Applicability of Multiplicative and Additive Hazards Regression Models in Survival Analysis

A Thesis submitted to the College of

Graduate studies and Research

In Partial Fulfillment of the Requirements

For the Master of Science Program

In the department of Community Health and Epidemiology

University of Saskatchewan

Saskatoon, Saskatchewan

Canada

By

Sabuj Sarker

PERMISSION TO USE

In presenting this thesis / dissertation in partial fulfillment of the requirements for a postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis/ dissertation in any manner, in whole or in part, for scholarly purposes may be granted by the professor or profession who supervised my thesis/ dissertation work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work has done. It is understood that any copying or publication or use of this thesis/ dissertation or parts there of for financial gain shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis/dissertation.

Requests for permission to copy or to make other uses of materials in this thesis/dissertation in whole or part should be addressed to:

Head of the Department of Community Health and Epidemiology

University of Saskatchewan

Saskatoon, Saskatchewan S7N 0W8

Abstract

Background: Survival analysis is sometimes called "time-to-event analysis". The Cox model is used widely in survival analysis, where the covariates act multiplicatively on unknown baseline hazards. However, the Cox model requires the proportionality assumption, which limits its applications. The additive hazards model has been used as an alternative to the Cox model, where the covariates act additively on unknown baseline hazards.

Objectives and methods: In this thesis, performance of the Cox multiplicative hazards model and the additive hazards model have been demonstrated and applied to the transfer, lifting and repositioning (TLR) injury prevention study. The TLR injury prevention study was a retrospective, pre-post intervention study that utilized a non-randomized control group. There were 1,467 healthcare workers from six hospitals in Saskatchewan, Canada who were injured from January 1, 1999 to December 1, 2006. De-identified data sets were received from the Saskatoon Health Region and Regina Qu'appelle Health Region. Time to repeated TLR injury was considered as the outcome variable. The models' goodness of fit was also assessed.

Results: Of a total of 1,467 individuals, 149 (56.7%) in the control group and 114 (43.3%) in the intervention group had repeated injuries during the study period. Nurses and nursing aides had the highest repeated TLR injuries (84.8%) among occupations. Back, neck and shoulders were the most common body parts injured (74.9%). These covariates were significant in both Cox multiplicative and additive hazards models. The intervention group had 27% fewer repeated injuries than the control group in the multiplicative hazards model (HR= 0.63; 95% CI=0.48-0.82; p-value=0.0002). In the additive model, the hazard difference between the intervention and the control groups was 0.002.

Conclusion: Both multiplicative and additive hazards models showed similar results, indicating that the TLR injury prevention intervention was effective in reducing repeated injuries. The additive hazards model is not widely used, but the coefficient of the covariates is easy to interpret in an additive manner. The additive hazards model should be considered when the proportionality assumption of the Cox model is doubtful.

Acknowledgements

I would like to thank my supervisor, Dr. Hyun J. Lim, for her continuous valuable advise, support and feedback throughout my thesis work and my degree. I am deeply grateful to Dr. Lim for teaching me the details of survival analysis and for guiding me through. Thanks also to my thesis committee members, Dr. Bonnie Janzen and Dr. Punam Pahwa for their guidance during all stages of my thesis. A special thank you to Dr. Punam Pahwa, who repeatedly provided helpful advice on this thesis. Thank you to my external examiner, Dr. Mobinul Huq for insightful feedback and valuable suggestions. My thanks go to Mr. Timothy Russell Black for sharing data and helping me to increase my grasp on the subject. Thank you to Dr. Xu Zhang for helping in the SAS additive macro and Mr. Hitesh Bhatt for helping me to understand the SAS Macro. I wish to acknowledge the Vaccine and Infectious Disease Organization (VIDO) for providing me the financial support during the study. A very special thank you to my family members and friends for their continuous support and encouragement.

Dedication

This work dedicated to my wife Shilpi and my son Shwapnil, to whom I owe an enormous debt of time, energy and gratitude. Without support of Shilpi, I would not able to do this study. She always encouraged me to challenge myself and to persevere.

Table of Contents

List of Tables

List of Figures

List of Abbreviation

MSI: Musculoskeletal Injuries

MSD: Musculoskeletal Disorder

RSI: Repetitive Strain Injuries

TLR: Transfer, Lifting and Repositioning

NIOSH: National Institute for Occupational Safety and Health

NORA: National Occupational Research Agenda

RN: Registered Nurses

GDN: General Duty Nurses

LPN: Licenced Parcticum Nurses

RPN: Registered Phychiatric Nurses

NNA: Nurses and nurses Aide

BNS: Back, Neck and Shoulder

CA: Care Aides

L-Y: Lin and Ying

PH: Proportional Hazards

SHR: Saskatoon Health Region

RQHR: Regina Qu'appelle Health Region

RUH: Royal University Hospital

SCH: Saskatoon City Hospital

PRC: Parkridge Centre

RGH: Regina General Hospital

WRC: Wascana Rehabilitation Centre

PH: Pasqua Hospital

1. Introduction

1.1 Background

1.1.1 Survival Analysis

“Survival analysis” describes the analysis of data (in units of time) culled from a well-defined time origin until the occurrence of some particular event or end-point (1). For that reason, survival analysis is often called “time-to-event analysis” and is used by many researchers in fields, such as medicine, public health, social science and engineering. Here, time is in units of years, months, weeks, or days. Time measures from the beginning of follow-up of a subject until an event occurs or a study ends. Here, an “event” is the occurrence of a death, disease incidence, relapse from remission, or recovery, which may happen to an individual (2). In the field of engineering, survival analysis is called reliability theory, reliability analysis, or failure time analysis because the main focus is in modeling the lifetimes of machines or electronic components (3). In sociology or economics, survival analysis is called event history analysis, duration analysis, or duration modeling. An example of time to an event modeling in the social sciences could be the rate or time at/to which former convicts commit a crime again after they have been released. Because analysis of time to event data arises in a number of applied fields, the developments from diverse fields have been consolidated into the field of "survival analysis" (4). Although survival analysis is called different names in various fields, such as event history analysis, reliability analysis, failure time analysis, or duration analysis, it uses the same analytic techniques (5).

1.1.2 Musculoskeletal Injuries

Injuries of the muscles, nerves, tendons, ligaments, joints, cartilage, or spinal discs are termed either musculoskeletal injuries (MSIs) or musculoskeletal disorders (MSDs). These injuries are not usually the result of any immediate or acute event, such as a slip, trip, or fall, but reflect a more gradual or chronic development. There are other terms that may be used to explain MSIs, such as repetitive strain injuries (RSIs), cumulative trauma disorders, overuse injuries, and repetitive motion disorders (6). According to the U.S. Bureau of Labor Statistics, MSIs include cases where the nature of the injury or illness is sprains, strains, back pain, hurt back, soreness, pain, hurt, except the back, carpal tunnel syndrome, or musculoskeletal system and connective tissue diseases and disorders, when the event or exposure leading to the injury or illness is bodily reaction/bending, climbing, crawling, reaching, twisting, overexertion, or repetition (7). To receive compensation from the workplace safety and insurance board (WSIB) for MSIs, the injury must belong to one of the following categories: sprains, strains, traumatic inflammation of, e.g., muscles, tendons, ligaments, joints; musculoskeletal system and connective tissue diseases and disorders; inflammation and irritation of joints, tendons, muscles and connective tissues; musculoskeletal system and connective tissue diseases and disorders, such as fibromyalgia, fibrosis, myofasciitis; back pain, hurt back, soreness, pain, hurt; carpal tunnel syndrome; symptoms involving nervous and musculoskeletal systems; or multiple symptoms involving the head and neck (8).

Some important causes of MSIs are (in combination or when one occurs at an extreme level of forceful exertion) repetitive movements, awkward postures, combination effects, or secondary risk factors (6). Here, “force” is defined as the amount of effort required to perform a task or job; this can be affected by one’s posture and the number of physical exertions performed. The stress

on the body is greater when more force is applied. Activities that require force exertion or muscle effort include, e.g., lifting, transferring, repositioning, pushing, pulling, and gripping a tool (6). When any activity is performed again and again, it is called a repetitive movement. However, an awkward posture (i.e., directed away from the body's natural position) held for long periods of time can also be of risk to the worker due to continual stress placed on one body part without sufficient muscle recovery time. The closer the joint is to its end of range of motion, the greater the stress placed on the soft tissues of the joint, such as muscles, nerves, and tendons (6). Secondary risk factors may be contact pressure, vibration, gloves, or temperature. When two or more risk factors combine in one job, the chance of injury is increased versus when there is only one risk factor (6). Many articles have discussed in detail the causes of MSI (9-16).

1.1.3 Summary

Survival analysis is a time-to-an-event analysis. In this study, survival analysis regards the time to repeated MSI injury. Multiplicative and additive hazards models will be compared, and their goodness of fit will be assessed.

1.2. Study Objectives and Rationale

1.2.1 Study Objectives

The objectives of this thesis are

(i) To demonstrate the performance of the multiplicative model (Cox hazards model) and the additive hazards model (Aalen's model and Lin & Ying's model);

(ii) To apply the multiplicative and additive hazards models to the transfer, lifting and repositioning (TLR) injury prevention study and to assess the models goodness of fit.

1.2.2 Rationale of Study

In survival analysis, the Cox proportional hazards regression (17) model is used widely. The Cox proportional hazards regression model (17) is one of the multiplicative models, which is also known as the Cox model, Cox proportional hazards model, Cox multiplicative hazards model, Cox hazards model, or Cox hazards regression model. In this model, the effect of the covariates acts multiplicatively on some unknown baseline hazard. The Cox models assume that the risk coefficients are unknown constants whose value does not change over time. If the baseline hazard has a particular parametric distribution, then it turns into a parametric model. When the proportionality assumption in the Cox model is not satisfied, this model can lead to potentially biased estimates and conclusions. When the proportionality assumption is violated in the Cox model, an alternative (additive hazards) model can be used, which assumes that the covariates act in an additive manner on an unknown baseline hazard. In this thesis, two additive hazards models, Aalen's model (18) and Lin & Ying's model (19-21) have been considered, which are sometimes called the "additive models". In Aalen's model, the unknown risk coefficients are allowed to be functions of time so that the effect of covariates may vary over time. In Lin & Ying's model, the unknown risk coefficients used in Aalen's model are replaced by a constant covariate effect. These additive models are not used widely due to a lack of availability of statistical software to estimate and test the models' adequacy.

Few studies have been published comparing the multiplicative hazards model and the additive hazards model; one recently compared these models in regard to pediatric firearm injuries (22). In this thesis, three models (the Cox hazards model, Aalen's additive hazards model and Lin & Ying's additive hazards model) will be compared and applied to data from the transfer, lifting and repositioning injury prevention study. We will also examine the goodness-of-fit analysis of the models.

2. Literature Review

2.1 Review of MSI Research

Musculoskeletal injuries (MSIs) at work result in considerable personal and societal burdens (23). The burden of MSIs is continuously high worldwide. Approximately 40% of all the occupational and work-related diseases has been due to MSIs (15, 24). One study showed that, for occupational exposure, the predominant portion of disease burden is due to ergonomic stressors. It has been estimated that 37% of workers in the healthcare services have lower back pain (15). A broad review of back pain prevalence studies reported a similar pattern of prevalence in several countries, such as the Netherlands (47%), Sweden (64%), and Greece (75%). However, there were some differences in measurement tools, back pain definitions, and occupational groups included in these studies (25). Another study in Great Britain indicated that 59% of nurses had MSI symptoms as well as high injury rates, while other studies indicated that 46% had lower back pain with no differences between healthcare occupations (26-28).

In North America, the National Institute for Occupational Safety and Health (NIOSH)'s National Occupational Research Agenda (NORA) indicates that MSIs are among the most costly healthcare problems facing society today (29). NIOSH research demonstrates that significant occupational risks for MSI exist and are the leading source of workers' compensation claims and costs in the healthcare setting (30, 31). Direct patient care workers have a fundamental role in healthcare. They provide basic patient care, assist patients with their daily activities, and provide emotional support to patients. In nursing homes, about 80-90% of care was provided by direct care workers (32, 33). Nursing aides, orderlies, and attendants have the highest rate of MSI

injuries and illness (465 and 449 per 10,000 in the years 2007 and 2008, respectively) (34). Among nurses, back injuries were associated with high physical loads involved in manual lifting and transferring of patients (35-37); thus, a major cause of MSIs was due to patient handling activities (38-43). Indeed, a study by the Duke Health and Safety Surveillance System (DHSSS) indicated that one third of all MSIs resulted from patient handling activities (44). The same study reported that inpatient nurses, nursing aides, and radiology technicians were the major groups that incurred MSIs (44).

One study of work-related injury among direct patient care occupations in British Columbia showed that MSIs constituted the highest proportion of total injuries in all occupations (45). In that study, the occupations considered were registered nurses (RN), licensed practical nurses (LPN), and care aides (CA) in three healthcare settings (i.e., acute care, nursing homes and community care). A study of ambulatory physician care for MSD in Canada showed that person-visit rates for MSD varied by province, were highest among older patients and were higher for women than for men (46).

2.2 Review of Methodology on Survival Analysis

Survival analysis is used to study how the survival experience of a group of patients depends on the values of explanatory variables. In the analysis, the values of explanatory variables have been recorded for each patient at the time of origin or are time-dependent. For that reason, the hazards regression model is used in survival analysis (1). Two broad reasons for modeling survival data are: (i) to determine which combination of potential explanatory variables affects the form of the hazard function and (ii) to obtain an estimate of the hazard function itself for an individual. The

most popular model in survival analysis is the proportional hazard model, which was proposed by Cox and is known as the Cox hazards regression model.

The Cox hazards regression model is one of the multiplicative hazards models and is the most widely used model in the field of Biostatistics. This Cox model has been used for several cases of musculoskeletal injuries. Crook et al. completed a study to determine specific clinical and behavioral factors that prognostically influence time to return to work following a musculoskeletal work-related injury (47); they used the Cox model for analysis for time-dependent covariates. Another retrospective cohort study was conducted, wherein a cohort of 3,769 healthcare workers in an acute care hospital in British Columbia, Canada, was considered (48). However, they used the Poisson model to study the relationship between work-organization factors and the risk of lower-body musculoskeletal injury among healthcare workers. The Cox regression model was also used to investigate the association between work-related risk factors and sickness levels (49).

Estimation of the Cox (17) model is based on the partial likelihood approach. The Cox model has the advantage of simple interpretation of the results and well established computer programs to conduct the parameter and variance estimations. However, there are some weaknesses in the Cox model. First, the proportionality assumptions may not be satisfied. Notably, the Cox model has been used without proper checks for model goodness of fit (50). Second, the influences of covariate changes over time are not easy to assess. Third, depending on (i) the modifications in the number of covariates modeled and (ii) the precision of their measurement, the proportionality assumption is vulnerable. The proportionality assumption might not be satisfied if the covariates

are deleted from a model or measured with a different level of precision (51, 52). The Cox model is thus different from the ordinary linear models of statistics due to the lack of consistency; this represents a conceptual weakness and may also have practical importance (52).

Considering all the above weaknesses in the Cox model, in 1989, Aalen suggested a simple linear model (18), which he originally suggested in 1978 (53). Aalen suggested using the counting process as a tool for formulating many of the statistical models encountered in the analysis of survival data and more general event history data (54). He suggested the multiplicative intensity model and, in 1980, he introduced a matrix version of the multiplicative intensity model (53). The multiplicative intensity model is primarily intended for the study of regression in life testing. This life testing situation along with covariate information has been not only the object of a number of studies, mostly parametric studies, but also a nonparametric/semi-parametric study by Cox (17). The model suggested by Aalen was not meant as a competitor to the others but as a supplementary approach to provide more detailed information (53). As a non-parametric approach, Aalen's model allows one to assess possible changes in the influence of the covariates over time. It is non-parametric in the sense that no assumption would be made about the functional form. Also, this intensity function will naturally be restricted by the fact that each component of intensity must be non-negative. In 1984, Buckley suggested the additive and multiplicative models for relative survival rates (55). This relative survival rate concept was introduced by Berkson (56). The relative or corrected survival rate for a group of patients is the ratio of an observed survival rate to an expected rate for the group for demographically similar individuals in a reference population. In his study, Buckley examined both the Cox and Aalen's models by using maximum likelihood estimates and related statistics for cancer (55). According

to that study, the choice of the models for the analysis was important. However, with smaller sample sizes and varying disease effects, the distinction would be less clear.

The application of Aalen's model has been described by several authors (57-59), and further development has been recommended by others (60-62). Aalen suggested that the new additive model may be useful in the medical field (18). Aalen's model specifies how the hazard rate depends on covariates in a linear way. In that article, the additive model was discussed in a broader context, and the estimators were presented in a less technical manner. The estimation procedure for Aalen's model was determined by the cumulative regression functions, which were mathematically defined. The main focus was the cumulative regression plots, where the slope of the plots at any given time should give information on the influence of the covariate at that moment. Also, a test procedure and goodness-of-fit plots were suggested.

Later, Aalen suggested further development of a nonparametric linear regression model in survival analysis (63). Three diagnostic methods were studied in his paper: (i) martingale residuals were introduced for the linear model to test the goodness of fit of the model; (ii) Aalen focused on the use of bootstrap replications to judge which features of the cumulative regression plots were likely to reflect real phenomena and not merely random variation. There were no existing formal tests for judging the significance of the cumulative regression plots. Thus, for judging which features were consistent throughout the curves and to reflect the real features, several bootstrap cumulative regression plots were used. Finally, (iii) because cumulative plots gave the information in an indirect way, the slopes of the curves needed to be interpreted; this

was not straight-forward. Thus, Aalen suggested the kernel smoothing procedure, which was generally applied in probability density estimation.

Although the various additive hazards models have been highly advocated and used successfully by numerous authors (18, 53, 55, 64-69), no satisfactory semi-parametric methods of estimation have been developed. Lin and Ying observed that this lack of progress is attributed to the fact that the partial likelihood approach cannot be used directly to eliminate the baseline hazard in estimating the intercept (20). They have developed simple procedures with high efficiencies for making inferences about the regression parameters under the additive hazards model with an unspecified baseline hazards function. In their study, a simple semi-parametric estimating function for the intercept was constructed, which imitated the martingale feature of the partial likelihood score function for baseline hazards. Still, there were some problems in that study in relation to generalizing estimating function to the case of multivariate failure time data as well as methods for checking the adequacy of the model. Ying and Lin further extended this model in two subsequent papers (19, 20). They suggested the semi-parametric analysis of general additive-multiplicative hazard models for the counting process (20) and the additive hazards regression models for survival data but compared these with the frailty model (19).

In consideration of this theoretical point of view, several studies have been done to fit both the Aalen and the L-Y additive hazards models. Martinussen & Scheike studied a flexible additive multiplicative hazard model, which was based on Aalen's and Cox's models (70). They considered a new additive-multiplicative hazard model that consists of two components: additive covariates from Aalen's additive model and multivariate covariates from Cox's regression

model. Martinussen & Scheike applied their model to the real tics data and discovered that the additive-multiplicative model for their study showed lower mortality than the additive model.

The Cox proportional hazards model and Aalen's additive hazards model were compared in a severe breast cancer study in 2004 (71). In that study, these two models provided the same results for some time periods, but for other time periods, they provided different results. Both the models indicated that the same covariates were significant for the model and were selected. The estimates of covariate effects were easily interpreted, but the assumption of proportionality was necessary to make that estimate valid. For the additive model, plots of the cumulative regression function provided an appealing explanation for how the hazards profiles were distributed. Those cumulative regression functions did not easily transform into a single numerical estimate of the covariate effect. Comparison of the additive and multiplicative hazards models was performed using simulation in the breast cancer study (72). According to that simulation, the two models should not be viewed as alternative to each other because they provide different kinds of information. They suggested that they may be used together to further the understanding of the data. Bhattacharya and Klein showed that Aalen's approach leads to weighted comparisons of the crude estimate of the hazards rate of each group as compared to a baseline group (73). In their study, they indicated that this weighting leads to inconsistent tests in the sense that the test statistic depends on which group someone picks as the baseline group. They showed that consistent tests were obtained by using common weight functions for all comparisons. If the weight functions are asymptotically equal under the null hypothesis, then the tests will lead to asymptotically equivalent results regardless of the choice of the baseline group.

3. Methods and Materials

3.1 Basic Concepts

The initial step of analyzing survival data is to present numerical or graphical summaries of the survival times of individuals in a particular group. However, standard statistical procedures in data analysis cannot be used in these analyses because the data are generally not symmetrically distributed (mainly when the survival time is censored). Generally, if someone constructs a histogram from the survival data, it will tend to be positively skewed. As a result, an assumption of normality does not satisfy because survival times are censored.

Censoring: The survival time of an individual is said to be censored when the end-point of interest has not been observed, the patient is lost to follow-up, or the individual withdraws from the study (1, 2). Suppose that a patient who entered a study at time t_0 dies at time t_0+t. Here, t is unknown because the patient may still be alive or lost to follow-up. However, if the patient was last known to be alive at time t_0+c, then the time c is called a censored survival time. The censoring indicator δ is

$\delta = 1$, if event

0, if censored.

The censoring is the key analytical problem in survival analysis. There are three types of censoring:

Right censoring: If the censoring occurs after the individual has been entered into a study, that is, to the right of the last known survival time, then this is called right censoring. The right-censored survival time is less than the actual, but unknown, survival time.

Non-parametric model: A model which does not require any specific assumptions about the underlying distribution of the survival times. This is also called the distribution free method. Common non-parametric methods for estimating the survival functions are Life-table (LT), Kaplan-Meier (KM) and Nelson-Aalen's (NA). For comparing two or more groups of survival times, non-parametric procedures such as, the log-rank test and the Wilcoxon test, are generally used.

Semi-parametric model: A model that has the components of both parametric and non-parametric models is called a semi-parametric model. A commonly used semi-parametric model is the multiplicative hazards model due to Cox (17), which is often called the proportional hazards model.

Parametric model: A model in which a specific probability distribution is assumed for the survival times is known as a parametric model. These models are chosen not only because of their popularity among researchers who analyze survival data but also because they offer insight into the nature of the various parameters and functions (4). Some of the important parametric models are: Exponential, Weibull, Gamma, Log-normal, Log-logistic, Gompertz, Inverse Gaussian, Pareto, and Generalized Gamma distributions.

Notations and Definitions

We will use the following notations throughout this thesis:

T: Here, *T* is a non-negative random variable and is the time until some specific event occurs. This event may be, e.g., death, the appearance of a tumor, the development of some disease, or recurrence of a disease.

t: Some values of time (non-negative values) of random variable *T*

In general, survival distribution is described by three functions: survivor or survival function, cumulative distribution function and hazard function (5). Survival data are summarized through estimates of the survivor function and hazard function (1).

3.1.1 Survival Function and Hazard Function

Cumulative distribution function (c.d.f.): Suppose the random variable *T* has a probability distribution with underlying probability density function (p.d.f.) *f(t)*. Then the c.d.f. of a variable *T*, denoted by *F(t)*, is a function that tells us the probability that the variable will be less than or equal to any value t that we choose. Thus,

$$F(t) = \mathrm{P}(T \le t) = \int_0^t f(u)du\,, \tag{3.1}$$

where

$$f(t) = \lim_{dt \to 0} \frac{P(t \le T < t + dt)}{dt}. \tag{3.2}$$

Survival function (*S(t)*): The main function used to describe time-to-event phenomena is the survival function. The survival function is defined as the probability that an individual survives to time *t*, which is denoted by *S(t)*. Thus,

$S(t) = P(T \ge t) = 1 - F(t)$. Note that *S(t)* is a monotone decreasing function, and $S(0)=1$, $S(\infty) = 0$.

Hazard function (*h(t)*): The hazard function *h(t)*, is defined as the risk or hazard of death at some *t* and is obtained from the probability that an individual dies at time *t*, conditional on that person having survived up to that time. Thus,

$$h(t) = \lim_{dt \to 0} \frac{P(t \le T < t + dt / T \ge t)}{dt}. \tag{3.3}$$

This function is also called the hazard rate, instantaneous death rate, the intensity rate, or the force of mortality, and *h(t)* is a non-negative function (i.e., $h(t) \ge 0$) and has no upper bound.

The relationship of *f(t), S(t)* and *h(t)*: The relationship between *f(t), F(t), S(t), h(t)* and *H(t)* can be expressed as follows (1):

$$f(t) = \frac{dF(t)}{dt} = -\frac{dS(t)}{dt} \tag{3.4}$$

$$h(t) = \frac{f(t)}{S(t)} = -\frac{d}{dt} \log S(t) \tag{3.5}$$

$$H(t) = - \log S(t) \tag{3.6}$$

$$S(t) = \exp\{-H(t)\} = \exp\left\{-\int_0^t h(u)du\right\} \tag{3.7}$$

$$f(t) = h(t) \exp\{-\int_0^t h(u)du\}. \tag{3.8}$$

Mathematical Notation

In this section, the following notations will be used for the multiplicative hazards model and the additive hazards models:

Let us suppose that:

X = time to some event at time *t*,

T_j = the time of study for the j^{th} patient, $j=1, 2, \ldots, n$,

δ_j = the event indicator for the j^{th} patient (δ_j =1 if the event has occurred and δ_j =0 if the life time is right-censored), and

$\boldsymbol{Z}_j(t) = (Z_{j1}(t), Z_{j2}(t), \ldots\ldots, Z_{jp}(t))'$ is the vector of *p* covariates or risk factors for the j^{th} individual at time *t*, which may affect the survival distribution of X.

The $Z_{jk}(t)$, $k=1,2, \ldots, p$ may be time-dependent covariates whose values change over time, such as current disease status and serial blood pressure measurements. They also may be constant values known at time 0, such as sex, treatment group, race, and initial disease state.

$\boldsymbol{\beta}=(\beta_1, \ldots\ldots\ldots, \beta_p)'$ is a parameter vector of $\boldsymbol{Z}$.

Define for the j^{th} individual:

$Y_j(t)$ = 1, if individual *j* is under observation (at risk) at time *t*

0, if individual *j* is not under observation (not at risk) at time *t*.

3.2 Estimate of Survival and Hazard Function

3.2.1 Non-parametric Estimate

The most widely used nonparametric methods for estimating and comparing survival distribution are the Kaplan-Meier (KM) product-limit (PL) estimates and the life-table (LT) or actuarial methods (56). The KM method is most suitable for smaller data sets with precisely measured event times, and the LT method may be better suited for large data sets (74). An alternative estimate of the survivor function is the Nelson-Aalen's (NA) estimate, which is based on the individual events time. The KM estimate can be regarded as an approximation of the NA estimate. For the non-parametric estimate of the survival and hazard function, in this section, only the Kaplan-Meier method will be discussed.

3.2.1.1 Kaplan-Meier Method

The earliest statistical method devised to study human mortality was the LT estimate (56), which is also known as the actuarial estimate of survival function. However, modern methods like the KM reduced its importance (75). The KM estimator of the survival function is usually used to analyze individual data. Suppose that the events occur at D distinct times $t_1<t_2<....<t_D$, and that at time t_i, there are d_i number of events. Let Y_i be the number of individuals who are at risk at time t_i. Note that Y_i is a count of the number of individuals with a time on study of t_i or more (i.e., the number of individuals who are alive at t_i or experience the event of interest at t_i). Then, the KM estimator is defined as

$$\widehat{S}(t) = 1 \text{ if } t < t_1$$

$$= \prod_{t_1 \le t} \left[1 - \frac{d_i}{Y_i} \right], \text{ if } t_1 \le t. \tag{3.9}$$

This estimator is a step function with jumps at the observed event times. The size of these jumps depends not only on the number of events observed at each time t_i but also on the pattern of the censored observations prior to t_i.

The KM estimator provides an efficient means of estimating the survival function for right-censored data. It can also be used to estimate the cumulative hazard function $H(t) = - ln[S(t)]$.

When the survival times of two or more groups of patients are being compared, the log-rank test and the Wilcoxon test can be used (1).

