is an area of great concern – the gradual education of a child, the health or wealth of a person, or the rise and fall of governments and economies. While many of these changes are slow and occur over time, in nearly every case there are distinct turning points or milestones that represent the culmination of these effects. The study of these distinct events is known as event history analysis. The graduation of a student from high school, the marriage of a couple, the birth of a child, all these are part of the growth of humans, but represent a distinct change, a move to the next stage of life.

Statement of the Problem

Though conceptually very simplistic, event research becomes more complicated when timing, number, and repeatability of events are all accounted for. In biological sciences one may only be interested in the mortality of a subject and the lifespan. In criminology one may be interested in not only the occurrence and timing of criminal activity, but in the number of occurrences. In substance abuse research one may be interested in the onset of drug use as well as the frequency and type of drug used.

Likely due to the complexity of modeling event histories, researchers often choose to omit information for the sake of simplicity and interpretability. Instead of examining the timing and type of drug use, a researcher may only be interested in if drug use occurs during a time period (Lee, 2012; Low et al., 2012; Ramirez, Hinman, Sterling, Weisner, & Campbell, 2012). Rather than examining when a student will graduate or drop out, a researcher may choose to simply investigate if a student graduates within six years (Astin & Osegura, 2005; Berkovitz & O'Quin, 2006; Zhang, Anderson, Ohland, & Thorndyke, 2004). While the omission of timing and/or event types simplifies the analysis and interpretation of these research questions, the lost information may play a substantial role in the true nature of the condition.

Purpose of the Study

The purpose of this study is to compare methods of analyzing competing risks models to investigate the effects that varying parameters have on model estimates. The parameters of data to be analyzed far exceed the research into the effects of these data parameters on the models being used. Varying data parameters will be analyzed to investigate the accuracy of four methods of analysis: Cox proportional hazards regression, Weibull regression, Fine and Gray's method, and Cox proportional hazards regression with frailty. College student retention and graduation data is presented as an example.

Primary Research Questions

How can event history best be modeled in the presence of competing events with varying data parameters? What are the effects of modeling event history when heterogeneity of variance is not accounted for and thus violating the assumption that all covariates are included in the model?

Significance of the Study

Many articles in current literature simplify multiple-event history data into simple dichotomies or single-event survival analysis and do not account for heterogeneity of variance. As methods for analyzing multiple-event data are relatively recent in development and statistical packages with programmed functions to perform the analyses are few, it is likely that many researchers are unfamiliar with the more complicated models. This study seeks to add to the knowledge-base on event history analysis by comparing multiple models using simulated data to manipulate various parameters.

Chapter II: Literature Review

In much of education research, the questions researchers ask lie in measuring outcomes. The most common of these outcomes are measurements of academic ability such as test scores, grades, and proficiencies. Also of importance are milestones – passing a grade, graduating high school, graduating college, etc. In these events often the question is not how well a person performed, but only if he or she performed well enough to satisfactorily complete a course of study. The outcome measure for the questions is therefore binary, either the event occurs or it does not. The most straightforward way to look at graduation is simply by examining if the event occurs. A logistic regression is a fine choice for such an analysis as it is able to predict membership in one of two groups, in this case graduates and non-graduates. Separating students into these two groups is very common in current literature (Arredondo & Knight, 2005; Berkovitz & O'Quin, 2006; Chimka, Reed-Rhoads, & Barker, 2007; Zhang et al., 2004) as well as one of the largest studies of graduation in higher education, the 2005 study out of the Higher Education Research Institute which provides formulas for predicting student degree attainment (Astin & Osegura, 2005). Other researchers have looked at the event of first-year retention with the two groups being those retained and those not retained. This method is also very common in literature (Braunstein, Lesser, & Pescatrice, 2008; Herzog, 2005; Robertson & Taylor, 2009; Szafran, 2001; Williams & Luo, 2010).

The most straightforward method of analyzing a question is not always the most thorough method however. While many researchers examining graduation data are satisfied with only the occurrence of graduation within six years, and researchers examining persistence to the second year are satisfied with only the occurrence of retention, both of these omit much useful information. For example, one may be interested in the timing of such an event. When will the student graduate? Will the event occur after four years? Will it occur within seven years? If the student does not graduate, how long will they have attempted before abandoning their educational pursuit? How many credit hours would they have completed? Universities are also very interested in being able to predict enrollment totals which require examining not just retention or graduation rates, but retention and graduation rates.

Survival Analysis

To answer these questions, a statistical test which incorporates the timing of such events is needed. Survival analysis, a subset of event history analysis, is a class of longitudinal statistical methods that involves both the timing and occurrence of events. As its name implies, the test's origins stem back to the life table data introduced by John Graunt.

When examining college graduation, much like logistic regression, the event in question in survival analysis is often student graduation. In logistic regression all subjects are classified into two groups: graduates and non-graduates. This is however, an over-simplification of the true nature of the situation. In reality, the two groups are: those who have graduated, and those who have not graduated yet. It is still possible for those students who have yet to complete a degree to eventually earn one. The difference between categorizing students who have not graduated as "not experiencing the event" and "not experiencing the event yet" is an important one. It means the outcome variable is not "no event," it is simply that the timing of the event is unknown, or missing for many students.

In logistic regression the unknown timing information is lost due to its categorization as a non-event. Students who have entered college and have completed less than six years, even if they have graduated, are excluded from the analysis as researchers cannot categorize the cohort until all students have had a full six years in which to graduate. In survival analysis, this missing information is retained and is considered censored. Censored data indicates that the observation of a subject is not yet complete as the event of interest has not yet occurred. It is also assumed that its incompleteness is independent of the event time. There are different forms of censoring including left censoring, in which a subject experiences an event before the onset of the study; interval censoring, in which the exact time of the event is unknown but a time interval is; and right censoring, in which censoring occurs with the end of subject observation. Right censoring is the most common form used and is the type of censoring discussed in this paper. A student studying at a university would be right censored for one of three reasons depending on the event of interest: either he or she graduates (if the event of interest is retention), leaves the university without completing a degree (if the event of interest is graduation), or the end of the study is reached. Each of these types of censoring are right censoring due to the fact that the study concludes at the end of the time period, the right side of a timeline with the occurrence of the event of interest, a competing event, or the end of the study.

The survival function. Survival analysis consists of two basic complementary parts: the survival function and the hazard function. The survival function models the probability of a subject experiencing the event before a certain time *t* and the hazard function is the event rate at time *t* given the subject has survived until time *t*. This survival function is defined as:

$$S(t) = P(T > t)$$

where *P* stands for probability and *T* is the time of the event. In other words, the survival function is the probability that a subject's event time T is later than some specified time *t*. At time 0 it is assumed that $S(0) = 1$. If modeling the survival of an animal, this would indicate that the moment the study begins the animal is known to be alive. In reference to the college student example, this would indicate that at the moment a student begins college he or she has a 100% chance of not graduating at that same moment, which would seem to be a tenable assumption for this example. Also assumed is that $\lim_{t \to \infty} S(t) = 0.$ Cox (1972, p. 188) explained "the only information available about the failure time of a censored individual is that it exceeds the censoring time." Again, if modeling the survival of an animal, this means that if one were to observe the animal indefinitely then it must at some point succumb to mortality. In the college student example, this would indicate that if one were to observe the student indefinitely then he or she must eventually graduate. This assumption is likely not tenable and is addressed later in this study.

The hazard function. For the first time period, the survival probability is simply the percentage of subjects who survived through the entire time period. The hazard function is the converse. The hazard function, as proposed by Cox (1972), is the instantaneous risk of experiencing the event given no previous event has occurred and is defined as:

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t}.$$

In other words, the hazard that a subject will experience the event at time T is the probability that the event time T is experienced during the time period from t to Δt given that the subject has not experienced the event at time t.

If we assume that time is measured in discrete intervals of time rather than continuously, which is discussed later in this study, we can simplify this formula to

$$h(t) = P\left(I_j \leq T < I_{j+1} \middle| T \geq I_j\right)$$

With I_j, j=1,…m-1 where m is the number of time intervals and T is the time of the event. Written another way, the formula is:

$$h(t) = \frac{\text{Number of subjects who experienced the event in interval j}}{\text{Number of subjects who did not experience an event prior to interval j}}$$

The hazard rate is therefore the ratio of people who experienced the event in a certain time period to the number of people who were still at risk for experiencing the event in that time period. This is the converse of the survival probability which is the number of people not experiencing an event in a certain time period and not having experienced the event prior. In the first time period this is computed simply by subtracting the hazard probability from 1 or $S(t) = 1 - h_1$. For all future time periods the survival probability is found by subtracting the hazard probability from 1 then multiplying times the probability of surviving through all previous intervals. The survival function can therefore be written as a function of the hazard h_k as:

$$S(t) = (1 - h_1) \cdot (1 - h_2) \cdot (1 - h_3) \cdots (1 - h_j)$$

$$= \prod_{k=1}^{j} (1 - h_k)$$

Let us examine a fictional example of the survival of ten subjects listed in Table 1 as an illustration.

Table 1

Example Survival Table

Subject	Time Period				
	1	2	3	4	5
1	0	0	0	1	.
2	0	0	0	0	0
3	0	1	.	.	.
4	0	0	0	0	1
5	0	0	1	.	.
6	1	.	.	.	.
7	0	0	0	0	1
8	0	0	0	1	.
9	0	1	.	.	.
10	0	0	0	0	.

In this table, a data value of "0" indicates the subject survives through the time period, a "1" indicates the subject has passed away during the time period, and a "." indicates that the subject is right-censored and is no longer being observed as they are no longer at risk for death. In the case of subject 10, who is censored without an event, it is assumed this subject was lost to follow-up. As nine of the ten subjects survived time period one, the survival probability would be $\frac{9}{10}$ and the hazard probability would be $\frac{1}{10}$. For time period two only nine subjects were at risk and two did not survive through the time period. The hazard probability would therefore be $\frac{2}{9}$ and the survival probability would be $\frac{7}{9}$ times the survival probability from the first time period $\frac{9}{10}$. S(2) is therefore $\frac{7}{9} \cdot \frac{9}{10} = \frac{7}{10}$. The remaining hazard ratios for intervals 3, 4 and 5 are $\frac{1}{7}, \frac{1}{3}, \frac{2}{3}$ respectively. The probability of surviving all five time periods is the product of the compliments of the hazard probabilities for each interval $\frac{9}{10} \cdot \frac{7}{9} \cdot \frac{6}{7} \cdot \frac{2}{3} \cdot \frac{1}{3} = \frac{2}{15}$. This probability is known as the Kaplan Meier product-limit estimate. This estimate, developed by Kaplan and Meier in their seminal 1958 paper, has the advantage of utilizing subject information even in the case of censoring. Subject 10 for example, though lost to follow-up with survival status unknown, still contributes to the hazard and survival probabilities for time periods 1-4. Since subjects do not all have to be studied for the exact same period of time, this also allows subjects to enter the study at different time points. Time 0 is able to represent a subject's entry into the study even if time 0 has a different calendar date for different subjects.

These calculations are of course quite simple when operating in the absence of covariates. The proportional hazard model proposed by Cox (1972) allows for covariates by modeling the hazard at time *t* as the product of the baseline hazard function – which is the hazard at time *t* in the absence of covariates – and an exponential equation of unknown parameters and covariates. The proportional hazards model is:

$$h(t|\boldsymbol{X}) = h_0(t)e^{\sum_{i=1}^{p} \beta_i X_i}$$

where $h_0(t)$ is the baseline hazard, β is a p × 1 vector of unknown parameters, and X is a p × 1 vector of independent variables. As seen in the formula, the baseline hazard portion involves no X's and is therefore the hazard unaffected by the covariates, while the exponential portion involves no time component *t* which therefore assumes the covariates act on subjects irrespective of time – hence the name of proportional hazards (Kleinbaum, 1995).

Discrete-time and continuous-time. In event history analysis, the event in question always represents some sort of a distinct change, a turning point, or a milestone. For each event, there is always an exact time that can be pointed to as the occurrence of that event: the birth of a child, the graduation of a student, the failure of a machine. Depending on the situation being measured, time data will be in one of two formats – discrete-time or continuous-time. When measuring time to injury of an athlete, for example, it is likely easy to pinpoint the exact moment an ACL tear occurs as it is a painful event that generally occurs during physical activity. However, in many other situations the only information known is a time period during which the event occurred. For example, trying to determine the exact date and time of catching a cold is quite difficult as the occurrence of symptoms begin gradually. Likely a subject can determine a small time period, a specific day, or even part of a specific day during which it became clear that he or she had caught a cold. The difference between these two types of measured data can be, in some situations, very small as all time measurements, no matter how small, are measured in discrete units. The exact hour, minute, and second of a day is, for most practical purposes, continuously measured, though of course even seconds are discrete-time units. In Allison's (1982, p. 70) paper on the subject of discrete-time, he summarizes the decision logic as such:

> "When these discrete units are very small, relative to the rate of event occurrence, it is usually acceptable to ignore the discreteness and treat time as if it were measured continuously. When the time units are very large – months, years, or decades – this treatment becomes problematic."

He discusses in a later work, however, that the choice will likely have little statistical impact as both methods give very similar results and "while there is some loss of information that comes from not knowing the exact time of the event, this loss will usually make little difference in the estimated standard errors (Allison, 1984, p. 22)." Masyn (2003) suggests the distinction that discrete-time data should be measured in large enough intervals such that multiple events occur during each interval and continuous-time data should be measured in small enough intervals such that no "ties" occur. The MPLUS statistical program requires this to be the case for analyses to be performed while other programs, such as R, allow for ties through exact partial likelihood, originally proposed by Cox (1972), exact marginal likelihood, or approximations using Efron or Breslow's methods.

Though discrete-time and continuous-time methods have been in use for approximately the same time period, most literature on event history analysis is focused on continuous-time survival models. The reason for this is not entirely clear, though it is likely influenced by the "less-than-ideal situation of not knowing the exact time-of-event" (Masyn, 2003, p. 4) and the fact that, until relatively recently, the additional computing time necessary for discrete-time survival analysis was prohibitive. With modern computing, this has essentially become a non-issue and analyses with discretely measured time are increasing in popularity (Allison, 1984).

For the college graduation example used in this paper, discrete-time is used. This is for a few reasons. First, though exact time of graduation and drop-out is known, because students often experience these events en masse, graduation occurs only at the end of the semester and most students who drop out also do so at the end of a semester, it makes more sense to measure time in discrete-time units (i.e., semesters). While not addressed in this paper, using discrete-time would also allow for time varying covariates such as GPA by term or class attendance. This adheres to the assumptions of discrete-time survival analysis that for each discrete-time interval each independent variable has one and only one value, and that if the discrete-time intervals are created from underlying continuous-time, that censoring occurs at the endpoint of the discrete-time interval. If time-varying covariates are present, discrete-time survival analysis also does not require the proportionality of hazard curves that is assumed in survival analysis methods (Masyn, 2003).

Assumptions of survival analysis. One of the assumptions of a survival analysis is that censoring is non-informative, that is, the censoring of a subject is independent of his or her likelihood of experiencing the event. At the end of the study it is assumed that if a student has not graduated that he or she still may or may not graduate, but the end of the study is independent of the occurrence of that event. This also means that if a subject is lost to follow-up that it is assumed that the subject still may or may not graduate, but in either case their graduation is independent of being lost to follow-up.

This presents the first of two main issues when using survival analysis to analyze events such as graduation. Under the model most commonly used, students who graduate are said to have experienced the event, and those who drop out or have not graduated by the end of the study are censored. Students who drop out are censored at the time at which they drop out; however, this censoring is informative. Once the student has dropped out, it can be reasonably assumed that the student has a lower likelihood of graduating, a zero percent chance to be exact unless they re-enroll. Censoring from drop-out is therefore far from independent of the subject experiencing the event of graduation. Also, because these subjects are censored in the same way that a student who is still enrolled at the end of the study, both are treated the same in the statistical test, though the researcher knows the students do not have the same chances of eventually graduating.

While the graduation of its students is certainly the end goal of all universities, it is not the only outcome measure universities are interested in. Present in much of literature regarding post-secondary student success is also the topic of retention – a measure of how well a school retains its students from year to year. Generally the most common metric is the first-year retention rate. While this is certainly closely linked to graduation, the timing of this occurrence is also of importance to a university. Improving graduation rates for a university that loses most students after their second year requires a completely different approach than if the university loses most students after the first year. This presents the second main issue with survival analysis as it is most commonly used in modeling college student graduation – the number of events. As survival analysis is most commonly used, the event of interest is student graduation which treats students who drop out and those who continue their studies but have not yet graduated as the same. This scenario is illustrated below in Table 2 with subjects 2 and 3 treated the same as both are non-graduates.

Table 2

Survival Data with Event of Graduation

Subject	Time Period					Description
	1	2	3	4	5	
1	0	0	0	1	.	Graduated Time Period 4
2	0	0	0	0	0	Did not Graduate (Enrollment Uncertain)
3	0	0	0	.	.	Missing Data (Non-Informative Censoring)

Conversely, if the event of interest was student drop-out, the analysis would treat those who graduate and those who continue their studies but have not yet graduated the same. This scenario is modeled below in Table 3 with subjects 2 and 3 treated the same as both have not dropped out.

Table 3
Survival Data with Event of Drop-Out

Subject	Time Period 1	2	3	4	5	Description
1	0	0	0	1	.	Drop-Out Time Period 4
2	0	0	0	0	0	Did not Drop Out (Graduated or Still Enrolled)
3	0	0	0	.	.	Missing Data (Non-Informative Censoring)

As mentioned previously, one of the assumptions of survival analysis is that $\lim_{t \to \infty} S(t) = 0$ which indicates that it is assumed that every subject is at risk for the event until they have experienced the event, or that every college student is assumed to graduate eventually. This is of course not the case. Nationally the average six-year graduation rate is only 59% (NCES 2013). In fact, for many of the previously mentioned uses of survival analysis – marriage, divorce, child birth, etc. – it cannot be reasonably assumed each subject will eventually experience the event.

Competing Risks

In Graunt's life tables, there existed only one event of interest, that of mortality. Though through his study, he examined many causes of mortality. This was crucial as when predicting lifespan it is indeed very important to distinguish between things like accidents and war and things like the plague and heart attack. Those who perish in workplace accidents, for example, are certainly removed from the risk of perishing from a heart attack, though certainly not by any measure of good fortune. The idea of cause-specific hazards was developed by Daniel Bernoulli, whose paper on the advantages of smallpox inoculation included a hypothetical life table mapping the mortality of individuals at different ages were smallpox to be eradicated (Klein &

Moeschberger, 2003). His methodology included examining subjects who moved from state A (those not having smallpox) to state B (those having smallpox) before perishing.

This idea of examining if a subject will experience one event before another event of interest occurs is particularly important in the medical field – especially cancer research (Belot, Abrahamowicz, Remontet, & Giorgi, 2010; Chapman et al., 2008; Chu-Ling et al., 2009; Ryberg et al., 2008; Schairer, Mink, Carroll, & Devesa, 2004). To study the mortality of cancer patients it is certainly necessary to distinguish between cancer patients who perish from cancer-related or non-cancer-related causes. In essence, researchers examine the likelihood that a subject is to perish from cancer before they would perish from some other cause. In the case of mortality, death from one cause precludes death from another cause and as such, these events are thought to be in competition with one other or, in other words, competing risks.

Cause specific hazards. The hazard function defined above in a single-event survival analysis must be amended to account for the competing risks. In contrast with a single-event survival analysis for which there is only one type of event of interest and therefore only one hazard function, a competing risks analysis will include a hazard function for each event type called a *cause-specific hazard function*. As defined earlier, the single-event survival analysis hazard function is the instantaneous risk of experiencing the event given no previous event has occurred. The competing risk cause-specific hazard is likewise the instantaneous risk experiencing a specific event given no previous event, of any type, has occurred and is defined as:

$$h_i(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t, \delta = i | T \geq t)}{\Delta t}$$

where i = 1,…,K with K (K≥2) event types and T =Min(X_1,…,X_K) survival time (Klein & Moeschberger, 2003). We are interested in only the minimum value for T due to the fact that, as defined above, the occurrence of a competing risk precludes the occurrence of any other risk. A cancer patient, for example, who perishes from heart disease, is of course no longer at risk of death from cancer. While it is certainly possible that had the patient not perished from heart disease they would have eventually perished from cancer-related causes, this cannot be known as the occurrence of the first event, death from heart disease, precluded any possibility of the occurrence from the second event, death from cancer.

The overall hazard rate, the total risk of experiencing an event of any type, at time t is:

$$h(t) = \sum_{i=1}^{k} h_i(t)$$

irrespective of whether the events are independent of one another, which is discussed at the end of this chapter. At the time of event occurrence of type i the probability of that event occurring is therefore $h_i(t)/h(t)$ (Cleves, 2010).

Cumulative incidence function. As with single-event survival analysis, competing risks is not only interested in the hazard present, but also in the probability of the occurrence of a competing risk. Due to the nature of competing risks, this cannot be summed up in a single-event survival probability as before but with three different probabilities: crude, net, and partial crude. The crude probability describes the probability of experiencing a certain event in a world where the subject is at risk of experiencing all competing risks. This probability is calculated by the cumulative incidence function. The net probability describes the probability of experiencing a certain event in a world where that certain event is the only event the subject is at risk for. The partial crude probability describes the probability of experiencing an event in a world where some event types have been eliminated (Klein & Moeschberger, 2003)

The survivor function at time *t* describes the probability of surviving, or not experiencing the event of interest, past time t and was defined previously as $S(t) = P(T > t)$. The creation of an opposing function that models the probability of failing or experiencing the event of interest, up to and including time t, would be calculated by simply subtracting the survivor function from 1. This failure function would therefore be:

$$F(t) = 1 - P(T > t) = P(T \leq t).$$

where *T* is the time of the event. While this is essentially modeling the same thing and simply changing the direction one is interpreting from, it becomes advantageous to the survivor function when discussing competing risks. This is due to the fact that, as mentioned previously, with competing events we are only interested in the time to the occurrence of the first event. To use the student retention and graduation example, one could ask "What is the probability that a student will drop out within the first two years?" or "What is the probability that a student will graduate within four years?" and both questions could be easily interpreted as the occurrence of either event indicates the other event did not occur previously. Both of these involve asking about the probability of experiencing an event, or "failing" before a certain time. Conversely, we could ask "What is the probability that a student will not drop out during the first two years and that after that they will not drop out but will graduate?" or "What is the probability that a student will not graduate within the first four years and after that will not graduate but will drop out?" Both of these questions involve asking about the probability of not experiencing the event or “surviving” past a certain time. Simply by the wording of the questions it is clear that asking and interpreting questions in the presence of competing risks is far easier when done from the perspective of “failing” or experiencing the event, than “surviving” or not experiencing the event. It is for this reason that with competing risks, rather than investigating a modified survivor function, we are instead interested in a modified failure function known as the cumulative incidence function (Cleves, 2010). The probabilities computed with this function are the crude probabilities mentioned earlier which model a world in which a subject is at risk for all possible competing events. With $i = 1,\ldots,K$ event types, this function is defined as

$$\mathrm{CIF}_i = P(T \leq t, \delta = i)$$

where T is the time of the event. This can be seen to be only a slightly modified version of the failure function above. Put simply, the cumulative incidence function is the probability of experiencing event i by time t (Klein & Moeschberger, 2003).

Models. Analyzing competing risks data involves utilizing one of two general methods – modeling the risks separately, or simultaneously. Modeling the risks separately involves analyzing time to a specific event of interest and treating all other events as censored utilizing either a parametric survival model, such as Weibull or Gompertz, or a semi-parametric survival model such as Cox Proportional Hazards. This process is then repeated by replacing the event of interest with a different competing risk and treating all others as censored. This will result in K different models, where K is the number of competing risks. Modeling the risks simultaneously involves utilizing a method proposed by Fine and Gray (1999). Deciding how to model the competing risks is not an issue when modeling hazard functions as the overall hazard function $h(t)$ is simply a sum of all cause-specific hazards and, as demonstrated by Cleves (2010), results from simultaneous modeling identically matches that of separate modeling. This, however, is not the case when modeling survival or failure probabilities, though both modeling risks separately and modeling them simultaneously can be seen to have advantages and disadvantages. While survival analysis via the Kaplan-Meier estimator or Cox proportional hazard regression is generally included in most commercial statistical software, this cannot be said of competing events analysis using Fine and Gray's method. Though different competing risks models can be run in programs like SAS or SPSS, it often requires intensive programming or is limited to modeling risks separately as there is no pre-programmed competing risks function for either program. Other programs, such as STATA and R, do have this ability, but are less widely used. This alone may be part of the reason that modeling risks separately, or just modeling one risk, appears to be more common in literature (Chimka et al., 2007; Heilig, 2011; Min, Zhang, Long, Anderson, & Ohland, 2011; Murtaugh, Burns, & Schuster, 1999).