3.2.2 Hazards Model

The hazards function is a useful way of describing the probability distribution for the time of event occurrence. Each hazards function has a corresponding probability distribution. However, the hazards function can be extremely complicated. One of the simplest hazards models is $h(t) = \lambda$, which is constant over time. This implies exponential distribution for the time until an event occurs (or the time between events).

Suppose we have the fixed-covariate $\boldsymbol{Z}_{j(t)} = \boldsymbol{Z}_j = (Z_{j1}, \ldots\ldots, Z_{jp})'$. Then, the exponential hazards model can be written as

$$h(t) = exp\{ \beta_0 + \beta_1 \boldsymbol{Z}_1 + \beta_2 \boldsymbol{Z}_2 + \ldots\ldots\ldots + \beta_p \boldsymbol{Z}_p \}. \qquad (3.10)$$

This method can be useful in the analysis of a single sample of survival data or in the comparison of two or more groups of survival times. However, in most medical studies, subjects in the groups have some additional characteristics that may affect their outcome. For example, subjects may have associated demographic variables, such as age, gender, socio-economic status, or education; behavioral variables, such as dietary habits, smoking history, physical activity level, or alcohol consumption; or physiological variables, such as blood pressure, blood glucose levels, hemoglobin levels, or heart rate. These variables may be used as covariates (i.e., explanatory variables, confounders, risk factors, or independent variables) to explain the response (dependent) variable. After adjusting for those potential explanatory variables, the comparison of survival times between groups should be less biased and more precise than a simple comparison. Another important problem is to predict the distribution of the time to some event from a set of explanatory variables. The interest is in predicting the risk factors for the event of interest (4). To explore the relationship between the survival experience of a patient and explanatory variables,

the models for survival data used are (i) the multiplicative hazards model and (ii) the additive hazards model.

3.2.2.1 Multiplicative Hazards Model

The Cox hazards model is one of the most commonly used multiplicative hazards models. This model is also known as the Cox model, Cox proportional hazards model, PH model, Cox multiplicative hazards model, proportional hazards model, Cox hazards model, or the hazards regression model The Cox model is based on the assumption of proportional hazards, that is, the hazard ratio is constant over time; i.e., the hazard for one individual is proportional to the hazard for any other individual, where the proportionality constant is independent of time. The Cox proportional hazard model is

$$h(t|\boldsymbol{Z})=h_0(t)c(\boldsymbol{\beta Z}), \tag{3.11}$$

where $h_0(t)$ is the baseline hazard, and $c(\boldsymbol{\beta Z})$ is a function of the values of the vector of explanatory variables.

The Cox model is the most widely used survival model in the health sciences, but it is not the only model available. There is a class of survival models, called parametric models, in which the distribution of the outcome (i.e., the time to an event) is specified in terms of unknown parameters (2). If we can assume a particular probability distribution for the data, inference based on such an assumption will be more precise.

We have been paying attention to the multiplicative regression model for the survival data based on the Cox hazards model. In the Cox model, the effect of the covariates was to act multiplicatively on some unknown baseline hazards rate. Covariates that do not act on the baseline hazards rate in this fashion were modeled either by the inclusion of a time-dependent covariate or by stratification. In a similar manner, the fully parametric models can be multiplicative (76). We know that the multiplicative models are very useful in practice because

either the estimated coefficients themselves or simple functions of them can be used to provide estimates of hazard ratios. In addition, statistical software is readily available, and it is easy to use it to fit models, check model assumptions, and assess model fit.

3.2.2.1.1 Cox Proportional Hazards Model

The Cox proportional hazards model is

$$h(t|\boldsymbol{Z}) = h_0(t)c(\boldsymbol{\beta Z}), \tag{3.12}$$

where $h_0(t)$ is an arbitrary baseline hazards rate, and $c(\boldsymbol{\beta' Z})$ is a known function. $\boldsymbol{Z}=(Z_1, \ldots, Z_p)$ is the covariate vector and $\boldsymbol{\beta}=(\beta_1, \ldots, \beta_p)'$ is the coefficient vector of $\boldsymbol{Z}$.

The Cox hazards model is also called a semi-parametric model because a parametric form is assumed only for the covariate effect. The baseline hazards rate is unspecified. Because $h(t|\boldsymbol{Z})$ must be positive, a common model for c($\boldsymbol{\beta Z}$) is

$$c(\boldsymbol{\beta Z}) = exp(\boldsymbol{\beta Z}) = \exp\left(\sum_{k=1}^{p}\beta_k Z_k\right), \tag{3.13}$$

which implies that

$$h(t|\boldsymbol{Z}) = h_0(t)\exp(\boldsymbol{\beta Z}) = h_0(t)\exp\left(\sum_{k=1}^{p}\beta_k Z_k\right) \tag{3.14}$$

The Cox hazards model is a proportional hazards model because if we look at two individuals with covariate values $\boldsymbol{Z}$ and $\boldsymbol{Z}^*$, the ratio of their hazard rates is

$$\frac{h(t/Z)}{h(t/Z^*)} = \frac{h_0(t)\exp\left\{\sum_{k=1}^{p}\beta_k Z_k\right\}}{h_0(t)\exp\left\{\sum_{k=1}^{p}\beta_k Z_k^*\right\}} = \exp\left\{\sum_{k=1}^{p}\beta_k (Z_k - Z_k^*)\right\}, \tag{3.15}$$

which is a constant. So the hazards rates are proportional. This is called the relative risk or hazards ratio of an individual with risk factor $\boldsymbol{Z}$ having the event as compared to an individual with risk factor $\boldsymbol{Z}^*$.

To fit the Cox hazards model, we need to estimate the unknown parameters $\boldsymbol{\beta}$-coefficients, which can be estimated using the maximum likelihood method. According to the maximum likelihood method, the likelihood that the sample data has been obtained first, which is the joint probability of the observed data, is regarded as a function of the unknown parameters in the assumed model. For the Cox hazards model, the maximum likelihood function is a function of the observed survival times, and the β-parameter is the linear component of the model. Estimates of the $\boldsymbol{\beta}$'s are then those values that are the most likely on the basis of the observed data. These maximum likelihood estimates are therefore the values that maximize the likelihood function. From a computational point of view, it is more convenient to maximize the logarithm of the likelihood function. Furthermore, approximations to the variance of maximum likelihood estimates can be obtained from the second derivatives of the log-likelihood function (1).

As indicated earlier, suppose the data is based on a sample of size n consisting of the triple (T_j, δ_j, $\boldsymbol{Z}_j$), $j=1, 2, ..., n$. Consider that the censoring is non-informative in that, given $\boldsymbol{Z}_j$, the event and censoring time for the j^{th} patient are independent, and there are no ties between the event times. Let $t_1<t_2<\ldots\ldots<t_D$ to denote the order event times and $Z_{(i)k}$ be the k^{th} covariate associated with the individual whose failure time is t_i. The set of individuals who are at risk at time t_i are denoted by $R(t_i)$, given the set of all individuals who are still under study and uncensored at a time just prior to t_i.

Then, the relevant likelihood function for the Cox hazards model is

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{D} \frac{\exp\left[\sum_{k=1}^{p} \beta_k Z_{(i)k}\right]}{\sum_{j \in R(t_i)} \exp\left[\sum_{k=1}^{p} \beta_k Z_{jk}\right]} \tag{3.16}$$

The log likelihood of the above likelihood, $LL(\boldsymbol{\beta}) = lnL(\boldsymbol{\beta})$, can then be written as

$$LL(\boldsymbol{\beta}) = \sum_{i=1}^{D}\sum_{k=1}^{p} \beta_k Z_{(i)k} - \sum_{i=1}^{D} \ln\left[\sum_{j \in R(t_i)} \exp\left(\sum_{k=1}^{p} \beta_k Z_{jk}\right)\right] \tag{3.17}$$

The partial maximum likelihood estimates of the β-parameter are found by maximizing (3.16), or equivalently, (3.17). The score equations are found by taking partial derivatives of (3.17) with respect to the ***β***'s as follows:

Let $U_k(\boldsymbol{\beta})=\frac{\delta LL(\beta)}{\delta \beta_k}$, *k=1, 2, ..., p.*

then,

$$U_k(\boldsymbol{\beta}) = \sum_{i=1}^{D} Z_{(i)k} - \sum_{i=1}^{D} \frac{\sum_{j \in R(t_i)} Z_{jk} \exp[\sum_{k=1}^{p} \beta_k Z_{jk}]}{\sum_{j \in R(t_i)} \exp[\sum_{k=1}^{p} \beta_k Z_{jk}]} \tag{3.18}$$

The partial maximum likelihood estimates are found by solving the set of *p* nonlinear equations $U_k(\boldsymbol{\beta})=0$, *k=1, 2, ..., p*. This can be done numerically by using a Newton-Raphson technique (4).

3.2.2.2 Additive Hazards Model

There may be times when a measure of the additive effect of a covariate is preferred over a relative measure. Several different forms of additive models are possible. The simple additive hazards model given by Cox and Oakes (69) is

$$h(t|\boldsymbol{Z}) = h_0(t) + \varphi(\boldsymbol{Z}) \tag{3.19}$$

where $\varphi(0)=0$ and $\varphi(\boldsymbol{Z})$ is constrained so that the right-hand side is non-negative, and $h_0(t)$ is the baseline hazard. Two additive models have been considered with great attention: Aalen's additive model (18) and Lin and Ying's (L-Y) Models (21). Aalen's model assumes that the covariates act in an additive manner on an unknown baseline hazards rate. The unknown risk coefficients in Aalen's model are allowed to be functions of time so that the effect of a covariate may vary over time. The least-squares approach is used to estimate the cumulative regression functions and the standard errors of these functions (4). In the L-Y model, the time-varying regression coefficients in Aalen's model are replaced by constants. For the L-Y model, the estimating equation is obtained from the score function to estimate the model (4). It is attractive

to study and utilize these additive hazards models for several reasons. The following two main justifications were described in great detail by (64, 67). First, the risk difference is complementary to and, from the public health point of view, more important than the risk ratio in describing the association between the risk factor and disease occurrence. Second, biological and empirical evidence suggests that the additive hazards model fits certain types of data better than the proportional hazards model (19). In the next section, two additive hazards models will be reviewed and compared. Furthermore, the models will be applied to the TLR injuries intervention data on musculoskeletal injury among healthcare workers.

3.2.2.2.1 Aalen's Additive Hazards Model

Aalen developed a more general additive model. In his model, he discussed the issues of estimation, testing and assessment of model fit (10, 11). The covariates perform in an additive manner on an unknown baseline hazards rate. The unknown risk coefficients in the model are the function of time so that the values of the regression coefficients are allowed to fluctuate over time.

Aalen's hazard model or the conditional hazards rate for the j^{th} individual at time t given $\boldsymbol{Z}_j(t)$ is defined as

$$h(t|\boldsymbol{Z}_j(t)) = h_0(t) + \sum_{k=1}^{p} \beta_k(t) Z_{jk}(t), \tag{3.20}$$

where $\boldsymbol{Z}_j(t) = Z_{j1}(t), \ldots, Z_{jp}(t)$ is a p-vector of , possibly time-dependent covariates.

Thus, the hazard at any time is a sum of a baseline hazard and a linear combination of the covariate values. Aalen's model measured the influence of the respective covariates. Because regression functions may vary with time, their analyses may reveal changes in the influence of the covariates over time, which is one of the main advantages of Aalen's model. Aalen's model is non-parametric in the sense that no assumption is made about the functional forms of the regression functions.

However, it is difficult to estimate $\beta_k(t)$ directly in the same way as the estimation of the hazards rate. The estimation of the risk coefficients is based on a least-squares technique (68), whereas the estimation in the proportional hazards model is based on a partial likelihood or conditional likelihood. We estimate the cumulative risk function $B_k(t)$, defined as

$$B_k(t) = \int_0^t \beta_k(u)du, \text{ k=0, 1, ..., } p \tag{3.21}$$

To estimate the $B_k(t)$, a least-squares technique has been used by Aalen. To obtain the estimates, let us define an n by $(p+1)$ design matrix, $\boldsymbol{X}(t)$, as follows: For the i^{th} row of $\boldsymbol{X}(t)$, we set $\boldsymbol{X}_i(t) = Y_i(t)(1,\boldsymbol{Z}_j(t))$. That is, $\boldsymbol{X}_i(t) = Y_i(t)(1, Z_{j1}(t), ... Z_{jp}(t))$, if the i^{th} individual is a member of the risk set at time t (event has not happened, and the individual is not censored). If the i^{th} individual is not in the risk set at time t, i.e., the event of interest has already occurred, or the individual has been censored, then the $\boldsymbol{X}_i(t)$ contains a $(p+1)$ vector of zeroes.

Suppose that $\boldsymbol{I}(t)$ be the n by 1 vector with the i^{th} element equal to 1 if subject i dies at time t and 0 otherwise. Then, the least-squares estimate of the vector $B(t)$ is

$$\hat{\mathrm{B}}(t) = \sum_{T_i \le t} \left[X'(T_i)X(T_i)\right]^{-1} X'(T_i)I(T_i) \tag{3.22}$$

The variance-covariance matrix of $B(t)$ by Aalen's, is

$$\hat{V}[\hat{B}(t)] = \sum_{T_i \le t} [X'(T_i)X(T_i)]^{-1}[X'(T_i)I^D(T_i)X(T_i)]\left\{[X'(T_i)X(T_i)]^{-1}\right\}', \tag{3.23}$$

where $I^D(t)$ is an n by n diagonal matrix with diagonal elements equal to $\boldsymbol{I}(t)$. The estimator $B(t)$ only exists up to time, t, which is the smallest time at which $X'(T_i)X(T_i)$ becomes singular.

We know from equation (3.21), that the estimators $\hat{B}_k(t)$ estimate the integral of the regression function $\beta_k(t)$. A crude estimate of the regression functions can be found by examining the slope of the fitted $\hat{B}_k(t)$s. Better estimates of the regression function can be found by using kernel smoothing techniques, which we do not pursue here (4).

One benefit of fitting Aalen's additive model is to provide graphical evidence of the effect of a covariate over time, rather than to provide an additive covariate-adjusted survivorship function.

This graphical representation is the plot of the estimate of $B_k(t)$ versus t along with the upper and lower endpoints of a point-wise confidence interval. For the 95% confidence interval, Aalen uses the plot of

$$\widehat{B}_k(t) \pm 1.96\sqrt{\widehat{V}\left[\widehat{B}_k(t)\right]}. \tag{3.24}$$

3.2.2.2.2 Lin and Ying's (L-Y) Additive Hazards Model

We know from Aalen's additive hazards model the conditional hazards rate of an individual, given a set of covariates, and that the regression coefficients are the function of time. Lin and Ying proposed an alternative additive hazards regression model (19-21). The L-Y additive hazards model for the conditional hazards rate for j^{th} individual with covariate vector $\boldsymbol{Z}_j(t)$ is

$$h(t|\boldsymbol{Z}_j(t)) = h_0(t) + \sum_{k=1}^{p} \beta_k Z_{jk}(t) \ . \tag{3.25}$$

When all the covariate values are fixed at time 0, it is easy to estimate the regression coefficient, β_k, $k=1, 2, ..., p$. In fact, as opposed to the estimates in the Cox model, an explicit formula is available for the estimates and their variances. In contrast to Aalen's model, we can directly estimate the regression coefficients. We will focus only on the case where all the covariates are fixed at time 0 (19, 21).

To estimate the coefficient β_k, we have to construct the vector $\overline{Z}(t)$ and define p by p matrix $\boldsymbol{A}$, p by 1 vector $\boldsymbol{B}$, and p by p matrix $\boldsymbol{C}$ in terms of $\overline{Z}(t)$. $\overline{Z}(t)$ is the average value of the covariates at time t. i.e.,

$$\overline{Z}(t) = \frac{\sum_{i=1}^{n} Z_i Y_i(t)}{\sum_{i=1}^{n} Y_i(t)}. \tag{3.26}$$

The p by p matrix $\boldsymbol{A}$ is given by

$$\boldsymbol{A} = \sum_{i=1}^{n}\sum_{j=1}^{i}\left(T_j - T_{j-1}\right)\left(Z_i - \bar{Z}(T_j)\right)'\left(Z_i - \bar{Z}(T_j\right), \tag{3.27}$$

the *p*-vector $\boldsymbol{B}$ is given by

$$\boldsymbol{B}' = \sum_{i=1}^{n}\delta_i\left(Z_i - \bar{Z}(T_i\right), \tag{3.28}$$

and the *p by p* matrix $\boldsymbol{C}$ is given by

$$\boldsymbol{C} = \sum_{i=1}^{n}\delta_i\left(Z_i - \bar{Z}(T_j\right)'\left(Z_i - \bar{Z}(T_j\right). \tag{3.29}$$

Then, the estimate of $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots \beta_p)$ is

$$\hat{\beta} = \boldsymbol{A}^{-1}\boldsymbol{B}', \tag{3.30}$$

and the estimate of variance of $\hat{\beta}$ is

$$\hat{V} = \hat{V}(\beta) = \boldsymbol{A}^{-1}\boldsymbol{C}\boldsymbol{A}^{-1}. \tag{3.31}$$

3.3. Model Goodness of Fit

In the last section, methods for analyzing the semi-parametric Cox hazards model and the additive hazards models are described. In this section, the focus is on estimating and testing effects assuming that the model is correctly chosen. In fact, the use of diagnostic procedures for model checking is an essential part of the modeling process. A series of regression diagnostics procedures will be performed to assess the adequacy of the Cox hazards model based on residual plots and a couple of methods for additive hazards models.

3.3.1 Diagnosis of Cox Hazards Model

We are generally interested in examining four aspects of the hazards model (4).

First, for a given covariate, we would like to see the best functional form by which to explain the influence of the covariate on survival, adjusting for other covariates. Second, we wish to check the adequacy of the proportional hazards assumption. If the assumption is not valid, then one

may be appreciably misled by the results of the analyses. While we have looked at the use of a time-dependent covariate to check this assumption, a graphical check may provide some additional insight into any departure from proportionality. Third, we wish to check its accuracy for predicting the survival of a given subject. Here, we are interested in patients who had the events either too early or too late as compared to what the fitted model predicts. This will tell us which patients are potential outliers and, perhaps, should be excluded from the analysis. The final and fourth aspect of the model to be examined is the influence or leverage each subjects has on the model fit. This will also provide some information on possible outliers.

Regarding the availability of computer software, the adequacy of the Cox model can be checked in several ways. Many model-checking procedures are based on quantities known as residuals. In general, for assessing the fit of a Cox model, diagnosis would occur via the following residual plots: Cox-Snell residuals, martingale residuals, Arjas plot, deviance residuals, and partial residuals or Score residuals. The Cox-Snell residuals are used widely in the analysis of survival data (77). These residuals are useful for checking the overall fit of the final model. The martingale residual is useful for determining the functional form of a covariate to be included in a proportional hazards regression model (78, 79). To check the proportional hazards assumptions, we can use the Score residuals, Arjas plot, and plots based on estimates of the cumulative hazards from a stratified model. The deviance residual is used for examining the accuracy of the model for each individual. We estimate the difference between an estimate of $\boldsymbol{\beta}$ based on a full sample and one based on a sample with the observation omitted due to the problem of determining leverage points. Approaches to determining leverage points are based on the partial residual or score residual.

3.3.1.1 Cox-Snell Residuals

The most widely used diagnostic procedure in survival data analysis is the Cox-Snell residual, so called because it is a particular example of the general definition of residuals given by Cox and Snell (77). To check the goodness of fit by using this process, the estimated cumulative hazards rate of the residual has to be plotted against these residuals. This gives the cumulative hazards plot of the residuals. If the fitted survival model is satisfactory, then the plot will be a straight

line with unit slope and zero intercept, i.e., if the Cox model fits the data, then the plot should follow the 45^0 line. However, a plot that displays a systematic departure from a straight line, or yields a line that does not have an approximately unit slope or zero intercept, might suggest that the model needs to be modified in some way. Equivalently, a log-cumulative hazard plot of the residuals may be used.

3.3.1.2 Martingale Residuals

The martingale residual is a slight modification of the Cox-Snell residual (1, 4), which has been defined as follows. Suppose that for the j^{th} individual in the sample, we have a vector $\boldsymbol{Z}_j(t)$ of possible time-dependent covariates. Let $N_j(t)$ have a value at time t if this individual has experienced the event of interest and 0 if the individual has yet to experience the event of interest. Let $Y_j(t)$ be the indicator that individual j is under study at a time just prior to time t. Finally, let $\boldsymbol{\beta}$ be the vector of regression coefficients and $\hat{H}_0(t)$ be the Breslow estimator of the cumulative baseline hazards rate. Then, the martingale residual is defined as

$$\hat{M}_j = N_j(\infty) - \int_0^{\infty} Y_j(t)\exp\left[\beta' Z_j(t)\right] d\hat{H}_0(t)\,,\ \text{j = 1, …, n.} \tag{3.32}$$

When the data is right-censored, and all the covariates are fixed at the start of the study, then the martingale residual reduces to

$$\hat{M}_j = \delta_j - \hat{H}_0(T_j)\exp\left(\sum_{k=1}^{p} Z_{jk}\beta_k\right) = \delta_j - r_j\,,\ \text{j= 1,…., n,} \tag{3.33}$$

where r_j is the Cox-Snell residual of the j^{th} individual. This residual has the property $\sum_{j=1}^{n} \hat{M}_j = 0$. Also, for large samples, the $\hat{M}_j$ s are uncorrelated samples from a population with a zero mean.

The martingale residuals can be interpreted as the difference over time of the observed number of events minus the expected number of events under the assumed Cox model; that is, the martingale residuals are an estimate of the excess number of events seen in the data but not

predicted by the model. In this study, these residuals will be used to examine the best functional form for a given covariate using an assumed Cox model for the remaining covariates. Suppose that the covariate vector $\boldsymbol{Z}$ is partitioned into a vector $\boldsymbol{Z}^*$, for which we know the proper functional form of the Cox model, and a single covariate Z_1 for which we are unsure of what functional form of Z_1 to use. We assume that Z_1 is independent of $\boldsymbol{Z}^*$. Let $f(Z_1)$ be the best function of Z_1 to explain its effect on survival. Then,

$$H(t|\boldsymbol{Z}^*,Z_1) = H_0(t)\exp(\beta^* Z^*)\exp[f(Z_1)] \tag{3.34}$$

is the optimal Cox model.

To find $f(Z_1)$, we fit a Cox model to the data based on $\boldsymbol{Z}^*$ and compute the martingale residuals, $\widehat{M}_j$, $j=1,\ldots,n$. These residuals are plotted against the value of Z_1 for the j^{th} observation. A smoothed fit of the scatter diagram is used. The smoothed-fitted curve gives an indication of the function f. If the plot is linear, then no transformation of Z_1 is needed. If there appears to be a threshold, then a discretized version of the covariate is indicated (4). If the plot is neither linear nor threshold, then we should use a transform, such as log, square or Z log Z.

Note that the martingale residuals are based on the fact that the process

$$M_j(t) = N_j(t) - \int_0^t Y_j(u)\exp\left[\beta' Z_j(u)\right] dH_0(u)$$

is a martingale when the proportional hazards model is correctly specified. The martingale residuals are obtained by substituting the estimates of $\boldsymbol{\beta}$ and $H_0(t)$ in this expression and evaluating the estimated martingale at time $t = \infty$.

A graphical plot of these residuals can be obtained by plotting martingale residuals versus survival time, index, the rank order of the survival times or explanatory variables. These residuals highlight individuals who, on the basis of the assumed model, have died too soon or lived too long. Large negative residuals will correspond to individuals who have a long survival time but covariate values that suggest they should have died earlier. On the other hand, a residual close to unity, the upper limit of a martingale residual, will be obtained when an individual has

an unexpectedly short survival time. An index plot of the martingale residuals will highlight individuals whose survival time is not well fitted by the model. Such observations may be termed outliers. The data from individuals for whom the residual is unusually large in absolute value will need to be subjected to further scrutiny (4).

3.3.1.3 Arjas Plots

To check the proportional hazards assumption, another method is use of the Arjas plot (50). By using this plot, one can also check the overall fit of the proportional hazards regression model. Let us suppose that a Cox model has been fitted with a covariate vector $\boldsymbol{Z}^*$ of p variables, and we wish to check if an additional covariate Z should be included in the model or if the new covariate has proportional hazards after adjustment for covariate $\boldsymbol{Z}^*$. Let $\hat{H}(t|\boldsymbol{Z}^*)$ be the estimated cumulative hazards rate for the j^{th} individual in the sample at time t. If the covariate Z_1 is continuous, then we have to group the values into K classes. At each event time for each level of Z_1, we compute the "total time on test (TOT)" of the estimated cumulative hazards rates up to this time and the observed number of events that have occurred up to this time. That is, at each event time t_i, we compute

$$TOT_g(t_i) = \sum_{Z_1 j=g} \hat{H}(\min(t_i, T_j)|\boldsymbol{Z}^*_j$$

and

$$N_g(t_i) = \sum_{Z1\, j=g} \delta_j I(T_j \le t_i).$$

If the covariate Z_1 does not need to be in the model, then, for each level of Z_1, the quantity $N_g(t_i) - TOT_g(t_i)$ is a martingale residual, and a plot of $N_g(t)$ versus $TOT_g(t_i)$ should be a roughly 45^0 line through the origin. Departures from this pattern provide evidence of a lack of fit of the model (4).

3.3.2 Diagnosis of Additive Hazards Model

There are several methods for testing the goodness of fit for the Cox model. However, because of the lack of software availability, the residuals plot for the additive models is limited. The Arjas

Plot and martingale residuals plot were used to assess the adequacy of the fit of the additive model (22, 71, 72, 80).

3.3.2.1 Arjas Plot

The Arjas Plot simply compares the observed and expected number of events as a function of time. In this method, the observed number of events is plotted against the expected number of events for various subgroups of covariate values.

Consider the additive model at time *t*, when the covariate ***Z(t)*** is time-independent. Then, based on the Klein and Moeschberger (4), the estimated cumulative hazard rate is

$$\hat{H}(t|\boldsymbol{Z}) = \hat{H}_0(t) + \sum_{k=1}^{p} B_k(t) Z_k ,$$

where $\hat{B}_k(t)$, are the least-squares estimates, *k=0,1,….p.*

Suppose that $N_j(t) = 1$ at time *t* if the individual *j* has been observed to experience the event of interest before or at time *t*, and $N_j(t) = 0$ if the individual has yet to experience the event of interest (until the event of interest has occurred). If the individual is censored, $N_j(t)$ will stay at 0.

To the check the goodness of fit, groups of individuals who might be expected to deviate from the proposed model were selected. Suppose there are q such groups. In the Arjas plot, we plot the sum of $N_j(t)$ over the g^{th} group against the values of $\hat{H}(t|\boldsymbol{Z}_j(t))$ summed over the group. For each group at each event time, a point would be produced, and the points are connected. If the model fits, then this plot should look like a 45^0 line through the origin for each group (4, 72).

3.3.2.2 Martingale Residuals Plot

The difference between $N_j(t)$ (the observed number of events) and $\hat{H}(t|\boldsymbol{Z}_j(t))$ (the expected number of events under the additive model) for the j^{th} individual is defined as the martingale residuals (81)

$$\hat{M}_j(t) = N_j(t) - \hat{H}[t/Z_j(t)], \; j=1, 2, \ldots, n.$$

These residuals are defined for $t \leq \tau$, where, τ is the maximal value of t for which the matrix $\boldsymbol{Z(t)}$ is a nonsingular matrix. The sum of these residuals over all n observations is zero at any event time. The martingale residuals plots give a picture of how accumulated hazard compares to events that occurred over time. The goal of the martingale residuals is to compare these residuals for a subgroup (suppose for the g^{th} group) within a dataset with different covariate values to find out if the model is valid for all subgroups. The martingale residual at time t for a given group is the sum of the martingale residuals at time t over the members of the group. Then, these sums are plotted against time. If the model fits the data, then the plotted curves should be close to zero (4, 72).

To determine if the martingale residual process is too far from zero for a model to be acceptable, we need to compute an estimate of the variance of the martingale residuals process. Let Q be the *n by q* matrix, which has as its j^{th} row a 1 in the column of the group in which the j^{th} observation belongs and 0 in other columns. Let $M_{res}(t)$ be the vector $[\widehat{M}_1(t),\ldots\ldots\ldots,\widehat{M}_n(t)]'$. The Q-vector of martingale residuals summed over groups is given by

$$M_{res}(t) = Q'M'.$$

Let D_i be the *n by n* matrix of all zeros at an event time t_i, except for the diagonal elements corresponding to individuals who die at time t_i, where the diagonal element has the value *1*. Let $\boldsymbol{X_i}$ be the *n by (p+1)* matrix whose j^{th} row is zero if the j^{th} individual is not at risk at time t_i and has the value $(1, Z_1(t_i), \ldots\ldots Z_p(t_i))$ if the individual j is at risk. Finally, let $\boldsymbol{I}$ be the *n by n* identity matrix. Then the covariance matrix for $M_{res}(t)$ is

$$Cov[M_{res}(t)] = \sum_{t_i \leq t} Q'[I - X_i(X_i'X_i)^{-1}X'_i]D_i[I - X_i(X_i'X_i)^{-1}X'_i]'Q.$$

The confidence interval for $M_{res}(t)$ can be calculated as

$$M_{res}(t) \pm z_{1-\alpha/2}(Cov[M_{res}(t)])^{1/2}.$$

A plot of $M_{res}(t)$ against time for various groups with 95% point wise confidence intervals construct using above equation is used to assess model fit. Both types of plots can be used to assess the fit of the additive mode. The Arjas plot gives a clearer indication of lack of model fit

than the martingale residuals plot. However, the martingale residuals plot, which explicitly involves time, gives a clear indication of where the problems may be arising from in the fit of the model.

3.4 Software

SAS version 9.2 was used for most of the analysis in this study. Other software that has been used in this study was R, MS-Word and Excel. For Aalen's additive hazard model, SAS Macro was obtained from the Statistical Software at the Medical College of Wisconsin (82). A SAS macro for the L-Y additive model was obtained from Dr. Xu Zhang in the Department of Mathematics and Statistics at Georgia State University (22). All of the SAS programs used for this thesis are provided in the Appendix.

3.5 Summary

A brief summary of the multiplicative and additive hazards models is given in Table 3.5.1

Table 3.5.1: Comparison of multiplicative and additive hazards models

Characteristics	**Multiplicative model**	**Aalen's additive model**	**L-Y additive model**
Basic model	$h(t\|\boldsymbol{Z})=h_0(t)exp\ (\beta_1 Z_1+ ..+ \beta pZp)$	$h(t\|\boldsymbol{Z})=h_0(t)+ \beta_1(t)Z_{j1}(t)+. .+ \beta_p(t)Z_{jp}\ (t)$	$h(t\|\boldsymbol{Z})=h_0(t)+\beta_1 Z_{j1}(t)+..+\beta_p Z_{jp}\ (t)$
Covariates	Covariates act in a multiplicative manner on an unknown baseline hazard rate.	Covariates act in additive manner on an unknown baseline hazard rate.	Covariates act in additive manner on an unknown baseline hazard rate.
Coefficients ($\boldsymbol{\beta}$)	Coefficient is constant, but it may be time-dependent.	$\boldsymbol{\beta}(t)$ might be dependent on time *t*.	Coefficient $\boldsymbol{\beta}$ is constant.
Interpretation of ($\boldsymbol{\beta}$)	Cox's model measures the relative risk or hazard ratio due to the effect of a covariate in relative terms.	Aalen's model measures the additional risk due to the effect of a covariate in absolute terms.	The L-Y model measures the excess risk due to the effect of a covariate in relative term.
Software	Algorithms for the estimation of $\boldsymbol{\beta}$ in the Cox model are available in many statistical packages. The	Algorithms for the estimation of $\boldsymbol{\beta}$ are not readily available in SAS. A SAS macro to fit Aalen's additive model is	Algorithms for the estimation of $\boldsymbol{\beta}$ are not readily available in SAS, S-Plus or R. A SAS macro

	procedure PHREG in SAS and Coxph in S-Plus provides estimates of $\boldsymbol{\beta}$, its standard error and the Wald, score and likelihood ratio test of the global hypothesis of no covariate effects. Also, STATA and R are used to estimate the effects.	available in www.mcw.edu. These days, R software can be used, which has a function named aareg to fit the Aalen's model.	is available at Georgia State University.
Goodness of fit	There are several methods to check the adequacy of the model, such as Cox-Snell, Martingale*, Deviance, Schoenfeld, Score residuals, Arjas Plot* etc.	For checking the adequacy of the model, Arjas Plot and Martingale plots are available, but not in SAS.	For checking the adequacy, Arjas plots and Martingale plots used.

* Martingale residual and Arjas plot has been used in this study.

4. Application to Injury Data

To compare the performance of the Cox hazards model and the additive hazards models, the TLR injury prevention study was used in this thesis. The TLR data have been collected by Timothy R. Black for his M.Sc. thesis (83) and further refined for this thesis. The data have been originally collected from a retrospective, pre-post intervention design utilizing a non-randomized, historical control group. In brief, these administrative data were obtained from the OH&S databases of the Saskatoon Health Region (SHR) and the Regina Qu'Appelle Health Region (RQHR). In this study, SHR was considered as an intervention group, and RQHR was considered as a control group because no training regarding patient handling had been provided. Three hospitals from each region were considered. The hospitals in the SHR were: Royal University Hospital (RUH), Saskatoon City Hospital (SCH) and Parkridge Centre (PRC). The hospitals in the RQHR were: Regina General Hospital (RGH), Pasqua Hospital (PH) and Wascana Rehabilitation Centre (WRC). All of the hospitals were categorized into three groups to investigate the effects of the intervention. Accordingly, RUH and RGH were considered as the largest hospitals, and SCH and PH were considered to be medium-sized hospitals. A Transfer, Lifting and Repositioning (TLR) program containing engineering and administrative ergonomic controls was implemented from 2002-2005 in three hospitals in Saskatoon, Saskatchewan, Canada. The TLR program was implemented at different times by the different hospitals in the SHR. The time frame was as follows:

SCH: September, 2002 – June, 2004

PRC: September, 2002 – September, 2004

RUH: January, 2005 – December, 2005

For this study, injury data was collected in the period from January 1, 1999 to December 1, 2006. The data collection time frame for different hospitals was as follows:

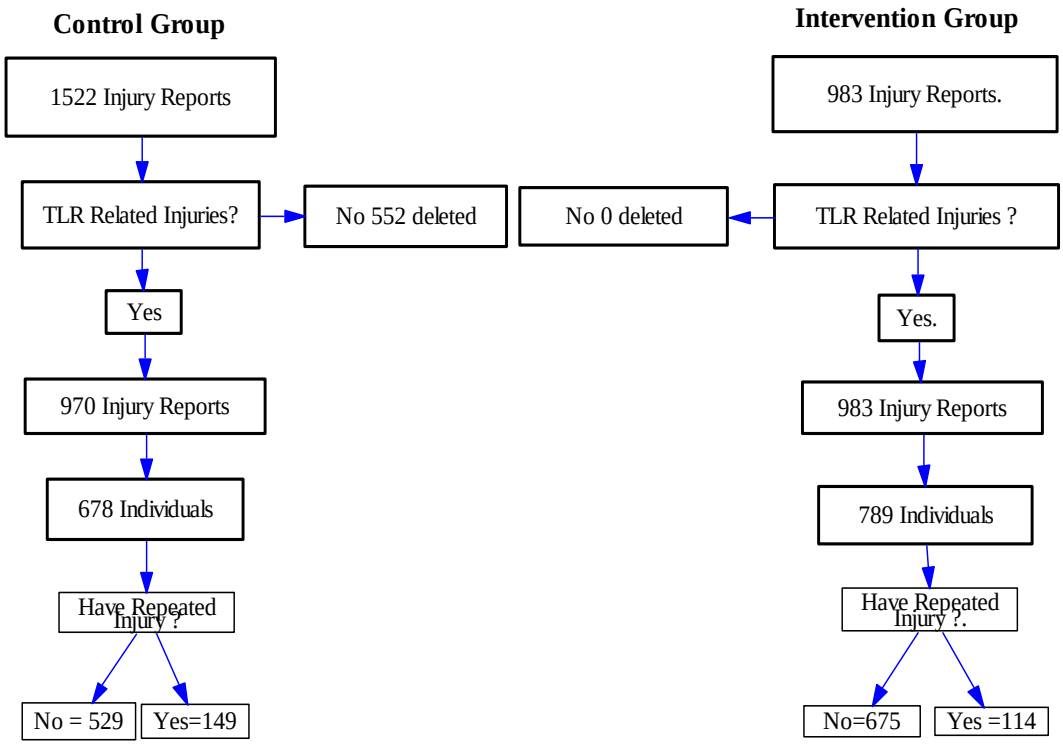
WRC and PRC: January 1, 1999 – September, 2005

PH and SCH: January 1, 1999 – June, 2005

RGH and RUH: January 1, 1999 – December, 2006

Time-loss and non-time-loss injury data, lost time days, and claims costs were collected from the intervention group (three hospitals) and the control group (three hospitals) for corresponding time periods one year pre- and one year post-intervention. The covariates that were selected were age, sex, date of birth (DOB), date of injury, body parts and occupation. This study considered repeated MSI injuries as the outcome variables. Because these administrative data contained only information about the injury and there were no identification numbers, it was hard to identify repeated injuries because the data contained the number of cases and not the number of individuals. The DOB, sex, occupation and body parts were used to identify repeated MSIs. If the DOB and sex were the same, then these may have been the same individual. However, if the occupation was different for a short period of time, then they were considered to be different individuals. In the last step, we checked the body parts involved. If the body parts were the same, then we considered the injury to be a repeated injury; otherwise, there was no repeated injury involved. Figure 4.1.1 is the data extraction flowchart

Figure 4.1.1: Data extraction flowchart



In this thesis, the outcome event was the TLR-related repeated injury. The survival time was calculated based on the time to the TLR-related repeated injury for event cases or time to last follow-up for censored cases depending on health regions. The univariate analysis was adopted to select the covariates. Then, the selected covariates were included for the additive hazards model and the multiplicative hazards model.

To do the survival analysis, we had to assess the repeated injuries from the 1,467 individuals who had injured. We also needed to ascertain the censor indicator. If any TLR injury occurred before January 1, 1999 and after December 1, 2006, it was censored. Additionally, any injury that was not related to a TRL injury was censored. Furthermore, if the identified individuals did not have a 2^{nd} injury till the end of data collection time frame, they were censored. In this study, only the first and 2^{nd} injuries were considered for calculating the survival time in months. Based on the analysis in this study, there were a total of 263 individuals who had a repeated injury. Among the repeatedly injured individuals were 114 from the intervention group and 149 from the control group. The survival time was calculated by subtracting the first injury date from the second injury date and converting it into months.

5. Results

5.1 Demographic Characteristics

5.1.1 Intervention/Control Group

Individual injury data were pooled for the intervention and control groups. As shown in the following Table, we have 789 individuals in the intervention group and 678 individuals in the control group. Compared to the control group, the intervention group contained more injured individuals.

Table 5.1.1: Number of individuals in the intervention and control groups

	Intervention group	Control group	Total
# of injured individuals	789 (54%)	678 (46%)	1467 (100 %)

5.1.2 Age

From each of the hospitals, injured individuals' ages were pooled for the intervention and control groups. From the available data, the means and standard deviations were calculated. We noticed from the following Table that the mean age for both groups was similar.

Table 5.1.2: Age of injured individuals

	Intervention group (N = 789)	Control group (N = 678)	Total (N = 1467)
Mean	41.09	38.99	40.04
SD*	10.08	10.14	10.11

* SD = Standard Deviation

5.1.3 Gender

Data regarding the gender of injured individuals was available for both groups. According to the available data, sex ratios were calculated and indicated in the following Table. The sex ratio was similar for both groups (p-value=0.103).

Table 5.1.3: Gender of injured individuals

	Intervention Group (N = 789)	Control Group (N = 678)	Total (N = 1467)
# of Female	734 (50 %)	615 (42 %)	1349 (92 %)
# of Male	55 (4 %)	63 (4 %)	118 (8 %)
Sex ratio F/M	13.35	9.97	11.54

* F: Female, M: Male

5.1.4 Hospital Size

The number of injured individuals at the different types of hospitals was calculated from the available data. Based on the analysis, that the result show that, overall, there were significantly different numbers of injuries among the differently sized hospitals between the two groups (p-value<0.0001).

Table 5.1.4 : Hospital size of injured individuals

	Intervention group (N = 789)	Control group (N = 678)	Total (N = 1467)
Large	379 (26%)	260 (18%)	639 (44%)
Medium	230 (16%)	182 (13%)	412 (28%)
Small	180 (12%)	236 (16%)	416 (28%)

5.1.5 Occupation

According to the available information, the majority of injured employees were nurses followed by attendants. Among the two groups, the intervention group had more injuries than the control group. Only licensed practicum nurses (LPN) and others (e.g., therapists, technicians, unit supporters, and paramedics) had higher injuries in the control group than the intervention group (p-value<0.001).

Table 5.1.5: Occupation of injured individuals

	Intervention group (N = 789)	Control group (N = 678)	Total (N = 1467)
Nurses: RN/GDN	453 (30%)	372 (25%)	825 (56%)
LPN	105 (7%)	112 (8%)	217 (15%)
Attendants	158 (11%)	37 (2%)	195 (13%)
Nurse-Aide/Attendants	13 (1%)	67 (5%)	80 (5%)
Clerks/Unit Assistants	35 (2%)	2 (0.14%)	37 (2%)
Others	25 (2%)	88 (6%)	113 (9%)

*Others include therapists, technicians, unit supporters, and paramedics.

* RN: Registered Nurse, GDN: General Duty Nurse, LPN: Licensed Practicum Nurse

5.1.6 Body Parts

According to the available information, injured body parts were identified based on the injured individual. We noticed that most of the individuals in the intervention group had back injuries followed by shoulder injuries and then all other body parts. However, in the control group, the 2nd highest injury involved all other body parts. Nevertheless, the control group had more injuries in the back and all other body parts (e.g., abdomen, chest, and face) than did the intervention group (p-value<0.001).

Table 5.1.6: Body parts of injured individuals

	Intervention group (N = 789)	Control group (N = 678)	Total (N = 1467)
All back injury (except neck)	413 (28%)	243 (16%)	656 (45%)
Shoulder	93 (6%)	30 (2%)	123 (8%)
Neck	41 (3%)	56 (4%)	97 (7%)
Multiple sites	82 (6%)	8 (1%)	90 (6%)
Extremity	77 (5%)	3 (0.2%)	80 (5%)
All other body parts	83 (6%)	338 (23%)	421 (29%)

* All other body parts include abdomen, chest, and face, etc.

5.1.7 Repeated Injury

In this study, there were 1,467 injured individuals. Among them, 263 individuals had repeated injures. The demographic information of the repeated injured individuals has been show in the Table 5.1.7.

Table 5.1.7: Demographic information of repeated injured individual

			Intervention group	Control group
Number of repeated injured individuals			114 *	149 **
Age			41.70 ± 8.89$	37.95 ± 9.90$
Gender:		Female	104 (91%)	134 (90%)
		Male	10 (9%)	15 (10%)
Hospital Size:		Large	58 (51%)	41 (28%)
		Medium	36 (32%)	54 (36%)
		Small	20 (17%)	54 (36%)
Occupation:	Nurses	RN/GDN	74 (65%)	100 (67%)
		LPN	16 (14%)	21 (14%)
	Attendants		21 (18%)	3 (2%)
	Nurse-Aide/Attendants		0 (0%)	12 (8%)
	Clerks/Unit Assistants		3 (3%)	0 (0%)
	Others+		0 (0%)	13 (9%)
Body Parts:	All back injury (except neck)		86 (75%)	88 (59%)
		Shoulder	10 (9%)	5 (3%)
		Neck	3 (3%)	5 (3%)
		Multiple sites	11 (10%)	0 (0%)
		Extremity	3 (3%)	0 (0%)
		All other body++	1 (1%)	51 (34%)

* 114 repeated injured individual from 789 injured individuals in intervention group

** 149 repeated injured individual from 678 injured individuals in control group

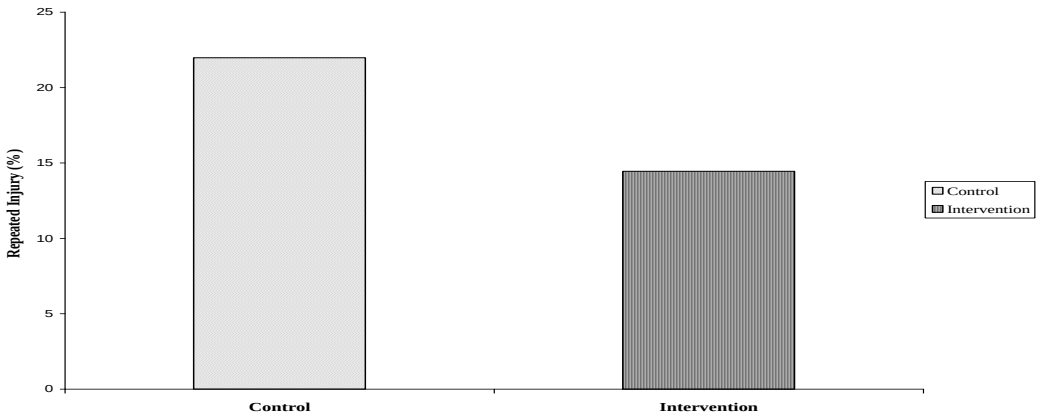
$ For Age: Mean ± Standard Deviation;

* RN: Registered Nurse, GDN: General Duty Nurse, LPN: Licensed Practicum Nurse

+Others include therapists, technicians, unit supporters, and paramedics; ++ All other body parts include abdomen, chest, and face, etc.

The proportion of repeated injuries for each group, occupation, and body part is presented in the following histograms (Figure 5.1.7.1 – 5.1.7.4).

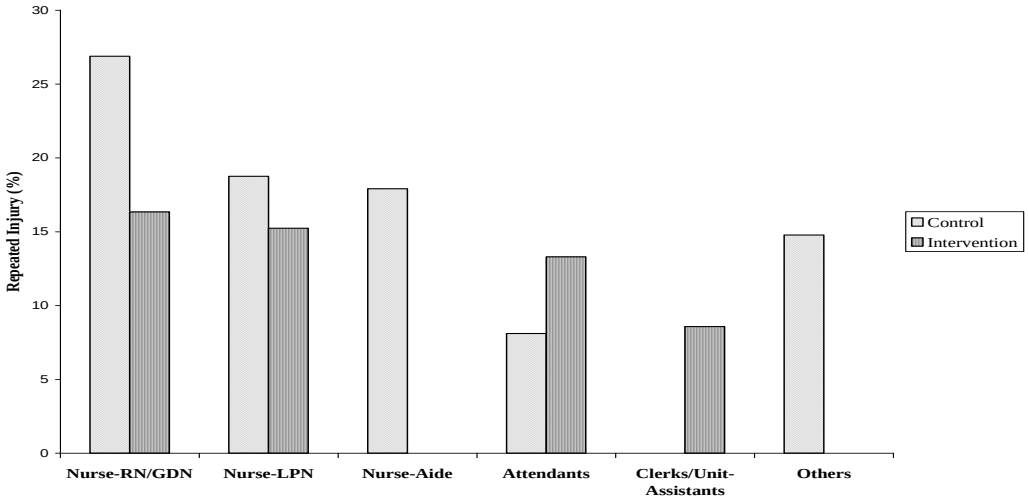
Figure 5.1.7.1: Proportion of repeated injury by group



Control: 149 (678)* Intervention: 114 (789) *

*Number of Repeated Injury (Number of Total Individual Injury)

Figure 5.1.7.2: Proportion of repeated injury by occupation

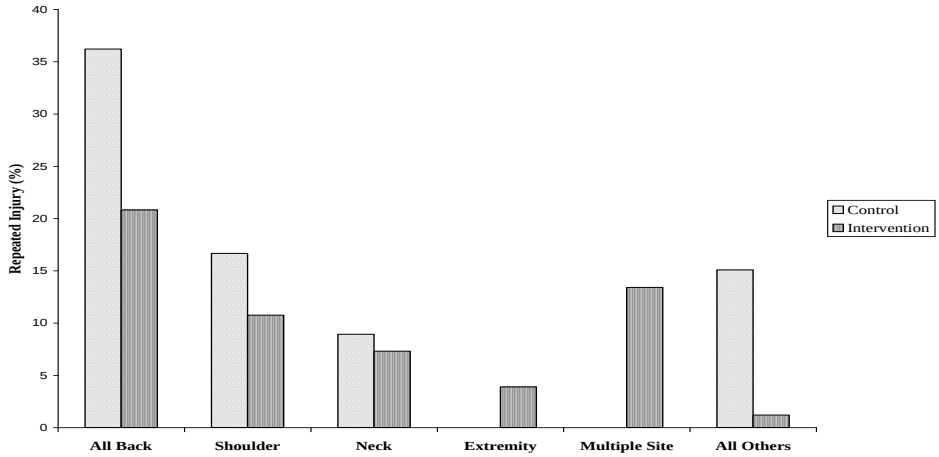


Control: 100(372)* 21(112) 12(67) 3(37) 0(2) 13(88)
Intervention: 74(453)* 16(105) 0(13) 21(158) 3(35) 0(25)

* Others includes physical therapists, occupational therapists, recreational therapists, paramedics, operation room technicians, and dispatch porters etc. The number of individuals belonging to each category is in parentheses.

*Number of Repeated Injury (Number of Total Individual Injury)

Figure 5.1.7.3: Proportion of repeated injury by body parts

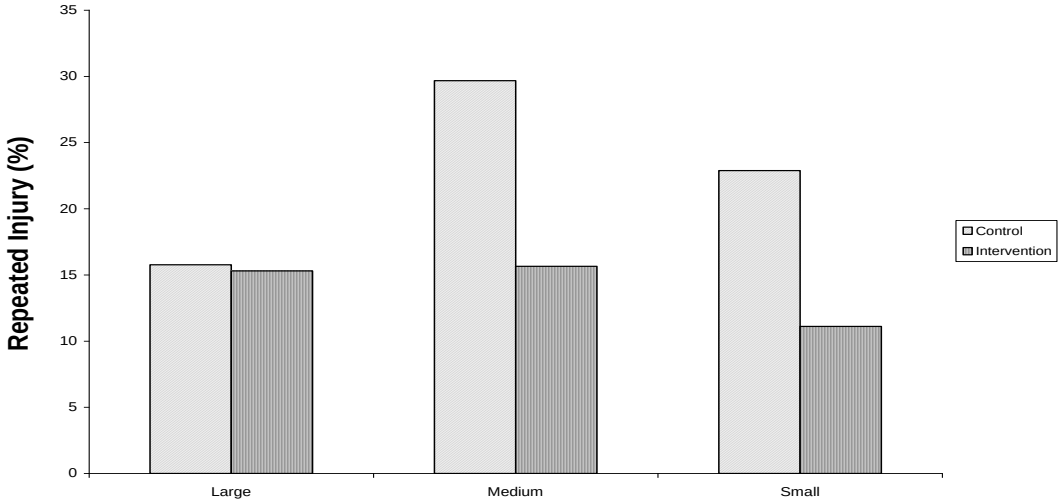


	All Back	Shoulder	Neck	Extremity	Multiple Site	All Others
Control:	88(243)*	5(30)	5(56)	0(3)	0(8)	51(338)
Intervention:	86(413)*	10(93)	3(41)	3(77)	11(82)	1(83)

* All Others includes abdomen, chest, face, head, arm, and eye, etc. The number of the individuals belonging to each category is in parentheses.