Modeling cumulative incidence functions using cause-specific hazards. Separate modeling has the benefit that estimating models for each risk separately allows for far more flexibility in parameters and models used. This is because each risk can be modeled in a different way having different independent variables used. It also allows for the exclusion of events that are not of interest, though certainly much caution must be taken if doing so (Allison, 1984). There are, however, downsides of this modeling process that are solved by modeling simultaneously. First, when modeling risks separately it is done so assuming that competing risks are uncorrelated, which is unlikely for many if not most scenarios. Second, the failure function $F(t) = 1 - S(t)$ does not account for the possibility of competing risks. As mentioned previously, it is assumed that $\lim_{t \to \infty} S(t) = 0$ and therefore it follows that $\lim_{t \to \infty} F(t) = 1$. However, while the failure function will indeed always approach one when modeling risks separately, the cumulative incidence function (which models risks simultaneously) for each event type must have a limit of strictly less than one due to the fact that each competing risk must have some probability of occurring (Cleves, 2010). If this were not the case, it would certainly be unnecessary to use a competing risks analysis as a traditional single-event survival analysis would suffice.

Modeling cumulative incidence functions using subdistribution hazards. Modeling the risks simultaneously has the advantage that it does not assume the competing risks are uncorrelated and does account for the possibility of more than one event; however, it also has its disadvantages. To describe these, it is necessary to backtrack to the earlier discussion of hazard probabilities for an illustrative comparison. In a single-event survival analysis the interpretation of a hazard probability in the presence of a covariate is quite straightforward. For simplicity sake, take an example of an analysis with a treatment group and a control group. If the hazard probability for treatment is .8, this would indicate that the hazard on the treatment group is 80% of the hazard on the control group or that the survival of the control group is 80% of the treatment. This could be translated into survivor functions for both control and treatment where the control survivor function is raised to the .8 power as such

$$S_C(t)^{.8} = S_T(t).$$

As such, the hazard probability is a clearly understood summary of the comparison. It is unfortunately not as clear when discussing cause-specific hazards. To return to the college student retention and graduation model, if one were to model time to the competing events of retention and graduation with the covariate of gender, for example, the resulting cumulative incidence functions would depend on the effect of gender on graduation and the effect of gender on retention, both the graduation baseline hazard and retention baseline hazard, and also on time itself (Cleves, 2010). This makes for a cumbersome and complicated interpretation. A model proposed by Fine and Gray (1999) sought to solve this issue by instead modeling an alternative *subdistribution hazard* (or *subhazard*) function that is more easily interpreted for cumulative incidence functions. This semi-parametric model for the hazard of experiencing event *i* is:

$$\bar{h}_i(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t, \delta = i | T > t \cup t \leq T, \delta \neq i)}{\Delta t}$$

The subhazard is therefore the instantaneous probability of experiencing event *i* at time *t* given that no event had been experienced before *t* or that an event other than event *i* had been experienced prior to time *t.* In other words, this subhazard function models the hazard of individuals experiencing event *i* but, unlike the previously defined cause-specific hazard, does not remove the subject from risk when a competing event occurs. Using our earlier college student example this would be akin to saying a student who has graduated from college is still at risk of dropping out, which is of course not the case. The authors do point out that this is indeed an “un-natural” risk set but also point out that this methodology is not unique to their model but is also used in the cure model also known as the long-term survivor model. This model is a type of survival analysis where individuals who have been “cured” are still included in the analysis of the survival model and thus are considered at risk (Fine & Gray, 1999, p. 497).

However, despite all this, the subhazard is indeed a very convenient way to model the cumulative incidence function for a specific event *i* because it is modeled as a function of solely event *i*. This has the benefit of allowing an easy interpretation of covariate effects on the cumulative incidence function due to the direct link between them.

Identifiability dilemma. In competing risks analysis, we observe only the event with minimum time *t* to event *i*. In many cases, the occurrence of event *i* precludes the occurrence of any other event, but does not indicate that no other event could have occurred. Take for example a competing risks analysis examining time to death of patients with competing risks of cancer, heart disease, and other. Should a patient perish due to cancer, it certainly cannot be said that that person would never have died of heart disease in the absence of cancer, only that in the presence of cancer it would be the cause of death. If there were a cure to cancer, this subject would surely have perished of another cause eventually, possibly of heart disease. This presents us with the identifiability dilemma. Since the separate modeling of cause-specific hazards assumes independence of event times, it is surely important to ensure this is the case. However, since we cannot observe the eventual event times of events that did not occur first, it is not possible to test the assumption that event times are independent (Klein & Moeschberger, 2003). We are therefore unable to distinguish between event times that are independent and infinitely many dependent event times that result in identical cause-specific hazards (Rodriguez, 2005).

Frailty

In a traditional linear regression, all sources of variation not accounted for by the independent variables are combined into a random error term. This error term is necessary as it is highly unlikely that in any empirical research that all sources of variation are accounted for by the independent variables. This is, however, the assumption of survival analysis, which includes no random error term and therefore assumes all sources of variation have been observed (Masyn, 2003). These sources of population heterogeneity may be covariates that were not included in the model or covariates that are simply unobservable. In any case, these sources of variation are, by definition, not estimated by the data. A frailty model is one that accounts for unobserved heterogeneity of variance by modeling a random-effects model that contains within-group correlation where the groups are unobserved. In any set of subjects it is likely that certain subgroups of the population – subgroups that are unknown to the researcher – not only exist but have hazards which differ from those in other subgroups. Defined in Klein and Moeschberger (2003, p. 51), a frailty is "an unobservable random effect shared by subjects within a subgroup." An easy example would be a study of mortality in which a subgroup of the population had an unknown heart defect. Certainly it is reasonable to assume that these subjects are likely to be more at risk of death than the general population. These subgroups would therefore have different frailties that may be able to be estimated by the model.

Vaupel (1979), who coined the term frailty, notes that the frailty of an individual does not vary over time, but remains the same from birth to death. The authors also defined frailty in relative terms where an individual with a frailty of 1 would be considered the standard, with frailty values less than 1 indicating the individual is less frail or less likely to experience the event, and frailty values greater than one indicating an individual is more frail or more likely to experience the event. An individual with a frailty value of 2 would be considered twice as frail as the standard individual. The frailties are a multiplicative effect that acts directly on the hazard of an individual. We can therefore account for these frailties in the hazard function as such

$$h\left(t_j|\boldsymbol{X},\alpha_j\right) = h_0(t)\alpha_j e^{\sum_{i=1}^{p}\beta_i X_i}$$

where α_j is the frailty of individual j, $h_0(t)$ is the baseline hazard, β is a p × 1 vector of unknown parameters, and X is a p × 1 vector of independent variables (Cleves, 2010).

Since the subgroups are, by definition, unobserved, it is necessary to note that it is often not possible to empirically distinguish between a heterogeneous population made up of unobserved subgroups and a homogeneous population that simply changes over time (Trussell & Richards, 1985). For example, should a study investigate time to first injury of mountain bikers and find a decreasing hazard over time, one of two things could be inferred. One could infer that the decreasing hazard occurs because the population contains two subgroups with one group composed of those not predisposed for mountain biking who would therefore fail early leaving those who are better suited and at less risk to comprise the population, or one could infer that the population is homogeneous and the decreasing hazard over time is due to the fact that the more experienced the mountain biker is, the less likely he or she is to be injured. It is therefore important that care is taken in applying these methods to certain situations.

Ignoring unobserved heterogeneity of variance. In literature modeling college student success, certain predictors are quite common. These are things like high school grade point average, high school rank, standardized test scores, early college grade point average, ethnicity, and gender. However, even if a researcher had access to the entirety of a university's data warehouse and included predictors such as family income, first-generation status, high school attended, college major, scholarship and other financial aid information, hours attempted, and standardized test sub-scores, it would still be far from including every source of variation in student success. If it were true that all covariates have been included, as is the assumption of survival analysis, it would indicate that given two students who are identical in academic scores, demographics, family background, etc., we would expect the exact same likelihood of success for these two students. However, intelligence and family background paint only a part of the picture. Work ethic and motivation play a large part in student success though are unlikely to be observed. Teachers and other educators could likely point to a litany of other unmeasured factors that play roles in how and why students succeed.

It is widely known that the omission of variables in a standard regression model leads to bias in the parameter estimates unless the omitted variables are uncorrelated with those variables that are included. This, however, is not the case with survival analysis. Suppose a population of 1,000 consists of two unobserved groups, A and B, of size 500 which, unknown to the researcher, each experience a constant hazard function with group A being at a lower risk of hazard $(h(t) = .1)$ and B being at a higher risk of hazard $(h(t) = .3)$ similar to the mountain biking example above. The population sizes over time are listed below in Table 4.

Table 4

Unobserved Group Population Size Over Time

Group	Time Period									
	1	2	3	4	5	6	7	8	9	10
A	500	450	405	365	328	295	266	239	215	194
B	500	350	245	172	120	84	59	41	29	20
Total	1000	800	650	536	448	379	325	280	244	214

During the first time period it would simply appear as though the entire population was at moderate risk of hazard. However, recall that the hazard probabilities are conditional on the proportion of the population that has yet to experience an event. As time proceeds and members of group B experience events at a higher frequency to members of group A due to their higher hazard, the overall hazard will be based increasingly on members of group A who experience a lower hazard. This can be seen in Table 5 below.

Table 5

Unobserved Group Hazard Rate Over Time

Group	Time Period									
	1	2	3	4	5	6	7	8	9	10
A	0.100	0.100	0.100	0.100	0.100	0.100	0.100	0.100	0.100	0.100
B	0.300	0.300	0.300	0.300	0.300	0.300	0.300	0.300	0.300	0.300
Overall	0.200	0.188	0.175	0.164	0.154	0.144	0.136	0.129	0.124	0.119

This will have the effect of the appearance of a time interaction where the hazard decreases with time. In other words, as time progresses and an increasing number of members of group B are censored, the overall hazard function will converge to the constant hazard of group A as eventually all members of group B will be censored leaving only group A members (Masyn, 2003; Trussell & Richards, 1985) as seen below in Figure 1.

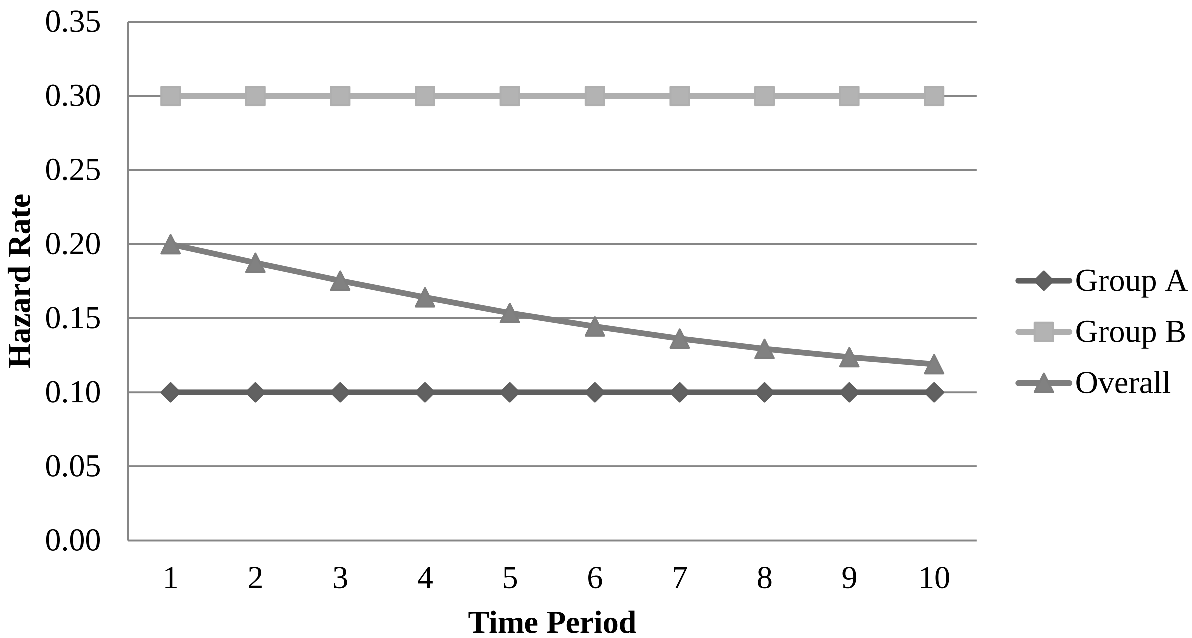

Figure 1. Unobserved group and overall hazard rate over time.

It can therefore be seen that in any population that contains two or more groups with differing hazard rates that the overall hazard rate will always converge to the rate of the lowest hazard group given a study of sufficient length. It is therefore important to account for such groups to avoid bias in the hazard function.

Summary

Literature on event analysis is widespread and present in a vast variety of fields. The methods of analysis used in these studies, however, are often simplified by omitting useful information. The inspiration for this paper was born, as many ideas are, out of necessity when trying to answer the question, "how likely is a college student to graduate, and if they do not, when will they drop out and why?" A review of literature on the topic found a significant portion of these studies employed logistic regression as the principal method of analysis (Berkovitz & O'Quin, 2006; Herzog, 2005; Robertson & Taylor, 2009; Szafran, 2001; Williams & Luo, 2010;

Zhang et al., 2004). However, using this methodology, event data are condensed to only a single dichotomy and time data are omitted entirely. Survival analysis addresses the timing issue and has the benefit of not restricting that all participants enter the study at the same time, subjects may enter and leave the study at will and still contribute useful information to the analysis, but competing risks are often omitted. Use of survival analysis to model retention and graduation data are present in literature and is becoming more common (Chimka et al., 2007; Min et al., 2011; Murtaugh et al., 1999), though likely due to the difficulty of obtaining all the necessary time data and the more complicated interpretation; it still appears to be less common than logistic regression in literature. However, it does have the benefits of wide availability in commercial statistical programs and frequent use in other fields in literature.

While survival analysis does have its advantages, it often presents issues when modeling time to events other than death. In its traditional use, survival analysis is used to model the lifetime of a subject with the single event death as the event of interest. Because this is the original purpose, it is logically assumed that all subjects are at risk of perishing until they eventually perish. This presents an issue when modeling time to an event that does not necessarily occur for all subjects. Since all subjects are assumed to eventually experience an event, there must be an end event for all subjects. If a subject is, for some reason, no longer at risk for the event of interest, the subject can then be considered "cured," which is identified as a second type of event as opposed to the non-occurrence of the event of interest. It is necessary to consider this as a second type of event due to the fact that the lack of occurrence of an event, also known as censoring, must be independent of a subject experiencing an event. This is known as non-informative censoring. Likely due to the complexity of dealing with a second event, much of the literature dealing with such circumstances treats those who will not experience the event as censored – which violates the assumption of non-informative censoring. This indeed does have its consequences. In an analysis of coronary heart disease, Wolbers et al. (2009) found that ignoring a competing risk resulted in a significant overestimation of coronary heart disease risk when compared to a model that accounted for the competing risk. Likewise, Kim (2007) found a substantial overestimation of cumulative incidence of relapse in examining allogeneic hematopoietic stem cell transplantation when time was modeled to relapse with transplant-related mortality treated as censored.

To account for multiple types of events, a competing risks analysis must be used. This type of analysis allows for the modeling of time to two or more competing events where the occurrence of one event precludes the occurrence of other events. There are two common methods of analyzing competing risks data: analyzing risks separately using cause-specific hazards, and analyzing risks simultaneously using subdistribution hazards. Each method has its advantages; however, any analysis of competing risks will violate at least one assumption regardless of the method used, or at the very least, will have an assumption that is not possible to prove has not been violated.

Modeling risks separately using cause-specific hazards assumes the competing risks are independent – an assumption which is impossible to prove outside of simulation given the nature of the data. Modeling risks simultaneously using subdistribution hazards has no such assumptions, but does assume that individuals are still at risk of a competing event even after experiencing an event – which is contrary to the definition of competing risks. Literature comparing these methods shows mixed results with no clear advantage of one method over the other for all circumstances. An article on coronary heart disease by Wolbers et al. (2009) found quite similar, though not identical, results when using the Fine and Gray subdistribution hazard model as when modeling separately using cause-specific hazards. A simulation study by Dignam et al. (2012) suggests that both models perform similarly when event times are uncorrelated and group membership is also uncorrelated to event time. The authors also found that when event times are uncorrelated and group membership *is* correlated to event time that modeling risks separately using cause-specific hazards appears more accurate. When event times are positively correlated however, the authors found mixed results, with cause-specific hazards performing better when one group experiences a higher hazard for both competing events, subdistribution hazards performing better when groups experience different hazards on both competing events, and similar performance when group membership was uncorrelated with event type. Williamson et al. (2007) also found mixed results with a simulation study comparing these two methods. They similarly found that the cause-specific hazard method is more accurate when event times are uncorrelated and group membership is correlated with the subdistribution hazard method - inflating type 1 error. The authors found that the subdistribution hazard method has greater power in detecting differences in treatment and improves in performance as event times are increasingly negatively correlated. The authors ultimately recommend that researchers utilize both methods in any analysis of competing risks data. A simulation study by Freidlin and Korn (2005) found the subdistribution hazard method to be unreliable in detecting group differences, especially when group affects both event types, and found the cause-specific hazard method to be more robust. These results, however, only examined positive and zero correlations between event times. The authors also found that when group membership affected only one event type and event types were highly correlated that the subdistribution hazard method had a higher power to detect group differences than the cause-specific hazard method.

Though competing risks addresses many of the issues of the previously mentioned methods, there is still one assumption that is likely not to hold in any survival analysis. Because survival analysis contains no error term, it is assumed that in any analysis that every source of variation, i.e. every covariate, has been included in the analysis. It is highly unlikely that any study outside of a laboratory can control for every possible source of variation. To account for the possibility of uncontrolled covariates requires the modeling of frailty. Literature on frailty suggests that indeed omitting these variables from the analyses creates bias in the results, though accounting for these unobserved sources of heterogeneity by modeling frailty requires caution. Using an application of the Cauchy-Schwartz theorem, Heckman and Singer (1984) were able to demonstrate that the presence of uncontrolled variables, whether they be observable or unobservable, creates bias in the estimated hazard function that has the effect of a sharper decline or slower rise in the overall hazard function than would be seen in the presence of the covariates. However, the authors do warn that the results can be sensitive to the parametric form chosen for the frailty model. Trussell and Richards (1985) confirm and extend this warning to nonparametric representations of heterogeneity, which are sensitive to hazard choice.

As interest has grown in modeling time-to-event data, survival analysis is increasing in occurrence in current literature in a variety of fields. However, when survival analysis is used to analyze data where more than one event is possible, at least one assumption is violated regardless of the method of analysis used. The effects of varying data parameters on survival analysis with competing risks, which lead to different violations of assumptions, while receiving some attention in current literature, are still largely unknown. This purpose of this study is to compare methods of analyzing competing risks models to investigate the effects that varying parameters have on model estimates.

Chapter III: Method

This study was designed to supplement current event history literature by investigating unresolved issues in modelling competing risks data. In current literature, survival methods are used in varying fields under a multitude of conditions. Many of these conditions do not satisfy the assumptions of these tests which may invalidate the results of these studies. These assumptions are: 1) $\lim_{t \to \infty} S(t) = 0$ which indicates that all subjects will experience the event eventually, 2) censoring is non-informative, and 3) all sources of variation have been observed and are included in the model. Issues arise with the first assumption when studies model time to an event that is not death such as relapse, recidivism, graduation, episode of drug abuse, and other such events where it is quite possible that the subject will never experience the event of interest. The second assumption requires attention when competing events are possible. When modeling time to an event such as death by cancer, high school drop-out, component failure or other such events where competing events are possible, it is common to see these competing events treated as censored or missing when the information these competing risks provide is likely informative. The third assumption is likely never met. Except in the most highly controlled laboratory experiments, it is highly unlikely that any study can claim to include even most sources of variation, let alone all sources.

To investigate the effects of these assumptions on different models, a simulation study was conducted using a variety of parameter combinations. In this simulation study four methods of survival analysis will be tested modeling two groups with two competing risks or events. The methods of survival analysis are: separate event modeling using cause-specific hazards and treating the competing risk as censored which will be modeled both in a parametric model and a semi-parametric model, simultaneous modeling of subdistribution hazards using Fine and Gray's method, and modeling separately using cause-specific hazards with frailty. The variables manipulated in this study are: sample size, correlation between event times, event time format, group failure rates by event type, event time distribution shape, and frailty.

Manipulated Model Parameters

Sample size. The first parameter manipulated is the sample size of the groups. In this study, two balanced groups are utilized for comparison. Simulation studies in current literature have not varied sample size, so implications are unknown. Sample size per group was selected as 250 and 500 for total sample sizes of 500 and 1000. Sample sizes were chosen to represent small and moderate size cohorts in universities with the belief that samples larger than 1000 will not significantly increase statistical power.

Correlation between event times. The next parameter manipulated is the correlation between event times. This parameter can only be theorized because in practice only one event will actually occur. It is therefore impossible to find the correlation between two events with one event having an unknown time. However, despite not knowing the timing of the second event, it is quite likely that two competing events could be correlated or could be independent depending on the type of event. If, for example, one were interested in time to death by heart disease a competing event could be death by auto accident. As death by auto accident is often unrelated to the age or health of the individual, it is likely that the timing of a possible death by auto accident would be uncorrelated with the timing of a possible death by heart disease. Alternatively if one were interested in time to first arrest of disadvantaged youths in the inner city with competing events of vandalism and theft, it could certainly be hypothesized that the timing of a possible arrest due to vandalism could be correlated with the timing of a possible arrest due to theft.

Likewise, in the student retention and graduation example, it is unlikely that the events of drop-out and graduation would be uncorrelated.

Although it is difficult to theorize possible correlations, since it is possible that the event times could be correlated, it is necessary to control for this possibility. In this simulation study two event times, one for each event type, are created with the minimum of these times being selected as the event time and the corresponding event selected as the event type. Simulation studies by Dignam et al. (2012) and Williamson et al. (2007) both suggest that when event times are uncorrelated, separate modeling using cause-specific hazards performs better. However, when event times are correlated, both papers found mixed results. As separate modeling using cause-specific hazards assumes event times are uncorrelated and simultaneous modeling using subdistribution hazards does not, different results are unsurprising.

Since correlation between event types is purely theoretical, it is impossible to determine what values are likely in practice. It is hypothesized that a very strong correlation between event times is unlikely in most situations and that moderate and weak correlations are more likely. To examine what the effects are of correlations on the stronger side of what is hypothesized to be possible, correlations between event times used in this study are -0.4 and 0.4 with 0 included to represent events that are uncorrelated. Weaker correlations would be expected to have less of an effect on results than ±0.4 but more of an effect than no correlation.

Event time format. The next parameter manipulated is the format the event times are recorded in. Most literature on survival analysis is focused on data that is recorded continuously, with exact times and dates for the event recorded. However, it is not always possible or economical to measure time data continuously. It is also sometimes advantageous, such as in the presence of time varying covariates, to have time data recorded in discrete intervals. In the student retention and graduation data used, data are recorded in discrete-time. This is due to the fact that while often the exact date is generally known for graduation, students do not always formally drop out but often simply stop coming or registering for classes. Data are first simulated as continuous then converted to a discrete format for comparison of analysis results.

Group failure rates by event type. The next parameter manipulated is the failure rate of each group for each event type. In this simulation group is the sole independent variable with differences between the groups being different likelihoods of experiencing the events. Results of these statistical analyses are displayed as hazard ratios of one group to the other. Failure rates of 0.5 and 1 are used to investigate the ability of the statistical tests when failure rates are: equal across groups and events, equal across events but double for one group, unequal across events with one event being more likely for one group and both events equally likely for the other. Failure rate combinations were chosen to represent: no covariate effect, equal covariate effect for both event times, and unequal covariate effect for event times respectively. Unequal covariate effect for event times is hypothesized for graduation and retention as it is likely that covariates would affect these events differently. The ratio of the failure rate of group 2 to the failure rate of group 1 is referred to as the hazard ratio.

Event time distribution shape. The next parameter manipulated is the event time distribution shape. The most common method of analyzing survival data, the Cox proportional hazard regression, is a semi-parametric test which makes no assumptions about the underlying hazard distribution. The Fine and Gray model also makes no such assumptions. Other parametric models, such as Weibull and exponential, make assumptions about the underlying hazard distribution. To evaluate the performance under different distributions, survival times are drawn from a Weibull distribution, as is commonly done in literature, using the quantile function

$$Q(p; k, \lambda) = \lambda(-\ln(1-p))^{1/k}$$

where λ is the scale parameter and *k* is the shape parameter. With *k* = 1 the Weibull distribution is the exponential distribution. Shape parameter *k* = 1 is used as a comparison due to its commonplace in recent literature. Shape parameter *k* = 2 is used due to its similarity with the student graduation distribution shown later in Figure 5. Both distributions are positively skewed though the graduation distribution is shifted further right. In Figures 2 and 3 below the Weibull distribution with both of the shape distributions are illustrated. A sample of 10,000 survival times was drawn from each of the distributions to create the following figures.

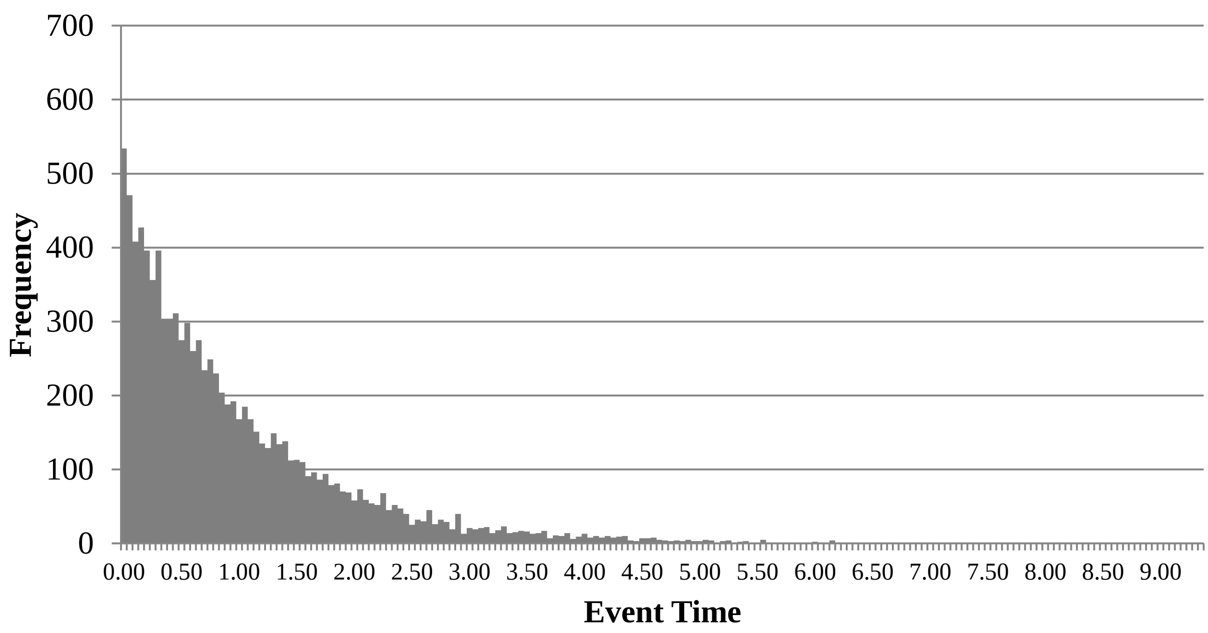

Figure 2. Weibull distribution with shape parameter k=1.

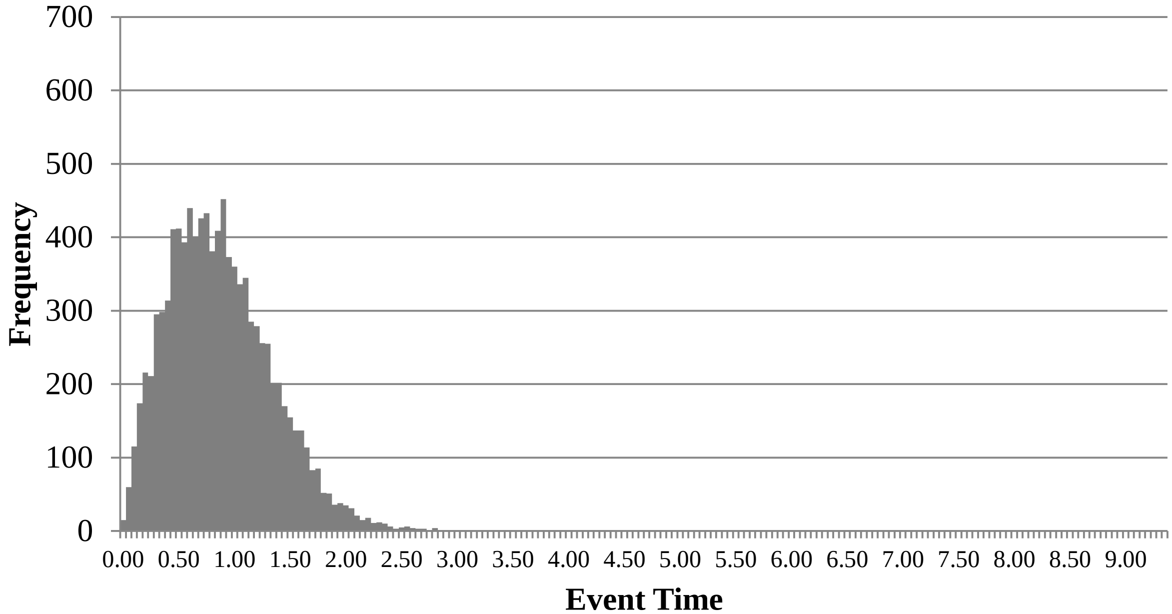

Figure 3. Weibull distribution with shape parameter k=2.

Frailty. The last parameter manipulated is frailty. Frailty accounts for unobserved heterogeneity of variance. Frailties are chosen such that the average frailty is constant across groups while the variance within group is acted on multiplicatively by the frailty parameter. Levels of the frailty parameter are set at no multiplicative frailty variance within group, moderate multiplicative frailty variance within group (multiplicative factor of 1.5 on standard deviation) and large multiplicative frailty variance within group (multiplicative factor of 2 on standard deviation).

Analyses

For each set of conditions, 1,000 data sets are generated and analyzed with each of four methods. The first method is the separate analysis of competing risks using two Cox proportional hazards regressions. The first analysis distinguishes event 1 as the event of interest and treats all event 2 times as if they are censored. The second analysis distinguishes event 2 as the event of interest and treats all event 1 times as if they are censored. These analyses are conducted using the coxph function included in the Survival package in R.

The second method is separate analysis of competing risks using two Weilbull regressions. The first analysis distinguishes event 1 as the event of interest and treats all event 2 times as if they are censored. The second analysis distinguishes event 2 as the event of interest and treats all event 1 times as if they are censored. These analyses are conducted using the survreg function included in the Survival package in R.

For the third method, two competing risks regressions are conducted in modeling the risks simultaneously using Fine and Gray's method. The first analysis distinguishes event 1 as the event of interest and treats event 2 as a competing event. The second analysis distinguishes event 2 as the event of interest and treats event 1 as a competing event. These analyses are conducted using the crr function included in the CMPRSK package in R.

The fourth method, like the first method, utilizes two Cox proportional hazard regressions to model competing risks separately with each analysis treating one event as the event of interest and the other as censored. This method however adds gamma frailty to the Cox proportional hazards regression. These analyses are conducted using the frailtypenal function included in the frailtypack package in R.

The results of these tests are averaged over the 1,000 repetitions. The model estimates are then compared with the set hazard ratios for the simulated data to investigate each method's ability to estimate the relative difference in failure rates between groups. The test of proportional hazards will be computed for each simulated data set. The proportion of these analyses found to have significantly non-proportional hazards at α=0.05 will be calculated. The test of proportional hazards is conducted using the cox.zph function included in the Survival package in R.

In total 216 combinations of conditions are evaluated. These conditions are summarized below in Table 6 and listed in detail in Table A1.

Table 6

Simulation Conditions

			Number of Levels	Conditions		
Sample Size			2	500	1000	
Event Time Correlation			3	-0.4	0	0.4
Event Time Format			2	Continuous	Discrete	
Failure Rate: Number of Failures per Time Period	Group 1	Event 1	3	1	0.5	0.5
		Event 2		1	1	0.5
	Group 2	Event 1		1	1	1
		Event 2		1	1	1
Shape: Weibull Shape Parameter *k*	Group 1	Event 1	2	1	2	
		Event 2		1	1	
	Group 2	Event 1		1	2	
		Event 2		1	1	
Frailty: Standard deviation Multiplicative Factor	Group 1	Event 1	3	1	1.5	2
		Event 2		1	1.5	2
	Group 2	Event 1		1	1	1
		Event 2		1	1	1

Simulation

Event times are simulated in r by sampling n event times for both event types from the standard normal distribution. A Cholesky decomposition is then multiplied by this $n \times 2$ matrix to obtain times that are correlated at the level set by the simulation. These times are then transformed into Weibull distributed data by calculating the cumulative density of these event times is then applying the Weibull quantile function. These Weibull times follow the function

$$e(x) = \frac{k}{\lambda}(\frac{x}{\lambda})^{k-1}e^{-(x/\lambda)^k} \quad x \geq 0$$

where k and λ are the Weibull shape and scale parameters respectively.

The shape parameter is assigned as 1 or 2 by the simulation and the scale parameter, the inverse of the failure rate, is likewise assigned as 1 or 2 by the simulation. An $n \times 1$ vector of normal distributed data was sampled with mean 1 and standard deviation assigned by the frailty factor and multiplied by the failure rate. The smallest event time is then selected as the "winning" competing risk. For discrete time simulations these event times are multiplied times four, shifted right one, and rounded to the nearest whole number to simulate student drop-out and graduation data.

The data are then analyzed by each of the four models and the resulting output, the estimated hazard ratio, was averaged over each of the 1000 repetitions. This average is then compared with the hazard ratio set by the simulation, by the formula below.

$$\frac{|Estimated\ Hazard\ Ratio - Set\ Hazard\ Ratio|}{Set\ Hazard\ Ratio}$$

The absolute value is taken of the difference to allow for average results by groups of similar conditions. The results prior to the absolute value can however be negative, indicating the error is an underestimate, or positive, indicating the error is an overestimate.

Student Retention and Graduation Data

As an illustrative example, each model will be fitted to a student data set to examine performance with real data. The student data are a sample of 5,419 students from a four-year Midwestern university. Definitions for events and other terms are chosen to best approximate the goals of a university administration interested in retaining and graduating students and minimizing drop-outs.

Enrollment. Student data spans six years with a record for each fall and spring semester indicating the student's enrollment status as a student for each semester. Semester 12 is therefore considered the conclusion of the study. The student census is taken at the 11th day of the semester with any subject who was enrolled until at least the 11th day counted as being enrolled for that semester. Enrollment was defined as registration for any university program of study with any non-zero number of credit hours.

Graduation. A student is considered graduated as long as he or she completes a program of study within six years of his or her first enrollment. A student who begins in a fall semester has until the end of the summer term of his or her sixth year to graduate to be counted among the graduates. This definition is used because six-year graduation rates are those published and of interest to universities.

Drop-out. A student is considered retained so long as he or she is enrolled for the eleventh day census for each subsequent semester. As students may skip a semester between enrollments, it is necessary to define a single, non-repeatable event as drop-out to permit evaluation as a competing risk. The event of drop-out is therefore defined as occurring directly before the first semester during which a student is not enrolled for the eleventh day census. A student is also said to experience the event of drop-out if he or she has not graduated within six years – even if the student stays enrolled. This is because a student who continues past six years without graduating is counted the same as a student who has dropped out in a university's six-year graduation rate. Students who drop out during or directly after their second semester, for example, would have an event time of 2. To account for students who stop out for a period of time before returning and completing a program of study within six years, the event of drop-out is ignored and replaced with graduate should a student complete his or her degree within six years. From the perspective of a university administration, for a student who drops out and later returns but does not graduate within six years, the first drop-out is likely to be the event of highest interest. For a student who drops out and later returns to graduate within six years, a university administration would have no reason to see that student as unsuccessful.

Covariate. One categorical variable is included in the model. The categorical variable is a socioeconomic status proxy variable indicating if a student is either a first-generation student – a student who does not have a parent who has graduated from an institution of higher education – or is classified as a low-income student. A student meeting either condition is coded as 1; students meeting neither condition are coded 0.

Event time distributions. Due to the nature of the issue being studied, the distributions for the two events are quite dissimilar. Students who experience event 1, drop-out, have an average event time of 5.34 with a standard deviation of 3.82. Drop-outs peak after the second semester and start a gradual downward trend ending with a sharp increase after semester 12 due to the conclusion of the study and the subsequent conversion of all who have not graduated to the event of “drop-out” as previously defined. This distribution is displayed below in Figure 4.

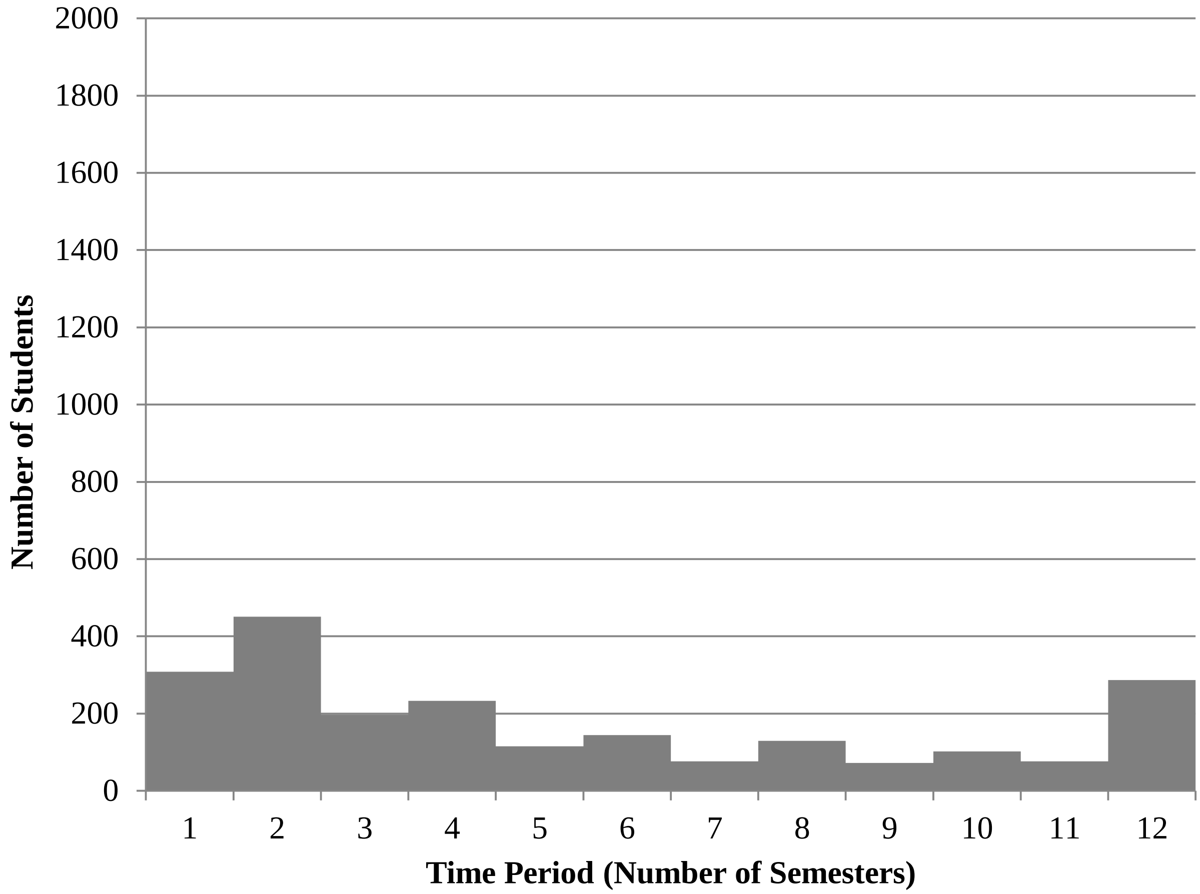

Figure 4. Event 1: Student drop-out occurrences by time.

Students experiencing event 2, graduation, have an average event time of 8.76 with a standard deviation of 1.23. Event times for graduation are highly clustered at time period 8, which represents the standard four-year graduation time for a bachelor’s degree. This distribution is displayed below in Figure 5.

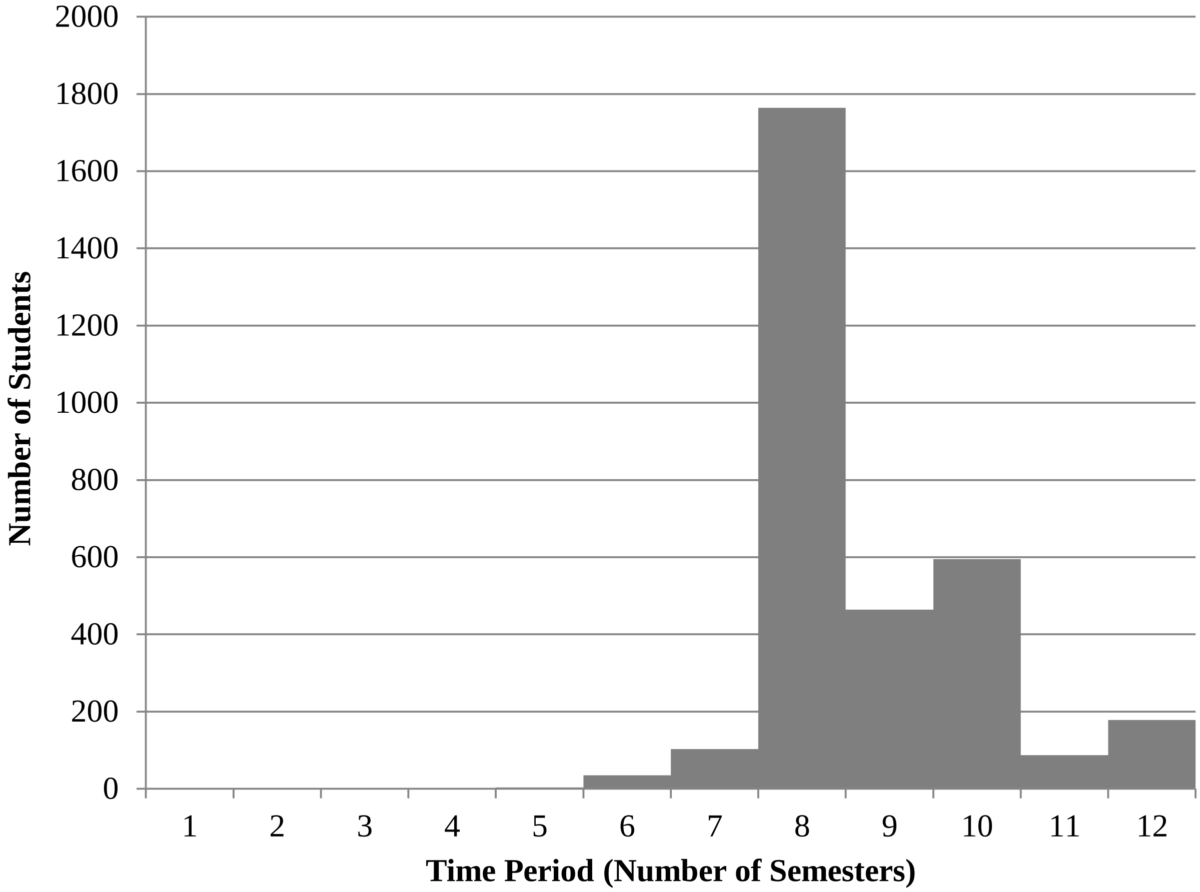

Figure 5. Event 1: Student graduation occurrences by time.

Chapter IV: Results

The following chapter presents a summary of the analyses conducted. The first section details the results of the simulation study which consists of 216 condition combinations each averaged over 1000 simulated data sets. Results are discussed as a comparison of the hazard ratio estimated by the model with the hazard ratio specified in the simulation. The estimated hazard ratio is computed by taking the exponential of the coefficient of the independent variable group while the simulation hazard ratio is computed by dividing the group 2 failure rate by the group 1 failure rate. The second section details the results of the student retention and graduation analyses. Full results are found in Table A2.

This paper will define the following combination of conditions as the "base combination" from which comparisons will be made: sample size of 500 per group, continuous event time format, no correlation between event times, homogeneous shape parameter for both events for both groups, homogeneous failure rates for both events for both groups, and no frailty. For this condition, the parameter estimate is very accurate for all models as seen below in Table 7.

Table 7

Percentage Difference Between Model Estimates and Set Hazard Ratios

Condition	Cox Proportional Hazards		Weibull Regression		Fine and Gray's Method		Cox Proportional Hazards Regression with Frailty	
	Event 1	Event 2	Event 1	Event 2	Event 1	Event 2	Event 1	Event 2
Base Condition	0.50%	0.56%	0.55%	0.58%	0.43%	0.44%	0.54%	0.57%

If any one parameter is changed from the base condition, most of the models are still able to estimate the parameter with good to fair accuracy as seen below in Table 8 – the main exception to this being Fine and Gray's method when failure rate varies by group or event.

Table 8

Percentage Difference Between Model Estimates and Set Failure Rates

Parameter Changed from Base Condition	Cox Proportional Hazards Regression		Weibull Regression		Fine and Gray's Method		Cox Proportional Hazards Regression with Frailty	
	Event 1	Event 2	Event 1	Event 2	Event 1	Event 2	Event 1	Event 2
n=250	0.57%	1.42%	0.52%	1.52%	0.17%	1.60%	0.54%	1.47%
Frailty = 1.5	2.43%	3.33%	2.46%	3.36%	1.58%	2.77%	2.45%	3.37%
Frailty = 2	4.27%	5.71%	4.30%	5.72%	2.84%	4.64%	4.37%	5.74%
Failure Rates 0.5,1	0.85%	0.38%	0.80%	0.31%	11.57%	30.75%	0.94%	0.33%
Failure Rates 0.5,0.5	0.68%	0.71%	0.55%	0.58%	43.48%	43.52%	0.76%	0.72%
Shape Parameter E1 k=1, E2 k=2	0.52%	0.63%	0.16%	0.66%	0.29%	0.53%	0.52%	0.70%
r = -0.4	0.60%	0.61%	0.54%	0.56%	0.44%	0.44%	0.61%	0.63%
r = 0.4	0.46%	0.52%	0.54%	0.57%	0.42%	0.44%	0.48%	0.52%
Discrete Time	0.49%	0.56%	0.30%	0.32%	0.37%	0.39%	0.72%	0.76%

Note. Cells are shaded darker as the values deviate further from zero.

The overall pattern of results is displayed below in Figure 6. Cells in this figure represent the deviation from zero of the absolute percentage difference between the estimated hazard ratio and the set hazard ratio. Values closer to zero are lighter in color and values further from zero are darker in color. From this it can be seen that the error in estimations by Cox proportional hazards are generally small but for certain blocks of parameter combinations these errors are quite large. A very similar pattern is seen in Cox proportional hazards with frailty. For the Weibull regression it can be seen that most errors are fairly small and for Fine and Gray's method it can be seen that the errors are either moderate or small. Detailed results from this figure are discussed in the following sections.

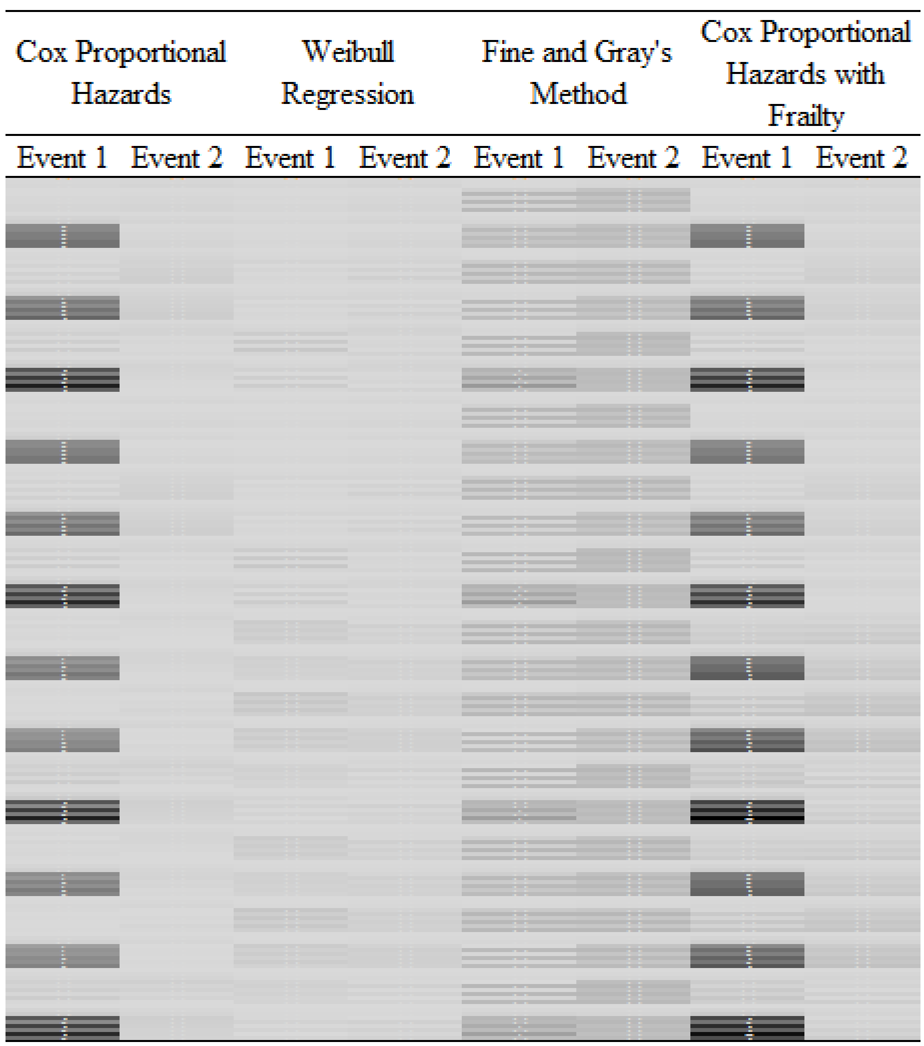

Figure 6. Parameter estimation error deviation from zero.