*Number of Repeated Injury (Number of Total Individual Injury)

Figure 5.1.7.4: Proportion of repeated injury by hospital size



	Large	Medium	Small
Control:	41(260)*	54(182)	54(236)
Intervention:	58 (379)*	36(230)	20(180)

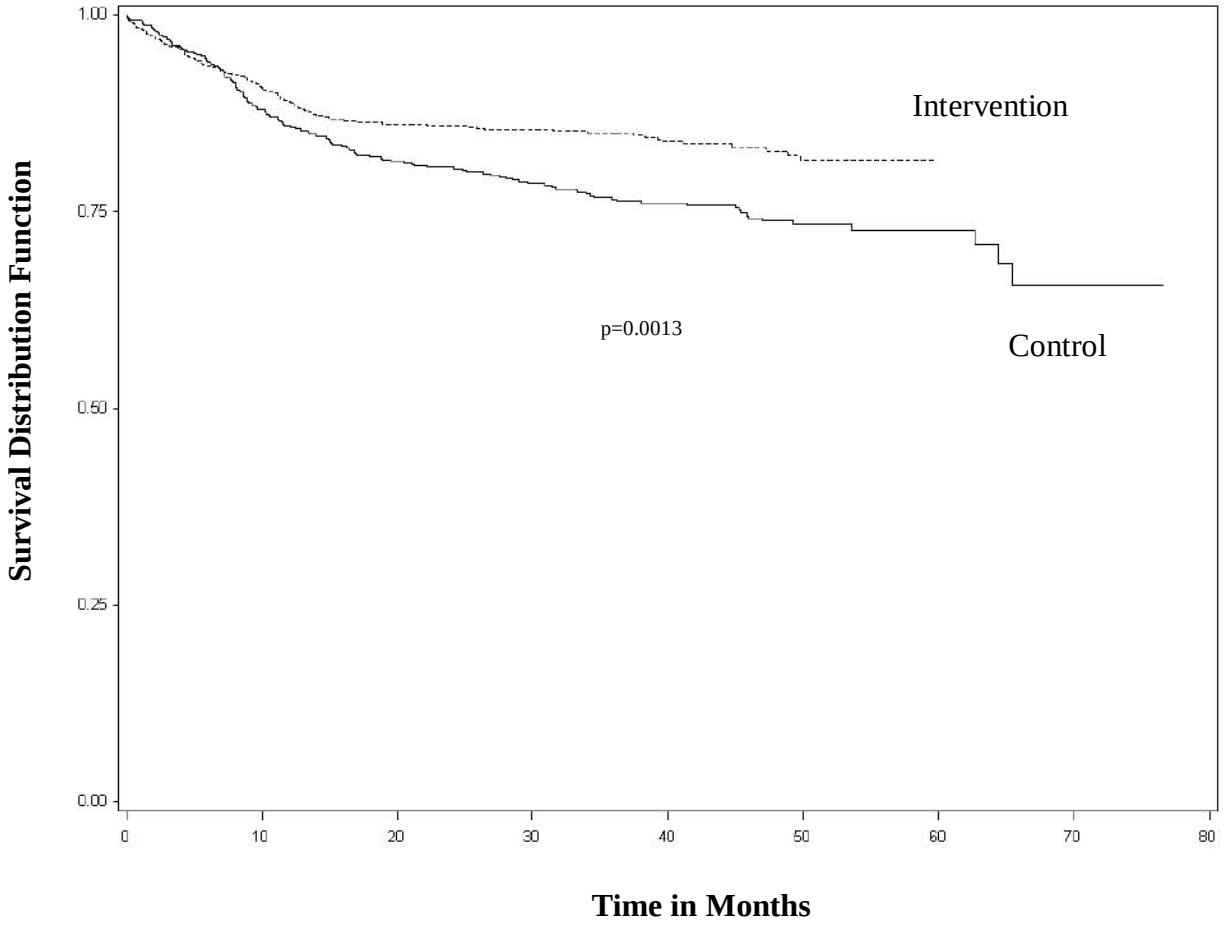
*Number of Repeated Injury (Number of Total Individual Injury)

5.2 Non-Parametric Model

5.2.1 Kaplan-Meier Method

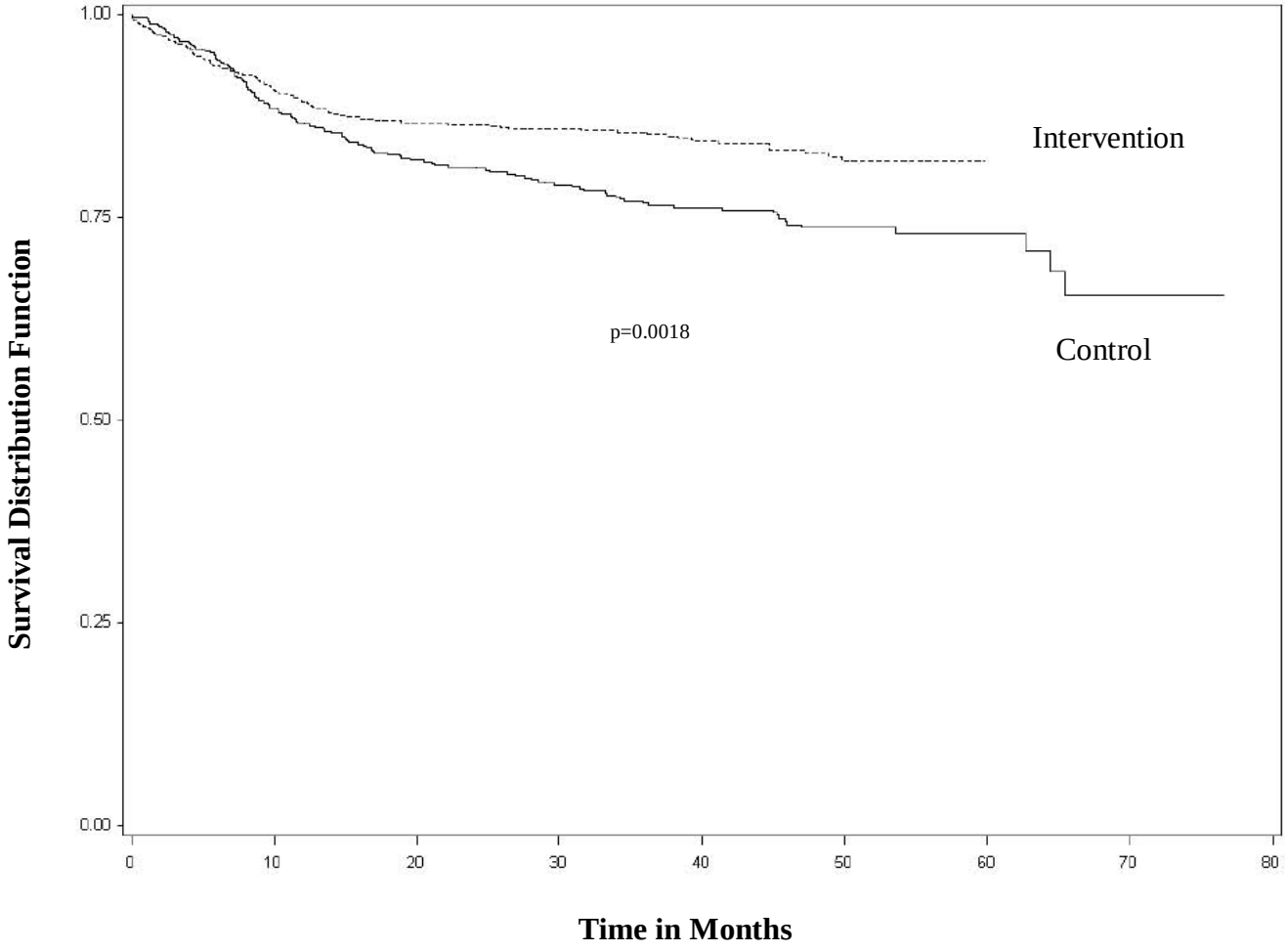
Using the KM analysis, the following results were obtained. The first analysis was performed to assess the overall difference among the intervention and control groups (Figure 5.2.1.1). This result indicated that before (approximately) 8 months, the two survival curves were close to identical. After 8 months, the intervention group had a higher probability of survival as compared to the control group. The log-rank and Wilcoxon test shows that there was a significant difference in survival function between the intervention and control groups (p-value=0.0013 and 0.0063, respectively).

Figure 5.2.1.1: Survival function by group



Because more than 90% of the individuals in this study were female, the impact of the treatment among the females was investigated. The KM analysis for females is given below (Figure 5.2.1.2). Females trended with the treatment group. Also, among females, the log-rank and Wilcoxon tests showed that there were significant differences in survival function between two groups (p-value=0.0018 and 0.0089, respectively).

Figure 5.2.1.2: Survival function among females by group



To see the effect of the TLR training intervention among the various hospital sizes, the analysis was done on three sizes of hospitals: Large (Figure 5.2.1.3), Medium (Figure 5.2.1.4) and Small (Figure 5.2.1.5). The KM estimates are presented in the following Figures.

From these three analyses (Figures 5.2.1.3, 5.2.1.4, and 5.2.1.5) based on the hospital size, the large hospitals showed differently than medium and small sized hospitals. The medium and small sized hospitals trended similarly to the treatment group. The *p*-values for the medium sized hospitals given by the log-rank and Wilcoxon tests are 0.0303 and 0.028, respectively. The p-values for the small hospitals given by the log-rank and Wilcoxon tests are 0.0042 and 0.0087, respectively. However, the large hospitals showed no significant intervention effect by the log-rank and Wilcoxon tests (0.5439 and 0.7418, respectively).

Figure 5.2.1.3: Survival function among large sized hospitals by group

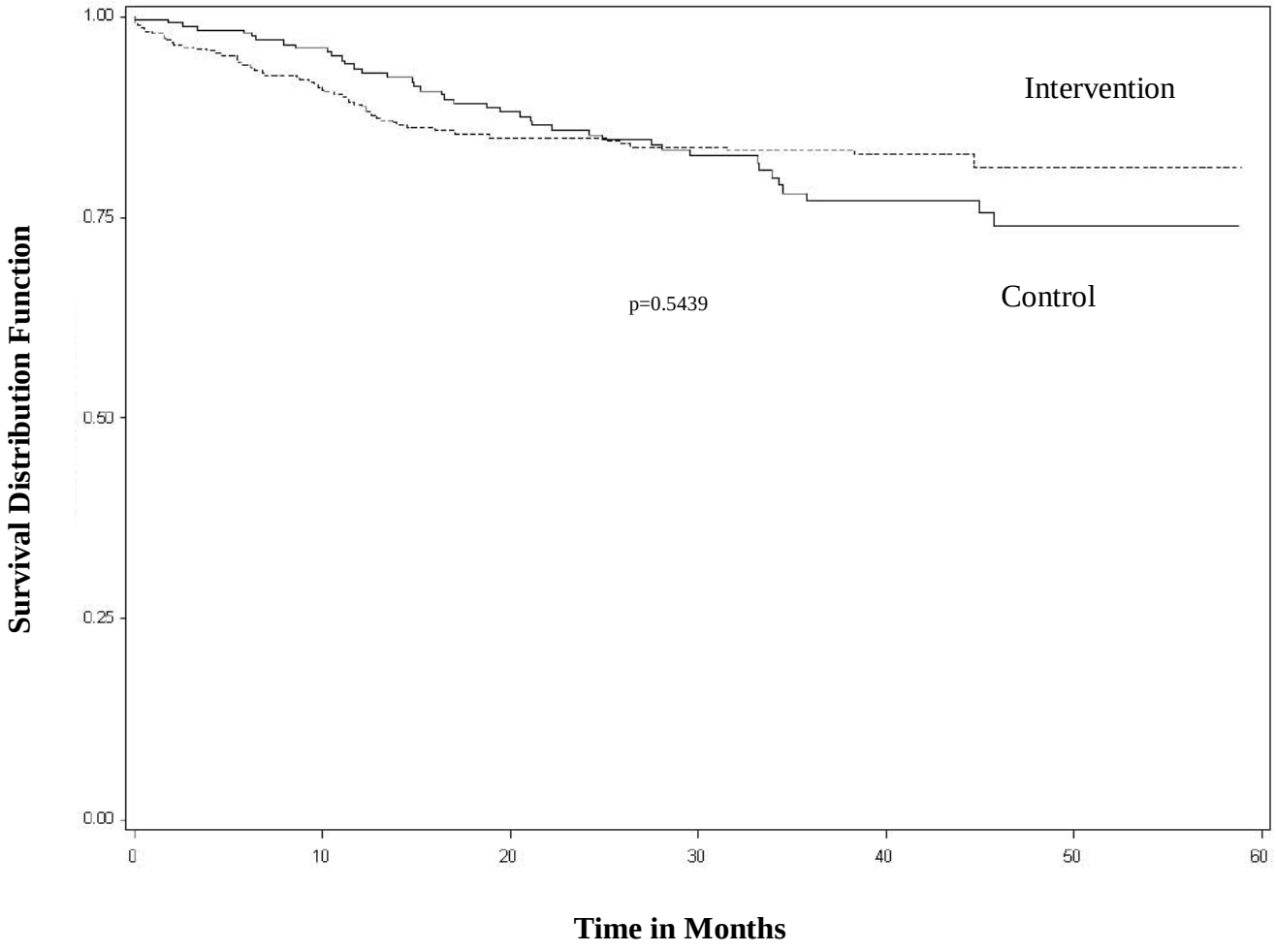


Figure 5.2.1.4: Survival function among medium sized hospitals by group

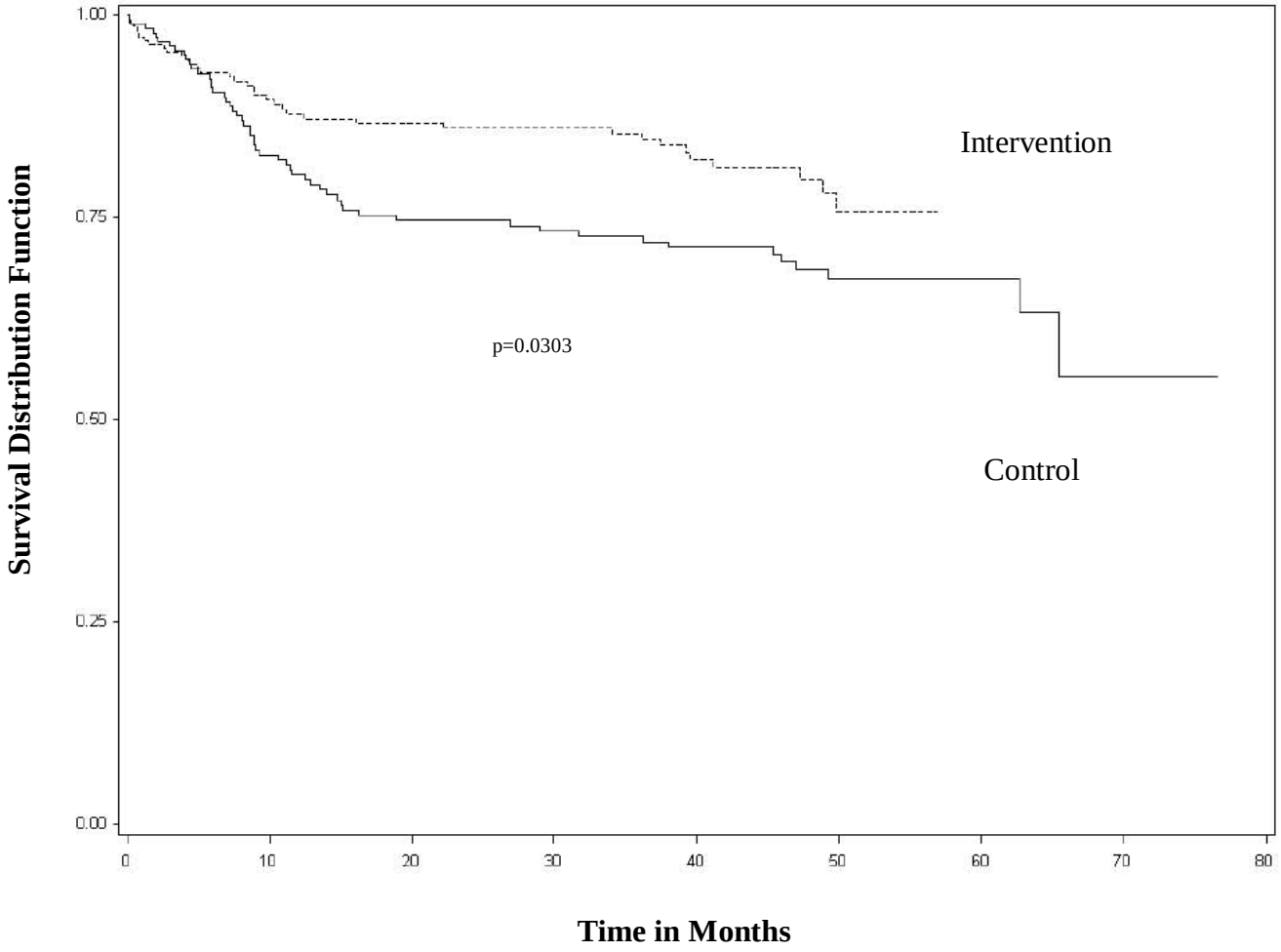
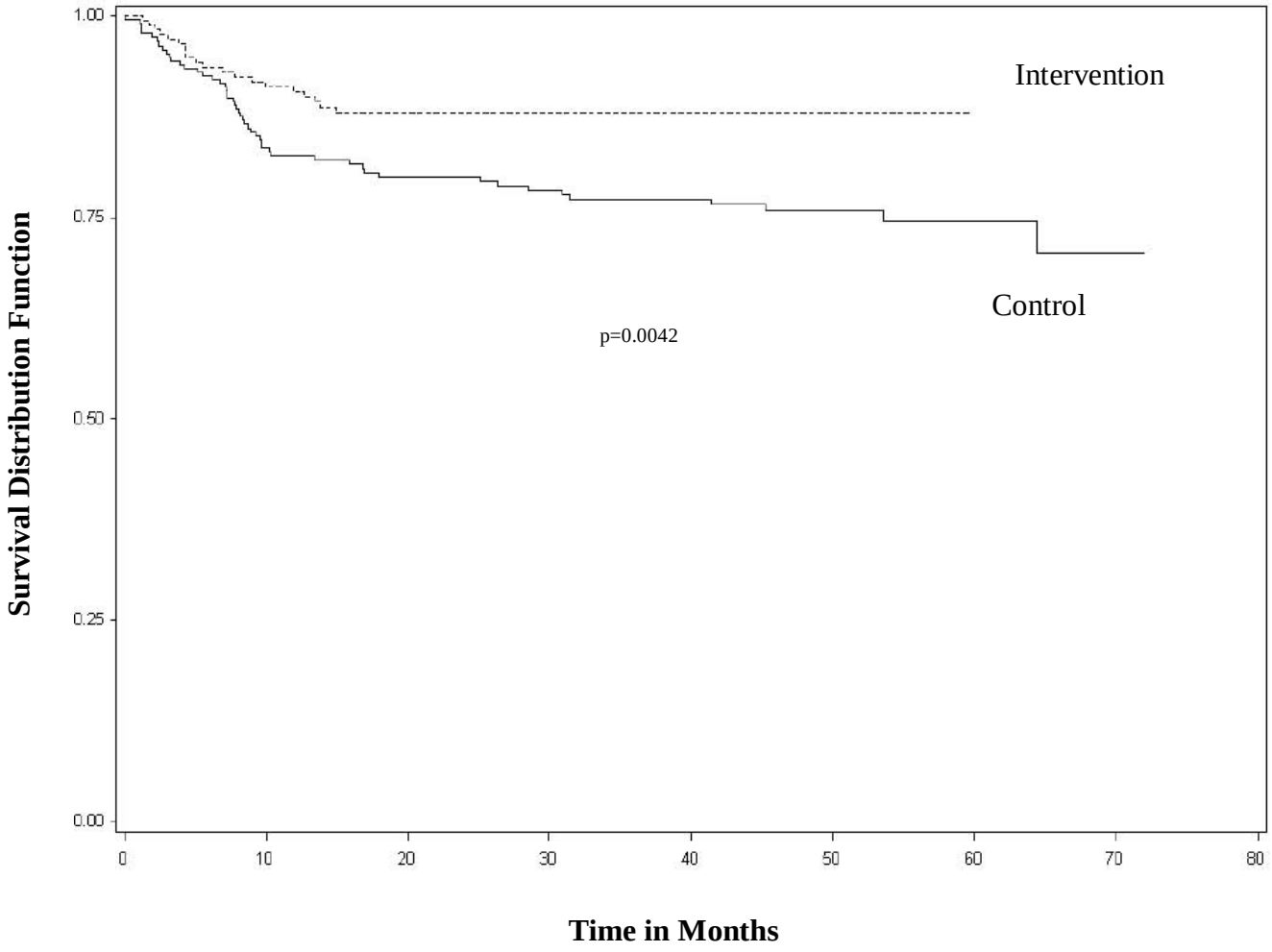
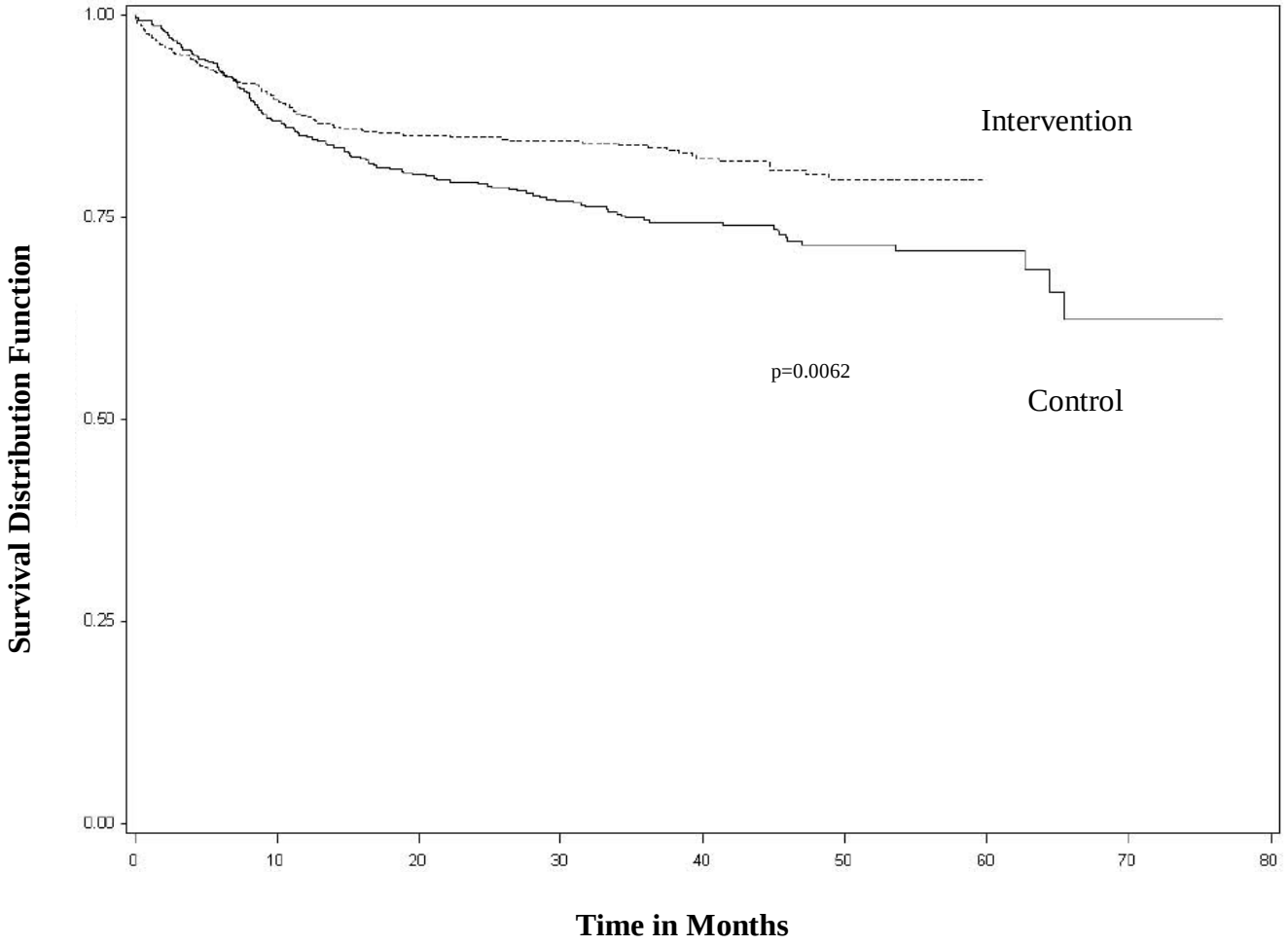


Figure 5.2.1.5: Survival function among small sized hospitals by group



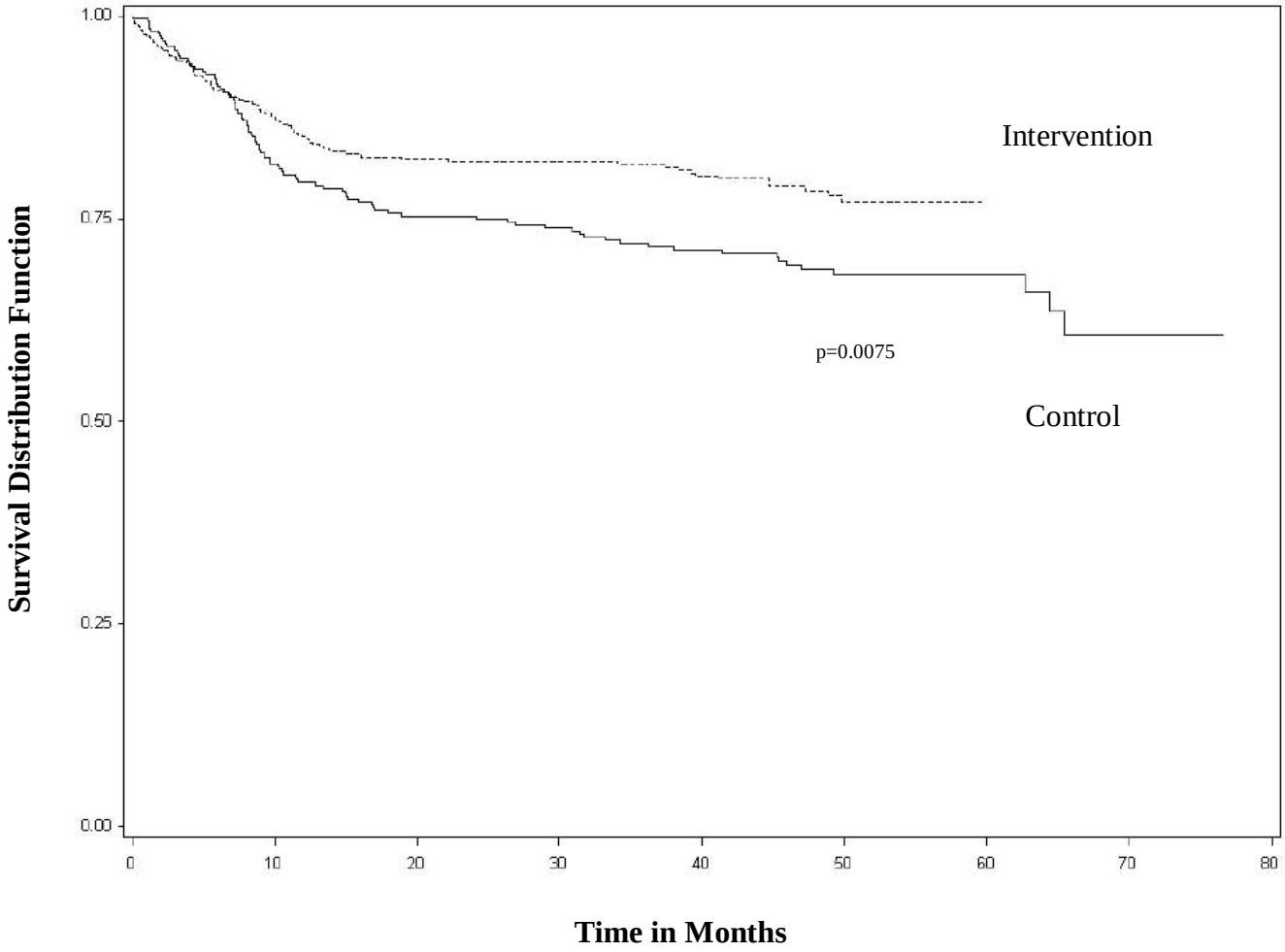
Because the majority of the individuals in this study had an occupation related to nursing, only the nursing and nursing aide occupations were considered for analysis. The KM analysis provided similar results (Figure 5.2.1.6). Also, the log-rank and Wilcoxon tests revealed a significant difference between the intervention and control groups (p-values=0.0062 and 0.0216, respectively).

Figure 5.2.1.6: Survival function among nurses/nursing aides by group



When body parts were considered, the same results pattern as that of the treatment group was shown in Figure 5.2.1.7. Because the majority of injuries were in the back, neck and shoulders, injuries to these three body parts were considered for the analysis. Test results showed a significant difference between the intervention and control groups (p-value=0.0075 and 0.0166, respectively for the log-rank and Wilcoxon tests).

Figure 5.2.1.7: Survival function among back, neck and shoulder by group



5.3 Cox Multiplicative Hazards Model

5.3.1 Cox Univariate Hazards Model

To consider the model, the covariate was selected for the univariate analysis of the multiplicative hazards model (Table 5.3.1). Age, gender, group (intervention versus control), occupation, body parts and hospital size (large and small) were selected for the univariate model. Anlysis showed that age and gender was not significant (Table 5.3.1) for the model. In this model, we have considered the dichotomous variable occupation as nurses and nurse's aide (NNA) and non-nurses (all except nurses and nurse aide, Non-NNA). In the same way, dichotomous variable body parts has been considered as back, neck & shoulders (BNS) and other body parts (except back, neck and shoulders; Non-BNS).

Table 5.3.1: Cox univariate hazards model

Covariates		**Estimates (S.E.*)**	**HR***	**P-Value**	**95% CI ***
Age		-0.0057 (0.0062)	0.994	0.3551	0.982 – 1.006
Gender	Female	-0.2533 (0.2104)	0.776	0.2285	0.514 – 1.172
Group	Intervention	-0.4019 (0.1254)	0.669	0.0014	0.523 – 0.856
Hospital Size	Large	-0.3141 (0.1461)	0.730	0.0316	0.548 – 0.973
	Small	-0.3284 (0.1583)	0.720	0.0380	0.528 – 0.982
Occupation+	NNA	0.5888 (0.1737)	1.802	0.0007	1.282 – 2.533
Body Parts++	BNS	0.6709 (0.1426)	1.956	<0.0001	1.956 – 2.587

*S.E.: Standard Error; * HR: Hazard Ratio; * CI: Confidence Interval

+ Others include, Therapists, Technicians, Unit Supporters, and Paramedics, etc.

++ Other Body Parts includes, Abdomen, Chest, Face, etc.

Note: In this analysis, the reference group: Male for gender, control for group, medium for hospital size, non-nurses for occupation (Non-NNA), and other body parts except back, neck and shoulder for body parts (Non-BNS)

5.3.2 Cox Multivariate Hazards Model

From the univariate analysis (Table 5.3.1), we noticed that age and gender were not significant (p-values=0.35 and 0.22, respectively); however, regarding their biological importance, they were considered in the multivariate multiplicative model. The analysis showed that these two covariates were not significant (p-values=0.86 and 0.26, respectively). Although hospital sizes were significant in the univariate analysis, size did not have a significant impact on the multivariate model for large and small sized hospitals (p-values=0.45 and 0.52, respectively). Thus, age, gender, and hospital sizes were not selected for the final model. Because it was observed that group, occupation and body parts were significant for the models, they were considered for the final Cox multivariate multiplicative hazards model.

Table 5.3.2: Cox multivariate hazards model

Covariates		Estimates (S.E.*)	HR**	P-value	95% CI***
Group	Intervention	-0.4686 (0.1276)	0.63	0.0002	0.497 - 0.804
Occupation	NNA	0.5401 (0.1745)	1.72	0.002	1.219 - 2.416
Body Parts	BNS	0.7643 (0.1444)	2.15	<0.0001	1.618 - 2.850

*S.E.: Standard Error; * HR: Hazard Ratio; * CI: Confidence Interval

+ Others include, Therapists, Technicians, Unit Supporters, and Paramedics, etc.

++ Other Body Parts includes Abdomen, Chest, Face, etc.

Note: In this analysis, the reference group: Control for group, , non-nurses for occupation (Non-NNA), and other body parts except back, neck and shoulder for body parts (Non-BNS)

Considering the group, occupation and body parts in the final Cox multivariate multiplicative hazards model, analysis has been done. The interaction was checked, but no significant interaction was observed. Analysis showed that all of the selected covariates were statistically significant on repeated injuries.

The group variable was significant for the Cox multivariate multiplicative hazards model. The results showed that the intervention group had a 27% lower risk of repeated injury as compared to the control group after the TLR intervention program (HR: 0.63, 95% CI: 0.49, 0.80) (Table

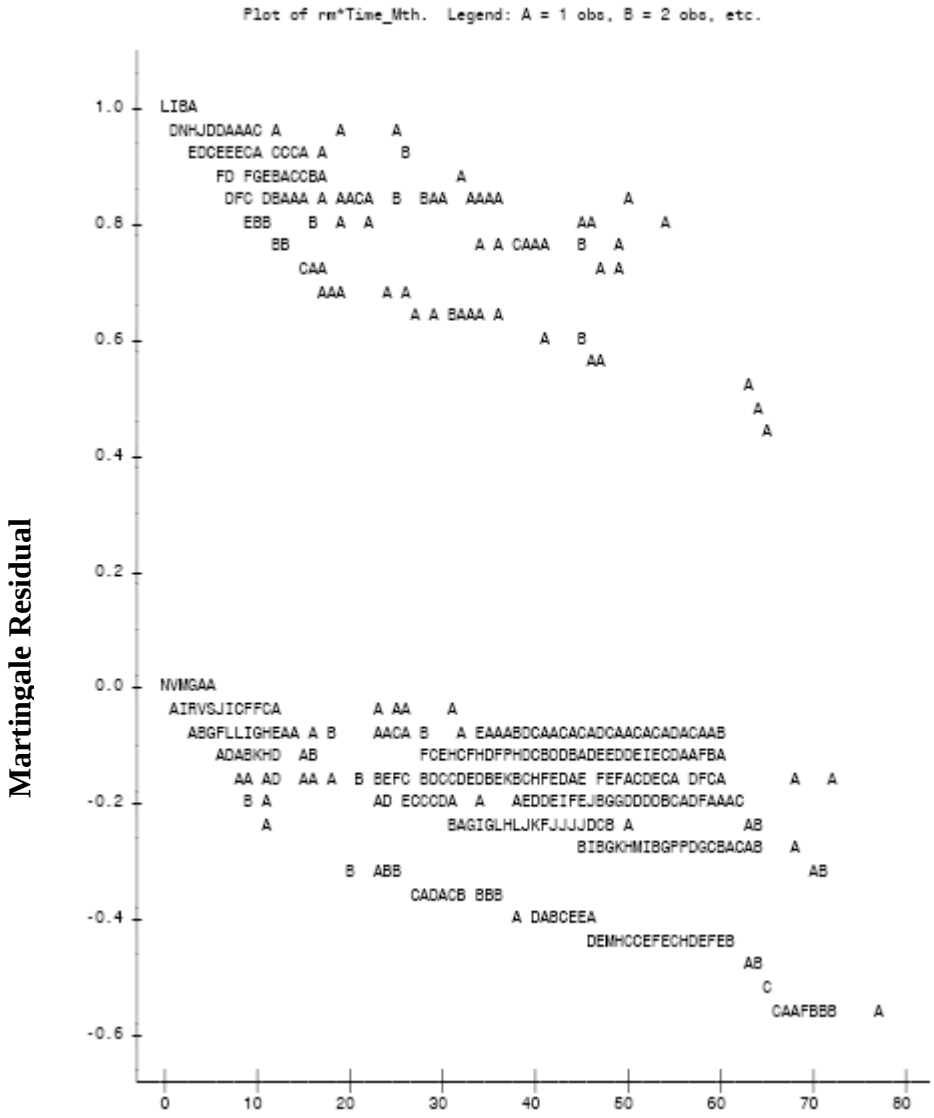
5.3.2). Similar to the group, occupation was significant for the Cox multivariate multiplicative hazards model. Compared to Non-NNA, nurses and nursing aides (NNA) had a 72% higher risk of repeated injury (HR: 1.72, 95% CI: 1.22, 2.42). As with group and occupation, body parts were significant for the Cox multivariate multiplicative hazards model. The back, neck and shoulder (BNS) were the most repeatedly injured body parts. Compared to other body parts (Non-BNS), the back, neck and shoulder (BNS) had a 115% increased risk of repeated injury (HR: 2.15, 95% CI: 1.62, 2.85).

5.3.3 Goodness of Fit

In this section, the adequacy of the Cox multivariate multiplicative hazards model was checked using the martingale residuals, deviance residuals, and Arjas plots.

5.3.3.1 Martingale residual plot:

Figure 5.3.3.1: Martingale residuals plotted against survival time

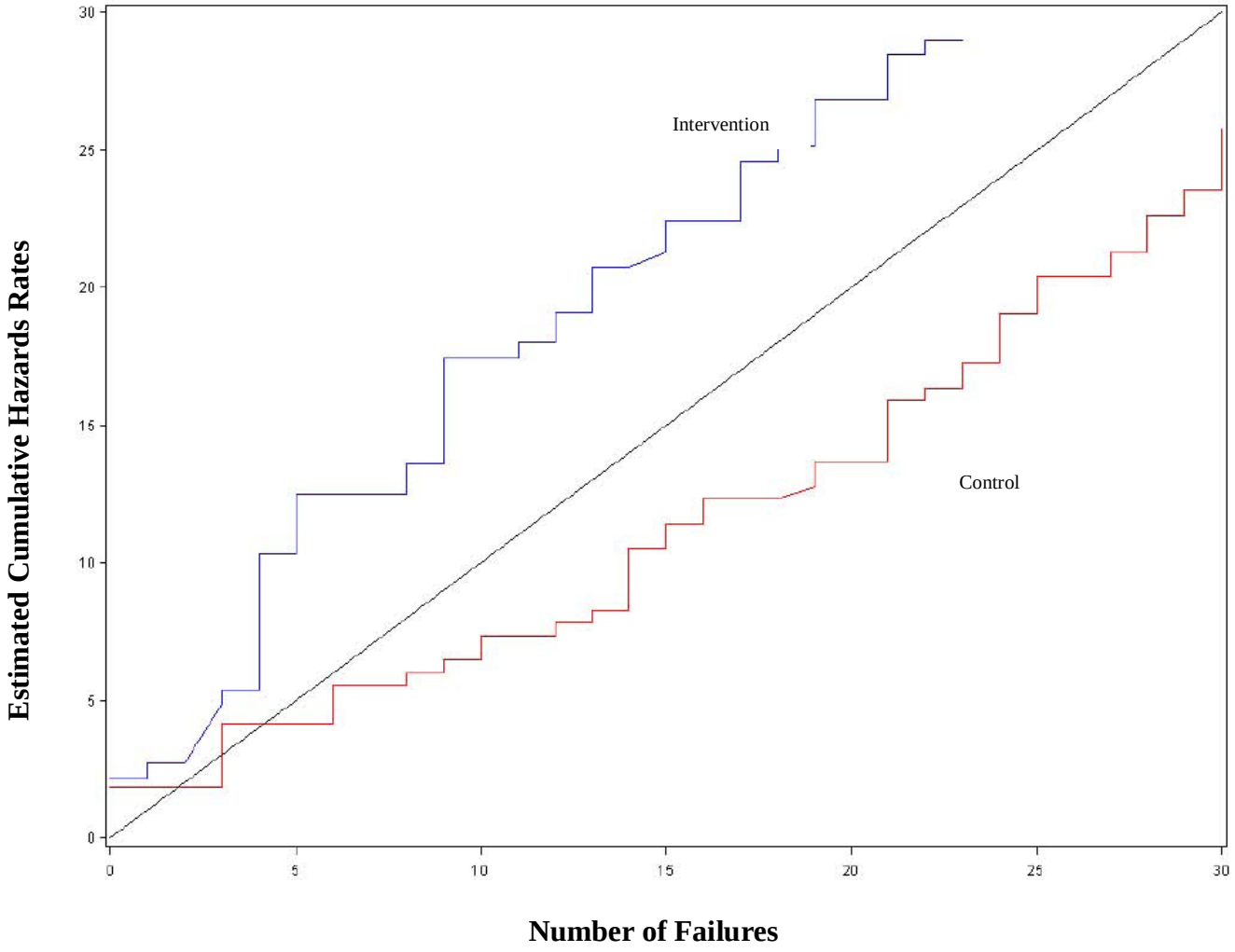


Time in Months

From the above martingale residual plots against time, we observed that some of the individuals had large negative values; thus, they had long survival times. This indicates that they may have the chance of getting injured again soon. Also, some individuals had residuals close to unity, which indicates that the individuals had the repeated injury within a short period of time. Particularly, one individual had too long a survival time in months, and some of the individuals did not fit the model well.

5.3.3.2 Arjas plot

Figure 5.3.3.2: Arjas residual plots of estimated cumulative hazards rates



An Arjas plot is used to check the overall fit of the multiplicative hazards model for the intervention and control groups (50). We noticed that the two curves were roughly close to the 45^0 line after the 5 failures. They were roughly parallel, which nearly satisfied the proportionality assumption. Thus, this may suggest that the multiplicative model is appropriate.

5.4 Additive Hazards Model

5.4.1 Aalen's Additive Hazards Model

The result of the analysis showed that the group, occupation and body parts had significant effect on the repeated MSIs (p-value<0.0001). To further examine the relationship between the group, occupation and body parts, an ANOVA Table was constructed using one degree of freedom. The intervention group had significantly different repeated injuries than the control group, adjusting for the other covariates (p-value=0.0002). Considering occupation and body parts, there were significant differences in the types of repeated injuries. In this model, we have considered the dichotomous variable occupation as nurses and nurse's aide (NNA) and non-nurses (all except nurses and nurse aide, Non-NNA). In the same way, dichotomous variable body parts has been considered as back, neck & shoulders (BNS) and other body parts (except back, neck and shoulders; Non-BNS).

Table 5.4.1: Aalen's additive hazards model

Global Test		
Chi-Square	D.F.[+]	P-value
50.72	3	<0.0001
ANOVA		
Covariates	Chi-Square	P-Value
Group: Intervention [*]	13.76	0.0002
Occupation: Nurses and Nurses' Aide [**]	11.33	0.0008
Body Parts: Back, Neck, and Shoulder [***]	31.77	0.0000

[+] D.F.: Degrees of Freedom

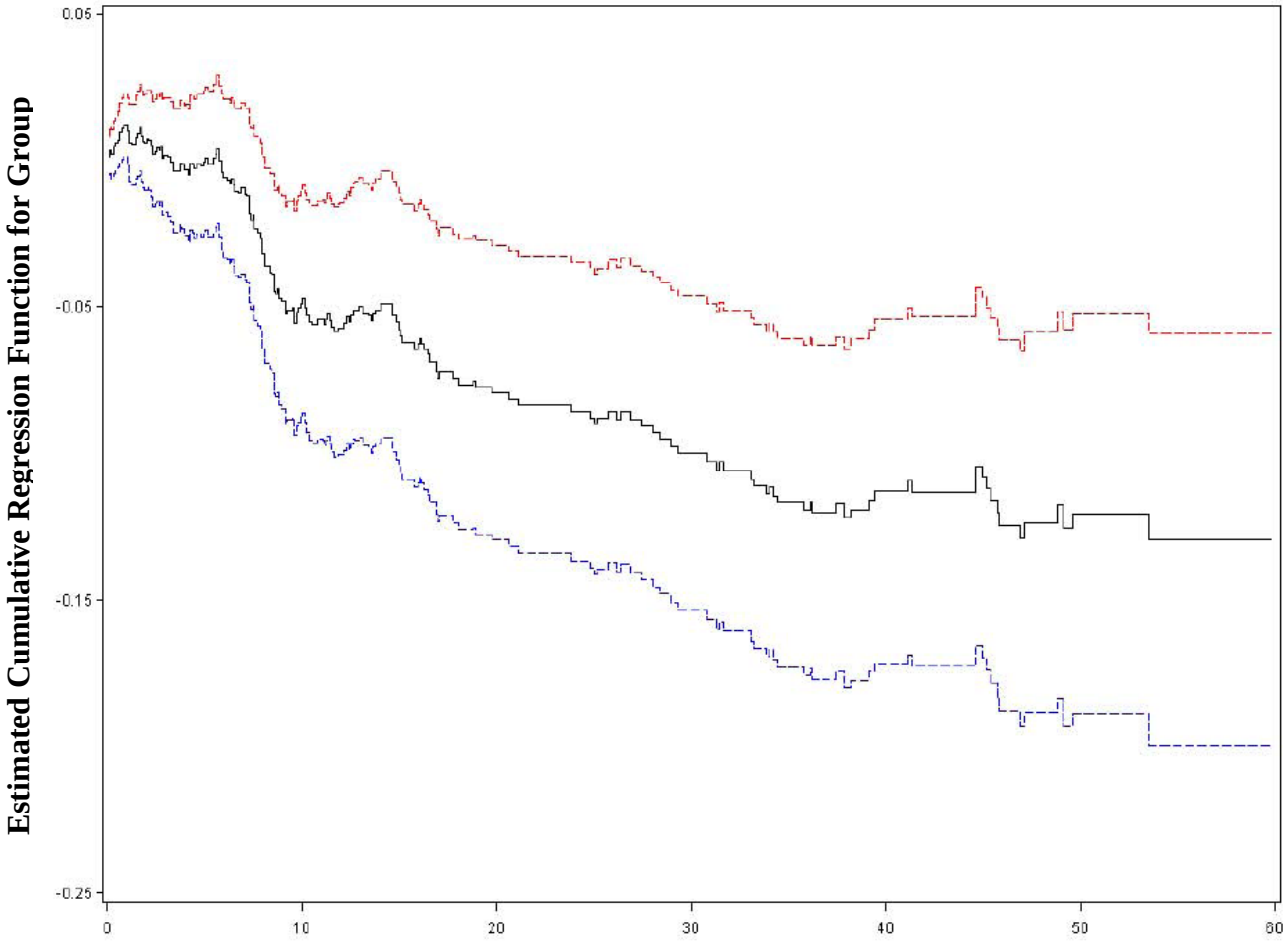
[*] Control group is the reference group

[**] Others (e.g., Therapists, Technicians, Unit Supporters, and Paramedics) is the reference occupation

[***] Other Body Parts (e.g., Abdomen, Chest, and Face) is the reference body parts

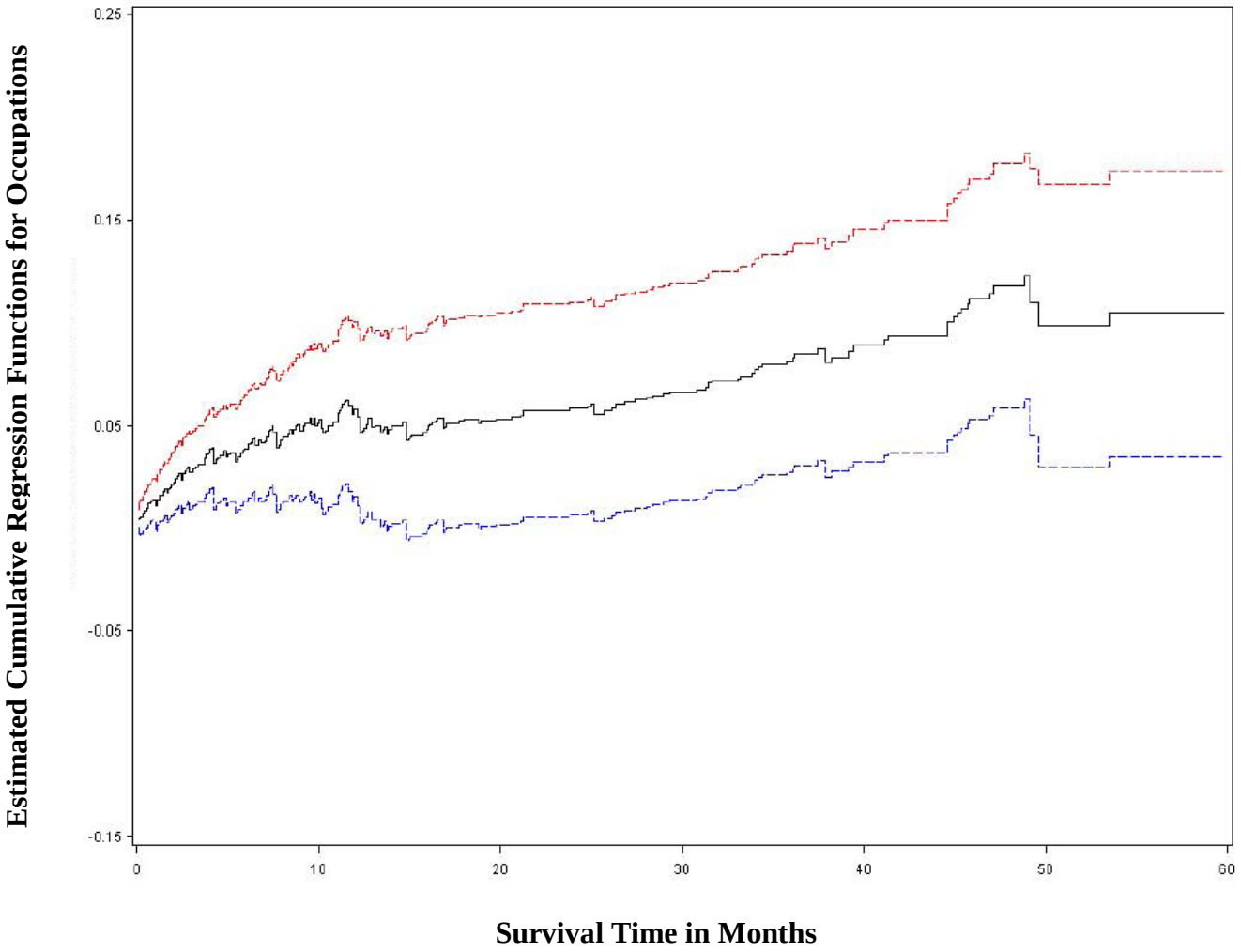
In the Aalen's model, in order to visualize a covariate effect over time, the estimated cumulative regression function has been examined, along with its upper and lower 95% point-wise confidence limits. The plot of the estimated cumulative regression functions for group (Figure 5.4.1.1) showed that there was no covariate effect on the hazard up to 8 months. The figure 5.4.1.1 showed that, for the period of 8-24 months, the slope for group was negative and clear effects of decreasing hazard, but after that it was approximately constant. So, based on the plot, it has been concluded that intervention group had the less risk of repeated injury as compared to the control group.

Figure 5.4.1.1: Estimated cumulative regression functions for group with 95% confidence interval



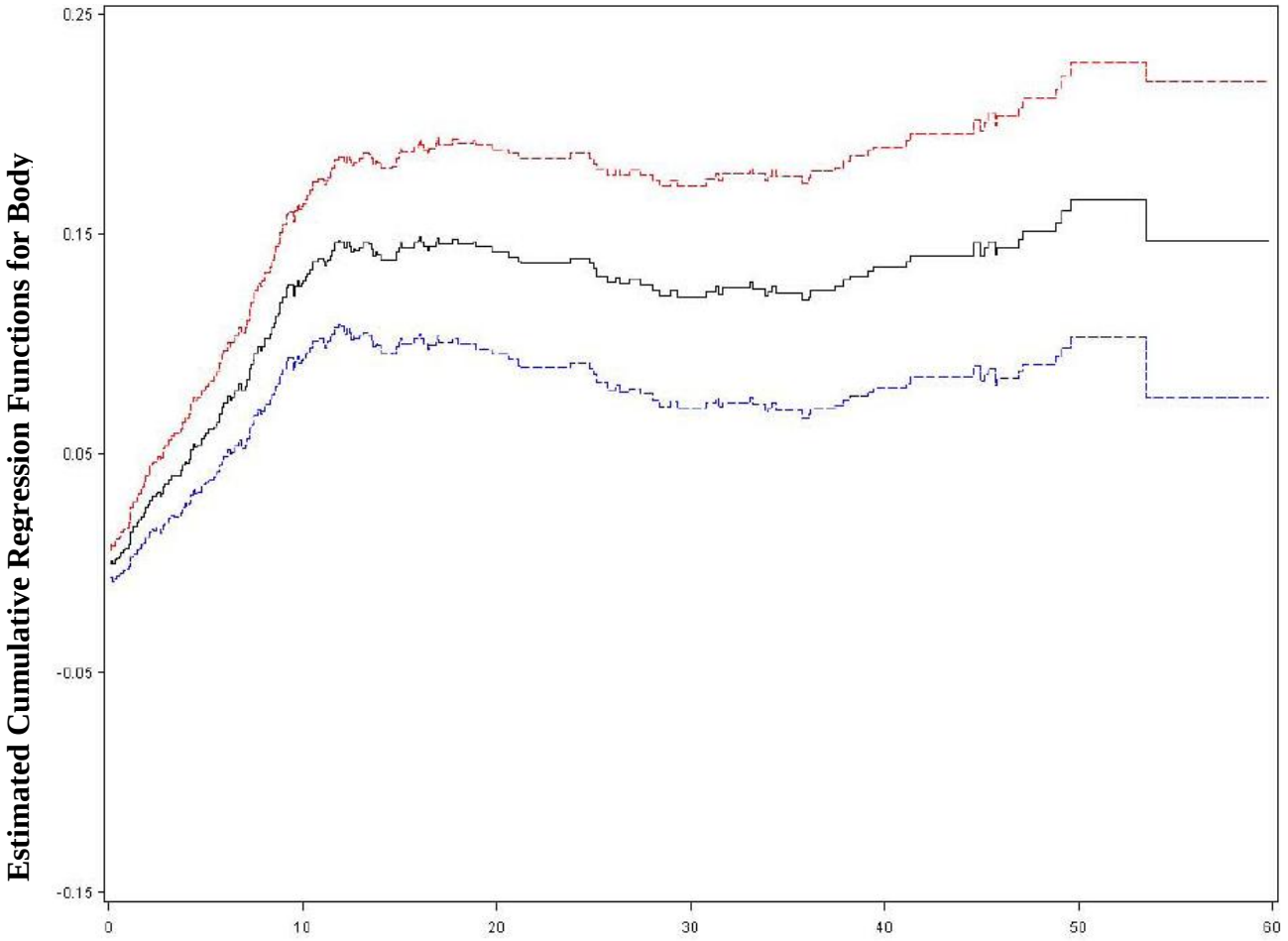
The estimated cumulative regression functions for occupation showed that there may be time varying occupation effect among nurses because its shows the non-zero slope over time (Figure 5.4.1.2). It has been observed that the covariate effects of occupation have been increased up to 10 months. The following figure shows that, after 10 months that it is constant (approximately).

Figure 5.4.1.2: Estimated cumulative regression functions for occupation with 95% confidence interval



The estimated cumulative regression functions for body parts showed that there may be time varying body parts effect because its shows the positive slope over time (Figure 5.4.1.3) and 95% confidence interval of the covariate effects did not includes zero. It has been observed that, up to 10 months, the covariate effects of body parts has been increased, after that it is approximately constant.

Figure 5.4.1.3: Estimated cumulative regression functions for body parts with 95% confidence interval



5.4.2 Lin and Ying Additive Hazards Model

All the variables deemed significant by the univariate analysis were considered in the univariate L-Y additive hazards model. The results are shown in Table 5.4.2.1.

The univariate L-Y additive hazards model indicates that the group, hospital size (large and small), occupation, and body parts were significant for the multivariate L-Y additive hazards model (Table 5.4.2.1). Considering all the significant covariates in the L-Y additive hazards model, multivariate analysis was conducted; this showed that hospital sizes were not significant for the L-Y additive hazards model (p-values=0.38 and 0.44, respectively). The final L-Y additive hazards model includes group, occupation and body parts. In this model, we have considered the dichotomous variable occupation as nurses and nurse's aide (NNA) and non-nurses (all except nurses and nurse aide, Non-NNA). In the same way, dichotomous variable body parts has been considered as back, neck & shoulders (BNS) and other body parts (except back, neck and shoulders; Non-BNS).

Table 5.4.2.1: Univariate Lin and Ying additive hazards model

Covariates		**Estimate (S.E. [*])**	**ER [*]**	**P-Value**	**95% CI [*]**
Age		-0.00002 (.00003)	0.0002	0.4555	(-0.00008, -0.00003)
Gender:	Female	-0.00164 (.00151)	- 0.001	0.2781	(-0.00461, 0.00132)
Group:	Intervention	-0.00226 (.00071)	- 0.002	0.0015	(-0.00366, -0.00086)
Hospital Size:	Large	-0.00188 (.00090)	- 0.001	0.0377	(-0.00366, -0.00010)
	Small	-0.00189 (.00091)	- 0.001	0.0388	(-0.00369, -0.00009)
Occupation:	NNA	0.00277 (.00069)	0.002	0.00006	(0.00141, 0.00413)
Body Parts:	BNS	0.00344 (.00067)	0.003	0.00001	(0.00212, 0.00476)

[*]S.E.: Standard Error; [*] ER: Excess Risk; [*] CI: Confidence Interval

[+] Others include Therapists, Technicians, Unit Supporters, Paramedics, etc.

[++] Other Body Parts includes Abdomen, Chest, Face, etc.

Note: In this analysis, the reference group: Male for gender, control for group, medium for hospital size, , non-nurses for occupation (Non-NNA), and other body parts (except back, neck and shoulder) for body parts (Non-BNS)

The final L-Y additive hazards model shows that other covariates, such as intervention group, had significantly different types of repeated injuries than the control group (p-value=0.0005) (Table 5.4.2.2). The estimate is negative (-0.0025), indicating that the intervention group had protection from repeated injury as compared to the control group, which had 0.002 more repeated TLR injuries than the intervention group, i.e. 2 person repeated injury can be prevented per 1000 person. Regarding occupation, nurses and nursing aides (NNA) had the most significantly different injuries than non-nurses (Non-NNA) occupations and 0.002 more repeated TLR injuries hazards (ER=0.002; p-value=0.0005; 95%CI=0.001 – 0.0038) (Table 5.4.2.2), which indicates that non-NNA had 2 less repeated injury compared to NNA per 1000. Regarding body parts, the back, neck and shoulders (BNS) had the most significantly different repeated injury instances than did other body parts (Non-BNS). Among the body parts, combined back, neck & shoulder had 0.003 more repeated TLR injuries hazards than other body parts (ER=0.003; p-value <0.0001; 95% CI=0.0025 – 0.0051) (Table 5.4.2.2), which indicates that Non-BNS had 3 less repeated injury compared to BNS per 1000.