Method 1: Cox Proportional Hazards Regression

In survival analysis, Cox proportional hazards regression is the standard. Its ease of use, straight forward interpretation, semi-parametric nature and inclusion in commercial statistical programs makes it a popular choice among researchers. Studies by Dignam et al. (2012) and Williamson et al. (2007) suggest it performs well under a variety of conditions. While this method does not assume a shape to the distribution, it does assume that the hazards are proportional. This assumption is tested and the results are reported below.

Overall model accuracy. Model accuracy is judged by the absolute percentage difference between the hazard ratio, set at either 1 or 2 by the simulation, and the model estimate of the same ratio. In general, this model's errors were overestimates with no clear pattern of conditions for the minority that were underestimates for event 1. For event 2 underestimates were less likely for data where failure rates were equal. Overall, the average percentage difference between the model estimate and the set hazard ratio is 45.91% for event 1 and 5.22% for event 2.

Examining each parameter independent of the others, in other words, changing just one parameter at a time from the base combination, Cox proportional hazards regression performs well with: a) both group sample sizes (250 and 500), b) positive, negative and zero correlation (0.4, -0.4, 0) between event times, c) both continuous and discrete data, d) varying failure rates between and within groups, e) different event time distribution shapes, and f) it performs moderately well with frailty. Looking at these parameters independent of the others, the percentage difference between the average model estimate and the set hazard ratio is 1.01% for event 1 and 1.34% for event 2. This method is also able to perform well with some combinations of conditions; however, some combinations cause the model estimate to deviate considerably from the set hazard ratio. Of some concern are data sets containing positively correlated continuous event times, and differing failure rates by group. The percentage difference between the model estimates and the set hazard ratios averaged 8.20% for event 1 and 3.97% for event 2 for these conditions.

However, the model performs especially poorly with data sets containing both differing event time distribution shapes by event and differing failure rates by group or event. The percentage difference between model estimates for data sets containing this combination and the set hazard ratios is 127.13% for event 1 and 6.38% for event 2.

Differences by parameter. For some parameters, the model estimates of the Cox proportional hazards regression are similar regardless of the condition level, for others, model estimates vary greatly by condition level. The differences between model estimates of condition levels are presented as a percentage of the set hazard ratios by the simulation. The robustness of the model is judged based on these differences – with smaller differences indicating the model is robust with respect to the condition levels.

***Sample size*.** Group sizes of 250 and 500 are simulated for each set of conditions for a total of 500 and 1000 respectively.

Examining the Cox proportional hazards results, the average percentage differences between the deviations of the model estimates from the set hazard ratios of identical combinations, save for sample size, are relatively small at 1.75% for event 1 and 0.94% for event 2. Percentage differences in model estimates between sample sizes appears to be the smallest when failure rates are the same for both events across both groups, regardless of what other conditions are present. Percentage differences are somewhat larger if failure rates are heterogeneous between group or within group or if event times are correlated and are largest when distribution shape differs by event type and failure rates are heterogeneous between group or within group. Average percentage differences between the deviations of the model estimates from the set hazard ratios of identical combinations, save for sample size is minimal between different event time formats and frailties.

Correlation between event times. Correlation between Event 1 and Event 2, for both groups, is set at either -0.4, 0, or 0.4 as seen in Table 9.

Table 9

Event Time Correlation Conditions

Even Time Correlation	Group 1	Group 2
Condition 1	-0.4	-0.4
Condition 2	0	0
Condition 3	0.4	0.4

Examining the Cox proportional hazards regression results, the average percentage differences between the deviations of the model estimates from the set hazard ratios of identical combinations, save for event time correlation, range from minimal to quite large. The average percentage differences between the deviations of the model estimates from the set hazard ratios between event time correlation conditions 1 and 3 is 24.56% for event 1 and 5.29% for event 2, between event time correlation conditions 1 and 2 the difference is 6.83% for event 1 and 4.07% for event 2, while the difference between event time correlation conditions 2 and 3 is 21.17% for event 1 and 3.07% for event 2. This trend in percentage differences, larger between event time correlation conditions 1 and 3, smaller between event time correlation conditions 2 and 3, and smaller still between event time correlation conditions 1 and 2, holds for most, but not all conditions. Percentage differences between event time correlation conditions 2 and 3 or 1 and 3 are smaller than between correlation conditions 1 and 2 for event 1 under two scenarios: when failure rates differ by group and event times are in continuous format, and when failure rates differ by group, event times are in discrete-time format, and event time distributions differ by event type. Event 2 differences exhibit a similar trend, though differ for some condition combinations when event time distributions or failure rates differ by event type.

The percentage difference between each pair of correlations is the smallest when failure rates are the same for both events across both groups, regardless of what other conditions are present with an average percentage difference of 1.34% for event 1 and 0.53% for event 2. Larger percentage differences are found in model estimates between each pair occur in combinations of condition levels containing differing event time distribution shapes and failure rates which differ by group and event type with an average percentage difference of 72.49% for event 1 and 6.63% for event 2. These differences are larger still when event time distribution shape differs by event type and frailty is present. Average percentage differences between the deviations of the model estimates from the set hazard ratios of identical combinations, save for event time correlation, are minimal between sample sizes and even time formats.

Event time format. Time data are simulated in two formats, discrete and continuous, with discrete data created as rounded continuous data. Examining the Cox proportional hazards regression results, the average percentage differences between the deviations of the model estimates from the set hazard ratios between event time formats are 5.88% for event 1 and 5.24% for event 2. These differences are the smallest when failure rates are the same for both events across both groups, regardless of what other conditions are present. These differences are slightly larger when the failure rate of one event for one group differs, but are much larger when failure rates for both events in one group differ from those of the other group. These differences are most pronounced when failure rates differ by group, event time distribution shape differs by event type, and frailty is present. Average percentage differences between the deviations of the model estimates from the set hazard ratios of identical combinations, save for event time format, are higher when a negative correlation between even times is present, smaller when event times have no correlation, and smaller still when event times are positively correlated. These differences are essentially the same for both sample sizes.

Group failure rates by event type. Group failure rates are simulated in three combinations with group 2 serving as the control with equal failure rates for both events. The combinations are: homogeneous failure rates within group 1 which equal those of group 2 (failure rate condition 1), homogeneous failure rates within both groups but heterogeneous between groups (failure rate condition 2), and heterogeneous failure rate within group 1 (failure rate condition 3). These can be seen in Table 10.

Table 10

Failure Rate Conditions: Number of Failures per Time Period

	Group 1		Group 2	
Failure Rate	Event 1	Event 2	Event 1	Event 2
Condition 1	1	1	1	1
Condition 2	0.5	0.5	1	1
Condition 3	0.5	1	1	1

Examining the Cox proportional hazards regression results, the average percentage differences between the deviations of the model estimates from the set hazard ratios of identical combinations, save for group failure rates by event type, range from minimal to very large. The average percentage difference between failure rate conditions 1 and 2 is 117.32% for event 1 and 8.08% for event 2, between failure rate conditions 1 and 3 the difference is 69.30% for event 1 and 4.04% for event 2, while the difference between failure rate conditions 2 and 3 is 22.16% for event 1 and 2.65% for event 2. These differences are smaller when distribution shape does not differ by event type, and smallest when that is combined with a zero correlation. In general an examination of differences in model estimates that includes failure rate condition 3, which has heterogeneous failure rates within group 1, will find much larger differences. The difference in model estimates between conditions 1 and 2 is generally small except when in the presence of the combination of non-zero correlation of event times and distribution shapes that differ by event type. Differences in model estimates are slightly smaller with larger sample sizes when the event time is discrete.

Shape parameter for event time distributions. The shape parameter is simulated such that it is either homogeneous within and between groups or heterogeneous within groups with the same heterogeneity for both groups as seen in Table 11.

Table 11

Shape Parameter Conditions: Weibull Shape Parameter k

	Group 1		Group 2	
Shape Parameter	Event 1	Event 2	Event 1	Event 2
Condition 1	1	1	1	1
Condition 2	2	1	2	1

Examining the Cox proportional hazards regression results, the average percentage differences between the deviations of the model estimates from the set hazard ratios of identical combinations, save for even time shape distribution, are 84.30% for event 1 and 1.37% for event

2. The large discrepancy between percentages for event 1 and event 2 is not unexpected as it is event 1 for both groups that the event shape differs; event 2 shape does not differ. The large percentage difference for event 1 is also not entirely unexpected. While Cox Proportional Hazards does not assume an event distribution, it does assume that the hazards are proportional – an assumption true for each of the other models as well. With many combinations including different failure rates by group and different event time distribution shapes, it was possible that this assumption might be violated for some combinations of conditions. It is standard procedure to test for proportional hazards prior to conducting a proportional hazards regression. This test of proportional hazards was conducted for each simulation, with an average of 3.73% of simulations indicating hazards were not proportional at an alpha of 0.05. This suggests that the simulated data sets, by and large, do not violate the proportionality assumption of the test.

The average percentage differences between the deviations of the model estimates from the set hazard ratios of identical combinations, save for even time shape distribution, are quite minimal, 0.15% for event 1 and 0.12% for event 2, if the failure rates do not vary between or within groups and there is no frailty. These differences are moderately larger with larger frailty and dramatically smaller when failure rates differ by group or within group. The percentage differences in model estimates are larger for positively correlated event times than zero or negatively correlated event times and are relatively equal between sample sizes and event time formats.

Frailty. The frailty parameter is simulated such that it is homogeneous within groups for both event types at three levels: no frailty, moderate frailty, and high frailty as seen in Table 12.

Table 12

Frailty Conditions: Standard Deviation Multiplicative Factor

	Group 1		Group 2	
Frailty Condition	Event 1	Event 2	Event 1	Event 2
Condition 1	1	1	1	1
Condition 2	1.5	1.5	1	1
Condition 3	2	2	1	1

Examining the Cox proportional hazards regression results, the average percentage differences between the deviations of the model estimates from the set hazard ratios of identical combinations, save for frailty, are smallest between no frailty and moderate frailty, largest between no frailty and high frailty, and in the middle between moderate frailty and high frailty for nearly all conditions with a different pattern occurring only when the differences between the conditions are less than 1%.

Overall the average percentage differences between the deviations of the model estimates from the set hazard ratios of identical combinations, save for frailty, are 7.63% for event 1 and 2.84% for event 2 between no frailty and moderate frailty, 8.34% for event 1 and 2.42% for event 2 between moderate frailty and high frailty and 15.96% for event 1 and 5.15% for event 2 between no frailty and high frailty. When event time shape distribution is the same for both events, differences are minimal and change very little with differing failure rates between or within groups. Average percentage differences between the deviations of the model estimates from the set hazard ratios of identical combinations, save for frailty, are larger with positively correlated event times than for negative or zero correlations. These differences are minimally affected with different sample sizes and time formats. These differences are larger than average when event time shape distributions differ by group and are larger still when event times are positively correlated and failure rates differ between or within group.

Method 2: Weibull Regression

The Weibull regression is a parametric test that assumes the distribution of event times is Weibull distributed. Like Cox proportional hazards regression, Weibull regression analyzes one event at a time while treating the competing event as censored. For a small subset of the simulations, the Weibull regression failed to converge. These failed analyses represent 0.14% of the Weibull analyses. Descriptive statistics were used to compare data whose analyses did converge to data whose analyses did not converge with. No discernable differences or anomalies were found. It is assumed that these failed analyses resulted due to limitations of the software.

Overall model accuracy. Model accuracy is judged by the absolute percentage difference between the ratio of failure rates of group 2 and group 1, the hazard ratio, set at either 1 or 2 by the simulation, and the model estimate of the same ratio. In general, this model's errors were overestimates with underestimates more likely with discrete data and heterogeneous failure rates. Overall, the average percentage difference between the model estimate and the set hazard ratio is 6.61% for event 1 and 5.75% for event 2.

Examining each parameter independent of the others, in other words, changing just one parameter at a time from the base combination, Weibull regression performs well with: a) both group sample sizes (250 and 500), b) positive, negative and zero correlation (0.4, -0.4, 0) between event times, c) both continuous and discrete data, d) varying failure rates between and within groups, e) different event time distribution shapes, and f) it performs moderately well with frailty. Looking at these parameters independent of the others, the percentage difference between the average model estimate and the set hazard ratio is 0.93% for event 1 and 1.31% for event 2. This method is also able to perform well with many combinations of conditions.

However, there are two combinations of parameters that appear to cause larger percentage differences between model estimates and the set hazard ratios. First is the combination of differing failure rates by group or event and discrete-time data. The average percentage difference between the model estimate and set hazard ratio for this combination is 12.33% for event 1 and 9.21% for event 2. Second is the combination of continuous time data, differing failure rates by group and event, and positively correlated event times. The average percentage difference between the model estimate and set hazard ratio for this combination is 17.79% for event 1 and 4.71% for event 2. Model estimates for this combination of conditions differs even further from the set hazard ratios when correlation between event times is negative. Omitting these combinations, the percentage difference between the average model estimate and the hazard ratio of the remaining 144 condition combinations is 2.47% for event 1 and 3.95% for event 2. The average percentage differences are similar between large and small sample size and homogeneous and heterogeneous event time shape distributions.

Differences by parameter. For some parameters, the model estimates of the Weibull regression are similar regardless of the condition level, for others, model estimates vary greatly by condition level. The differences between model estimates of condition levels are presented as a percentage of the set hazard ratios by the simulation. The robustness of the model will be judged based on these differences – with smaller differences indicating the model is robust with respect to the condition levels.

***Sample size*.** Group sizes of 250 and 500 are simulated for each set of conditions for a total of 500 and 1000 respectively.

Examining the Weibull regression model estimates of identical conditions, save for sample size, the average percentage differences between the deviations of the model estimates from the set hazard ratios are very small with an average percentage difference in model estimates of 0.16% for event 1 and 0.72% for event 2. These differences change very little when examining different levels of conditions with all but two differences under 1% for event 1.

***Correlation between event times*.** Correlation between Event 1 and Event 2, for both groups, is set at either -0.4, 0, or 0.4. Correlation conditions are listed above in Table 9.

Examining the Weibull regression model estimates of identical conditions, save for event time correlation, average percentage differences between the deviations of the model estimates from the set hazard ratios range from minimal to quite large. The average percentage differences between the deviations of the model estimates from the set hazard ratios between event time correlation conditions 1 ($r = -0.4$) and 3 ($r = 0.4$) is 7.68% for event 1 and 3.57% for event 2, between event time correlation conditions 1 and 2 ($r = 0$) the difference is 2.80% for event 1 and 2.24% for event 2, while the difference between event time correlation conditions 2 and 3 is 4.88% for event 1 and 2.17% for event 2. This trend in percentage differences, larger between event time correlation conditions 1 and 3, smaller between event time correlation conditions 2 and 3, and smaller still between event time correlation conditions 1 and 2, holds for most combination conditions for event 1. For event 2, differences are still larger between correlation conditions 1, 3 for the most part but are mixed between conditions 2, 3 and 1, 2.

This percentage difference between each pair of correlations are small when failure rates are the same for both groups, even if they are different between groups, regardless of what other conditions are present with an average percentage difference of 1.04% for event 1 and 1.26% for event 2. Also notable for this model, the average percentage difference between model estimates when event time distribution shape differs by event type is actually smaller than when event time distribution shape does not differ by event type. The largest average percentage differences between the deviations of the model estimates from the set hazard ratios of identical combinations, save for event time correlation, occur in combinations containing failure rates which differ by group and event type with an average percentage difference of 13.27% for event 1 and 5.47% for event 2. Percentage differences in model estimates differ only slightly between sample sizes, event time formats, event time distribution shapes and frailties.

Event time format. Time data are simulated in two formats, discrete and continuous, with discrete data created as rounded continuous data. Examining the Weibull regression model, the average percentage differences between the deviations of the model estimates from the set hazard ratios of identical combinations, save for event time format vary greatly between condition combinations with an average percentage difference of 10.91% for event 1 and 6.99% for event 2. These differences are the smallest, 0.97% on average for event 1 and 1.78% for event 2, when failure rates are the same for both events and groups, regardless of what other conditions are present. These differences are considerably larger when the failure rates differ by group but not event within each group. The average percentage difference is larger at 13.39% for event 1 and 15.61% for event two when failure rates differ by group and event type. These differences are largest when differing failure rates by group and event type are combined with positively correlated event times and frailty. Percentage differences in model estimates between event time formats are not notably different between the different simulated sample sizes.

Group failure rates by event type. Group failure rates are simulated in three combinations with group 2 serving as the control with equal failure rates for both events. The combinations are: homogeneous failure rates within group 1 which equal those of group 2 (failure rate condition 1), homogeneous failure rates within both groups but heterogeneous between groups (failure rate condition 2), and heterogeneous failure rate within group 1 (failure rate condition 3). These can be seen above in Table 10.

Examining the Weibull regression, the average percentage differences between the deviations of the model estimates from the set hazard ratios of identical combinations, save for group failure rates by event type range from minimal to moderately large. The average percentage difference in model estimates between failure rate conditions 1 and 2 is 13.02% for event 1 and 12.56% for event 2, between failure rate conditions 1 and 3 the difference is 11.28% for event 1 and 4.57% for event 2, while the difference between failure rate conditions 2 and 3 is 6.51% for event 1 and 5.51% for event 2. For event 1 the percentage differences in model estimates between failure rate conditions 2 and 3 are largest when event time is continuous and event times are positively correlated while differences between failure rate conditions 1, 3 and 1, 2 are largest when event time is discrete and events are negatively correlated. For event 2 these differences are largest when event time is discrete and event times are negatively correlated. Differences in model estimates of identical conditions, save for group failure rates by event type, are similar between the different simulated sample sizes, event time distribution shapes, and frailties.

***Shape parameter for event time distributions*.** The shape parameter is simulated such that it is either homogeneous within and between groups or heterogeneous within groups with the same heterogeneity for both groups. Shape parameter conditions are listed above in Table 11.

Examining the Weibull regression results, the average percentage differences between the deviations of the model estimates from the set hazard ratios of identical combinations, save for even time shape distribution, are relatively small with an average percentage difference of 2.08% for event 1 and 0.85% for event 2.

The average percentage differences between the deviations of the model estimates from the set hazard ratios of identical combinations, save for even time shape distribution, are largest, 6.05% for event 1 and 2.94% for event 2, when failure rates vary both between and within groups and events are positively correlated. These differences are larger for positively correlated event times and discrete-time format but are minimally different between the different simulated sample sizes and frailties.

Frailty. The frailty parameter is simulated such that it is homogeneous within groups for both event types at three levels: no frailty, moderate frailty, and high frailty. Frailty conditions are listed above in Table 12.

Examining the Weibull regression results, average percentage differences between the deviations of the model estimates from the set hazard ratios of identical combinations, save for frailty, are similar between frailty conditions 1 and 2 (no frailty and moderate frailty) and frailty conditions 2 and 3 (moderate and high frailty) and largest between frailty conditions 1 and 3 (no frailty and high frailty).

Overall average percentage differences between the deviations of the model estimates from the set hazard ratios of identical combinations, save for frailty, are 1.68% for event 1 and 2.44% for event 2 between no frailty and moderate frailty, 1.73% for event 1 and 1.97% for event 2 between moderate frailty and high frailty and 3.41% for event 1 and 4.38% for event 2 between no frailty and high frailty. These differences are largest when event times are in continuous time format and events are positively correlated. The average percentage differences between the deviations of the model estimates from the set hazard ratios of identical combinations, save for frailty, are minimal with different sample sizes, event time distribution shapes and failure rates.

Method 3: Fine and Gray's Method

Fine and Gray's method is the only method tested that is able to model the competing events simultaneously by treating one event as the event of interest and the other as a competing event. All other methods simply treat the competing risk as censored and instead analyze only the event of interest. Because of this, it is also the only method that does not assume event times are uncorrelated.

Overall model accuracy. Model accuracy is judged by the absolute percentage difference between the ratio of failure rates of group 2 and group 1, the hazard ratio, set at either 1 or 2 by the simulation, and the model estimate of the same ratio. In general, this model's errors were underestimates with no clear pattern of conditions for the minority that were overestimates. Overall, the average percentage difference between the model estimate and the set hazard ratio is 21.49% for event 1 and 25.18% for event 2.

Examining each parameter independent of the others, in other words, changing just one parameter at a time from the base combination, Fine and Gray's method performs well with: a) both group sample sizes (250 and 500), b) positive, negative and zero correlation (0.4, -0.4, 0) between event times, c) both continuous and discrete data, d) different event time distribution shapes, and e) with frailty. It performs less well with varying failure rates between and within groups. Looking at these parameters independent of the others, omitting failure rate, the percentage difference between the average model estimate and the set hazard ratio is 0.71% for event 1 and 1.21% for event 2.

Looking at a varying failure rate between groups independent of the other parameters, the percentage difference between the average model estimate and the set ratio of failure rates is 43.48% for event 1 and 43.52% for event 2 when failure rates differ by group and event. This method is also able to perform well with most combinations of conditions that do not include varying failure rates.

Omitting condition combinations with varying failure rates, the percentage difference between the average model estimate and the set ratio of failure rates of the remaining 72 condition combinations is 2.50% for event 1 and 3.52% for event 2. Examining only the condition combinations with varying failure rates, the average percentage difference is 30.98% for event 1 and 36.01% for event 2. Differences are larger when event times are positively correlated or event time distribution shape differs by event. These differences are larger when failure rate differs only between groups than when they differ both between and within groups.

Differences by parameter. For some parameters, the model estimates of Fine and Gray's method are similar regardless of the condition level, for others, model estimates vary greatly by condition level. The differences between model estimates of condition levels are presented as a percentage of the set hazard ratios by the simulation. The robustness of the model will be judged based on these differences – with smaller differences indicating the model is robust with respect to the condition levels.

Sample size. Group sizes of 250 and 500 are simulated for each set of conditions for a total of 500 and 1000 respectively.

Examining Fine and Gray's method, the average percentage differences between the deviations of the model estimates from the set hazard ratios, save for sample size, are very small with an average percentage difference in model estimates of 0.23% for event 1 and 0.96% for event 2. These differences change very little when examining different levels of conditions with nearly all under 1%. The relatively larger differences, the largest of which being 3.03%, occur when event times are positively correlated and have differing distribution shapes, failure rates differ by group and event type, and when frailty is present. Percentage differences between identical conditions, save for sample size, are relatively unchanged between event time formats.

Correlation between event times. Correlation between Event 1 and Event 2, for both groups, is set at either -0.4, 0, or 0.4. Correlation conditions are listed above in Table 9.

Examining Fine and Gray's method results, the average percentage differences between the deviations of the model estimates from the set hazard ratios of identical combinations, save for event time correlation, range from minimal to quite large. The average percentage differences between the deviations of the model estimates from the set hazard ratios between event time correlation conditions 1 ($r = -0.4$) and 3 ($r = 0.4$) is 14.38% for event 1 and 4.12% for event 2, between event time correlation conditions 1 and 2 ($r = 0$) the difference is 4.06% for event 1 and

1.64% for event 2, while the difference between event time correlation conditions 2 and 3 is 10.37% for event 1 and 2.51% for event 2. This trend in percentage differences, larger between event time correlation conditions 1 and 3, smaller between event time correlation conditions 2 and 3, and smaller still between event time correlation conditions 1 and 2, holds for most, but not all condition combinations. Those that do not hold to this trend have minimal differences of less than 1%.

The percentage difference between each pair of correlations are the smallest when failure rates are the same for both events across both groups, regardless of what other conditions are present with an average percentage difference of 1.30% for event 1 and 1.05% for event 2. The average difference is only moderately larger at 1.78% for event 1 and 0.97% for event 2 when event time distribution shape differs by event type. Larger percentage differences in model estimates of identical conditions, save for event time correlation, occur in combinations containing failure rates which differ by group and event type with an average percentage difference of 26.91% for event 1 and 6.76% for event 2. These differences are larger still when event time distribution shape differs by event type and frailty is present. Percentage differences in model estimates differ only slightly between the different sample sizes and event time formats.

Event time format. Time data are simulated in two formats, discrete and continuous, with discrete data created as rounded continuous data. Examining Fine and Gray's method, the average percentage differences between the deviations of the model estimates from the set hazard ratios of identical combinations, save for event time format are relatively small at 1.60% for event 1 and 0.88% for event 2. These differences are the smallest, only 0.17% on average for event 1 and 0.31% for event 2, when failure rates are the same for both events across both groups, regardless of what other conditions are present. These differences are only slightly larger when the failure rates differ by group but are the same for both events within each group. The average percentage difference increases to 3.53% for event 1 and 0.99% for event 2 when failure rates differ by group and event type. These differences are largest when differing failure rates by group and event type are combined with differing event time distribution shapes. Percentage differences in model estimates between event time formats are not notably different between the different simulated correlations or sample sizes.

***Group failure rates by event type*.** Group failure rates are simulated in three combinations with group 2 serving as the control with equal failure rates for both events. The combinations are: homogeneous failure rates within group 1 which equal those of group 2 (failure rate condition 1), homogeneous failure rates within both groups but heterogeneous between groups (failure rate condition 2), and heterogeneous failure rate within group 1 (failure rate condition 3). These can be seen above in Table 10.

Examining Fine and Gray's method model estimates of identical conditions, save for group failure rates by event type, percentage differences in estimations between the correlations range from minimal to very large. The average percentage differences between the deviations of the model estimates from the set hazard ratios between failure rate conditions 1 and 2 is 42.55% for event 1 and 39.61% for event 2, between failure rate conditions 1 and 3 the difference is 39.68% for event 1 and 27.12% for event 2, while the difference between failure rate conditions 2 and 3 is 51.75% for event 1 and 52.11% for event 2. The differences in model estimates between conditions are smaller when distribution shape does not differ by event type. Differences in model estimates of identical conditions, save for group failure rates by event type, are similar between the different simulated event time correlations when event time distribution shape is the same for both groups. If event time distribution shape differs by group these differences are largest for positively correlated event times and smallest for the negatively correlated event times.

Shape parameter for event time distributions. The shape parameter is simulated such that it is either homogeneous within and between groups or heterogeneous within groups with the same heterogeneity for both groups. Shape parameter conditions are listed above in Table 11.

Examining Fine and Gray's method, the average percentage differences between the deviations of the model estimates from the set hazard ratios of identical combinations between homogeneous shape parameters within groups and heterogeneous shape parameters within group are 14.78% for event 1 and 1.63% for event 2. The discrepancy between percentages for event 1 and event 2 is not unexpected as it is event 1 for both groups that the event shape differs; event 2 shape does not differ.

The average percentage difference for the model estimates are quite minimal, 1.85% for event 1 and 0.46% for event 2, if the failure rates do not vary between or within groups and frailty is zero. These differences are moderately larger when failure rates differ by group but not by event type within group and are dramatically larger when failure rates differ by group and by event type. The differences also increase moderately as frailty increases. The percentage differences in model estimates are larger for positively correlated event times than zero or negatively correlated event times and are relatively equal between sample sizes and event time formats.

Frailty. The frailty parameter is simulated such that it is homogeneous within groups for both event types at three levels: no frailty, moderate frailty, and high frailty. Frailty conditions are listed above in Table 12.

Examining Fine and Gray's method model estimates of identical conditions, save for frailty, percentage differences in estimations are smallest between no frailty and moderate frailty, largest between no frailty and high frailty, and in the middle between moderate frailty and high frailty for all 216 differences for events 1 and 2.

Overall the average percentage differences between the deviations of the model estimates from the set hazard ratios of identical combinations, save for frailty, are 2.42% for event 1 and 1.90% for event 2 between no frailty and moderate frailty, 2.79% for event 1 and 1.49% for event 2 between moderate frailty and high frailty and 5.21% for event 1 and 3.38% for event 2 between no frailty and high frailty. These differences are smallest when failure rates differ by group but not by event type and event times are negatively correlated. Percentage differences in model estimations are larger than average when event time shape distributions differ by group. These differences are minimal with comparing condition combinations with different sample sizes and event time formats.

Method 4: Cox Proportional Hazards Regression with Frailty

Similar to Method 1, Method 4 utilizes Cox proportional hazards regression as the method of analysis. In this method, a shared gamma frailty model is added. Like Method 1, hazards are assumed to be proportional and events are modeled independent of one another.

Overall model accuracy. Model accuracy is judged by the absolute percentage difference between the ratio of failure rates of group 2 and group 1, the hazard ratio, set at either 1 or 2 by the simulation, and the model estimate of the same ratio. This model's errors were almost exclusively overestimates. Overall, the average percentage difference between the model estimate and the set hazard ratio is 53.69% for event 1 and 8.66% for event 2.

Examining each parameter independent of the others, in other words, changing just one parameter at a time from the base combination, Cox proportional hazards regression with frailty performs well with: a) both group sample sizes (250 and 500), b) positive, negative and zero correlation (0.4, -0.4, 0) between event times, c) both continuous and discrete data, d) varying failure rates between and within groups, e) different event time distribution shapes, and f) it performs moderately well with frailty.

Looking at these parameters independent of the others, the average percentage difference between the model estimates and the set hazard ratio is 1.07% for event 1 and 1.39% for event 2. This method is also able to perform well with some combinations of conditions.

There are however, some combinations cause the model estimate to deviate drastically from the set hazard ratio. Most different are combinations of conditions including different event time distribution shapes and differing failure rates by group or event. The percentage difference between the model estimates and the set hazard ratios averaged 145.02% for event 1 and 11.53% for event 2 for these conditions. When these combinations are not present, these differences are considerably smaller at 8.02% for event 1 and 7.23% for event 2. These differences are moderately larger than the overall average when event times are in discrete format, frailty is present, or event times are positively correlated. Differences are not substantially different between the different sample sizes.

Differences by parameter. For some parameters, the model estimates of the Cox proportional hazards regression with frailty are similar between the condition levels. However, for others, model estimates vary greatly between the condition levels. The differences between model estimates of condition levels are presented as a percentage of the set hazard ratios by the simulation. The robustness of the model is judged based on these differences – with smaller differences indicating the model is robust with respect to the condition levels.

Sample size**.** Group sizes of 250 and 500 are simulated for each set of conditions for a total of 500 and 1000 respectively.

Examining the model estimates of identical conditions, save for sample size, the average percentage differences between the deviations of the model estimates from the set hazard ratios between groups of 250 and 500 are relatively small for Cox proportional hazards with frailty with an average percentage differences of 2.40% for event 1 and 1.01% for event 2. Between sample sizes these differences appear to be the smallest when failure rates are the same for both events across both groups, regardless of what other conditions are present. Percentage differences are somewhat larger if failure rates are heterogeneous between group or within group or if event times are positively correlated and are largest when distribution shape differs by event type and failure rates are heterogeneous between group or within group. The average percentage differences between the deviations of the model estimates from the set hazard ratios are minimal between different event time formats and frailties.

Correlation between event times**.** Correlation between Event 1 and Event 2, for both groups, is set at either -0.4, 0, or 0.4. These can be seen above in Table 9.

Examining the Cox proportional hazards regression with frailty, the average percentage differences between the deviations of the model estimates from the set hazard ratios of identical combinations, save for event time correlation, range from minimal to quite large. The average percentage differences between event time correlation conditions 1 and 3 is 23.11% for event 1 and 7.49% for event 2, between event time correlation conditions 1 and 2 the difference is 7.74%

for event 1 and 5.15% for event 2, while the difference between event time correlation conditions 2 and 3 is 20.20% for event 1 and 3.36% for event 2. This trend in percentage differences, larger between event time correlation conditions 1 and 3, smaller between event time correlation conditions 2 and 3, and smaller still between event time correlation conditions 1 and 2, holds for most conditions combination comparisons. However, when failure rates differ by group but not event within group and event time distribution shape does not differ by event, the differences are smallest between correlation conditions 2 and 3. When failure rates differ by group but not event within group and event time distribution shape does differ by event, the differences are smallest between correlation conditions 1 and 3 for event 1 and between correlation conditions 2 and 3 for event 2.

The percentage difference between each pair of correlations are the smallest when failure rates are the same for both events across both groups, regardless of what other conditions are present with an average percentage difference of 1.20% for event 1 and 0.43% for event 2. Large percentage differences in model estimates between each pair occur in combinations of condition levels containing differing event time distribution shapes and failure rates which differ by group and event type with an average percentage difference of 69.92% for event 1 and 8.49% for event 2. These differences are larger still when frailty is present. The average percentage differences between the deviations of the model estimates from the set hazard ratios of identical combinations, save for correlation, are minimal for different sample sizes and even time formats.

Event time format. Time data are simulated in two formats, discrete and continuous, with discrete data created as rounded continuous data. Examining the Cox proportional hazards regression with frailty results, the average percentage differences between the deviations of the model estimates from the set hazard ratios of identical combinations, save for event time format, are 11.75% for event 1 and 7.87% for event 2. These percentage differences are smallest when failure rates are the same for both events across both groups, regardless of what other conditions are present. These differences are slightly larger with frailty and positively correlated event times, but are notably larger when failure rates differ by group or event within group, event time distribution shape differs by event type, events are positively correlated, and frailty is present. These differences are essentially the same between sample sizes.

Group failure rates by event type**.** Group failure rates are simulated in three combinations with group 2 serving as the control with equal failure rates for both events. The combinations are: homogeneous failure rates within group 1 which equal those of group 2 (failure rate condition 1), homogeneous failure rates within both groups but heterogeneous between groups (failure rate condition 2), and heterogeneous failure rate within group 1 (failure rate condition 3). These can be seen above in Table 10.

Examining the Cox proportional hazards regression results, the average percentage differences between the deviations of the model estimates from the set hazard ratios of identical combinations, save for group failure rates by event type, range from minimal to very large. This average percentage difference between failure rate conditions 1 and 2 is 144.76% for event 1 and 21.67% for event 2, between failure rate conditions 1 and 3 the difference is 77.29% for event 1 and 6.31% for event 2, while the difference between failure rate conditions 2 and 3 is 22.39% for event 1 and 8.42% for event 2. The differences between conditions are considerably smaller when distribution shape does not differ by event type with overall averages of 12.43% for event 1 and 11.57% for event 2, and smallest when that is combined with a zero correlation. The average percentage differences between the deviations of the model estimates from the set hazard ratios of identical combinations, save for group failure rates, are largest for event 1 when event time distribution shapes differ and correlation is non-zero. This does not hold true for event 2 however, where the largest percentage differences occur with discrete time format and homogeneous event time distribution shapes. These percentage differences are minimal between sample sizes and relatively small between differing frailties.

***Shape parameter for event time distributions*.** The shape parameter is simulated such that it is either homogeneous within and between groups or heterogeneous within groups with the same heterogeneity for both groups. These can be seen above in Table 11.

Examining the Cox proportional hazards regression with frailty results, the average percentage differences between the deviations of the model estimates from the set hazard ratios of identical combinations, save for event time distribution shape, are between 91.91% for event 1 and 1.57% for event 2. The percentage differences are quite minimal, 0.22% for event 1 and 0.16% for event 2, if the failure rates do not vary between or within groups and there is no frailty. These differences increase moderately as frailty increases and are dramatically larger when failure rates differ by group or within group. The percentage differences are larger for positively correlated event times and smaller with zero correlated event times with negatively correlated event times falling in between. These differences are slightly larger for discrete-time format and relatively minimal for the different sample sizes.

Frailty. The frailty parameter is simulated such that it is homogeneous within groups for both event types at three levels: no frailty, moderate frailty, and high frailty. These can be seen above in Table 12.

Examining the Cox proportional hazards regression with frailty results, the average percentage differences between the deviations of the model estimates from the set hazard ratios of identical combinations, save for frailty, are, for event 1, smallest between no frailty and moderate frailty, largest between no frailty and high frailty, and in the middle between moderate frailty and high frailty for nearly all conditions with a different pattern occurring only when the differences between the conditions are less than 1%. For event 2 however, the percentage differences between moderate frailty and high frailty are smaller than between no frailty and moderate frailty.

Overall average percentage differences between the deviations of the model estimates from the set hazard ratios of identical combinations, save for frailty, are 8.72% for event 1 and 3.72% for event 2 between no frailty and moderate frailty, 9.49% for event 1 and 2.75% for event 2 between moderate frailty and high frailty and 18.21% for event 1 and 5.79% for event 2 between no frailty and high frailty. When event time shape distribution is the same for both events, these differences are minimal and change very little with differing failure rates between or within groups. These percentage differences are moderately larger with positively correlated event times than for negative or zero correlations. Percentage differences in model estimations are larger than average when event time shape distributions differ by group. These differences are larger still when combined with event times that are positively correlated and failure rates differ between or within group. The average percentage differences between the deviations of the model estimates from the set hazard ratios of identical combinations, save for frailty, are minimal between different sample sizes and event time formats.

Student Retention and Graduation Results

Each of the four methods was used to analyze a student data set containing retention and graduation data. Like the simulations, this data contains two events, graduation and drop-out, which are defined above in Chapter 3. The resulting model estimates are listed in Table 13.

Table 13

Student Data Model Estimates

	Model Estimates	
	Event 1	Event 2
Cox Proportional Hazards Regression	1.420	0.685
Weibull Regression	1.390	0.929
Fine and Gray's Method	1.603	0.655
Cox Proportional Hazards Regression with Frailty	1.530	0.654

The model estimates represent the hazard ratio – a ratio of the failure rate of group 2 to the failure rate of group 1. As such a model estimate of 1 would indicate equal failure rates by group, or in other words, no group effect. A model estimate greater than 1 would indicate that group 2 is at a higher risk of the event while a model estimate less than 1 would indicate that the reference group is at a lower risk of the event. For the student data, group 2 is composed of students who are low socio-economic status (either first-generation or low-income) and group 1 is composed of students who are high socio-economic status (neither first-generation nor low-income). Event 1 is drop-out while event 2 is graduation.

For each method, the model estimates suggest that students who are of low socio-economic status are significantly more likely to experience the event of drop-out. The model estimates of each method also suggest that students who are of low socio-economic status are significantly less likely to experience the event of graduation. While the direction of the effect is agreed upon by the methods, the magnitude of the effect is not. To determine which method likely has the closest model estimates, the simulation results are considered. Though no simulation condition combination perfectly matches the real data, it is hypothesized that the closest combination of conditions is large sample size (n=500), time data in discrete form, negatively correlated event times, heterogeneous event time distribution shape, and heterogeneous failure rates between groups and events. Large sample size is chosen as the data set contains approximately 5400 subjects. Discrete-time form is not chosen but simply is the format the data are in. Negatively correlated event times is chosen due to the likelihood that, should it be possible to both drop out and graduate, students whose drop-out time is earlier would have a later graduation time giving the event times a negative correlation. Heterogeneous event time distribution shape is based on the time distributions for the events shown above in Figure 4 and Figure 5 which appear to be heterogeneous. Heterogeneous failure rates between groups and events is chosen based on descriptive statistics suggesting that each group experiences each event at different rates. Frailty is, by definition, unknown and therefore not hypothesized.

Based on the analyses of the simulated data sets under this combination of conditions, the most accurate method is likely Fine and Gray's method. The simulations suggest this model is more accurate regardless of the level of frailty. Though this method does not estimate the second event rate as well as the Weibull regression under these conditions, it does estimate the first event, which has a Weibull shape parameter of 2, more accurately. As this event shape is more closely related to the student data events than a shape parameter of 1, the accuracy in estimating the first event is likely more crucial.

The percentage differences between model estimates and the set hazard ratios for this combination of conditions are listed below in Table 14 for no frailty, moderate frailty, and high frailty respectively.

Table 14

Percentage Difference Between Model Estimates and Set Hazard Ratios

Condition	Cox Proportional Hazards		Weibull Regression		Fine and Gray's Method		Cox Proportional Hazards Regression with Frailty	
	Event 1	Event 2	Event 1	Event 2	Event 1	Event 2	Event 1	Event 2
193	*77.15%*	1.15%	19.06%	10.02%	0.78%	28.43%	**103.01%**	18.17%
195	*86.72%*	1.60%	18.09%	11.74%	3.16%	27.13%	**115.51%**	21.82%
197	*96.71%*	3.85%	17.16%	13.08%	5.97%	26.18%	**129.71%**	24.76%

Note. Percentages greater than 50% are in italics and those greater than 100% are in bold.

While Fine and Gray's method appears to be the most accurate for this data, it is possible that the results may be inaccurate as a test of proportional hazards, an assumption of each method, concluded that the hazards are not proportional. However, as the proportional hazards test statistic is a chi-square, it is sensitive to sample size and thus may be significant based only on the large sample size. A second method of testing for proportional hazards is with Kaplan-Meier survival curves. For hazards to be proportional, the effect on each group should remain relatively constant over time. Graphically this would appear as parallel lines where the effects may be stronger on one group, but not different over time. If the survival curves do not cross, this would indicate the lines are approximately parallel and that the hazards are approximately proportional. An examination of these curves for event 1 and event 2 can be found in Figure 7 and Figure 8 respectively. As the curves do not cross, the hazards appear to be proportional.

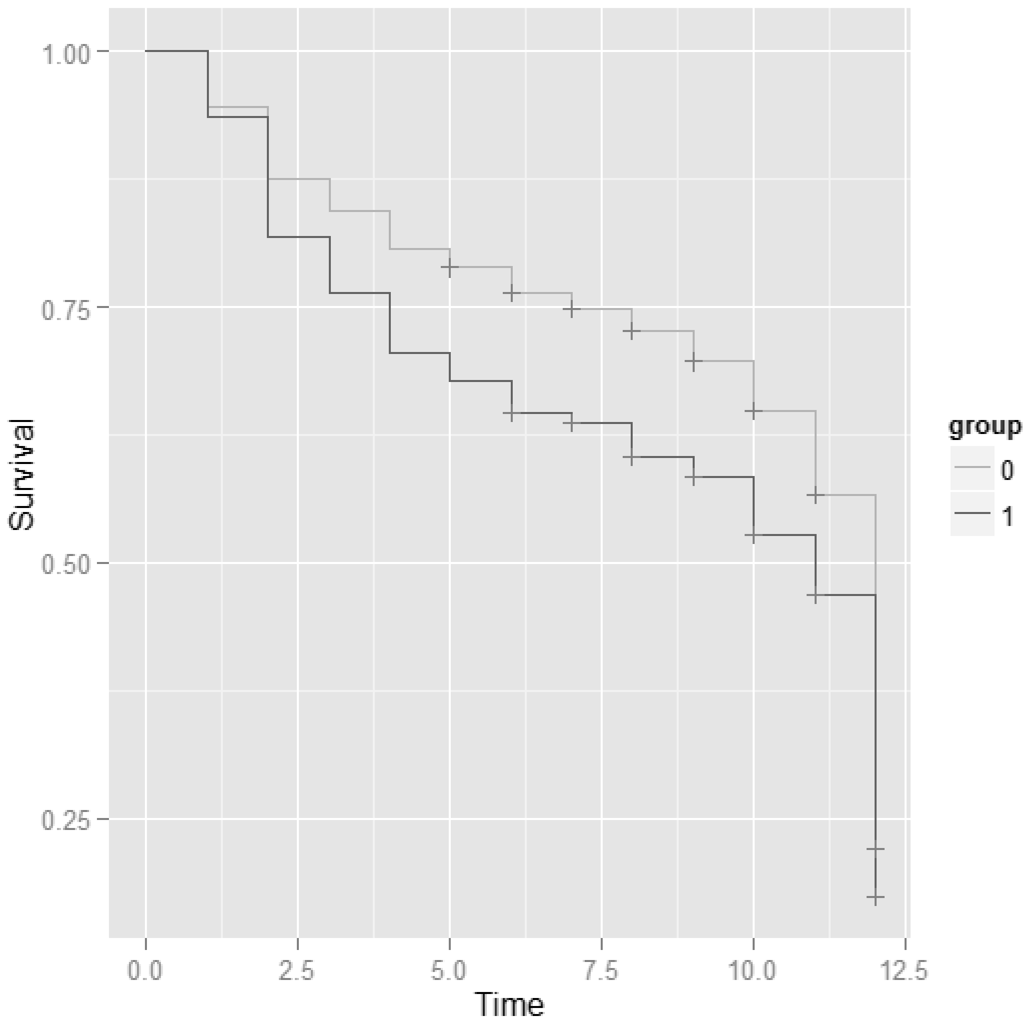

Figure 7. Kaplan-Meier survival curves for event 1 – drop-out.

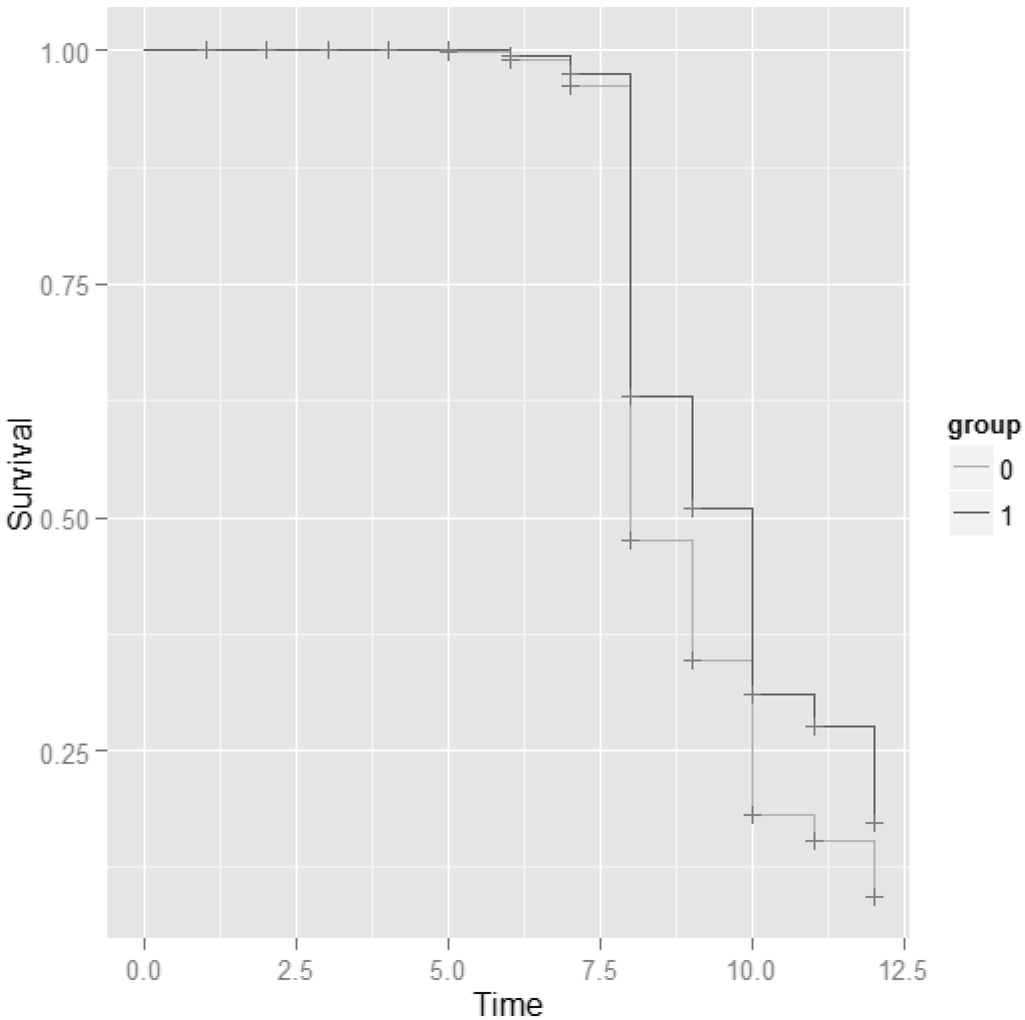

Figure 8. Kaplan-Meier survival curves for event 2 – graduation.

Chapter V: Discussion

The purpose of this study was to compare methods of analyzing competing risks models to investigate the effects that varying parameters have on model estimates. Specifically, this study sought to add to current literature an analysis of simulated data sets which contain differing sample sizes, event time distribution shape parameters, time data in discrete form, and frailty. Other parameters, correlation between event time, and failure rate by group and event, occur in current literature and were also included in this study. The results from this simulation study were used to aid in the selection of a model for student retention and graduation research.