Table 5.4.2.2: Multivariate Lin and Ying additive hazards model

Covariates		**Estimate (S.E. *)**	**ER ***	**P-Value**	**95% CI ***
Group:	Intervention	-0.0025 (0.0007)	- 0.002	0.0005	-.0039 - -.0010
Occupation:	NNA	0.0024 (0.0006)	0.002	0.0005	.0010 –.0038
Body Parts:	BNS	0.0038 (0.0006)	0.003	<0.0001	.0025 - .0051

*S.E.: Standard Error; * ER: Excess Risk; * CI: Confidence Interval

+ Others include Therapists, Technicians, Unit Supporters, Paramedics, etc.

++ Other Body Parts includes Abdomen, Chest, Face, etc.

Note: In this analysis, the reference group: Control for group, non-nurses for occupation (Non-NNA), and other body parts (except back, neck and shoulder) for body parts (Non-BNS).

5.5 Goodness of Fit for the L-Y Additive Hazards Model

To check the adequacy of the model, Arjas and martingale plots were used for the final selected covariates. The plots are presented below.

5.5.1: Arjas plot

Figure 5.5.1(a) shows that the lines are approximately close to 45^0, which indicates that the group fit the model well. Figure 5.5.1(b) also shows that nurses, nursing aides and other occupations are approximately close to the 45^0 line. However, nurses and nursing aides had the smallest number of repeated injuries as compared to others. Thus, the follow-up time was shorter. Notably, the Arjas plot of nurses and nursing aides is not long enough but reasonably satisfies the model. However, for the body parts, we observed that the back, neck and shoulders have concave downwards curves that are far from the 45^0 line [Figure 5.5.1(c)]. Thus, we may conclude that the model may not be appropriate in regard to the covariates body parts.

Figure 5.5.1(a): Expected cumulative hazards rate for group

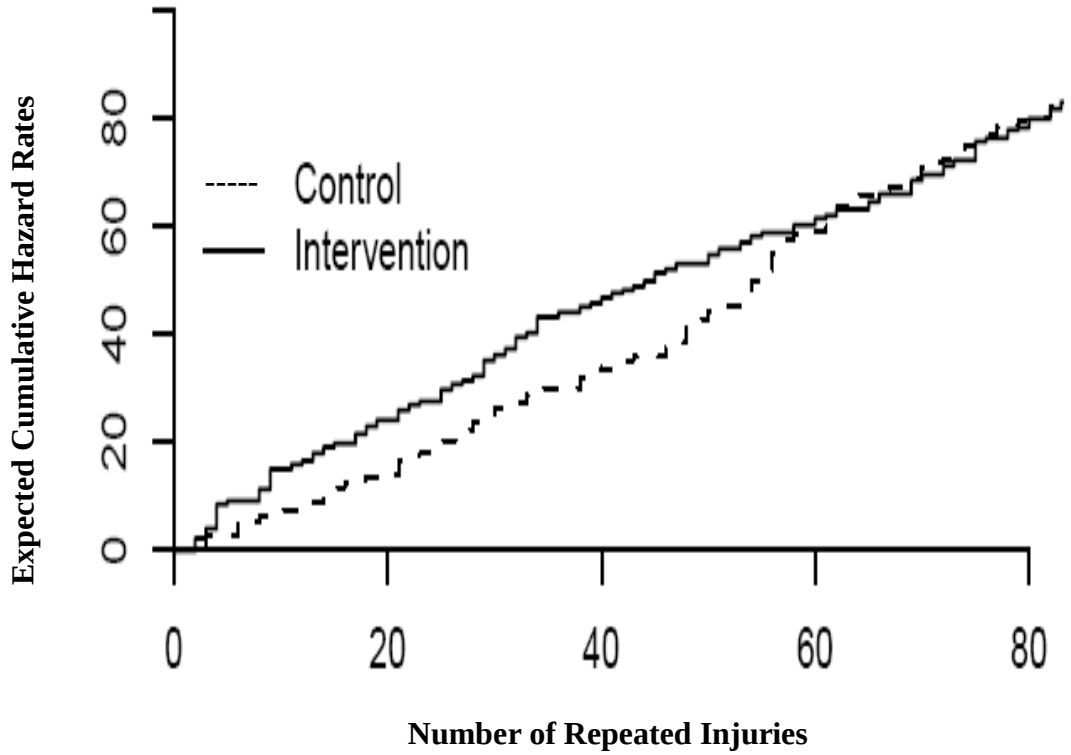
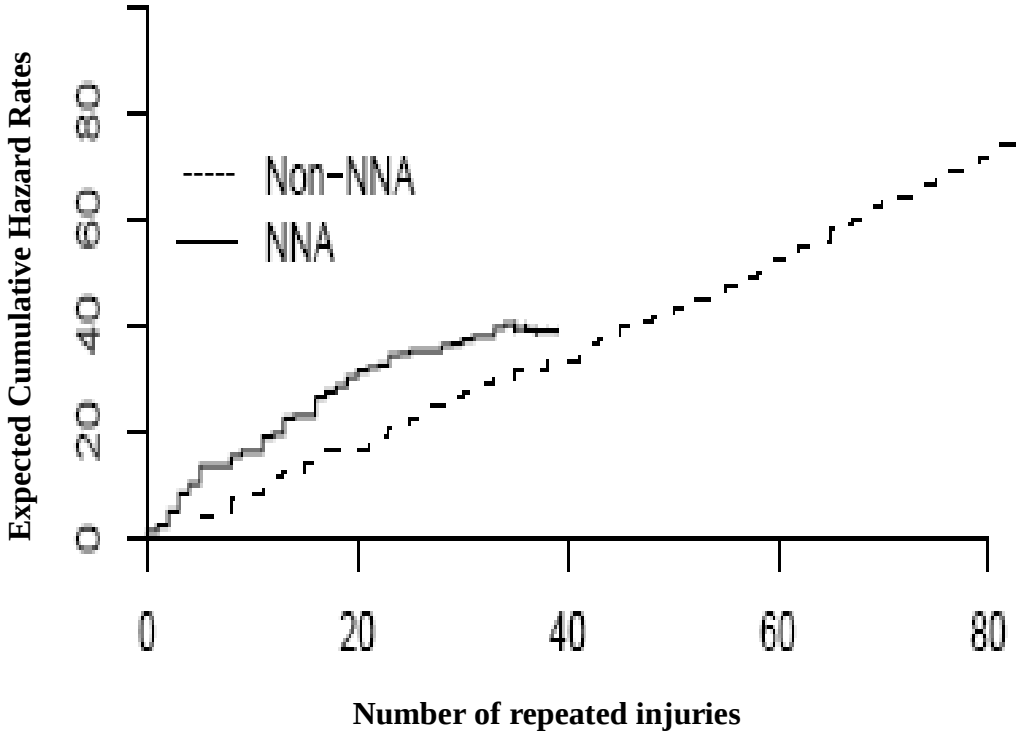
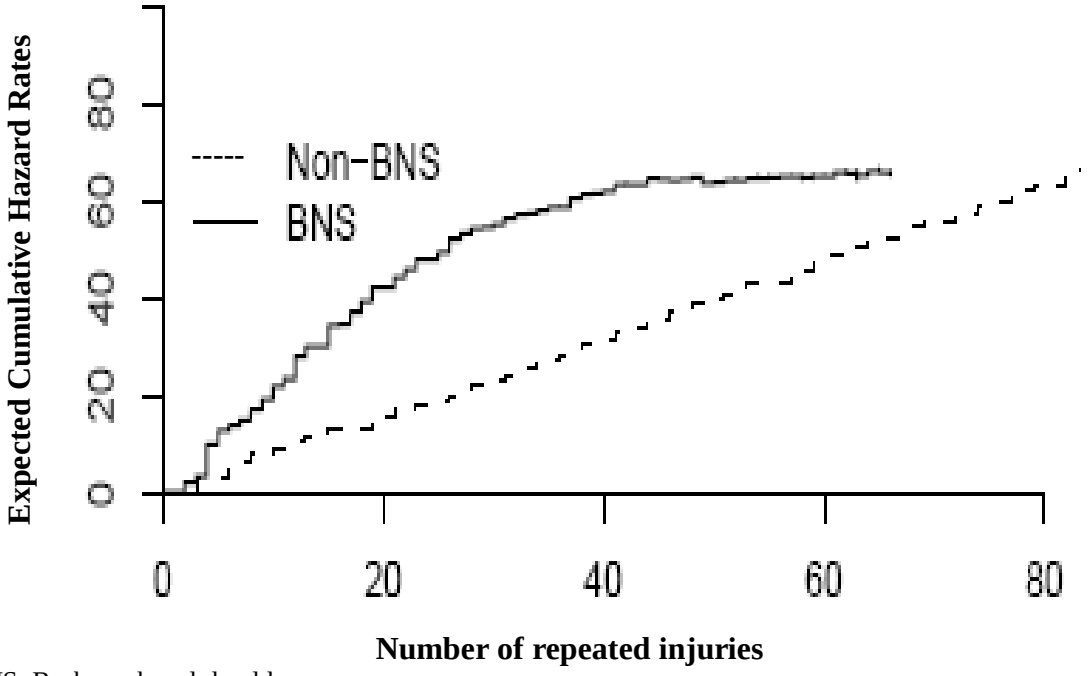


Figure 5.5.1 (b): Expected cumulative hazards rate for occupation



*NNA: Nurses and nursing aides
*Non-NNA: Other than NNA

Figure 5.5.1 (c): Expected cumulative hazards rate for body parts



* BNS: Back, neck and shoulder
* Non-BNS: Other than BNS

5.5.2 Martingale Residuals Plots

Figure 5.5.2 a-c show the Martingale residuals plots for the L-Y additive model. Figure 5.5.2(a) does not show any distinguishable pattern, but because the intervention group was protected from repeated injuries, the plots are below the origin line, which indicates that it is acceptable for the model. Figure 5.5.2(b) indicates that up to four months, there are some increasing trends, which disappear after that and move approximately constantly. Thus, we approximate that occupations reasonably fit the model. Figure 5.5.2(c) shows significantly increasing trends, which is consistent with Arjas plot. Thus, the body parts raise some doubt for the model.

Figure 5.5.2(c): Martingale residuals for body parts was plotted against survival time

Based on the above Arjas residual plots and Martingale residuals plots, we may conclude that the main effects of group and occupations are reasonably fit for the Lin and Ying additive hazards model, but the body parts are questionable.

Figure 5.5.2 (a): Martingale residuals plot for group

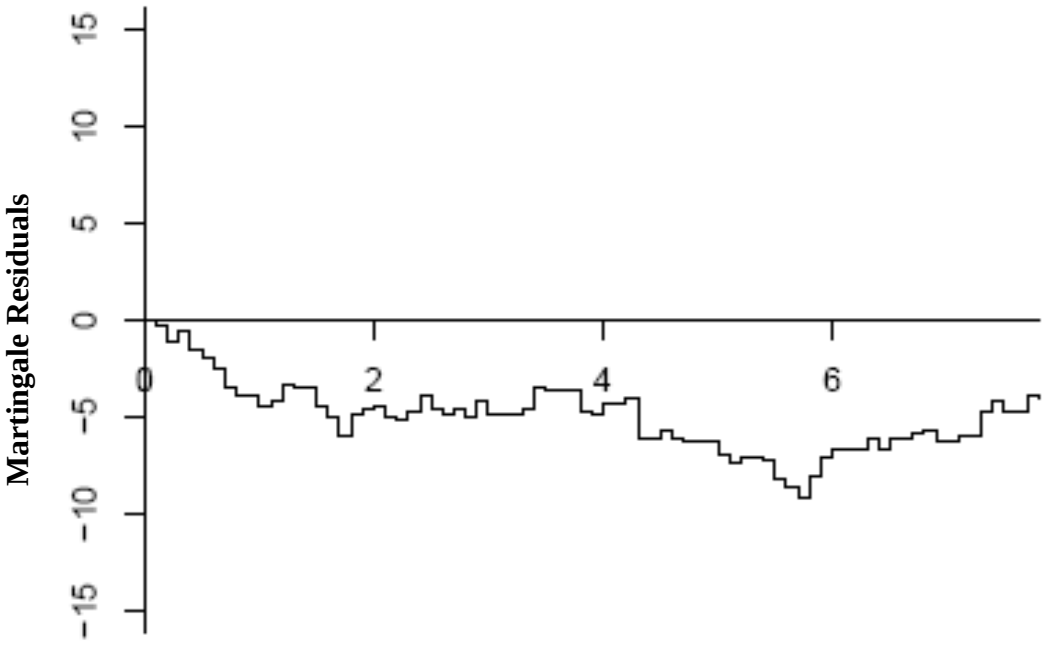


Figure 5.5.2 (b): Martingale residuals plot for occupation

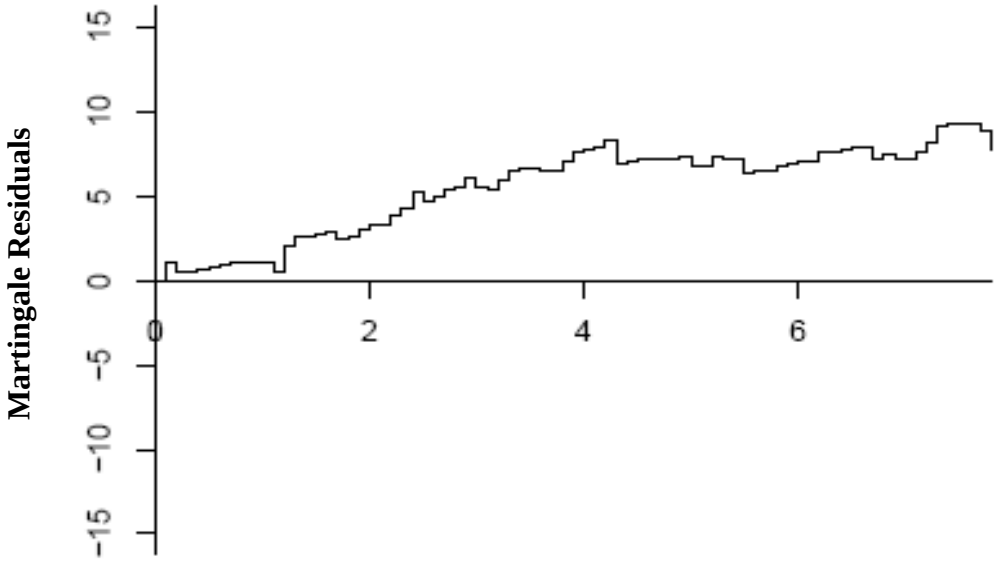
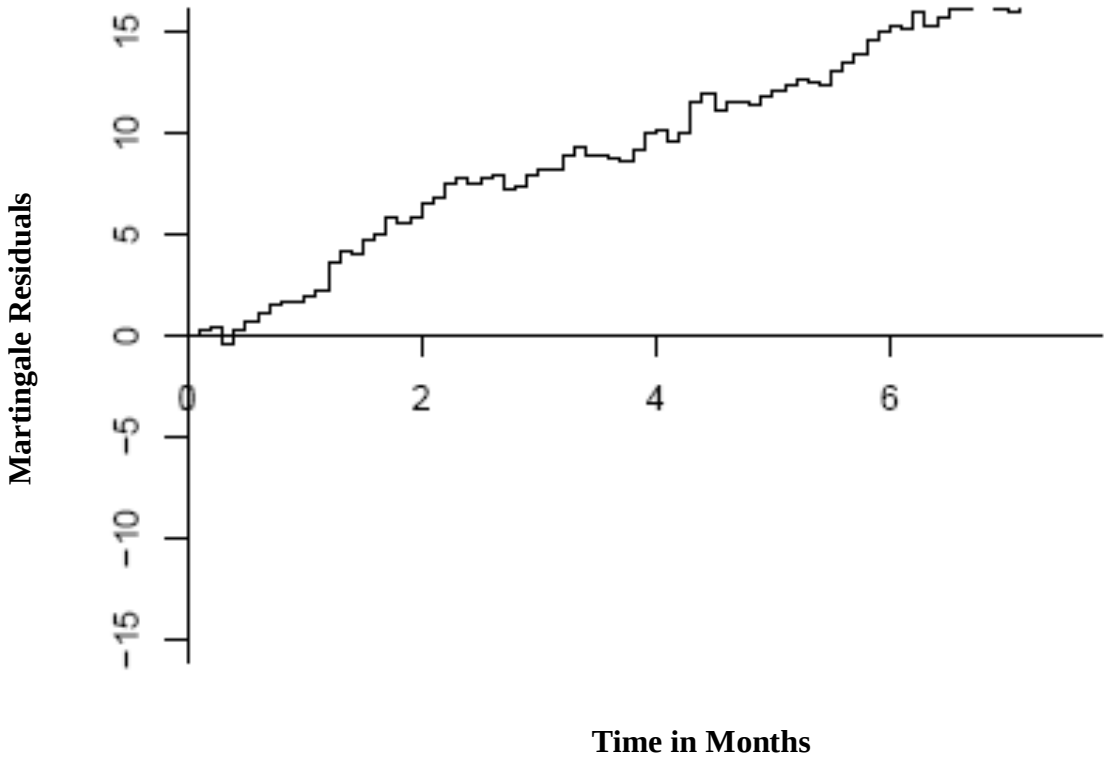


Figure 5.5.2 (c): Martingale residuals plot for body parts



5.6 Summary of Results for the Final Multivariate Cox and L-Y Models

Table 5.6.1 shows comparison of the results from the final Cox multivariate multiplicative hazards model and Lin and Ying additive hazards model. Two models provide the same significant factors: group, occupations, and body parts, as expected (22). The estimates that were obtained from the analysis were different. Because one model is multiplicative and the others are additive, they are not comparable. Based on the goodness-of-fit assessment for both models, group and occupations are reasonably adequate for the model, but the body parts variable is questionable for both models.

Table 5.6.1: Comparison of Cox and Lin and Ying additive hazards models

Model	**Covariates**	**Estimates (S.E.*)**	**HR*/ER+**	**P-value**	**95% CI***
Cox	Group	-0.4686 (0.1276)	0.63	0.0002	0.497 - 0.804
	Occupation	0.5401 (0.1745)	1.72	0.002	1.219 - 2.416
	Body Parts	0.7643 (0.1444)	2.15	<0.0001	1.618 - 2.850
Additive	Group	-0.0025 (0.0007)	-0.002	0.0005	-0.0039 - -0.0010
	Occupation	0.0024 (0.0006)	0.002	0.0005	0.0010 –0.0038
	Body Parts	0.0038 (0.0006)	0.003	<0.0001	0.0025 - 0.0051

*S.E.: Standard Error; ** HR: Hazard Ratio; *** CI: Confidence Interval

+ ER: Excess Risk

6. Discussion

To reduce the risk of transfer, lifting and repositioning (TLR)-related repeated injuries from patient handling, an ergonomic injury prevention program was implemented for the healthcare workers in the intervention group. However, no intervention program was implemented for the control group. Both the intervention and control groups were contained within three hospital sizes: large, medium and small. The goal of this study was to investigate the effectiveness of the injury prevention program using the multiplicative Cox and additive hazards models. Based on our analysis, the results indicate that the TLR intervention program was effective and sustained for healthcare workers by reducing repeated injuries induced by patient handling. Therefore, the risks of patient handling-related repeated injuries among healthcare workers can be lowered by implementing a multi-factor TLR intervention program with the right equipment and training.

The TLR intervention program was implemented for the healthcare workers in the intervention group as compared to the control group. The multivariate Cox model showed that the intervention group had 27% fewer repeated injuries than the control group, which indicates the effectiveness of the TLR intervention program. The intervention group also showed protection from repeated TLR injures by the L-Y additive hazards model. The intervention group had 0.002 fewer hazards for repeated injuries than the control group, which supports the result of the Cox model. On the other hand, nurses and nursing aides had the most repeated injuries by occupation; the Cox model showed a 72% higher hazard of repeated injuries than other occupations. According to the L-Y model, nurses and nursing aides had a 0.0024 higher risk of repeated injuries than other occupations. Nurses and nursing aides are directly involved in patient handling. Although the TLR intervention program was implemented for the intervention group, they still had a higher risk of repeated injuries regarding TLR, which showed significance in the model. Among all of the healthcare workers, body parts were the most significant risk factor for repeated injuries. The Cox model indicated that the back, neck and shoulders had a 115% increased risk of repeated injuries as compared to other body parts. The L-Y additive hazards model also showed that the back, neck and shoulder had a 0.0038 increased risk of repeated injures than the others body parts. Both the Cox and L-Y models, as well as Aalen's additive

hazards model, showed that the TLR intervention program had a significant impact on reducing repeated injures among healthcare workers.

The Cox model is the most widely used model for the analysis of survival data in clinical research. However, the proportional hazard assumption may not always be satisfied in the data. In such cases, there are various solutions to consider; for example, inclusion of a time-dependent covariate. While the coefficients in the Cox model act in a multiplicative way on unknown baseline hazards, coefficients in the additive hazards models act in an additive way on unknown baseline hazards. Because the coefficients act in different ways in the multiplicative and additive hazards models, it is very difficult to compare them directly. In this thesis, the multiplicative and additive hazards models similarly identified the significant covariates of the repeated injuries among healthcare workers. However, the different models interpreted the coefficients in different ways. The association between the covariates and the time to repeated injuries in the additive hazards models was explained in terms of the risk difference or excess risk rather than the risk ratio. However, if one would like to estimate the cumulative hazard of an event for more extreme values, the additive and the Cox hazards models estimates are remarkably different. By using the time varying covariates effect, this can be settled on by which are taken into account by the additive hazards model but not by a multiplicative Cox hazards model. Moreover, when using a more compromised covariate profile, the multiplicative model gives a higher estimate than the additive model, probably because of the multiplicative effect of fixed covariates on baseline function (21).

In this thesis, administrative data was used that had been supplied by two Health Regions. The data acquisition, injury classification criteria, and data extraction process could not be controlled or evaluated. The lack of information on the subjects' demographic and injury characteristics and the total number of direct care workers employed at each site weakened the study. Another drawback of the data was that there was no identification number for the control group. Thus, in case of identified the repeated injury there would be personal selection bias. The additive models have some limitations. Aalen's model may provide more in-depth information on the effect of a prognostic factor over time. However, one has to visualize all covariates' effects over time, and a

simple interpretation of the effects is not possible, which makes Aalen's model less appealing in real applications than other models. A theoretical limitation of the Lin and Ying additive hazards model is that the linear predictors in the model constrain to be positive (19). A very practical limitation of the additive hazards models is the availability of computer programs. For the Cox hazards model, various statistical software packages are available, and it is easy to fit the models. However, for the additive hazards model, any standard procedure is limited to SAS, R, and Stat. Few macros are available for the analysis of goodness of fit (22, 84). Because the different macros are not used globally, it will be difficult to make a real comparison.

In many applications, the additive hazards models are plausible and often attractive in epidemiologic applications, where the baseline hazards is taken to be the baseline mortality of the population and the coefficient measures the excess risk of the patients under study. As an example, in a study of diabetic patients (85), if the measured covariates predict the severity of disease and its downstream mortality/morbidity, but have no impact on independent causes of death, such as malignancy, then the multiplicative hazards model might not be appropriate. In such cases, the additive hazards model may be better for patients with more severe clinical profiles, which is relevant to the development of patient management and care (86). The risk difference can be more important than the risk ratio in understanding an association between a risk factor and disease occurrence (19). The results of this study are also consistent with another published study (22).

7. Conclusion

Generally, the preference between the Cox hazards model and the additive hazards model will normally be a practical matter. Although in theory, either model can provide adequate fit to a given time to event data set, the more parsimonious one will unquestionably be preferable to clinical investigators. An overall conclusion is that the multiplicative and additive hazards models describe different features of the association between the risk factors and the study outcomes. Practitioners may benefit from the use of statistical models, which help in predicting the effect of one or more variables and in verifying their influence on the study outcomes. It seems desirable to use them together as complementary methods so as to give a more comprehensive understanding of the data. Furthermore, the additive hazards model can be expanded to a competing risks setting in survival analysis.

8. References

1. Collet D. Modelling Survival Data in Medical Research. Second ed: Chapman & Hall/CRC CRC Press LLC; 2003.
2. Kleinbaum DG, Klein M. Survival Analysis: A Self-Learning Text. Second ed: Springer; 2005.
3. Lawless JF. Statistical Models and Methods for Lifetime Data Analysis. 2nd ed. New York: Wiley; 2002.
4. Klein JP, & Moeschberger, M.L. Survival Analysis: Techniques for Censored and Truncated Data. 2nd Edition ed. K. Dietz MG, K. Krickeberg, J. Samet, A. Tsiatis, editor: Springer; 2003.
5. Lee ET, Wang JW. Statistical Methods for Survival Data Analysis. Third ed. New York: Wiley-Interscience; 2003.
6. Vi P. Musculoskeletal Disorders (MSDs). Construction Safety Association of Ontario; Available from: http://www.csao.org/t.tools/t6.news/msd.cfm.
7. Statistics USBoL. Occupational Safety and Health Definitions. In: Labor USDo, editor.
8. Disorders CoREftPoM. How well do we understand MSDs and their burdens? Waterloo, ON, Canada; Available from: http://cre-msd.uwaterloo.ca/Understanding_MSDs.aspx.
9. Semple JC. Tenosynovitis, Repetitive Strain Injury, Cumulative Trauma Disorder, and Overuse Syndrome, Et-cetera. Journal of Bone and Joint Surgery-British Volume1991; 73(4):536-8.
10. Barlow Y, Willoughby J. Pathophysiology of Soft-tissue Repair. British Medical Bulletin 1992; 48(3):698-711.
11. Barr AE, Barbe MF. Pathophysiological tissue changes associated with repetitive movement: A review of the evidence. Physical Therapy2002;82(2):173-87.
12. Szabo RM. Determining causation of work-related upper extremity disorders. Clin Occup Environ Med2006;5(2):225-34, v.
13. Crumpton-Young LL, Killough MK, Parker PL, Brandon KM. Quantitative analysis of cumulative trauma risk factors and risk factor interactions. Journal of Occupational and Environmental Medicine2000;42(10):1013-20.
14. Bernard BP, Putz-Anderson V, Burt SE, Cole LL, Fairfield-Estill C, Fine LJ, et al. Musculoskeletal Disorders and Workplace Factors. In: U.S. DoHaHS, editor.: National Institute for Occupational Safety and Health (NIOSH); July 1997.
15. Concha-Barrientos M ND, Driscoll T, Steenland N, Punnett L, Fingerhut M, Prüss-Ustün A, Corvalan C, Leigh J, Punnett L, Tak S. Selected occupational risk factors. In: Ezzati M, Lopez A, Rodgers A, Murray C editors. Comparative quantification of health risks: global and regional burden of disease attributable to selected major risk factors. Geneva: World Health Organization; 2004. p. 1651-802.
16. Magee DJ. Orthopedic Physical Assessment 5th Edition ed: A Saunders Co. ; 2007.
17. Cox DR. Rregression Models and Life-Tables. Journal of the Royal Statistical Society Series B-Statistical Methodology1972;34(2):187-&.
18. Aalen OO. A Linear-Rregression Model for the Analysis of Life Times. Statistics in Medicine. [Article]. 1989 Aug;8(8):907-25.
19. Lin DY, and Ying, Z., Additive Hazards Regression Models for Survival Data: Springer; 1997.