Summary of Results

It is clear from the results that there is no one best model for all scenarios. Instead, it appears the selection of the method of analysis should be carefully considered by the researcher based on the characteristics of the data. While the more traditional Cox proportional hazards regression estimates model estimates accurately under many scenarios, there are many combinations of conditions that cause the model estimates to deviate drastically from the actual ratio of failure rates. The Weibull Regression appears to be the most accurate model for most combination of conditions; however, it is important to consider that the event time data was simulated as Weibull distributed data, and as such these results can only be generalized if the event time data are in fact Weibull distributed. It is notable, however, that the Weibull regression was able to accurately estimate model parameters for various versions of the Weibull distribution with different shape and scale parameters. Fine and Gray's method was often ranked last by accuracy, but there are some situations where its accuracy is quite good. In addition, its more conservative estimates result in less drastic errors than the other methods. Cox proportional hazards regression with frailty performed very similarly to the Cox regression without frailty. It is unclear if this is a limitation to the software or if the simulated conditions did not deviate drastically enough to produce different results.

Method selection. In this section, the advantages and disadvantages of each method will be discussed in comparison to other methods. Its aim is to aid empirical researchers in selecting the optimal method based on the type of data they are analyzing.

Cox proportional hazards regression. The results of this study, and previously cited research, lend credence to Cox being the default method of survival analysis for statistical software packages as well as for many researchers. For data where the event time distribution shape does not differ by event, the accuracy of the model was, for the most part, quite similar to that of the Weibull regression. While the model estimates often deviated slightly more from the set hazard ratio than the Weibull regression, as a semi-parametric test the Cox regression is attractive for data where event times do not follow a Weibull distribution. It is advantageous over the Weibull distribution under most condition combinations involving homogeneous event time distribution shapes, and heterogeneous failure rates between groups and events.

Great care should be taken when selecting the Cox model should the event time distribution shapes differ by event. Should failure rates also differ by event or by group, which is often not easily discernable, results deviate considerably from the true parameters. This is illustrated below in Table 15. These condition combinations include different event time distribution shapes by event and different failure rates by group and/or event.

Table 15

Percentage Difference Between Model Estimates and Set Hazard Ratios

Condition	Cox Proportional Hazards		Weibull Regression		Fine and Gray's Method		Cox Proportional Hazards Regression with Frailty	
	Event 1	Event 2	Event 1	Event 2	Event 1	Event 2	Event 1	Event 2
13	**104.55%**	1.27%	0.22%	1.16%	18.11%	33.09%	**104.79%**	1.20%
14	**103.57%**	1.63%	0.12%	1.31%	40.41%	40.88%	**103.89%**	1.61%
15	**118.08%**	4.31%	2.07%	4.20%	22.57%	31.44%	**118.75%**	4.25%
16	**118.00%**	4.71%	2.02%	4.30%	39.16%	39.34%	**118.82%**	4.65%
17	**134.03%**	6.74%	3.95%	6.73%	27.79%	30.16%	**134.09%**	6.72%
18	**133.30%**	7.18%	3.63%	6.92%	37.77%	38.12%	**134.12%**	7.16%

Note. Percentages greater than 50% are in italics and those greater than 100% are in bold.

Should data contain differing event time distribution shapes by event and Cox proportional hazards be chosen as the method of analysis, it is recommended that the researcher analyze additional methods in addition to ensure these drastic deviations have not occurred.

Weibull regression. As the data was generated to be Weibull distributed, it is not entirely unexpected for the Weibull regression to most accurately estimate the model parameters. Out of 216 condition combinations the Weibull regression proved most accurate, in terms of the percentage difference between model estimates and set ratios of failure rates, in 149 condition combinations. The method's main drawback appears with the combination of homogeneous event time distribution shape, heterogeneous failure rate by group and/or event, and discrete format time. For this combination of conditions, the Weibull regression is more accurate than Fine and Gray's method, but is easily surpassed by the Cox proportional hazards regression. Under this combination of conditions Cox is the clear choice. When instead event time distribution shape is heterogeneous with the above combination of conditions, the accuracy of the Weibull method is not optimal, but it is more accurate than the other models tested. This can be seen below in Table 16.

Table 16

Percentage Difference Between Model Estimates and Set Hazard Ratios

Condition	Cox Proportional Hazards		Weibull Regression		Fine and Gray's Method		Cox Proportional Hazards Regression with Frailty	
	Event 1	Event 2	Event 1	Event 2	Event 1	Event 2	Event 1	Event 2
175	*98.17%*	7.44%	12.92%	6.24%	14.37%	32.57%	**120.71%**	9.99%
176	*84.55%*	8.26%	11.21%	13.43%	41.44%	42.85%	**120.82%**	16.31%
193	*77.15%*	1.15%	19.06%	10.02%	0.78%	28.43%	**103.01%**	18.17%
194	*90.54%*	2.41%	12.46%	14.50%	41.00%	42.64%	**137.65%**	24.70%
211	**169.95%**	14.48%	3.55%	0.98%	45.83%	37.96%	**192.01%**	1.57%
212	**100.83%**	13.09%	9.96%	12.44%	41.42%	42.76%	**131.16%**	10.13%

Note. Percentages greater than 50% are in italics and those greater than 100% are in bold.

Caution should be taken when selecting this method if the data does not display a clear Weibull distribution shape. The accuracy of this method outside of Weibull distributed event time data is beyond the scope of this study, and as such these results should not be generalized to other event time distributions.

Fine and Gray’s Method. When failure rates do not vary by group or event, the accuracy of Fine and Gray’s method, in terms of the percentage difference between model estimates and set ratios of failure rates, is generally on par with or even better than Cox proportional hazards or Weibull regression. When failure rate varies between groups but not within groups this accuracy falls considerably. This can be seen below in Table 17.

Table 17

Percentage Difference Between Model Estimates and Set Hazard Ratios

Condition	Cox Proportional Hazards		Weibull Regression		Fine and Gray's Method		Cox Proportional Hazards Regression with Frailty	
	Event 1	Event 2	Event 1	Event 2	Event 1	Event 2	Event 1	Event 2
68	**101.75%**	0.67%	0.06%	0.66%	40.39%	41.40%	**101.95%**	0.75%
86	**117.03%**	9.83%	0.10%	0.54%	39.68%	41.07%	**117.74%**	6.35%
104	**114.26%**	6.13%	0.08%	0.75%	40.57%	41.37%	**111.44%**	3.60%
176	*84.55%*	8.26%	11.21%	13.43%	41.44%	42.85%	**120.82%**	16.31%
194	*90.54%*	2.41%	12.46%	14.50%	41.00%	42.64%	**137.65%**	24.70%
212	**100.83%**	13.09%	9.96%	12.44%	41.42%	42.76%	**131.16%**	10.13%

Note. Percentages greater than 50% are in italics and those greater than 100% are in bold.

It is noteworthy, however, that while the accuracy of Fine and Gray's model under this combination of conditions is not optimal, the model estimates deviate from the set hazard ratios considerably less than Cox proportional hazards. Should the event time data deviate drastically from a Weibull distribution, it may be that Fine and Gray's method would produce the most accurate model estimates, though certainly still less than optimal.

As indicated in the student data example, there are also combinations of conditions for which Fine and Gray's method appears to be well suited. These combinations are few, but they present interesting results. For data sets containing discrete event times, negative event time correlation, heterogeneous event time shape distributions by event, and failure rates that differ both by group and by event, Fine and Gray's method is able to accurately estimate the model parameters far more accurately than the other methods, but only for event 1. The estimates produced by Cox proportional hazards regression, in comparison, are quite accurate for event 2 but not for event 1. This comparison can be seen below in Table 18.

Table 18

Percentage Difference Between Model Estimates and Set Hazard Ratios

Condition	Cox Proportional Hazards		Weibull Regression		Fine and Gray's Method		Cox Proportional Hazards Regression with Frailty	
	Event 1	Event 2	Event 1	Event 2	Event 1	Event 2	Event 1	Event 2
139	*78.93%*	0.59%	19.13%	10.38%	0.80%	27.88%	**106.72%**	19.03%
141	*89.71%*	2.32%	18.10%	12.13%	3.21%	26.57%	**120.70%**	22.89%
143	**100.82%**	4.71%	17.15%	13.57%	5.75%	25.58%	**136.57%**	26.10%
193	*77.15%*	1.15%	19.06%	10.02%	0.78%	28.43%	**103.01%**	18.17%
195	*86.72%*	1.60%	18.09%	11.74%	3.16%	27.13%	**115.51%**	21.82%
197	*96.71%*	3.85%	17.16%	13.08%	5.97%	26.18%	**129.71%**	24.76%

Note. Percentages greater than 50% are in italics and those greater than 100% are in bold.

These results may indicate that, under similar combinations of conditions, Fine and Gray's method is better suited to analyze event time shape distributions more similar to event 2 – a Weibull distribution with shape parameter $k = 2$ while Cox proportional hazards regression is better suited to analyze event time shape distributions more similar to event 1 – a Weibull distribution with shape parameter $k = 1$. As simulations where event time distribution shapes for both events are distributed Weibull with $k > 1$are beyond the scope of this study, this postulation is just that and cannot be directly inferred from the results.

Cox proportional hazards with frailty. Results from this method closely mirror those of method 1, the Cox proportional hazards without frailty. While differences in accuracy, in terms of the percentage difference between model estimates and set ratios of failure rates, do exist, they are generally minimal and do not display a clear pattern of increased accuracy for the model with frailty over the model without. It is unclear if this is due to the software or due to the data being modeled in a way that does not take full advantage of the test. What is clear is that the Cox model with frailty is noticeably less accurate when analyzing discrete-time data than the Cox model without frailty. For continuous time data the accuracy is nearly identical between the Cox models. Based on the results of this study, there are no combinations of conditions for which this clearly is the best method.

Limitations

Originally included in this study was a fifth method of analyzing competing risks – Weibull regression with gamma frailty. Like the Weibull regression, this was planned to be analyzed using the survreg package in R. However, it was found that the results of the Weibull regression with gramma frailty were exactly identical to the Weibull regression without frailty. Further investigation revealed that the gamma frailty model in survreg is an invalid model. For this reason this model was omitted.

Recommendations for Future Research

Studies involving simulations on the effects of varying conditions on competing risks analyses are few in current literature. This study sought to expand upon current literature by including additional conditions and analyses; however, the combinations of conditions presented in this study comprise a mere fraction of the combinations likely to be encountered by applied researchers. Previous research has focused on events with an exponential event time which is equivalent to the Weibull distribution with shape parameter $k = 1$. Event times with this distribution, seen in figure 2, occur most frequently at the beginning of the time period and occur in declining frequency from there. For the events of drop-out and graduation, as seen in figures 4 and 5, it is clear that neither of these events follow this distribution. While this study expanded research to include event time data with a Weibull distribution with shape parameter $k = 2$ for one event, further research is needed to examine results of analyses where both events have Weibull distributed event times with shape parameter $k = 2$. Further analyses involving combinations of event times with a Weibull distribution and shape parameters $k > 1$ are also needed. Specifically, research is needed into competing risks data where events 1 and/or 2 have shape parameters of $k = 2$ and $k = 3$. Research into non-Weibull data is also needed.

Additionally, this study is focused on compete or uncensored data, that is, data in which the event time is known for every subject. One benefit of survival analysis over logistic regression is that the event of interest need not have occurred prior to analysis. At least six years had passed for all students in the retention and graduation example, so all could be coded as graduated or drop-out, though this required omitting many students who had begun their studies less than six years ago. Research is needed to examine the effects of varying proportions of censoring on these methods under these conditions.

Event time correlation for this study was set equally by group between the event types at either -0.4, 0, or 0.4. Though impossible to prove in true competing events, is certainly likely that it is possible that events could be correlated differently for different groups. In addition to analyzing these possibilities, stronger correlations should also be examined for their effects on these methods. In the same manner, it is also likely that failure rates could differ by event but in the same manner for both groups.

Lastly, and likely most importantly for practitioners, is a need for study on the effects of numerical covariates, both time dependent and time independent, tested as a part of combinations of conditions included in this study. While categorical variables are common in education, continuous and finite numerical variables in the form of test scores, GPAs, and credit hours, are essential in analyzing student data.

Implications for Educational Researchers

For educational researchers searching for a better method to analyze competing risks data, this study is unfortunately not a panacea. Though many questions were researched here, there remain far more. Until more research is done, it is recommended that the methods in this study should be used to supplement traditional methods of analysis, especially outside of tested conditions. However, researchers should not shy away from these methods as they are able to utilize more of the available data and allow for greater flexibility in analysis. Care should be taken and results should be compared with survival curves, traditional methods, and descriptive statistics to ensure results are likely valid. It is the opinion of this researcher that these types of analyses should become the future of educational research. Current methods in use in current education literature are insufficient in modeling student progress through educational institutions. Though this study merely scratches the surface on the topic of using these analyses in education, it is hoped that future research will enable them to become a part of the mainstream for educators and educational researchers interested in student success.

References

Allison, P. D. (1982). Discrete-Time Methods for the Analysis of Event Histories. *Sociological Methodology*, *13*, 61–98.

Allison, P. D. (1984). *Event history analysis: regression for longitudinal event data*. Newbury Park, CA: Sage Publications.

Arredondo, M., & Knight, S. (2005). Estimating degree attainment rates of freshmen: A campus perspective. *Journal of College Student Retention: Research, Theory and Practice*, *7*(1-2), 91–115. http://doi.org/10.2190/EG5X-LN6A-JL8X-CJ0W

Astin, A. W., & Osegura, L. (2005). *Degree Attainment Rates at American Colleges and Universities: Effects of Race, Gender, and Institutional Type*. Los Angeles, CA: Higher Education Research Inst. Inc.

Belot, A., Abrahamowicz, M., Remontet, L., & Giorgi, R. (2010). Flexible modeling of competing risks in survival analysis. *Statistics in Medicine*, *29*(23). Retrieved from http://0-search.proquest.com.library.uark.edu/pqrl/docview/755548624/7AFD0EC5D24B4DE1PQ/28?accountid=8361

Berkovitz, R. A., & O’Quin, K. (2006). Predictors of Graduation of Readmitted “at Risk” College Students. *Journal of College Student Retention*, *8*(2), 199–214.

Braunstein, A. W., Lesser, M. N., & Pescatrice, D. R. (2008). The Impact of a Program for the Disadvantaged on Student Retention. *College Student Journal*, *42*(1), 36–40.

Chapman, J.-A. W., Meng, D., Shepherd, L., Parulekar, W., Ingle, J. N., Muss, H. B., … Goss, P. E. (2008). Competing Causes of Death From a Randomized Trial of Extended Adjuvant Endocrine Therapy for Breast Cancer. *Journal of the National Cancer Institute*, *100*(4), 252–60. http://doi.org/http://0-dx.doi.org.library.uark.edu/10.1093/jnci/djn014

Chimka, J. R., Reed-Rhoads, T., & Barker, K. (2007). Proportional Hazards Models of Graduation. *Journal of College Student Retention*, *9*(2), 221–232.

Chu-Ling, Y., Tucker, M. A., Abramson, D. H., Furukawa, K., Seddon, J. M., Stovall, M., … Kleinerman, R. A. (2009). Cause-Specific Mortality in Long-Term Survivors of Retinoblastoma. *Journal of the National Cancer Institute*, *101*(8), 581–591. http://doi.org/http://0-dx.doi.org.library.uark.edu/10.1093/jnci/djp046

Cleves, M. A. (2010). *An introduction to survival analysis using Stata*. College Station, Tex.: Stata Press.

Cox, D. R. (1972). Regression Models and Life-Tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, *34*(2), 187–220.

Dignam, J. J., Zhang, Q., & Kocherginsky, M. (2012). The use and interpretation of competing risks regression models. *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, *18*(8), 2301–2308. http://doi.org/10.1158/1078-0432.CCR-11-2097

Fine, J. P., & Gray, R. J. (1999). A Proportional Hazards Model for the Subdistribution of a Competing Risk. *Journal of the American Statistical Association*, *94*(446), 496–509. http://doi.org/10.2307/2670170

Freidlin, B., & Korn, E. L. (2005). Testing treatment effects in the presence of competing risks. *Statistics in Medicine*, *24*(11), 1703–1712.

Graunt, J. (1662). *Natural and Political Observations mentioned in a following index, and made upon the Bills of Mortality With reference to the Government, Religion, Trade, Growth, Ayre, diseases, and the several Changes of the said City*. London: Printed by Tho. Roycroft, for John Martin, James Allestry and Tho. Dicas.

Heckman, J. J., & Singer, B. (1984). Econometric Duration Analysis. *Journal of Econometrics, 24*(1,2), 63.

Heilig, J. V. (2011). As GOOD AS ADVERTISED? TRACKING URBAN STUDENT PROGRESS THROUGH HIGH SCHOOL IN AN ENVIRONMENT OF ACCOUNTABILITY. *American Secondary Education, 39*(3), 17–41.

Herzog, S. (2005). Measuring Determinants of Student Return VS. Dropout/Stopout VS. Transfer: A First-to-Second Year Analysis of New Freshmen. *Research in Higher Education*, *46*(8), 883–928. http://doi.org/10.1007/s11162-005-6933-7

Hinchli_ffe, S. (2012). *Competing Risks - What, Why, When and How?*. Presented at the Survival Analysis for Junior Researchers. Retrieved from http://www2.le.ac.uk/departments/health-sciences/research/biostats/youngsurv/pdf/SHinchliffe.pdf/view

John Graunt. (2004). In *Encyclopedia of World Biography*. Retrieved from http://www.encyclopedia.com/doc/1G2-3404707807.html

Kim, H. T. (2007). Cumulative Incidence in Competing Risks Data and Competing Risks Regression Analysis. *Clinical Cancer Research*, *13*(2), 559–565. http://doi.org/10.1158/1078-0432.CCR-06-1210

Kleinbaum, D. G. (1995). *Survival analysis: a self learning text*. New York [etc.]: Springer.

Klein, J. P., & Moeschberger, M. L. (2003). *Survival analysis: techniques for censored and truncated data*. New York: Springer.

Lee, R. (2012). Community Violence Exposure and Adolescent Substance Use: Does Monitoring and Positive Parenting Moderate Risk in Urban Communities? *Journal of Community Psychology*, *40*(4), 406–421. http://doi.org/10.1002/jcop.20520

Low, N. C., Dugas, E., O'Loughlin, E., Rodriguez, D., Contreras, G., Chaiton, M., & O'Loughlin, J. (2012). Common stressful life events and difficulties are associated with mental health symptoms and substance use in young adolescents. *BMC Psychiatry*, *12*(1), 116–125. http://doi.org/10.1186/1471-244X-12-116

Masyn, K. E. (2003). *Discrete-time survival mixture analysis for single and recurrent events using latent variables* (Ph.D.). University of California, Los Angeles, United States -- California. Retrieved from http://0-search.proquest.com.library.uark.edu/pqdtft/docview/305351843/abstract/63FA190A07D84D83PQ/1?accountid=8361

Min, Y., Zhang, G., Long, R. A., Anderson, T. J., & Ohland, M. W. (2011). Nonparametric Survival Analysis of the Loss Rate of Undergraduate Engineering Students. *Journal of Engineering Education*, *100*(2), 349–373.

Murtaugh, P. A., Burns, L. D., & Schuster, J. (1999). Predicting the Retention of University Students. *Research in Higher Education*, *40*(3), 355–371. http://doi.org/http://0-dx.doi.org.library.uark.edu/10.1023/A:1018755201899

Ramirez, R., Hinman, A., Sterling, S., Weisner, C., & Campbell, C. (2012). Peer Influences on Adolescent Alcohol and Other Drug Use Outcomes. *Journal of Nursing Scholarship*, *44*(1), 36–44. http://doi.org/10.1111/j.1547-5069.2011.01437.x

Robertson, L. J., & Taylor, C. A. (2009). Student Persistence in the Human Sciences: Freshman to Sophomore Year. *Journal of Family and Consumer Sciences*, *101*(1), 36–43.

Rodriguez, G. (2005). Competing Risks. Retrieved February 9, 2014, from http://data.princeton.edu/pop509/CompetingRisks.pdf

Ryberg, M., Nielsen, D., Cortese, G., Nielsen, G., Skovsgaard, T., & Andersen, P. K. (2008). New Insight Into Epirubicin Cardiac Toxicity: Competing Risks Analysis of 1097 Breast Cancer Patients. *Journal of the National Cancer Institute*, *100*(15), 1058–67. http://doi.org/http://0-dx.doi.org.library.uark.edu/10.1093/jnci/djn206

Schairer, C., Mink, P. J., Carroll, L., & Devesa, S. S. (2004). Probabilities of Death From Breast Cancer and Other Causes Among Female Breast Cancer Patients. *Journal of the National Cancer Institute*, *96*(17), 1311–21.

Szafran, R. F. (2001). The Effect of Academic Load on Success for New College Students: Is Lighter Better? *Research in Higher Education*, *42*(1), 27–50. http://doi.org/http://0-dx.doi.org.library.uark.edu/10.1023/A:1018712527023

Trussell, J., & Richards, T. (1985). Correcting for Unmeasured Heterogeneity in Hazard Models Using the Heckman - Singer Procedure. *Sociological Methodology*, *15*, 242–276. http://doi.org/10.2307/270852

Vaupel, J. W., Manton, K. G., & Stallard, E. (1979). The Impact of Heterogeneity in Individual Frailty on the Dynamics of Mortality. *Demography*, *16*(3), 439–454. http://doi.org/10.2307/2061224

Williams, J. E., & Luo, M. (2010). Understanding First-Year Persistence at a Micropolitan University: Do Geographic Characteristics of Students' Home City Matter? *College Student Journal*, *44*(2), 362–376.

Williamson, P. R., Kolamunnage-Dona, R., & Smith, C. T. (2007). The influence of competing-risks setting on the choice of hypothesis test for treatment effect. *Biostatistics*, *8*(4), 689–94. http://doi.org/http://0-dx.doi.org.library.uark.edu/10.1093/biostatistics/kxl040

Wolbers, M., Koller, M. T., Witteman, J. C. M., & Steyerberg, E. W. (2009). Prognostic Models With Competing Risks: Methods and Application to Coronary Risk Prediction. *Epidemiology*, *20*(4), 555–561.

Zhang, G., Anderson, T. J., Ohland, M. W., & Thorndyke, B. R. (2004). Identifying Factors Influencing Engineering Student Graduation: A Longitudinal and Cross-Institutional Study. *Journal of Engineering Education*, *93*(4), 313–320.

Appendix

Table A1

Simulation Conditions

Condition Number	Group Size	Event Time Format	Correlation Between Event Times	Shape Parameter Weibull *k*				Failures per Time Period				Frailty Multiplicative Factor			
				Group 1		Group 2		Group 1		Group 2		Group 1		Group 2	
				Event 1	Event 2	Event 1	Event 2	Event 1	Event 2	Event 1	Event 2	Event 1	Event 2	Event 1	Event 2
001	250	Continuous	0	1	1	1	1	1	1	1	1	1	1	1	1
002	250	Continuous	0	1	1	1	1	1	1	1	1	1.5	1.5	1	1
003	250	Continuous	0	1	1	1	1	1	1	1	1	2	2	1	1
004	250	Continuous	0	1	1	1	1	0.5	1	1	1	1	1	1	1
005	250	Continuous	0	1	1	1	1	0.5	0.5	1	1	1	1	1	1
006	250	Continuous	0	1	1	1	1	0.5	1	1	1	1.5	1.5	1	1
007	250	Continuous	0	1	1	1	1	0.5	0.5	1	1	1.5	1.5	1	1
008	250	Continuous	0	1	1	1	1	0.5	1	1	1	2	2	1	1
009	250	Continuous	0	1	1	1	1	0.5	0.5	1	1	2	2	1	1
010	250	Continuous	0	2	1	2	1	1	1	1	1	1	1	1	1
011	250	Continuous	0	2	1	2	1	1	1	1	1	1.5	1.5	1	1
012	250	Continuous	0	2	1	2	1	1	1	1	1	2	2	1	1
013	250	Continuous	0	2	1	2	1	0.5	1	1	1	1	1	1	1
014	250	Continuous	0	2	1	2	1	0.5	0.5	1	1	1	1	1	1
015	250	Continuous	0	2	1	2	1	0.5	1	1	1	1.5	1.5	1	1
016	250	Continuous	0	2	1	2	1	0.5	0.5	1	1	1.5	1.5	1	1
017	250	Continuous	0	2	1	2	1	0.5	1	1	1	2	2	1	1
018	250	Continuous	0	2	1	2	1	0.5	0.5	1	1	2	2	1	1
019	250	Continuous	-0.4	1	1	1	1	1	1	1	1	1	1	1	1
020	250	Continuous	-0.4	1	1	1	1	1	1	1	1	1.5	1.5	1	1
021	250	Continuous	-0.4	1	1	1	1	1	1	1	1	2	2	1	1

Table A1 (Cont.)

Simulation Conditions

Condition Number	Group Size	Event Time Format	Correlation Between Event Times	Shape Parameter Weibull *k*				Failures per Time Period				Frailty Multiplicative Factor			
				Group 1		Group 2		Group 1		Group 2		Group 1		Group 2	
				Event 1	Event 2	Event 1	Event 2	Event 1	Event 2	Event 1	Event 2	Event 1	Event 2	Event 1	Event 2
022	250	Continuous	-0.4	1	1	1	1	0.5	1	1	1	1	1	1	1
023	250	Continuous	-0.4	1	1	1	1	0.5	0.5	1	1	1	1	1	1
024	250	Continuous	-0.4	1	1	1	1	0.5	1	1	1	1.5	1.5	1	1
025	250	Continuous	-0.4	1	1	1	1	0.5	0.5	1	1	1.5	1.5	1	1
026	250	Continuous	-0.4	1	1	1	1	0.5	1	1	1	2	2	1	1
027	250	Continuous	-0.4	1	1	1	1	0.5	0.5	1	1	2	2	1	1
028	250	Continuous	-0.4	2	1	2	1	1	1	1	1	1	1	1	1
029	250	Continuous	-0.4	2	1	2	1	1	1	1	1	1.5	1.5	1	1
030	250	Continuous	-0.4	2	1	2	1	1	1	1	1	2	2	1	1
031	250	Continuous	-0.4	2	1	2	1	0.5	1	1	1	1	1	1	1
032	250	Continuous	-0.4	2	1	2	1	0.5	0.5	1	1	1	1	1	1
033	250	Continuous	-0.4	2	1	2	1	0.5	1	1	1	1.5	1.5	1	1
034	250	Continuous	-0.4	2	1	2	1	0.5	0.5	1	1	1.5	1.5	1	1
035	250	Continuous	-0.4	2	1	2	1	0.5	1	1	1	2	2	1	1
036	250	Continuous	-0.4	2	1	2	1	0.5	0.5	1	1	2	2	1	1
037	250	Continuous	0.4	1	1	1	1	1	1	1	1	1	1	1	1
038	250	Continuous	0.4	1	1	1	1	1	1	1	1	1.5	1.5	1	1
039	250	Continuous	0.4	1	1	1	1	1	1	1	1	2	2	1	1
040	250	Continuous	0.4	1	1	1	1	0.5	1	1	1	1	1	1	1
041	250	Continuous	0.4	1	1	1	1	0.5	0.5	1	1	1	1	1	1
042	250	Continuous	0.4	1	1	1	1	0.5	1	1	1	1.5	1.5	1	1
043	250	Continuous	0.4	1	1	1	1	0.5	0.5	1	1	1.5	1.5	1	1
044	250	Continuous	0.4	1	1	1	1	0.5	1	1	1	2	2	1	1

Table A1 (Cont.)

Simulation Conditions

Condition Number	Group Size	Event Time Format	Correlation Between Event Times	Shape Parameter Weibull *k*				Failures per Time Period				Frailty Multiplicative Factor			
				Group 1		Group 2		Group 1		Group 2		Group 1		Group 2	
				Event 1	Event 2	Event 1	Event 2	Event 1	Event 2	Event 1	Event 2	Event 1	Event 2	Event 1	Event 2
045	250	Continuous	0.4	1	1	1	1	0.5	0.5	1	1	2	2	1	1
046	250	Continuous	0.4	2	1	2	1	1	1	1	1	1	1	1	1
047	250	Continuous	0.4	2	1	2	1	1	1	1	1	1.5	1.5	1	1
048	250	Continuous	0.4	2	1	2	1	1	1	1	1	2	2	1	1
049	250	Continuous	0.4	2	1	2	1	0.5	1	1	1	1	1	1	1
050	250	Continuous	0.4	2	1	2	1	0.5	0.5	1	1	1	1	1	1
051	250	Continuous	0.4	2	1	2	1	0.5	1	1	1	1.5	1.5	1	1
052	250	Continuous	0.4	2	1	2	1	0.5	0.5	1	1	1.5	1.5	1	1
053	250	Continuous	0.4	2	1	2	1	0.5	1	1	1	2	2	1	1
054	250	Continuous	0.4	2	1	2	1	0.5	0.5	1	1	2	2	1	1
055	500	Continuous	0	1	1	1	1	1	1	1	1	1	1	1	1
056	500	Continuous	0	1	1	1	1	1	1	1	1	1.5	1.5	1	1
057	500	Continuous	0	1	1	1	1	1	1	1	1	2	2	1	1
058	500	Continuous	0	1	1	1	1	0.5	1	1	1	1	1	1	1
059	500	Continuous	0	1	1	1	1	0.5	0.5	1	1	1	1	1	1
060	500	Continuous	0	1	1	1	1	0.5	1	1	1	1.5	1.5	1	1
061	500	Continuous	0	1	1	1	1	0.5	0.5	1	1	1.5	1.5	1	1
062	500	Continuous	0	1	1	1	1	0.5	1	1	1	2	2	1	1
063	500	Continuous	0	1	1	1	1	0.5	0.5	1	1	2	2	1	1
064	500	Continuous	0	2	1	2	1	1	1	1	1	1	1	1	1
065	500	Continuous	0	2	1	2	1	1	1	1	1	1.5	1.5	1	1
066	500	Continuous	0	2	1	2	1	1	1	1	1	2	2	1	1
067	500	Continuous	0	2	1	2	1	0.5	1	1	1	1	1	1	1

Table A1 (Cont.)
Simulation Conditions

Condition Number	Group Size	Event Time Format	Correlation Between Event Times	Shape Parameter Weibull *k*				Failures per Time Period				Frailty Multiplicative Factor			
				Group 1		Group 2		Group 1		Group 2		Group 1		Group 2	
				Event 1	Event 2	Event 1	Event 2	Event 1	Event 2	Event 1	Event 2	Event 1	Event 2	Event 1	Event 2
068	500	Continuous	0	2	1	2	1	0.5	0.5	1	1	1	1	1	1
069	500	Continuous	0	2	1	2	1	0.5	1	1	1	1.5	1.5	1	1
070	500	Continuous	0	2	1	2	1	0.5	0.5	1	1	1.5	1.5	1	1
071	500	Continuous	0	2	1	2	1	0.5	1	1	1	2	2	1	1
072	500	Continuous	0	2	1	2	1	0.5	0.5	1	1	2	2	1	1
073	500	Continuous	-0.4	1	1	1	1	1	1	1	1	1	1	1	1
074	500	Continuous	-0.4	1	1	1	1	1	1	1	1	1.5	1.5	1	1
075	500	Continuous	-0.4	1	1	1	1	1	1	1	1	2	2	1	1
076	500	Continuous	-0.4	1	1	1	1	0.5	1	1	1	1	1	1	1
077	500	Continuous	-0.4	1	1	1	1	0.5	0.5	1	1	1	1	1	1
078	500	Continuous	-0.4	1	1	1	1	0.5	1	1	1	1.5	1.5	1	1
079	500	Continuous	-0.4	1	1	1	1	0.5	0.5	1	1	1.5	1.5	1	1
080	500	Continuous	-0.4	1	1	1	1	0.5	1	1	1	2	2	1	1
081	500	Continuous	-0.4	1	1	1	1	0.5	0.5	1	1	2	2	1	1
082	500	Continuous	-0.4	2	1	2	1	1	1	1	1	1	1	1	1
083	500	Continuous	-0.4	2	1	2	1	1	1	1	1	1.5	1.5	1	1
084	500	Continuous	-0.4	2	1	2	1	1	1	1	1	2	2	1	1
085	500	Continuous	-0.4	2	1	2	1	0.5	1	1	1	1	1	1	1
086	500	Continuous	-0.4	2	1	2	1	0.5	0.5	1	1	1	1	1	1
087	500	Continuous	-0.4	2	1	2	1	0.5	1	1	1	1.5	1.5	1	1
088	500	Continuous	-0.4	2	1	2	1	0.5	0.5	1	1	1.5	1.5	1	1
089	500	Continuous	-0.4	2	1	2	1	0.5	1	1	1	2	2	1	1
090	500	Continuous	-0.4	2	1	2	1	0.5	0.5	1	1	2	2	1	1

Table A1 (Cont.)

Simulation Conditions

Condition Number	Group Size	Event Time Format	Correlation Between Event Times	Shape Parameter Weibull *k*				Failures per Time Period				Frailty Multiplicative Factor			
				Group 1		Group 2		Group 1		Group 2		Group 1		Group 2	
				Event 1	Event 2	Event 1	Event 2	Event 1	Event 2	Event 1	Event 2	Event 1	Event 2	Event 1	Event 2
091	500	Continuous	0.4	1	1	1	1	1	1	1	1	1	1	1	1
092	500	Continuous	0.4	1	1	1	1	1	1	1	1	1.5	1.5	1	1
093	500	Continuous	0.4	1	1	1	1	1	1	1	1	2	2	1	1
094	500	Continuous	0.4	1	1	1	1	0.5	1	1	1	1	1	1	1
095	500	Continuous	0.4	1	1	1	1	0.5	0.5	1	1	1	1	1	1
096	500	Continuous	0.4	1	1	1	1	0.5	1	1	1	1.5	1.5	1	1
097	500	Continuous	0.4	1	1	1	1	0.5	0.5	1	1	1.5	1.5	1	1
098	500	Continuous	0.4	1	1	1	1	0.5	1	1	1	2	2	1	1
099	500	Continuous	0.4	1	1	1	1	0.5	0.5	1	1	2	2	1	1
100	500	Continuous	0.4	2	1	2	1	1	1	1	1	1	1	1	1
101	500	Continuous	0.4	2	1	2	1	1	1	1	1	1.5	1.5	1	1
102	500	Continuous	0.4	2	1	2	1	1	1	1	1	2	2	1	1
103	500	Continuous	0.4	2	1	2	1	0.5	1	1	1	1	1	1	1
104	500	Continuous	0.4	2	1	2	1	0.5	0.5	1	1	1	1	1	1
105	500	Continuous	0.4	2	1	2	1	0.5	1	1	1	1.5	1.5	1	1
106	500	Continuous	0.4	2	1	2	1	0.5	0.5	1	1	1.5	1.5	1	1
107	500	Continuous	0.4	2	1	2	1	0.5	1	1	1	2	2	1	1
108	500	Continuous	0.4	2	1	2	1	0.5	0.5	1	1	2	2	1	1
109	250	Discrete	0	1	1	1	1	1	1	1	1	1	1	1	1
110	250	Discrete	0	1	1	1	1	1	1	1	1	1.5	1.5	1	1
111	250	Discrete	0	1	1	1	1	1	1	1	1	2	2	1	1
112	250	Discrete	0	1	1	1	1	0.5	1	1	1	1	1	1	1
113	250	Discrete	0	1	1	1	1	0.5	0.5	1	1	1	1	1	1

Table A1 (Cont.)

Simulation Conditions

Condition Number	Group Size	Event Time Format	Correlation Between Event Times	Shape Parameter Weibull *k*				Failures per Time Period				Frailty Multiplicative Factor			
				Group 1		Group 2		Group 1		Group 2		Group 1		Group 2	
				Event 1	Event 2	Event 1	Event 2	Event 1	Event 2	Event 1	Event 2	Event 1	Event 2	Event 1	Event 2
114	250	Discrete	0	1	1	1	1	0.5	1	1	1	1.5	1.5	1	1
115	250	Discrete	0	1	1	1	1	0.5	0.5	1	1	1.5	1.5	1	1
116	250	Discrete	0	1	1	1	1	0.5	1	1	1	2	2	1	1
117	250	Discrete	0	1	1	1	1	0.5	0.5	1	1	2	2	1	1
118	250	Discrete	0	2	1	2	1	1	1	1	1	1	1	1	1
119	250	Discrete	0	2	1	2	1	1	1	1	1	1.5	1.5	1	1
120	250	Discrete	0	2	1	2	1	1	1	1	1	2	2	1	1
121	250	Discrete	0	2	1	2	1	0.5	1	1	1	1	1	1	1
122	250	Discrete	0	2	1	2	1	0.5	0.5	1	1	1	1	1	1
123	250	Discrete	0	2	1	2	1	0.5	1	1	1	1.5	1.5	1	1
124	250	Discrete	0	2	1	2	1	0.5	0.5	1	1	1.5	1.5	1	1
125	250	Discrete	0	2	1	2	1	0.5	1	1	1	2	2	1	1
126	250	Discrete	0	2	1	2	1	0.5	0.5	1	1	2	2	1	1
127	250	Discrete	-0.4	1	1	1	1	1	1	1	1	1	1	1	1
128	250	Discrete	-0.4	1	1	1	1	1	1	1	1	1.5	1.5	1	1
129	250	Discrete	-0.4	1	1	1	1	1	1	1	1	2	2	1	1
130	250	Discrete	-0.4	1	1	1	1	0.5	1	1	1	1	1	1	1
131	250	Discrete	-0.4	1	1	1	1	0.5	0.5	1	1	1	1	1	1
132	250	Discrete	-0.4	1	1	1	1	0.5	1	1	1	1.5	1.5	1	1
133	250	Discrete	-0.4	1	1	1	1	0.5	0.5	1	1	1.5	1.5	1	1
134	250	Discrete	-0.4	1	1	1	1	0.5	1	1	1	2	2	1	1
135	250	Discrete	-0.4	1	1	1	1	0.5	0.5	1	1	2	2	1	1
136	250	Discrete	-0.4	2	1	2	1	1	1	1	1	1	1	1	1

Table A1 (Cont.)

Simulation Conditions

Condition Number	Group Size	Event Time Format	Correlation Between Event Times	Shape Parameter Weibull *k*				Failures per Time Period				Frailty Multiplicative Factor			
				Group 1		Group 2		Group 1		Group 2		Group 1		Group 2	
				Event 1	Event 2	Event 1	Event 2	Event 1	Event 2	Event 1	Event 2	Event 1	Event 2	Event 1	Event 2
137	250	Discrete	-0.4	2	1	2	1	1	1	1	1	1.5	1.5	1	1
138	250	Discrete	-0.4	2	1	2	1	1	1	1	1	2	2	1	1
139	250	Discrete	-0.4	2	1	2	1	0.5	1	1	1	1	1	1	1
140	250	Discrete	-0.4	2	1	2	1	0.5	0.5	1	1	1	1	1	1
141	250	Discrete	-0.4	2	1	2	1	0.5	1	1	1	1.5	1.5	1	1
142	250	Discrete	-0.4	2	1	2	1	0.5	0.5	1	1	1.5	1.5	1	1
143	250	Discrete	-0.4	2	1	2	1	0.5	1	1	1	2	2	1	1
144	250	Discrete	-0.4	2	1	2	1	0.5	0.5	1	1	2	2	1	1
145	250	Discrete	0.4	1	1	1	1	1	1	1	1	1	1	1	1
146	250	Discrete	0.4	1	1	1	1	1	1	1	1	1.5	1.5	1	1
147	250	Discrete	0.4	1	1	1	1	1	1	1	1	2	2	1	1
148	250	Discrete	0.4	1	1	1	1	0.5	1	1	1	1	1	1	1
149	250	Discrete	0.4	1	1	1	1	0.5	0.5	1	1	1	1	1	1
150	250	Discrete	0.4	1	1	1	1	0.5	1	1	1	1.5	1.5	1	1
151	250	Discrete	0.4	1	1	1	1	0.5	0.5	1	1	1.5	1.5	1	1
152	250	Discrete	0.4	1	1	1	1	0.5	1	1	1	2	2	1	1
153	250	Discrete	0.4	1	1	1	1	0.5	0.5	1	1	2	2	1	1
154	250	Discrete	0.4	2	1	2	1	1	1	1	1	1	1	1	1
155	250	Discrete	0.4	2	1	2	1	1	1	1	1	1.5	1.5	1	1
156	250	Discrete	0.4	2	1	2	1	1	1	1	1	2	2	1	1
157	250	Discrete	0.4	2	1	2	1	0.5	1	1	1	1	1	1	1
158	250	Discrete	0.4	2	1	2	1	0.5	0.5	1	1	1	1	1	1
159	250	Discrete	0.4	2	1	2	1	0.5	1	1	1	1.5	1.5	1	1

Table A1 (Cont.)

Simulation Conditions

Condition Number	Group Size	Event Time Format	Correlation Between Event Times	Shape Parameter Weibull *k*				Failures per Time Period				Frailty Multiplicative Factor			
				Group 1		Group 2		Group 1		Group 2		Group 1		Group 2	
				Event 1	Event 2	Event 1	Event 2	Event 1	Event 2	Event 1	Event 2	Event 1	Event 2	Event 1	Event 2
160	250	Discrete	0.4	2	1	2	1	0.5	0.5	1	1	1.5	1.5	1	1
161	250	Discrete	0.4	2	1	2	1	0.5	1	1	1	2	2	1	1
162	250	Discrete	0.4	2	1	2	1	0.5	0.5	1	1	2	2	1	1
163	500	Discrete	0	1	1	1	1	1	1	1	1	1	1	1	1
164	500	Discrete	0	1	1	1	1	1	1	1	1	1.5	1.5	1	1
165	500	Discrete	0	1	1	1	1	1	1	1	1	2	2	1	1
166	500	Discrete	0	1	1	1	1	0.5	1	1	1	1	1	1	1
167	500	Discrete	0	1	1	1	1	0.5	0.5	1	1	1	1	1	1
168	500	Discrete	0	1	1	1	1	0.5	1	1	1	1.5	1.5	1	1
169	500	Discrete	0	1	1	1	1	0.5	0.5	1	1	1.5	1.5	1	1
170	500	Discrete	0	1	1	1	1	0.5	1	1	1	2	2	1	1
171	500	Discrete	0	1	1	1	1	0.5	0.5	1	1	2	2	1	1
172	500	Discrete	0	2	1	2	1	1	1	1	1	1	1	1	1
173	500	Discrete	0	2	1	2	1	1	1	1	1	1.5	1.5	1	1
174	500	Discrete	0	2	1	2	1	1	1	1	1	2	2	1	1
175	500	Discrete	0	2	1	2	1	0.5	1	1	1	1	1	1	1
176	500	Discrete	0	2	1	2	1	0.5	0.5	1	1	1	1	1	1
177	500	Discrete	0	2	1	2	1	0.5	1	1	1	1.5	1.5	1	1
178	500	Discrete	0	2	1	2	1	0.5	0.5	1	1	1.5	1.5	1	1
179	500	Discrete	0	2	1	2	1	0.5	1	1	1	2	2	1	1
180	500	Discrete	0	2	1	2	1	0.5	0.5	1	1	2	2	1	1
181	500	Discrete	-0.4	1	1	1	1	1	1	1	1	1	1	1	1
182	500	Discrete	-0.4	1	1	1	1	1	1	1	1	1.5	1.5	1	1

Table A1 (Cont.)

Simulation Conditions

Condition Number	Group Size	Event Time Format	Correlation Between Event Times	Shape Parameter Weibull *k*				Failures per Time Period				Frailty Multiplicative Factor			
				Group 1		Group 2		Group 1		Group 2		Group 1		Group 2	
				Event 1	Event 2	Event 1	Event 2	Event 1	Event 2	Event 1	Event 2	Event 1	Event 2	Event 1	Event 2
183	500	Discrete	-0.4	1	1	1	1	1	1	1	1	2	2	1	1
184	500	Discrete	-0.4	1	1	1	1	0.5	1	1	1	1	1	1	1
185	500	Discrete	-0.4	1	1	1	1	0.5	0.5	1	1	1	1	1	1
186	500	Discrete	-0.4	1	1	1	1	0.5	1	1	1	1.5	1.5	1	1
187	500	Discrete	-0.4	1	1	1	1	0.5	0.5	1	1	1.5	1.5	1	1
188	500	Discrete	-0.4	1	1	1	1	0.5	1	1	1	2	2	1	1
189	500	Discrete	-0.4	1	1	1	1	0.5	0.5	1	1	2	2	1	1
190	500	Discrete	-0.4	2	1	2	1	1	1	1	1	1	1	1	1
191	500	Discrete	-0.4	2	1	2	1	1	1	1	1	1.5	1.5	1	1
192	500	Discrete	-0.4	2	1	2	1	1	1	1	1	2	2	1	1
193	500	Discrete	-0.4	2	1	2	1	0.5	1	1	1	1	1	1	1
194	500	Discrete	-0.4	2	1	2	1	0.5	0.5	1	1	1	1	1	1
195	500	Discrete	-0.4	2	1	2	1	0.5	1	1	1	1.5	1.5	1	1
196	500	Discrete	-0.4	2	1	2	1	0.5	0.5	1	1	1.5	1.5	1	1
197	500	Discrete	-0.4	2	1	2	1	0.5	1	1	1	2	2	1	1
198	500	Discrete	-0.4	2	1	2	1	0.5	0.5	1	1	2	2	1	1
199	500	Discrete	0.4	1	1	1	1	1	1	1	1	1	1	1	1
200	500	Discrete	0.4	1	1	1	1	1	1	1	1	1.5	1.5	1	1
201	500	Discrete	0.4	1	1	1	1	1	1	1	1	2	2	1	1
202	500	Discrete	0.4	1	1	1	1	0.5	1	1	1	1	1	1	1
203	500	Discrete	0.4	1	1	1	1	0.5	0.5	1	1	1	1	1	1
204	500	Discrete	0.4	1	1	1	1	0.5	1	1	1	1.5	1.5	1	1
205	500	Discrete	0.4	1	1	1	1	0.5	0.5	1	1	1.5	1.5	1	1

Table A1 (Cont.)

Simulation Conditions

Condition Number	Group Size	Event Time Format	Correlation Between Event Times	Shape Parameter Weibull *k*				Failures per Time Period				Frailty Multiplicative Factor			
				Group 1		Group 2		Group 1		Group 2		Group 1		Group 2	
				Event 1	Event 2	Event 1	Event 2	Event 1	Event 2	Event 1	Event 2	Event 1	Event 2	Event 1	Event 2
206	500	Discrete	0.4	1	1	1	1	0.5	1	1	1	2	2	1	1
207	500	Discrete	0.4	1	1	1	1	0.5	0.5	1	1	2	2	1	1
208	500	Discrete	0.4	2	1	2	1	1	1	1	1	1	1	1	1
209	500	Discrete	0.4	2	1	2	1	1	1	1	1	1.5	1.5	1	1
210	500	Discrete	0.4	2	1	2	1	1	1	1	1	2	2	1	1
211	500	Discrete	0.4	2	1	2	1	0.5	1	1	1	1	1	1	1
212	500	Discrete	0.4	2	1	2	1	0.5	0.5	1	1	1	1	1	1
213	500	Discrete	0.4	2	1	2	1	0.5	1	1	1	1.5	1.5	1	1
214	500	Discrete	0.4	2	1	2	1	0.5	0.5	1	1	1.5	1.5	1	1
215	500	Discrete	0.4	2	1	2	1	0.5	1	1	1	2	2	1	1
216	500	Discrete	0.4	2	1	2	1	0.5	0.5	1	1	2	2	1	1

Table A2

Percentage Difference Between Model Estimates and Set Failure Rates

Condition	Cox Proportional Hazards		Weibull Regression		Fine and Gray's Method		Cox Proportional Hazards Regression with Frailty	
	Event 1	Event 2	Event 1	Event 2	Event 1	Event 2	Event 1	Event 2
1	0.57%	1.42%	0.52%	1.52%	0.17%	1.60%	0.54%	1.47%
2	2.71%	4.55%	2.54%	4.57%	1.31%	4.04%	2.75%	4.54%
3	4.71%	6.97%	4.50%	7.01%	2.57%	5.88%	4.72%	7.06%
4	1.10%	1.30%	0.78%	1.25%	11.80%	29.96%	1.20%	1.27%
5	1.09%	1.70%	0.52%	1.52%	43.63%	42.85%	1.10%	1.76%
6	3.26%	4.34%	2.76%	4.28%	10.68%	28.64%	3.37%	4.31%
7	3.27%	4.84%	2.54%	4.57%	43.07%	41.61%	3.26%	4.86%
8	5.31%	6.80%	4.62%	6.76%	9.46%	27.63%	5.42%	6.80%
9	5.34%	7.32%	4.50%	7.01%	42.44%	40.68%	5.37%	7.33%
10	0.76%	1.30%	0.11%	1.37%	0.29%	1.40%	0.68%	1.32%
11	7.81%	4.34%	2.08%	4.35%	3.00%	4.41%	7.88%	4.33%
12	15.21%	6.86%	3.99%	6.94%	5.90%	6.83%	15.11%	6.94%
13	**104.55%**	1.27%	0.22%	1.16%	18.11%	33.09%	**104.79%**	1.20%
14	**103.57%**	1.63%	0.12%	1.31%	40.41%	40.88%	**103.89%**	1.61%
15	**118.08%**	4.31%	2.07%	4.20%	22.57%	31.44%	**118.75%**	4.25%
16	**118.00%**	4.71%	2.02%	4.30%	39.16%	39.34%	**118.82%**	4.65%
17	**134.03%**	6.74%	3.95%	6.73%	27.79%	30.16%	**134.09%**	6.72%
18	**133.30%**	7.18%	3.63%	6.92%	37.77%	38.12%	**134.12%**	7.16%
19	0.57%	1.45%	0.39%	1.32%	0.16%	1.61%	0.50%	1.47%
20	2.72%	4.44%	2.14%	3.88%	0.92%	3.54%	2.75%	4.40%
21	4.78%	6.81%	3.83%	5.90%	1.89%	4.91%	4.64%	6.79%
22	1.32%	9.40%	7.80%	8.16%	18.30%	25.40%	1.94%	8.90%
23	8.40%	9.05%	0.39%	1.32%	43.10%	42.30%	7.12%	7.82%
24	0.82%	12.70%	6.30%	11.06%	17.69%	24.16%	0.08%	12.17%

Note. Percentages greater than 50% are in italics and those greater than 100% are in bold.