20. Lin DY, Ying ZL. Semiparametric analysis of general additive-multiplicative hazard models for counting processes. Annals of Statistics1995;23(5):1712-34.
21. Lin DY, Ying ZL. Semiparametric Analysis of the Additive Risk Model. Biometrika1994;81(1):61-71.
22. Lim HJ, Zhang X. Semi-parametric additive risk models: Application to injury duration study. Accident Analysis and Prevention2009;41(2):211-6.
23. Wells R. Why have we not solved the MSD problem? Work-a Journal of Prevention Assessment & Rehabilitation2009;34(1):117-21.
24. Takala J. Introductory report of the International Labor Office. International Occupational Safety and Health Information Center, Geneva, International Labor Office. 1999.
25. Menzel NN. Back pain prevalence in nursing personnel: measurement issues. Aaohn J 2004;52(2):54-65.
26. Turnbull N, Dornan J, Fletcher B, Wilson S. Prevalence of Spinal Pain among the Staff of a District Health Authority. Occupational Medicine-Oxford1992;42(3):143-8.
27. Heap DC. Low-Back Injuries in Nursing Staff. Journal of the Society of Occupational Medicine1987;37(2):66-9.
28. Cunningham C, Flynn T, Blake C. Low back pain and occupation among Irish health service workers. Occupational Medicine-Oxford2006;56(7):447-54.
29. Marras WS, Cutlip RG, Burt SE, Waters TR. National occupational research agenda (NORA) future directions in occupational musculoskeletal disorder health research. Applied Ergonomics2009;40(1):15-22.
30. Council NHCaSAS. State of the Sector/ Healthcare and Social Assistance: Identification of Research Opportunities for the Next Decade of NORA 2009.
31. Council NHaSAS. National Healthcare and Social Assistance Sector Agenda 2009.
32. Zinn JS. The Influence of Nurse Wage Differentials on Nursing-home Staffing and Resident Care Decisions. Gerontologist1993;33(6):721-9.
33. Zontek TL, Isernhagen JC, Ogle BR. Psychosocial factors contributing to occupational injuries among direct care workers. Aaohn J2009;57(8):338-47.
34. Statistics BoL. Nonfatal Occupational Injuries and Illness Requiring Days away from work, 2007. 2008; Available from: www.bls.gov/iif/home.htm.
35. Engkvist IL, Hagberg M, Linden A, Malker B. Over-exertion Back Accidents among Nurses Aides in Sweden. Safety Science1992;15(2):97-108.
36. Engkvist IL. Back injuries among nurses - A comparison of the accident processes after a 10-year follow-up. Safety Science. [Proceedings Paper]. 2008 Feb;46(2):291-301.
37. Engkvist IL, Hagberg M, Hjelm EW, Menckel E, Ekenvall L, group Ps. The accident process preceding overexertion back injuries in nursing personnel. Scandinavian Journal of Work Environment & Health. [Article]. 1998 Oct;24(5):367-75.
38. Smedley J, Egger P, Cooper C, Coggon D. Manual Handling Activities and Risk of Low-back-pain in Nurses. Occupational and Environmental Medicine1995;52(3):160-3.
39. Koppelaar E, Knibbe JJ, Miedema HS, Burdorf A. Determinants of implementation of primary preventive interventions on patient handling in healthcare: a systematic review. Occupational and Environmental Medicine2009;66(6):353-60.
40. Smedley J, Trevelyan F, Inskip H, Buckle P, Cooper C, Coggon D. Impact of ergonomic intervention on back pain among nurses. Scandinavian Journal of Work Environment & Health2003;29(2):117-23.

41. Smedley J, Inskip H, Trevelyan F, Buckle P, Cooper C, Coggon D. Risk factors for incident neck and shoulder pain in hospital nurses. Occupational and Environmental Medicine2003;60(11):864-9.
42. Trinkoff AM, Brady B, Nielsen K. Workplace prevention and musculoskeletal injuries in nurses. Journal of Nursing Administration2003;33(3):153-8.
43. Byrns G, Reeder G, Jin G, Pachis K. Risk factors for work-related low back pain in registered nurses, and potential obstacles in using mechanical lifting devices. Journal of Occupational and Environmental Hygiene2004;1(1):11-21.
44. Pompeii LA, Lipscomb HJ, Schoenfisch AL, Dement JM. Musculoskeletal Injuries Resulting From Patient Handling Tasks Among Hospital Workers. American Journal of Industrial Medicine. [Article]. 2009 Jul;52(7):571-8.
45. Alamgir H, Cvitkovich Y, Yu S, Yassi A. Work-related injury among direct care occupations in British Columbia, Canada. Occupational and Environmental Medicine2007;64:769-75.
46. Power JD, Perruccio AV, Desmeules M, Lagace C, Badley EM. Ambulatory physician care for musculoskeletal disorders in Canada. Journal of Rheumatology2006;33(1):133-9.
47. Crook J, Moldofsky H. Prognostic indicators of disability after a work-related musculoskeletal injury. Fibromyalgia, Chronic Fatigue Syndrome, and Repetitive Strain Injury1995:155-9.
48. Koehoorn M, Demers PA, Hertzman C, Village J, Kennedy SM. Work organization and musculoskeletal injuries among a cohort of health care workers. Scandinavian Journal of Work Environment & Health2006;32(4):285-93.
49. Bot SDM, Terwee CB, van der Windt D, van der Beek AJ, Bouter LM, Dekker J. Work related physical and psychosocial risk factors for sick leave in patients with neck or upper extremity complaints. International Archives of Occupational and Environmental Health 2007 Aug;80(8):733-41.
50. Arjas E. A Graphical Method for Assessing Goodness of Fit in Cox Proportional Hazards Model. Journal of the American Statistical Association1988;83(401):204-12.
51. Struthers CA, Kalbfleisch JD. Misspecified Proportional Hazard Models. Biometrika 1986; 73(2):363-9.
52. Bretagnolle J, Hubercarol C. Effects of Omitting Covariates in Cox Model for Survival Data. Scandinavian Journal of Statistics1988;15(2):125-38.
53. Aalen O. A Model for Nonparametric Regression Analysis Of Counting Processes. Lecture Notes in Statistics. [Lecture Notes]. 1980;2:1-25.
54. Aalen O. Nonparametric Inference for a Family of Counting Processes. Annals of Statistics 1978; 6(4):701-26.
55. Buckley JD. Additive and Multiplicative Models for Relative Survival Rates. Biometrics 1984; 40(1):51-62.
56. Berkson J, Gage RP. Calculation of Survival Rates for Cancer. Proceedings of the Staff Meetings of the Mayo Clinic 1950; 25(11):270-86.
57. Andersen PK, Vaeth M. Simple Parametric and Nonparametric Models for Excess and Relative Mortality. Biometrics 1989; 45(2):523-35.
58. Mau J. A comparison of counting process models for complicated life histories,. Applied Stochastic Models and Data Analysis,1988 December, ;Volume 4 (Issue 4): Pages 283-98.

59. Mau J. On a Graphical Method for the Detection of Time-dependent Effects of Covariates in Survival Data. Applied Statistics-Journal of the Royal Statistical Society Series C1986;35(3):245-55.
60. McKeague IW, Utikal KJ. Goodness-of-Fit Tests for Additive Hazards and Proportional Hazards Models. Scandinavian Journal of Statistics 1991;18(3):177-95.
61. Huffer FW, McKeague IW. Survival analysis using additive risk models, technical report 396, : Department of Statistics, Stanford University,1987.
62. McKeague IW. Estimation for a Semimartingale Regression Model Using the Method of Sieves. Annals of Statistics 1986; 14(2):579-89.
63. Aalen OO. Further Results on the Nonparametric Linear Regression Model in Survival Analysis. Statistics in Medicine1993;12(17):1569-88.
64. Breslow NE, Day NE. Statistical methods in cancer research. Volume I - The analysis of case-control studies. IARC Sci Publ1980(32):5-338.
65. Pocock SJ, Gore SM, Kerr GR. Long term survival analysis: the curability of breast cancer. Stat Med1982;1(2):93-104.
66. Pierce DA, Preston DL. Hazard Function Modeling for Dose-Response Analysis of Incidence in the Atomic Bomb Survivor Data. Prentice, R L And D J Thompson Atomic Bomb Survivor Data: Utilization and Analysis; Meeting, Alta, Utah, USA, Sept 12-16, 1983 Ix+289p Siam (Society for Industrial and Applied Mathematics): Philadelphia, Pa, USA Illus Paper1984:51-66.
67. Breslow NE, Day NE. International Agency for Research on Cancer (IARC) Scientific Publications No. 82. Statistical Methods in Cancer Research Vol. II The Design and Analysis of Cohort Studies. Breslow, N E And N E Day Iarc (International Agency for Research on Cancer) Scientific Publications, No 82 Statistical Methods in Cancer Research, Vol Ii: The Design and Analysis of Cohort Studies Xii+406p Iarc: Lyon, France; (Dist In the USA by Oxford University Press: New York, New York, USA) Illus1987:XII+406P.
68. Huffer FW, McKeague IW. Weighted Least-Squares Estimation for Aalen Additive Risk Model. Journal of the American Statistical Association 1991;86(413):114-29.
69. Cox DR, Oakes D. Analysis of Survival Data,. London: Chapman & Hall; 1984.
70. Martinussen T, Scheike TH. A flexible additive multiplicative hazard model. Biometrika2002;89(2):283-98.
71. Torner A. Proportional Hazards and Additive Regression Analhysis of Survival for Severe Breast Cancer. Technical Report. Stockholm: Stockholm University, Mathematical Statistics DoM;2004.
72. Cao H. A Comparison Between the Additive and Multiplicative Risk Models. Que'bec: Universite' Laval; 2005.
73. Bhattacharyya M, Klein JP. A note on testing in Aalen's additive hazards regression models. Statistics in Medicine2005;24(14):2235-40.
74. Allison PD. Survival Analysis Using SAS: A Practical Guide. First ed: SAS Publishing; 1995.
75. Kaplan EL, Meier P. Nonparametric Estimation From Incomplete Observations. Journal of the American Statistical Association1958;53(282):457-81.
76. Hosmer DW, Lemeshow S, May S. Applied Survival Analysis: Regression Modeling of Time to Event Data. Second ed: Wiley-Interscience; 2008.

77. Cox DR, Snell EJ. A General Definition of Residuals. Journal of the Royal Statistical Society Series B-Statistical Methodology1968;30(2):248-&.
78. Barlow WE, Prentice RL. Residuals for Relative Risk Regression. Biometrika1988; 75(1):65-74.
79. Therneau TM, Grambsch PM, Fleming TR. Martingale-Based Residuals for Survival Models. Biometrika1990; 77(1):147-60.
80. Schaubel DE, Wei GH. Fitting semiparametric additive hazards models using standard statistical software. Biometrical Journal2007;49:719-30.
81. Kim JH, Lee SY. A Goodness-of-fit test based on martingale residuals for the additive risk model. The Korean Journal of Applied Statistics1996;9:75-89.
82. Howell AM. SAS Macro for the Additive Hazards Model: Medical College of Wisconsin, Biostatistics; 2007.
83. Black TR. The effect of a transfer, lifting and repositioning (TLR) injury prevention program on musculoskeletal injury rates among direct care workers [Thesis]. Saskatoon: University of Saskatchewan; 2009.
84. Howell AM, Klein JP. A SAS macro for the additive hazards regression model. American Statistical Association - 1996 Proceedings of the Statistical Computing Section1996: 282-7.
85. Rosato R, Ciccone G, Bo S PG, Merletti F, D. G. Evaluating cardiovascular mortality in type 2 diabetes patients: an analysis based on competing risks Markov chains and additive regression models. J Eval Clin Pract2007;13(3):422-8.
86. Therneau TM, Grambsch PM. Modeling Survival Data: Springer-Veriag, New York; 2000.

9. Appendix

Macro for the Aalen's Additive Hazards Model

```
/*

This macro is written by Alicia M. Howell, MS
For further information contact her at:
alicia@hp06.biostat.mcw.edu

---------------------------------CUT HERE ----------------------------------
*/

%macro additive(dataset,siglevel,timeunit,effects,option,contrast,outdat1,
outdat2);

* dataset: data set that contains time, censor indicator, covariates;
* siglevel: significance level;
* timeunit: unit of time;
* effects: the covariates;
* option: character vector defining which options are chosen;
* contrast: contrast vector (or matrix);
* outdat1: first output data set;
* outdat2: second output data set;

    use &dataset;
    read all var _num_ into imldat;
    goodrow=LOC(((imldat=.)  [,+])=0);   * searches for missing values;
    missing=LOC(((imldat=.)  [,+])^=0); * identifies rows w/missing values;
    imldat=imldat[goodrow,];           * deletes rows with missing values;

    n=nrow(imldat);     * number of rows in Y;
    col=ncol(imldat);
    pprime=col-1;       * number of columns in Y (baseline + covariates);
    p=pprime-1;         * number of covariates;
    t=imldat[,1];       * death or censored time;
    c=imldat[,2];       * 1=uncensored, 0=censored;
    zero=j(n,1,0);
    A=j(n,1,0);         * n-vector whose ith element is 1 if subject i
                            experiences event;
    B=j(pprime,1,0);    * estimates for Betas;
    s=j(pprime,1,0);    * decoy for Betas;
    betatime=j(n,pprime,0);   * Betas over time;
    cov=j(pprime,pprime,0);   * covariance for Betas;
    cm=j(pprime,pprime,0);    * decoy for covariance;
```

```
    var=j(pprime,1,0);        * variance of Betas;
    stdev=j(pprime,1,0);      * stdev of Betas;
    LCI=j(pprime,1,0);        * lower confidence bound for each estimate;
    UCI=j(pprime,1,0);        * upper confidence bound for each estimate;
    sdtime=j(n,pprime,0);   * stdev of Betas over time;
    Lcontime=j(n,pprime,0);   * lower conf bound over time for each est;
    Ucontime=j(n,pprime,0);   * upper conf bound over time for each est;
    GLTEST=I(p);
    constant=j(p,1,0);
    GLTEST=constant || GLTEST;
    K=j(p,p,0);         * weight for U;
    U=j(p,1,0);         * test stat based on this vector;
    V=j(p,p,0);          * variance matrix for U;
    xy=j(n,2,0);


*Check to see if data is sorted in ascending order;
    do i=2 to n;
     if t[i-1]>t[i] then do;
      print 'Data not sorted by time in ascending order!';
      abort;
     end;
    end;

* Option [2,1] is for testing contrasts;
    if &option[2,1]={y} then do;
     rowcon=nrow(&contrast);
     conK=j(rowcon,rowcon,0);
     conU=j(rowcon,1,0);
     conV=j(rowcon,rowcon,0);
    end;

* Macro that checks if Y is singular;
    %macro rankmat(time,estimtes);
    rank=round(trace(ginv(Y)*Y));
    if rank^=min(n,pprime) then do;
     fintime=&time;            * final time for estimates;
     stop;
    end;
    %mend;

* Macro that creates confidence intervals for untied observations;
    %macro confint(est,covarian,alpha);
    %global stdev sdtime LCI UCI Lcontime Ucontime;
    zscore=probit(1-&alpha/2);
    do f=1 to pprime;
```

```
 stdev[f,]=sqrt(&covarian[f,f]);
 sdtime[i,f]=sqrt(&covarian[f,f]);   * stdev for betas thru time;
 LCI[f,]=&est[f,]-(zscore#stdev[f,]);
 UCI[f,]=&est[f,]+(zscore#stdev[f,]);
 Lcontime[i,f]=LCI[f,1];
 Ucontime[i,f]=UCI[f,1];
end;
%mend;

* Macro that creates confidence intervals for tied observations;
   %macro conftied(est,covarian,alpha);
   %global stdev sdtime LCI UCI Lcontime Ucontime;
   zscore=probit(1-&alpha/2);
   do ff=i to jj;
    do f=1 to pprime;
     stdev[f,]=sqrt(cov[f,f]);
     sdtime[ff,f]=sqrt(&covarian[f,f]);
     LCI[f,]=&est[f,]-(zscore#stdev[f,]);
     UCI[f,]=&est[f,]+(zscore#stdev[f,]);
     Lcontime[ff,f]=LCI[f,1];
     Ucontime[ff,f]=UCI[f,1];
    end;
   end;
   %mend;

* Creating Y matrix;
   Y=j(n,pprime,0);
   Y[,1]=1;          * baseline column is 1s;
   do q= 2 to pprime;
    Y[,q]=imldat[,q+1];
   end;
* Computing the estimates:;
   do i=1 to (n-1);

    if t[i]=t[1] then do;
     s[,1]=0;cm[,1]=0;
    end;
    else do;
     s[,1]=B[,1];cm=cov;
    end;

* If time(i) is not equal to time(i+1):;
     if t[i]^=t[i+1] then do;
      if c[i]=0 then do;          * for censored observation;
      B=s;
       do f=1 to pprime;
```

```
          betatime[i,f]=B[f,1];
           end;
         cov=cm;
         K=K;
         U=U;
         V=V;
         if &option[2,1]={y} then do;
          conK=conK;
          conU=conU;
          conV=conV;
         end;
         %confint(B,cov,&siglevel);
         Y[i,]=0;
         %rankmat((t[i]),B);
         end;

         if c[i]>0 then do;            * for uncensored observation;
         A[i]=1;
         X=inv(Y`*Y)*Y`;
         B=s+(X*A);
          do f=1 to pprime;
           betatime[i,f]=B[f,1];
            end;
         cov=cm+(X*(diag(A))*X`);
         K=inv(diag(GLTEST*inv(Y`*Y)*GLTEST`));
         U=U+(K*GLTEST*X*A);
         V=V+(K*GLTEST*X*diag(A)*X`*GLTEST`*K`);
         if &option[2,1]={y} then do;
          conK=inv(diag(&contrast*inv(Y`*Y)*&contrast`));
          conU=conU+(conK*&contrast*X*A);
          conV=conV+(conK*&contrast*X*diag(A)*X`*&contrast`*conK`);
         end;
         %confint(B,cov,&siglevel);
         Y[i,]=0;
         A[i]=0;
         %rankmat((t[i]),B);
         end;

        end;

* If time(i) is equal to time(i+1):;
        if t[i]=t[i+1] then do;
         d=c[i]+c[i+1];
         do j=i+2 to n;
          if t[i]=t[j] then do;
           d=d+c[j];    * d is the # of uncensored cases at time(i);
```

```
 end;
 else if t[i]^=t[j] then do;
  jj=j-1;        * jj is the last case number that is tied at time(i);
  j=n;
 end;
end;
if d=0 then do;                  * for the censored tied times;
 B=s;
 do ff=i to jj;
  do f=1 to pprime;
   betatime[ff,f]=B[f,1];
    end;
   end;
 cov=cm;
 K=K;
 U=U;
 V=V;
 if &option[2,1]={y} then do;
  conK=conK;
  conU=conU;
  conV=conV;
 end;
 %conftied(B,cov,&siglevel);
 do m=i to jj;
  Y[m,]=0;
 end;
 i=jj;
 %rankmat((t[i]),B);
end;
if d>0 then do;                  * for the uncensored tied times;
 do dd=i to jj;
   if c[dd]=1 then A[dd]=1;
 end;
 X=inv(Y`*Y)*Y`;
 B=s+(X*A);
 do ff=i to jj;
  do f=1 to pprime;
   betatime[ff,f]=B[f,1];
    end;
   end;
 cov=cm+(X*(diag(A))*X`);
 K=inv(diag(GLTEST*inv(Y`*Y)*GLTEST`));
 U=U+(K*GLTEST*X*A);
 V=V+(K*GLTEST*X*diag(A)*X`*GLTEST`*K`);
 if &option[2,1]={y} then do;
  conK=inv(diag(&contrast*inv(Y`*Y)*&contrast`));
```

```
        conU=conU+(conK*&contrast*X*A);
        conV=conV+(conK*&contrast*X*diag(A)*X`*&contrast`*conK`);
       end;
       %conftied(B,cov,&siglevel);
       do m=i to jj;
        Y[m,]=0;
        A[m]=0;
       end;
       i=jj;
       %rankmat((t[i]),B);
      end;
     end;

   end;

   fincase=i;  * final case number, where estimates are still estimable;
   restime=t[1:fincase,];       * restricted time interval for estimates;
   Lcontime=Lcontime[1:fincase,];
   Ucontime=Ucontime[1:fincase,];
   betatime=betatime[1:fincase,];
   sdtime=sdtime[1:fincase,];

* Create BStime which contains parameter estimates & standard deviations
over time;
   BStime=j(fincase,2*pprime+1,0);
   BStime[,1]=restime;          * first column is time;
   BStime[,2]=betatime[,1];     * second column is BO estimate;
   BStime[,3]=sdtime[,1];       * third column is standard deviation(BO);
   do i=2 to pprime;
    BStime[,i*2]=Betatime[,i];  * even columns are B estimates;
    BStime[,i*2+1]=sdtime[,i];  * odd columns are st.dev;
   end;
   BStime=BStime[1:fincase,];   * eliminates final rows where estimate is
                          not estimable (YprimeY not full rank);

* GLOBAL TEST;
   gltstat=U`*inv(V)*U;
   zstat=sqrt(gltstat);
   dfgltest=p;
   pval=1-probchi(gltstat,dfgltest);
   col1={"Chi-Square"};
   col2={"d.f"};
   col3={"p-value"};
   timecol={"Time"};
   lab={" "};
   blankcol={" "};
```

```
      mattrib restime colname=timecol label=lab;
      mattrib fintime colname=blankcol label=lab format=4.2;
      mattrib gltstat colname=col1 label=lab format=10.4;
      mattrib dfgltest colname=col2 label=lab format=3.;
      mattrib pval colname=col3 label=lab format=6.4;

* INDIVIDUAL effects;
      indchi=j(p,1,0);
      indpval=j(p,1,0);
      do i=1 to p;
       indchi[i]=U[i]##2/V[i,i];
       indpval[i]=1-probchi(indchi[i],1);
      end;
      inddf=j(p,1,1);
      col4={"Effect"};
      mattrib indchi colname=col1 label=lab format=10.4;
      mattrib inddf colname=col2 label=lab format=3.;
      mattrib indpval colname=col3 label=lab format=6.4;
      mattrib &effects colname=col4 label=lab;


* contrasts;
      if &option[2,1]={y} then do;
       statcon=conU`*inv(conV)*conU;
       dfcon=rowcon;
       pvalcon=1-probchi(statcon,dfcon);
       mattrib statcon colname=col1 label=lab format=10.4;
       mattrib dfcon colname=col2 label=lab format=3.;
       mattrib pvalcon colname=col3 label=lab format=6.4;
       row1={"Contrast Matrix"};
       mattrib &contrast rowname=row1 label=lab;
      end;


      print '          Additive Hazards Model          ',,,;
      mattrib n colname=lab label=lab;
      misnames={"Case #"};
      mattrib missing colname=misnames label=lab;
      if nrow(missing)>0 then do;
       missing=t(missing);
       print 'The following observations have missing values and are excluded
from analysis:', missing,;
       print n 'observations used in analysis.';
      end;
      else do;
       print 'No missing data: all observations were used in analysis.';
```

```
      print n 'observations used.';
     end;
     print 'Estimates are restricted to the time interval 0 to' fintime,;
     print '              Global Test               ';
     print  gltstat dfgltest pval;
     print "    ", "  ", "    ";
     print '          Analysis of Variance          ';
     print &effects indchi inddf indpval;
     print "    ", "  ", "    ";


* Printing contrast output;
     if &option[2,1]={y} then do;
      print '       Test of Linear Combinations     ';
      print &contrast;
      print statcon dfcon pvalcon;
     end;

* Printing beta estimate, standard deviation at each time;
     if &option[1,1]={y} then do;
      endcount=2*pprime;
      c1={"Beta"};
      c2={"Standard Deviation"};
      mattrib parname colname=blankcol label=lab;
      mattrib betaspr colname=c1 label=lab format=10.4;
      mattrib stdevspr colname=c2 label=lab format=10.4;
      j=0;
      temp=t(&effects);
      tempnew='Baseline' || temp;
      do i=2 to endcount by 2;
       betaspr=BStime[,i];
       stdevspr=BStime[,i+1];
       j=j+1;
       parname=tempnew[,j];
       print 'Cumulative estimate and standard deviation for:' parname;
       print restime betaspr stdevspr;
      end;
     end;


* Line Plots;
     if &option[4,1]={y} then do;
      xy=j(fincase,2,0);
      xy[,1]=restime;
      temp=t(&effects);
      names='Baseline' || temp;
```

```
      do gg=1 to pprime;
       xy[,2]=betatime[,gg];
       call pgraf(xy,'*',&timeunit,names[,gg]);
      end;
     end;


* Creating first output dataset;
     if &option[3,1]={y} then do;
     NEW1= restime || betatime || sdtime || Lcontime || Ucontime;
       create &outdat1 from NEW1;
      append from NEW1;
     end;

* Creating second output dataset;
     if &option[5,1]={y} then do;
      NEW2=U || V;
      create &outdat2 from NEW2;
      append from NEW2;
     end;



%mend;
```

Macro for the Lin and Ying Additive Hazards Model

```
/*****************************************/
/*                                       */
/*   Ling & Ying's's additive model      */
/*     This macro also produces survival */
/*     estimation for a given patient.   */
/*                                       */
/*****************************************/



%macro est(indata, time, event, covlist, zvec);


/*   Find out unique event times    */
proc sort data=&indata out=tempdata; by &time descending &event; run;


data etime;
     set tempdata; by &time descending &event;
     if first.&time;
     if &event;
     keep &time;
run;


data otime;
     set tempdata; by &time;
     if first.&time;
     keep &time;
run;


proc iml;
     use &indata;
     read all var {&time} into time;
     read all var {&event} into event;
     read all var {&covlist} into zmat;
     close &indata;

     use etime;
     read all var {&time} into etime;
     close etime;
```

```
etime=0//etime;

use otime;
read all var {&time} into otime;
close otime;

otime=0//otime;

numobs=nrow(time);
numetime=nrow(etime);
numotime=nrow(otime);
numcov=ncol(zmat);


Amat=j(numcov,numcov,0);
Bmat=j(numcov,numcov,0);
Uvec=j(1,numcov,0);
ybar=j(numetime,1,0);
do i=2 to numetime;

   sumy=0;
   sumyz=j(1,numcov,0);
   do j=1 to numobs;
         if time[j]>=etime[i] then do;
            sumy=sumy+1;
            sumyz=sumyz+zmat[j,];
         end;
   end;
   ybar[i]=sumy;

   do j=1 to numobs;
         if time[j]=etime[i] & event[j]=1 then do;
            ztemp=zmat[j,]-sumyz/sumy;
            Uvec=Uvec+zmat[j,]-sumyz/sumy;
            Bmat=Bmat+t(ztemp)*ztemp;
         end;
   end;
end;

ybar=j(numotime,1,0);
do i=2 to numotime;
   sumy=0;
   sumyz=j(1,numcov,0);
   do j=1 to numobs;
         if time[j]>=otime[i] then do;
```

```
            sumy=sumy+1;
            sumyz=sumyz+zmat[j,];
         end;
      end;
      ybar[i]=sumy;

      zz=j(numcov,numcov,0);
      do j=1 to numobs;
         if time[j]>=otime[i] then do;
            ztemp=zmat[j,]-sumyz/sumy;
            zz=zz+t(ztemp)*ztemp;
         end;
      end;
      Amat=Amat+zz*(otime[i]-otime[i-1]);
   end;

print Amat Bmat;
*print Uvec;

   beta=Uvec*inv(Amat);
   sigma=inv(Amat)*Bmat*inv(Amat);

   create best from beta;
   append from beta;
   close best;


   create cov from sigma;
   append from sigma;
   close cov;


   se=j(1,numcov,0);
   do i=1 to numcov;
      se[1,i]=sqrt(sigma[i,i]);
   end;

   out=beta//se;

   create estout from out;
   append from out;
   close estout;


   /*   Find cumulative baseline hazard  */
   bzmat=j(numobs,1,0);
```

```
    do i=1 to numobs;
       bzmat[i]=zmat[i,]*t(beta);
    end;

    chaz=j(numotime,1,0);
    ctemp=0;
    do i=2 to numotime;

       yhaz=0;
       dn=0;
       do j=1 to numobs;
            if time[j]>=otime[i] then yhaz=yhaz+bzmat[j];
            if time[j]=otime[i] & event[j]=1 then dn=dn+1;
       end;
       yhaz=yhaz*(otime[i]-otime[i-1]);

       ctemp=ctemp+(dn-yhaz)/ybar[i];
       chaz[i]=ctemp;
    end;

/*
    out=otime||chaz;
    create haz from out[colname={'time' 'base_chaz'}];
    append from out;
    close haz;
*/

    zvec={&zvec};

    bz=beta*t(zvec);
    bzt=otime*bz;

    hazz=chaz+bzt;


    surv=j(numotime,1,0);
    do i=1 to numotime;
       surv[i]=exp(-hazz[i]);
    end;

    yz=j(numotime,numcov,0);
    ybar=j(numotime,1,0);
    do i=2 to numotime;

       yhaz=0;
       dn=0;
```

```
        do j=1 to numobs;
            if time[j]>=otime[i] then do;
                ybar[i]=ybar[i]+1;
                yz[i,]=yz[i,]+zmat[j,];
            end;
        end;
    end;


    Gtz=j(numotime,numcov,0);
    Dt=j(numotime,numcov,0);
    dtemp=j(1,numcov,0);
    term1=j(numotime,1,0);
    t1temp=0;
    do i=2 to numotime;
        Gtz[i,]=Gtz[i-1,]+(zvec-yz[i,]/ybar[i])*(otime[i]-otime[i-1]);

        do j=1 to numobs;
            if time[j]=otime[i] & event[j]=1 then do;
                dtemp=dtemp+(zmat[j,]-yz[i,]/ybar[i])/ybar[i];
                t1temp=t1temp+1/ybar[i]/ybar[i];
*obst=time[j]; obse=event[j];
*print t1temp obst obse j;
            end;
        end;
        Dt[i,]=dtemp;
        term1[i]=t1temp;
    end;


    svar=j(numotime,1,0);
    term2=j(numotime,1,0);
    term3=j(numotime,1,0);
    do i=2 to numotime;
        term2[i]=Gtz[i,]*inv(Amat)*Bmat*inv(Amat)*t(Gtz[i,]);
        term3[i]=Gtz[i,]*inv(Amat)*t(Dt[i,])#2;
        svar[i]=surv[i]#surv[i]#(term1[i]+term2[i]+term3[i]);
    end;

    survSE=svar##0.5;

*print otime ybar surv term1 term2 term3 svar;

    out=otime||chaz||surv||survSE;
    create adjsurv from out[colname={'time' 'baseline_chaz' 'surv' 'SE'}];
    append from out;
```

```
        close adjsurv;
    quit;

    proc print data=best; run;
    proc print data=cov; run;
    proc print data=adjsurv; run;


    %mend;
```