Table A2 (Cont.)

Percentage Difference Between Model Estimates and Set Failure Rates

Condition	Cox Proportional Hazards		Weibull Regression		Fine and Gray's Method		Cox Proportional Hazards Regression with Frailty	
	Event 1	Event 2	Event 1	Event 2	Event 1	Event 2	Event 1	Event 2
25	10.80%	12.37%	2.14%	3.88%	42.74%	41.33%	9.42%	11.02%
26	2.86%	15.33%	4.80%	13.31%	16.85%	23.32%	2.11%	14.79%
27	13.12%	14.92%	3.83%	5.90%	42.27%	40.65%	11.63%	13.53%
28	0.73%	1.31%	0.11%	1.19%	0.35%	1.38%	0.72%	1.28%
29	7.44%	4.28%	1.85%	3.72%	2.27%	3.85%	7.41%	4.19%
30	14.43%	6.66%	3.62%	5.73%	4.50%	5.71%	14.41%	6.53%
31	*88.85%*	9.64%	7.05%	7.26%	4.61%	28.87%	*89.54%*	7.84%
32	**119.23%**	10.77%	0.30%	1.12%	39.72%	40.52%	**119.82%**	7.28%
33	**101.71%**	12.91%	5.46%	10.01%	7.19%	27.44%	**102.32%**	11.01%
34	**134.25%**	14.04%	1.74%	3.83%	38.89%	39.27%	**135.19%**	10.38%
35	**114.11%**	15.55%	4.01%	12.31%	9.88%	26.36%	**114.73%**	13.53%
36	**150.04%**	16.65%	3.14%	5.63%	37.86%	38.34%	**151.01%**	12.85%
37	0.60%	1.45%	0.60%	1.64%	0.17%	1.60%	0.56%	1.49%
38	3.06%	5.20%	3.22%	5.76%	2.10%	5.29%	3.18%	5.31%
39	5.73%	8.25%	6.13%	9.20%	4.41%	8.26%	5.80%	8.35%
40	12.34%	7.87%	17.92%	8.15%	3.40%	37.15%	12.88%	7.94%
41	2.09%	1.49%	0.60%	1.64%	43.91%	43.14%	1.92%	1.28%
42	14.99%	4.65%	20.76%	4.77%	5.57%	35.45%	15.48%	4.73%
43	0.31%	2.16%	3.22%	5.76%	42.94%	41.25%	0.52%	2.35%
44	17.93%	2.04%	24.04%	1.95%	8.28%	34.15%	18.51%	2.11%
45	2.74%	5.05%	6.13%	9.20%	41.77%	39.74%	3.10%	5.30%
46	0.88%	1.27%	0.13%	1.47%	0.37%	1.31%	0.76%	1.29%
47	10.26%	4.79%	2.41%	5.43%	4.82%	5.31%	10.15%	4.76%
48	20.58%	7.80%	4.80%	8.83%	9.64%	8.65%	20.53%	7.80%

Note. Percentages greater than 50% are in italics and those greater than 100% are in bold.

Table A2 (Cont.)

Percentage Difference Between Model Estimates and Set Failure Rates

Condition	Cox Proportional Hazards		Weibull Regression		Fine and Gray's Method		Cox Proportional Hazards Regression with Frailty	
	Event 1	Event 2	Event 1	Event 2	Event 1	Event 2	Event 1	Event 2
49	**173.08%**	7.21%	11.51%	6.65%	*50.34%*	38.73%	**168.15%**	6.40%
50	**116.09%**	5.34%	0.07%	1.48%	40.47%	40.95%	**113.71%**	2.88%
51	**204.95%**	4.23%	14.77%	3.34%	*65.08%*	36.82%	**200.72%**	3.39%
52	**136.26%**	2.10%	2.45%	5.38%	38.35%	38.91%	**133.69%**	0.44%
53	**242.57%**	1.85%	17.95%	0.56%	*82.87%*	35.28%	**236.33%**	0.92%
54	**157.72%**	0.55%	4.58%	8.76%	36.01%	37.20%	**155.61%**	3.25%
55	0.50%	0.56%	0.55%	0.58%	0.43%	0.44%	0.54%	0.57%
56	2.43%	3.33%	2.46%	3.36%	1.58%	2.77%	2.45%	3.37%
57	4.27%	5.71%	4.30%	5.72%	2.84%	4.64%	4.37%	5.74%
58	0.85%	0.38%	0.80%	0.31%	11.57%	30.75%	0.94%	0.33%
59	0.68%	0.71%	0.55%	0.58%	43.48%	43.52%	0.76%	0.72%
60	2.71%	3.21%	2.71%	3.17%	10.53%	29.42%	2.82%	3.20%
61	2.58%	3.47%	2.46%	3.36%	42.91%	42.33%	2.68%	3.51%
62	4.57%	5.61%	4.55%	5.54%	9.30%	28.41%	4.65%	5.60%
63	4.45%	5.83%	4.30%	5.72%	42.29%	41.38%	4.50%	5.93%
64	0.52%	0.63%	0.16%	0.66%	0.29%	0.53%	0.52%	0.70%
65	6.69%	3.52%	1.99%	3.51%	3.01%	3.54%	6.69%	3.58%
66	13.36%	5.93%	3.84%	5.91%	6.06%	5.92%	13.40%	5.93%
67	**102.44%**	0.49%	0.35%	0.33%	18.17%	33.76%	**102.63%**	0.36%
68	**101.75%**	0.67%	0.06%	0.66%	40.39%	41.40%	**101.95%**	0.75%
69	**114.68%**	3.34%	2.24%	3.16%	22.96%	32.18%	**115.15%**	3.19%
70	**114.01%**	3.52%	2.01%	3.49%	39.11%	39.85%	**114.19%**	3.60%
71	**127.80%**	5.74%	3.95%	5.59%	28.02%	30.86%	**127.76%**	5.67%
72	**127.27%**	5.87%	3.98%	5.88%	37.64%	38.64%	**127.36%**	6.04%

Note. Percentages greater than 50% are in italics and those greater than 100% are in bold.

Table A2 (Cont.)

Percentage Difference Between Model Estimates and Set Failure Rates

Condition	Cox Proportional Hazards		Weibull Regression		Fine and Gray's Method		Cox Proportional Hazards Regression with Frailty	
	Event 1	Event 2	Event 1	Event 2	Event 1	Event 2	Event 1	Event 2
73	0.60%	0.61%	0.54%	0.56%	0.44%	0.44%	0.61%	0.63%
74	2.54%	3.27%	2.19%	2.86%	1.26%	2.24%	2.55%	3.27%
75	4.39%	5.56%	3.75%	4.81%	2.18%	3.66%	4.36%	5.52%
76	1.55%	8.54%	7.76%	7.37%	18.12%	26.18%	2.19%	7.98%
77	8.02%	8.04%	0.54%	0.56%	42.93%	42.99%	6.83%	6.74%
78	0.35%	11.55%	6.28%	10.00%	17.50%	25.02%	0.36%	10.96%
79	10.11%	10.92%	2.19%	2.86%	42.55%	42.08%	8.83%	9.56%
80	2.08%	14.12%	4.87%	12.18%	16.73%	24.15%	1.36%	13.48%
81	12.13%	13.40%	3.75%	4.81%	42.10%	41.36%	10.76%	11.98%
82	0.73%	0.64%	0.24%	0.53%	0.42%	0.46%	0.72%	0.60%
83	6.46%	3.43%	1.83%	2.93%	2.35%	2.99%	6.56%	3.33%
84	12.50%	5.76%	3.40%	4.88%	4.51%	4.88%	12.60%	5.63%
85	*86.78%*	9.01%	6.97%	6.56%	4.61%	29.50%	*87.52%*	7.03%
86	**117.03%**	9.83%	0.10%	0.54%	39.68%	41.07%	**117.74%**	6.35%
87	*97.36%*	12.04%	5.52%	9.19%	7.12%	28.08%	*98.04%*	10.03%
88	**129.73%**	12.91%	1.61%	2.91%	38.83%	39.79%	**130.28%**	9.30%
89	**108.46%**	14.61%	4.06%	11.19%	10.10%	27.05%	**109.02%**	12.47%
90	**143.07%**	15.42%	3.35%	4.90%	37.82%	38.83%	**143.35%**	11.72%
91	0.46%	0.52%	0.54%	0.57%	0.42%	0.44%	0.48%	0.52%
92	2.74%	4.04%	3.08%	4.41%	2.40%	4.03%	2.80%	4.10%
93	5.14%	7.01%	5.81%	7.74%	4.66%	6.95%	5.18%	7.08%
94	11.79%	8.65%	17.60%	9.00%	3.42%	37.77%	12.26%	8.75%
95	2.50%	2.48%	0.54%	0.57%	43.76%	43.79%	2.30%	2.33%
96	14.45%	5.71%	20.85%	5.87%	5.85%	36.17%	14.87%	5.78%

Note. Percentages greater than 50% are in italics and those greater than 100% are in bold.

Table A2 (Cont.)

Percentage Difference Between Model Estimates and Set Failure Rates

Condition	Cox Proportional Hazards		Weibull Regression		Fine and Gray's Method		Cox Proportional Hazards Regression with Frailty	
	Event 1	Event 2	Event 1	Event 2	Event 1	Event 2	Event 1	Event 2
97	0.35%	0.78%	3.08%	4.41%	42.77%	41.95%	0.09%	1.05%
98	17.06%	3.16%	23.93%	3.12%	8.34%	34.84%	17.46%	3.18%
99	1.89%	3.57%	5.81%	7.74%	41.62%	40.46%	2.21%	3.95%
100	0.37%	0.67%	0.04%	0.72%	0.16%	0.63%	0.36%	0.73%
101	9.05%	4.00%	2.34%	4.55%	4.80%	4.55%	8.82%	3.99%
102	19.06%	6.70%	4.80%	7.62%	10.19%	7.62%	18.66%	6.72%
103	**170.06%**	7.90%	11.70%	7.44%	*50.32%*	39.32%	**165.00%**	7.12%
104	**114.26%**	6.13%	0.08%	0.75%	40.57%	41.37%	**111.44%**	3.60%
105	**197.14%**	5.07%	14.59%	4.22%	*63.78%*	37.35%	**191.00%**	4.18%
106	**132.39%**	3.15%	2.38%	4.48%	38.32%	39.36%	**128.99%**	0.44%
107	**229.26%**	2.71%	17.80%	1.50%	*79.84%*	35.80%	**221.75%**	1.69%
108	**153.09%**	0.70%	5.65%	7.53%	35.67%	37.79%	**149.02%**	2.17%
109	0.48%	1.44%	0.14%	0.75%	0.12%	1.44%	0.69%	1.63%
110	2.49%	4.49%	1.37%	2.60%	1.19%	3.71%	3.23%	5.20%
111	4.51%	6.97%	2.57%	4.07%	2.37%	5.44%	5.61%	7.89%
112	0.40%	4.44%	19.09%	3.87%	14.27%	29.03%	6.96%	6.41%
113	5.23%	4.57%	13.52%	12.96%	44.62%	43.92%	13.48%	14.30%
114	2.50%	1.60%	18.08%	5.92%	13.22%	27.79%	9.69%	9.89%
115	3.19%	1.60%	12.30%	11.17%	44.10%	42.74%	16.39%	18.30%
116	4.50%	0.74%	17.11%	7.61%	12.07%	26.83%	12.22%	13.08%
117	1.30%	0.71%	11.18%	9.78%	43.51%	41.87%	19.02%	21.44%
118	0.59%	1.26%	0.14%	0.63%	0.22%	1.20%	0.98%	1.36%
119	7.29%	4.41%	1.43%	2.26%	2.82%	3.96%	8.74%	4.64%
120	14.44%	7.03%	2.74%	3.69%	5.58%	6.19%	16.85%	7.45%

Note. Percentages greater than 50% are in italics and those greater than 100% are in bold.

Table A2 (Cont.)

Percentage Difference Between Model Estimates and Set Failure Rates

Condition	Cox Proportional Hazards		Weibull Regression		Fine and Gray's Method		Cox Proportional Hazards Regression with Frailty	
	Event 1	Event 2	Event 1	Event 2	Event 1	Event 2	Event 1	Event 2
121	*99.33%*	6.77%	13.05%	6.68%	14.31%	31.97%	**124.02%**	10.98%
122	*85.64%*	7.42%	11.30%	13.16%	41.46%	42.39%	**122.83%**	17.38%
123	**112.25%**	3.92%	11.88%	8.64%	18.57%	30.44%	**141.31%**	14.57%
124	*98.53%*	4.59%	9.91%	11.61%	40.26%	40.96%	**141.87%**	21.22%
125	**127.11%**	1.68%	10.71%	10.25%	23.52%	29.25%	**161.11%**	17.54%
126	**112.10%**	2.36%	8.57%	10.36%	38.94%	39.83%	**161.92%**	24.36%
127	0.37%	1.38%	0.02%	0.56%	0.08%	1.41%	0.69%	1.46%
128	2.29%	4.25%	1.12%	2.13%	0.78%	3.19%	3.23%	4.97%
129	4.21%	6.51%	2.15%	3.34%	1.67%	4.47%	5.90%	7.73%
130	3.41%	1.76%	24.70%	8.30%	20.82%	24.57%	4.56%	15.45%
131	0.77%	0.12%	15.26%	14.78%	44.31%	43.59%	21.80%	22.72%
132	1.56%	4.74%	23.84%	10.20%	20.28%	23.43%	7.25%	19.37%
133	1.26%	2.87%	14.21%	13.27%	43.99%	42.68%	25.03%	26.96%
134	0.29%	7.10%	23.06%	11.62%	19.50%	22.64%	9.80%	22.54%
135	3.25%	5.13%	13.22%	12.12%	43.55%	42.04%	28.06%	30.34%
136	0.72%	1.27%	0.02%	0.49%	0.28%	1.18%	0.96%	1.34%
137	6.56%	4.18%	1.25%	1.94%	2.07%	3.43%	8.13%	4.66%
138	12.90%	6.52%	2.45%	3.06%	4.17%	5.13%	15.97%	7.44%
139	*78.93%*	0.59%	19.13%	10.38%	0.80%	27.88%	**106.72%**	19.03%
140	*92.33%*	1.60%	12.59%	14.29%	41.04%	42.16%	**140.49%**	25.83%
141	*89.71%*	2.32%	18.10%	12.13%	3.21%	26.57%	**120.70%**	22.89%
142	**104.31%**	1.27%	11.38%	12.94%	40.25%	41.00%	**159.50%**	29.94%
143	**100.82%**	4.71%	17.15%	13.57%	5.75%	25.58%	**136.57%**	26.10%
144	**117.27%**	3.52%	10.15%	11.85%	39.28%	40.15%	**179.97%**	33.30%

Note. Percentages greater than 50% are in italics and those greater than 100% are in bold.

Table A2 (Cont.)

Percentage Difference Between Model Estimates and Set Failure Rates

Condition	Cox Proportional Hazards		Weibull Regression		Fine and Gray's Method		Cox Proportional Hazards Regression with Frailty	
	Event 1	Event 2	Event 1	Event 2	Event 1	Event 2	Event 1	Event 2
145	0.57%	1.48%	0.26%	0.92%	0.13%	1.47%	0.75%	1.66%
146	3.03%	5.35%	1.74%	3.35%	1.97%	4.97%	3.57%	5.79%
147	5.78%	8.52%	3.38%	5.31%	4.18%	7.79%	6.50%	9.16%
148	12.88%	12.54%	9.09%	2.99%	0.64%	36.06%	18.30%	4.03%
149	6.83%	6.20%	11.86%	11.24%	44.74%	44.01%	9.29%	10.09%
150	15.56%	9.40%	7.86%	0.56%	2.69%	34.44%	21.52%	0.35%
151	4.55%	2.66%	10.40%	8.89%	43.83%	42.22%	12.42%	14.58%
152	18.56%	6.90%	6.48%	1.39%	5.27%	33.20%	24.87%	2.88%
153	2.17%	0.14%	8.85%	7.03%	42.70%	40.77%	15.69%	18.27%
154	0.80%	1.26%	0.10%	0.73%	0.31%	1.15%	1.25%	1.33%
155	10.14%	5.05%	1.92%	2.76%	4.59%	4.83%	11.31%	4.94%
156	20.32%	8.25%	7.88%	4.46%	9.20%	7.95%	22.84%	8.14%
157	**172.16%**	13.83%	3.77%	1.50%	45.83%	37.41%	**195.60%**	2.33%
158	**102.32%**	12.37%	9.94%	12.11%	41.34%	42.38%	**133.92%**	11.07%
159	**203.99%**	11.00%	1.92%	3.67%	*59.96%*	35.64%	**234.50%**	5.87%
160	**120.87%**	9.30%	8.23%	10.22%	39.29%	40.49%	**159.39%**	15.05%
161	**242.33%**	8.70%	0.04%	5.47%	*76.99%*	34.19%	**280.14%**	8.79%
162	**141.11%**	6.72%	6.63%	8.56%	37.04%	38.90%	**188.92%**	18.50%
163	0.49%	0.56%	0.30%	0.32%	0.37%	0.39%	0.72%	0.76%
164	2.35%	3.37%	1.43%	2.02%	1.45%	2.57%	2.90%	3.85%
165	4.16%	5.76%	2.50%	3.39%	2.63%	4.33%	4.96%	6.45%
166	0.32%	5.29%	18.98%	3.34%	14.05%	29.74%	6.60%	5.41%
167	5.47%	5.47%	13.37%	13.35%	44.48%	44.52%	13.03%	12.93%
168	2.08%	2.64%	18.04%	5.27%	13.09%	28.50%	8.82%	8.73%
169	3.72%	2.85%	12.28%	11.73%	43.96%	43.39%	15.41%	16.52%

Note. Percentages greater than 50% are in italics and those greater than 100% are in bold.

Table A2 (Cont.)

Percentage Difference Between Model Estimates and Set Failure Rates

Condition	Cox Proportional Hazards		Weibull Regression		Fine and Gray's Method		Cox Proportional Hazards Regression with Frailty	
	Event 1	Event 2	Event 1	Event 2	Event 1	Event 2	Event 1	Event 2
170	3.90%	0.40%	17.16%	6.86%	11.93%	27.54%	10.99%	11.63%
171	2.01%	0.68%	11.26%	10.41%	43.37%	42.50%	17.86%	19.56%
172	0.52%	0.64%	0.12%	0.33%	0.23%	0.46%	0.90%	0.76%
173	6.42%	3.66%	1.44%	1.92%	2.85%	3.23%	7.47%	3.82%
174	12.88%	6.15%	2.75%	3.16%	5.74%	5.43%	15.24%	6.36%
175	*98.17%*	7.44%	12.92%	6.24%	14.37%	32.57%	**120.71%**	9.99%
176	*84.55%*	8.26%	11.21%	13.43%	41.44%	42.85%	**120.82%**	16.31%
177	**109.87%**	4.81%	11.75%	8.07%	18.94%	31.10%	**136.98%**	13.44%
178	*95.64%*	5.61%	9.90%	11.94%	40.21%	41.41%	**136.39%**	19.89%
179	**122.47%**	2.57%	10.68%	9.58%	23.75%	29.88%	**152.31%**	16.43%
180	**107.52%**	3.49%	8.62%	10.79%	38.81%	40.30%	**153.76%**	22.90%
181	0.59%	0.63%	0.30%	0.32%	0.39%	0.39%	1.15%	1.10%
182	2.40%	3.17%	1.31%	1.72%	1.14%	2.05%	3.28%	3.87%
183	4.11%	5.35%	2.24%	2.88%	1.98%	3.37%	5.33%	6.57%
184	3.55%	0.97%	24.59%	7.91%	20.68%	25.26%	4.15%	14.41%
185	1.04%	1.01%	15.05%	15.03%	44.17%	44.20%	21.43%	21.42%
186	1.81%	3.72%	23.76%	9.67%	20.10%	24.19%	6.61%	18.00%
187	0.78%	1.54%	14.09%	13.70%	43.81%	43.35%	24.16%	24.94%
188	0.22%	5.99%	23.03%	11.06%	19.39%	23.39%	8.76%	21.06%
189	2.54%	3.74%	13.18%	12.58%	43.39%	42.68%	26.82%	28.26%
190	0.70%	0.63%	0.08%	0.30%	0.37%	0.39%	1.25%	0.93%
191	5.92%	3.45%	1.29%	1.67%	2.19%	2.72%	7.48%	4.00%
192	11.35%	5.69%	2.37%	2.74%	4.21%	4.43%	14.26%	6.67%
193	*77.15%*	1.15%	19.06%	10.02%	0.78%	28.43%	**103.01%**	18.17%

Note. Percentages greater than 50% are in italics and those greater than 100% are in bold.

Table A2 (Cont.)

Percentage Difference Between Model Estimates and Set Failure Rates

Condition	Cox Proportional Hazards		Weibull Regression		Fine and Gray's Method		Cox Proportional Hazards Regression with Frailty	
	Event 1	Event 2	Event 1	Event 2	Event 1	Event 2	Event 1	Event 2
194	*90.54%*	2.41%	12.46%	14.50%	41.00%	42.64%	**137.65%**	24.70%
195	*86.72%*	1.60%	18.09%	11.74%	3.16%	27.13%	**115.51%**	21.82%
196	**101.23%**	0.26%	11.35%	13.24%	40.19%	41.46%	**153.42%**	28.45%
197	*96.71%*	3.85%	17.16%	13.08%	5.97%	26.18%	**129.71%**	24.76%
198	**112.47%**	2.47%	10.26%	12.21%	39.23%	40.58%	**170.73%**	31.71%
199	0.47%	0.53%	0.29%	0.32%	0.38%	0.40%	0.57%	0.62%
200	2.77%	4.18%	1.70%	2.58%	2.25%	3.81%	3.05%	4.35%
201	5.26%	7.26%	3.18%	4.44%	4.40%	6.59%	5.67%	7.61%
202	12.41%	13.28%	9.16%	3.55%	0.64%	36.63%	17.49%	4.92%
203	7.15%	7.15%	11.81%	11.78%	44.60%	44.63%	8.75%	8.71%
204	15.11%	10.46%	7.83%	1.32%	2.95%	35.11%	20.67%	1.57%
205	5.08%	3.96%	10.44%	9.60%	43.66%	42.87%	11.47%	12.76%
206	17.79%	7.98%	6.62%	0.57%	5.31%	33.84%	23.50%	1.48%
207	2.94%	1.26%	9.03%	7.83%	42.57%	41.45%	14.30%	16.39%
208	0.34%	0.70%	0.43%	0.39%	0.14%	0.56%	1.00%	0.71%
209	9.03%	4.27%	1.51%	2.30%	4.61%	4.18%	9.83%	4.12%
210	19.20%	7.16%	3.13%	3.87%	9.79%	7.03%	20.80%	6.95%
211	**169.95%**	14.48%	3.55%	0.98%	45.83%	37.96%	**192.01%**	1.57%
212	**100.83%**	13.09%	9.96%	12.44%	41.42%	42.76%	**131.16%**	10.13%
213	**197.26%**	11.79%	1.88%	3.16%	*58.75%*	36.11%	**224.38%**	4.91%
214	**117.76%**	10.21%	8.29%	10.62%	39.25%	40.89%	**153.53%**	13.88%
215	**229.43%**	9.55%	0.06%	4.85%	*74.13%*	34.67%	**262.30%**	7.82%
216	**137.26%**	7.91%	6.59%	9.22%	36.70%	39.44%	**180.28%**	17.02%

Note. Percentages greater than 50% are in italics and those greater than 100% are in bold.