Code for the Lin and Ying Additive Hazards model

```
Options mprint=on notes source ls=85 nocenter nodate;

* For SHR & RQHR;
* Creaet Permanent data set for Saskatoon Health Region (SHR) & Regina Qu'Appelle Health
Region;
libname SHR_RQHR 'C:\Documents and Settings\sas862\My
Documents\MSc_Thesis\InjuryData\NewSelectedData';
* libname SHR 'C:\Documents and Settings\sas862\My
Documents\MSc_Thesis\InjuryData\NewSelectedData';
* libname RQHR 'C:\Documents and Settings\sas862\My
Documents\MSc_Thesis\InjuryData\NewSelectedData'; *For Department;
* libname SHR_RQHR 'F:\Msc_Thesis\Injury-Data-Tim\Selected Injury
Data\NewSelectedData'; * For USB Port;
Run;

%include 'C:\Documents and Settings\sas862\My Documents\MSc_Thesis\Survival
Analysis\Additive Model\LY\LY_Surv.txt';

Title "Lin and Ying's Additive Model";

Data ly_One;
        set SHR_RQHR.Additive_072010_04;
        if time=0 then time=0.1;
        if time=. then delete;
        if Age=0 then Age=0.1;
        if Age=. then Age=0.1;
Run;

Data SHR_RQHR.ly_One;
        Set ly_One;
Run;

* options mprint mlogic symbolgen;
```

```
%est (SHR_RQHR.ly_One, Time, Censor, SHR1 Nurses_Nurse_Aide Back_Neck_Shld, 1 1 1);
quit;

Data Name;
        Item='Estimate'; output;
        Item='SE'; output;
Run;

Data estout;
        merge Name estout (rename = (Col1=SHR Col2=Nurses Col3=BackInjuries));
Run;

proc transpose data=estout out= estout_tran;
        id item;
Run;

ods listing close;
ods rtf file= 'C:\Documents and Settings\sas862\My Documents\MSc_Thesis\Survival
Analysis\Additive Model\LY\LYOutput\LYFinal_111310.rtf';
Title 'Estimate of the L-Y Model-Final';
Data estout_new;
        set estout_tran (rename=(_name_=Variable));
        Chisq=(estimate/se)**2;
        Pr=1-probchi(chisq,1);
        Llt=estimate-1.96*se; * For Lower CI;
        Ult=estimate+1.96*se; * for Upper CI;

Proc print;
Run;

ods rtf close;
ods listing;
Title;


Data _null_;
        set adjsurv;
        slow=surv-1.96*se;
        sup=surv+1.96*se;
        file 'C:\Documents and Settings\sas862\My Documents\MSc_Thesis\Survival
Analysis\Additive Model\LY\LY_Output_111310.txt';
        put time surv slow sup;
        format _all_ 7.4;
run;
```

```
Data LY_Surv_CI;
        set adjsurv;
        slow=surv-1.96*se;
        sup=surv+1.96*se;
        * file 'C:\Documents and Settings\sas862\My Documents\MSc_Thesis\Survival
Analysis\Additive Model\LY\LY_Output_111310_01.txt';
        put time surv slow sup;
        format _all_ 7.4;
run;

/*

Data LY_Plot;
        Set LY_Surv_CI;
        If time>65 then delete;
Run;

Title "Fugure Lin and Ying's Additive Model";

axis1 label=(j=c 'Month') minor=none;
axis2 label=(a=90 j=c "Estimated Cumulative Hazard Rate for SHR") minor=none;

Symbol1 interpol=stepjr c=black l=1 value=none;
Symbol2 interpol=stepjr c=blue l=3 value=none;
Symbol3 interpol=stepjr c=red l=3 value=none;

proc gplot data=LY_Plot;
        plot surv*time slow*time sup*time/overlay haxis=axis1 vaxis=axis2;
Run;
Quit;

/*
proc print; run;
*
*/

/* For Univariate L-Y Model */

Proc contents data=SHR_RQHR.ly_one;
Run;

/* For Intervention and Control (SHR1) */
%est (SHR_RQHR.ly_One, Time, Censor, SHR1, 1);
quit;

Data LY_SHR;
```

```
        Item='Estimate'; output;
        Item='SE'; output;
Run;

Data estout_SHR;
        merge LY_SHR estout (rename = (Col1=SHR));
Run;

proc transpose data=estout_SHR out= estout_SHR_tran;
        id item;
Run;

/* For  Age */
%est (SHR_RQHR.ly_One, Time, Censor, Age , 18);
quit;

Data LY_Age;
        Item='Estimate'; output;
        Item='SE'; output;
Run;

Data estout_Age;
        merge LY_Age estout (rename = (Col1=Age));
Run;

proc transpose data=estout_Age out= estout_Age_tran;
        id item;
Run;

/* For Female */
%est (SHR_RQHR.ly_One, Time, Censor, Female, 1);
quit;

Data LY_Female;
        Item='Estimate'; output;
        Item='SE'; output;
Run;

Data estout_Female;
        merge LY_SHR estout (rename = (Col1=Female));
Run;

proc transpose data=estout_Female out= estout_Female_tran;
        id item;
Run;
```

```
/* For Occupation: Nurses and Nurses Aide (NNA) */
%est (SHR_RQHR.ly_One, Time, Censor, Nurses_Nurse_Aide, 1);
quit;

Data LY_NNA;
        Item='Estimate'; output;
        Item='SE'; output;
Run;

Data estout_NNA;
        merge LY_NNA estout (rename = (Col1=NNA));
Run;

proc transpose data=estout_NNA out= estout_NNA_tran;
        id item;
Run;

/* For Body Parts: Back, Neck & Shoulders (BNS) */
%est (SHR_RQHR.ly_One, Time, Censor, Back_Neck_Shld, 1);
quit;

Data LY_BNS;
        Item='Estimate'; output;
        Item='SE'; output;
Run;

Data estout_BNS;
        merge LY_BNS estout (rename = (Col1=BNS));
Run;

proc transpose data=estout_BNS out= estout_BNS_tran;
        id item;
Run;

/* For Hospital Size: Large */
%est (SHR_RQHR.ly_One, Time, Censor, Large, 1);
quit;

Data LY_Large;
        Item='Estimate'; output;
        Item='SE'; output;
Run;

Data estout_Large;
        merge LY_Large estout (rename = (Col1=Large));
Run;
```

```
proc transpose data=estout_Large out= estout_Large_tran;
        id item;
Run;

/* For Hospital Size: Medium */
%est (SHR_RQHR.ly_One, Time, Censor, Medium, 1);
quit;

Data LY_Medium;
        Item='Estimate'; output;
        Item='SE'; output;
Run;

Data estout_Medium;
        merge LY_Medium estout (rename = (Col1=Medium));
Run;

proc transpose data=estout_Medium out= estout_Medium_tran;
        id item;
Run;

/* For Hospital Size: Small */
%est (SHR_RQHR.ly_One, Time, Censor, SMall, 1);
quit;

Data LY_Small;
        Item='Estimate'; output;
        Item='SE'; output;
Run;

Data estout_Small;
        merge LY_Small estout (rename = (Col1=Small));
Run;

proc transpose data=estout_Small out= estout_Small_tran;
        id item;
Run;

/* For Hospital Size: Large & Small */
%est (SHR_RQHR.ly_One, Time, Censor, Large Small, 1 1);
quit;

Data LY_Large_Small;
        Item='Estimate'; output;
        Item='SE'; output;
```

```
Run;

Data estout_Large_Small;
        merge LY_Large_Small estout (rename = (Col1=Large Col2=Small));
Run;

proc transpose data=estout_Large_Small out= estout_Large_Small_tran;
        id item;
Run;


ods listing close;
ods rtf file= 'C:\Documents and Settings\sas862\My Documents\MSc_Thesis\Survival
Analysis\Additive Model\LY\LYOutput\LY_Uni_111310.rtf';
Title 'Estimate of the L-Y Model-Univariate';
Data estout_SHR_new;
        set estout_Age_tran estout_Female_tran estout_SHR_tran estout_NNA_tran
estout_BNS_tran estout_Large_tran estout_Medium_tran estout_Small_tran
estout_Large_Small_tran;
        rename _name_=Variable;
        Chisq=(estimate/se)**2;
        Pr=1-probchi(chisq,1);
        Llt=estimate-1.96*se; * For Lower CI;
        Ult=estimate+1.96*se; * for Upper CI;

Proc print;
Run;

Title;

ods rtf close;
ods listing;

/* L-Y Multivariate with Siginificant Variable: SHR1 NNA BNS Large Small */

%est (SHR_RQHR.ly_One, Time, Censor,SHR1 Nurses_Nurse_Aide Back_Neck_Shld Large
Small, 1 1 1 1 1);
quit;

Data LY_All5;
        Item='Estimate'; output;
        Item='SE'; output;
Run;

Data estout_All5;
```

```
        merge LY_All5 estout (rename = (Col1=SHR Col2=NNA Col3=BNS Col4=Large
Col5=Small));
Run;

proc transpose data=estout_All5 out= estout_All5_tran;
        id item;
Run;

ods listing close;
ods rtf file= 'C:\Documents and Settings\sas862\My Documents\MSc_Thesis\Survival
Analysis\Additive Model\LY\LYOutput\LY_Multi_All5_111310.rtf';
Title 'Estimate of the L-Y Model-Multivariate All 5 Variable';
Data estout_SHR_new;
        set estout_All5_tran;
        rename _name_=Variable;
        Chisq=(estimate/se)**2;
        Pr=1-probchi(chisq,1);
        Llt=estimate-1.96*se; * For Lower CI;
        Ult=estimate+1.96*se; * for Upper CI;

Proc print;
Run;

Title;

ods rtf close;
ods listing;
```

R Code for Good-ness of Fit (Ling and Ying Additive Hazards Models)

```
# Arjas Plot
# At the University
# postscript("d:/sabuj/arjas.ps",horizontal=F)
postscript ("C:/Documents and Settings/sas862/My Documents/MSc_Thesis/Survival
Analysis/SAS-Macro-Additive
Model/LY/GOF/Additive_LY/Output/arjas_shr.ps",horizontal=F)
par(mfrow=c(3,2),oma=c(1,1,3,1),font=1,font.main=1)

# Arjas for Health Region
#At the University
mat1<-matrix(scan("C:/Documents and Settings/sas862/My
Documents/MSc_Thesis/Survival Analysis/SAS-Macro-Additive
Model/LY/GOF/Additive_LY/Output/SHR1.txt"),ncol=7,byrow=TRUE)

# mat1 <- matrix(scan("d:/sabuj/SHR1.txt"),ncol=7,byrow=TRUE)
```

```
#At Home
# mat1 <- matrix(scan("d:/sabuj/SHR1.txt"),ncol=7,byrow=TRUE)

r1t <- mat1[,1]

nt0 <- mat1[,2]
htz0 <- mat1[,3]
mart0 <- mat1[,4]

nt1 <- mat1[,5]
htz1 <- mat1[,6]
mart1 <- mat1[,7]

xlim1 <- c(0,80)
ylim1 <- c(0,180)

plot(nt0,htz0,type="s",bty="l",lty=1,lwd=1,xlim=xlim1,ylim=ylim1,xlab="",ylab="",xaxt="n
",yaxt="n",axes=F)
axis(1,pos=0,at=c(0,20,40,60,80),cex=0.3)
axis(2,pos=0,at=c(0,20,40, 60, 80,100,120,150,180),cex=0.3)
par(new=T)
plot(nt1,htz1,type="s",bty="l",lty=2,lwd=1,xlim=xlim1,ylim=ylim1,xlab="",ylab="",xaxt="n
",yaxt="n",axes=F)
title(xlab="Number of Repeated Injuries",ylab="Expected Cumulative Hazard Rates",
cex=0.5)
title(main="Arjas Plot for Intervention and Control Group",cex=0.4)
charvec <- c("Control", "Intervention")
legend(0, 80, charvec, lty=c(0,1),lwd=c(1,1),bty="n", adj=0,cex=1)

# Arjas for Back, Neck and Shoulders
postscript ("C:/Documents and Settings/sas862/My Documents/MSc_Thesis/Survival
Analysis/SAS-Macro-Additive
Model/LY/GOF/Additive_LY/Output/arjas_BNS.ps",horizontal=F)
par(mfrow=c(3,2),oma=c(1,1,3,1),font=1,font.main=1)

# mat1 <- matrix(scan("d:/sabuj/BNS.txt"),ncol=7,byrow=TRUE)
# At the University
mat1 <- matrix(scan("C:/Documents and Settings/sas862/My
Documents/MSc_Thesis/Survival Analysis/SAS-Macro-Additive
Model/LY/GOF/Additive_LY/Output/BNS.txt"),ncol=7,byrow=TRUE)

r1t <- mat1[,1]

nt0 <- mat1[,2]
htz0 <- mat1[,3]
mart0 <- mat1[,4]
```

```
nt1 <- mat1[,5]
htz1 <- mat1[,6]
mart1 <- mat1[,7]

plot(nt0,htz0,type="s",bty="l",lty=1,lwd=1,xlim=xlim1,ylim=ylim1,
   xlab="",ylab="",xaxt="n",yaxt="n",axes=F)
axis(1,pos=0,at=c(0,20,40,60,80),cex=0.3)
axis(2,pos=0,at=c(0,20,40, 60, 80,100,120,150,200),cex=0.3)
par(new=T)
plot(nt1,htz1,type="s",bty="l",lty=2,lwd=1,xlim=xlim1,ylim=ylim1,
   xlab="",ylab="",xaxt="n",yaxt="n",axes=F)
title(xlab="Number of Repeated Injury",ylab="Expected Cumulative Hazard Rates",
cex=0.5)
title(main="Back, Neck Shoulders with Other Body Parts",cex=0.4)
charvec <- c("Others","BNS")
legend(0, 80, charvec, lty=c(0,1),lwd=c(1,1),bty="n", adj=0,cex=1)

# Arjas for Nurses and Nurses Aide
# mat1 <- matrix(scan("d:/sabuj/NNA.txt"),ncol=7,byrow=TRUE)
# At the University

postscript ("C:/Documents and Settings/sas862/My Documents/MSc_Thesis/Survival
Analysis/SAS-Macro-Additive
Model/LY/GOF/Additive_LY/Output/arjas_NNA.ps",horizontal=F)
par(mfrow=c(3,2),oma=c(1,1,3,1),font=1,font.main=1)

# At the University
mat1 <- matrix(scan("C:/Documents and Settings/sas862/My
Documents/MSc_Thesis/Survival Analysis/SAS-Macro-Additive
Model/LY/GOF/Additive_LY/Output/NNA.txt"),ncol=7,byrow=TRUE)

r1t <- mat1[,1]

nt0 <- mat1[,2]
htz0 <- mat1[,3]
mart0 <- mat1[,4]

nt1 <- mat1[,5]
htz1 <- mat1[,6]
mart1 <- mat1[,7]

plot(nt0,htz0,type="s",bty="l",lty=1,lwd=1,xlim=xlim1,ylim=ylim1,
   xlab="",ylab="",xaxt="n",yaxt="n",axes=F)
axis(1,pos=0,at=c(0,20,40,60,80),cex=0.3)
axis(2,pos=0,at=c(0,20,40, 60, 80,100,120,150,200),cex=0.3)
```

```
par(new=T)
plot(nt1,htz1,type="s",bty="l",lty=2,lwd=1,xlim=xlim1,ylim=ylim1,
  xlab="",ylab="",xaxt="n",yaxt="n",axes=F)
title(xlab="Number of Repeated Injury",ylab="Expected Cumulative Hazard Rates",
cex=0.5)
title(main="Nurses and Nurses Aide with Other Occupation",cex=0.4)
charvec <- c("Others","NNA")
legend(0, 80, charvec, lty=c(0,1),lwd=c(1,1),bty="n", adj=0,cex=1)

mtext("Arjas Plot - common beta, different baselines",
     NORTH<-3, line=1, adj=0.5, cex=1.0, font=1,col="black", outer=TRUE)

dev.off()


# Martingale Plots
# postscript("d:/sabuj/mart.ps",horizontal=F)
# par(mfrow=c(3,2),oma=c(1,1,3,1),font=1,font.main=1)
# mat1 <- matrix(scan("d:/sabuj/SHR1.txt"),ncol=7,byrow=TRUE)

# Martingale Plots For Health Region - Intervention/Control

postscript ("C:/Documents and Settings/sas862/My Documents/MSc_Thesis/Survival
Analysis/SAS-Macro-Additive
Model/LY/GOF/Additive_LY/Output/Mart_shr.ps",horizontal=F)
par(mfrow=c(3,2),oma=c(1,1,3,1),font=1,font.main=1)

#At the University
mat1<-matrix(scan("C:/Documents and Settings/sas862/My
Documents/MSc_Thesis/Survival Analysis/SAS-Macro-Additive
Model/LY/GOF/Additive_LY/Output/SHR1.txt"),ncol=7,byrow=TRUE)

r1t <- mat1[,1]

nt0 <- mat1[,2]
htz0 <- mat1[,3]
mart0 <- mat1[,4]

nt1 <- mat1[,5]
htz1 <- mat1[,6]
mart1 <- mat1[,7]

xlim1 <- c(0,7.5)
ylim1 <- c(-15,15)

plot(r1t,mart0,type="s",bty="l",lty=1,lwd=1,xlim=xlim1,ylim=ylim1,
```

```
   xlab="",ylab="",xaxt="n",yaxt="n",axes=F)
axis(1,pos=-.04,at=c(0,2,4,6,8,10),cex=0.3)
axis(2,pos=0,at=c(-20,-15, -10, -5, 0,5,10,15,20),cex=0.3)
title(xlab="Month",ylab="Martingale Residuals", cex=0.5)
title(main="Repeated Injury for Intervention and Control Group",cex=0.4)

# Martingale Plots For Bak, Neck & Shoulders
# mat1 <- matrix(scan("d:/sabuj/BNS.txt"),ncol=7,byrow=TRUE)

postscript ("C:/Documents and Settings/sas862/My Documents/MSc_Thesis/Survival
Analysis/SAS-Macro-Additive
Model/LY/GOF/Additive_LY/Output/Mart_BNS.ps",horizontal=F)
par(mfrow=c(3,2),oma=c(1,1,3,1),font=1,font.main=1)

#At the University
mat1<-matrix(scan("C:/Documents and Settings/sas862/My
Documents/MSc_Thesis/Survival Analysis/SAS-Macro-Additive
Model/LY/GOF/Additive_LY/Output/BNS.txt"),ncol=7,byrow=TRUE)

r1t <- mat1[,1]

nt0 <- mat1[,2]
htz0 <- mat1[,3]
mart0 <- mat1[,4]

nt1 <- mat1[,5]
htz1 <- mat1[,6]
mart1 <- mat1[,7]

plot(r1t,mart1,type="s",bty="l",lty=1,lwd=1,xlim=xlim1,ylim=ylim1,
   xlab="",ylab="",xaxt="n",yaxt="n",axes=F)
axis(1,pos=0,at=c(0,2,4,6,8,10),cex=0.3)
axis(2,pos=0,at=c(-20,-15, -10, -5, 0,5,10,15,20),cex=0.3)
title(xlab="Months",ylab="Martingale Residuals", cex=0.5)
title(main="BNS - Back, Neck and Shoulders with Otheers Body Parts",cex=0.4)

#Martingale Plots For Nurses and Nurses Aide
# mat1 <- matrix(scan("d:/sabuj/NNA.txt"),ncol=7,byrow=TRUE)

postscript ("C:/Documents and Settings/sas862/My Documents/MSc_Thesis/Survival
Analysis/SAS-Macro-Additive
Model/LY/GOF/Additive_LY/Output/Mart_NNA.ps",horizontal=F)
par(mfrow=c(3,2),oma=c(1,1,3,1),font=1,font.main=1)

#At the University
```

```
mat1<-matrix(scan("C:/Documents and Settings/sas862/My
Documents/MSc_Thesis/Survival Analysis/SAS-Macro-Additive
Model/LY/GOF/Additive_LY/Output/NNA.txt"),ncol=7,byrow=TRUE)

r1t <- mat1[,1]

nt0 <- mat1[,2]
htz0 <- mat1[,3]
mart0 <- mat1[,4]

nt1 <- mat1[,5]
htz1 <- mat1[,6]
mart1 <- mat1[,7]

plot(r1t,mart1,type="s",bty="l",lty=1,lwd=1,xlim=xlim1,ylim=ylim1,
   xlab="",ylab="",xaxt="n",yaxt="n",axes=F)
axis(1,pos=0,at=c(0,2,4,6,8,10),cex=0.3)
axis(2,pos=0,at=c(-20,-15, -10, -5, 0,5,10,15,20),cex=0.3)
title(xlab="Months",ylab="Martingale Residuals", cex=0.5)
title(main="NNA - Nurses and Nurses Aide with Others Occupation",cex=0.4)

 mtext("Martingale Residual Process Plot - common beta, different baselines",
     NORTH<-3, line=1, adj=0.5, cex=1.0, font=1,col="black", outer=TRUE)

dev.off()
```

UNIVERSITY OF SASKATCHEWAN

Research Ethics Office

Box 5000 RPO University
1607 – 110 Gymnasium Place
NRC/PBI Building
Saskatoon SK S7N 4J8 Canada
Telephone: (306) 966-2975
Facsimile: (306) 966-2069

February 24, 2010

Sabuj Sarker
M.Sc. Candidate
Department of Community Health and Epidemiology
RUH-2708
107 Wiggins Road
Saskatoon, SK, S7N 5E5

Dear Mr. Sarker:

Thank you for your email requesting information about using data from a previous study conducted by Mr. Timothy R. Black for your master's thesis. Accordingly, as outlined in Article 3.3, 3.3 and 3.4 of the Tri-Council Policy Statement: Ethical Conduct for the Research involving Humans, 1998 (with 2000, 2002 updates) states that if personally identifiable information is accessible through any linkage with the data sample, REB approval shall be sought for the "secondary use" of data. Since you have assured our office that the data sets have been de-identified, this project is not subject to further ethics review.

It should be noted that although your project is exempt of ethics review; your project should be conducted in an ethical manner and in accordance with the information that you submitted.

It should be further noted that any deviation from the original methodology and/or research question should be brought to the attention of the Biomedical Research Ethics Board for further review.

Notwithstanding research investigators must ensure that the project is carried out in keeping with the Saskatchewan Health Information Protection Act (HIPA).

Sincerely,

Gordon McKay, Ph.D., Vice-Chair
Biomedical Research Ethics Board
University of Saskatchewan

/ber

Saskatoon Health Region

Research Services Unit
Strategic Health Information & Planning Services (SHIPS)
Joanne Franko, Manager
Suite 300 Saskatoon Square
410 22nd St E
Saskatoon, SK S7K 5T6
Phone: 306.655.3356 Fax: 306.655.3373

DATE: November 22, 2005

TO: Tim Black, P.T., Community Health and Epidemiology, U of S

FROM: Joanne Franko
Manager, Research Services Unit

RE: RESEARCH PROJECT ETHICS COMMITTEE (EC)#: 2005-168
PROJECT NAME: The Effect of a Transfer, Lifting and Repositioning Program on Musculoskeletal Injury Rates among Healthcare Workers in Selected Facilities within the Saskatoon Health Region.
PROTOCOL #: N/A

Saskatoon Health Region is pleased to provide you with operational approval of the above-mentioned research project.

Please advise me when the data collection phase of the research project is completed. I would also appreciate receiving a summary of the results for this research project. As well, any publications or presentations that result from this research should include a statement acknowledging the assistance of Saskatoon Health Region.

I would like to wish you every success with your project. If you have any questions, please contact our office at 655-3351.

Yours truly,

Joanne Franko, M.Sc.
Manager, Research Services Unit

cc: Dr. Syed Shah, Thesis Supervisor, Community Health and Epidemiology, U of S
Judy Metcalf, Manager, Occupational Health and Safety, SPH
Father Marc Miller, Mission Office, SPH

Regina Qu'Appelle
HEALTH REGION

Certificate of Approval
Research Ethics Board

PRINCIPAL INVESTIGATOR	Mr. Timothy R. Black	*Mailing Address:* Community Health & Epidemiology University of Saskatchewan Health Sciences Building 107 Wiggins Road Saskatoon SK S7N 5E5
APPROVAL DATE	May 29, 2006	
RQHR PROJECT #	REB-06-47	
TITLE	The effect of a Transfer, Lifting, and Repositioning (TLR) Program on musculoskeletal injury rates among healthcare workers in selected facilities within the Saskatoon Health Region	

CERTIFICATION

The protocol and consent form for the above named project have been reviewed by the Chair of the Regina Qu'Appelle Health Region Research Ethics Board and the experimental procedures were found to be acceptable on ethical grounds for research involving human subjects.

The Regina Qu'Appelle Health Region Research Ethics Board meets the standards outlined by Canada's Tri-Council Policy Statement for Ethical Conduct for Research Involving Humans.

The Regina Qu'Appelle Health Region Research Ethics Board has met the criteria for purposes of Section 29 of the *Health Information Protection Act*.

Please note that all future correspondence regarding this project must include the RQHR project number.

Best wishes in your continuing research endeavours.

for Dr. Elan Paluck, Chair
Regina Qu'Appelle Health Region
Research Ethics Board

/lgp
cc. Ms. C. Klassen, Corporate Services, WRC

This Certificate of Approval is valid provided there is no change in the experimental procedures. Any significant changes to the protocol must be reported to the Chair for the Board's consideration, in advance of implementation of such changes. You are required to provide a status report on an annual basis.

Research and Performance Support
2180 - 23rd Avenue • Regina • SK • S4S 0A5 • www.rqhealth.ca
Phone 306-766-5451 • Fax 306-766-5530 • Email linda.picot@rqhealth.ca • web www.mad-fax.com/rhc/crdp