Spring 5-8-2023

An Analysis of All-Cause Mortality on Patients with Sickle Cell Disease and Kidney Disease using Propensity Score Matching

Adam Garrison
garrisonam1@jacks.sfasu.edu

An Analysis of All-Cause Mortality on Patients with Sickle Cell Disease and Kidney Disease using Propensity Score Matching

An Analysis of All-Cause Mortality on Patients with Sickle Cell Disease and Kidney Disease using Propensity Score Matching

by

Adam Garrison, B.S.

Presented to the Faculty of the Graduate School of

Stephen F. Austin State University

In Partial Fulfillment

of the Requirements

For the Degree of

Master of Science

STEPHEN F. AUSTIN STATE UNIVERSITY

May 2023

An Analysis of All-Cause Mortality on Patients with Sickle Cell Disease and Kidney Disease using Propensity Score Matching

by

Adam Garrison, B.S.

APPROVED:

Jacob Turner, Ph.D., Thesis Director

Robert Henderson, Ph.D., Committee Member

Jeremy Becnel, Ph.D., Committee Member

Kent Riggs, Ph.D., Committee Member

Sheryll Jerez, Ph. D.
Interim Dean of Research and Graduate Studies

ABSTRACT

In this work, we provide an overview of the Cox proportional hazards model for time to event or survival analysis and the notion of propensity score matching to deal with confounding factors. A full analysis is reported in Chapter 2 concerning mortality for in-center dialysis patients with sickle cell disease to demonstrate the application of a general analysis strategy that has some logistical benefits over more traditional approaches to accounting for confounding variables. We also provide some insight and discussions on the challenges and future research questions that will emerge when trying to implement this strategy as a monitoring tool over time.

ACKNOWLEDGEMENTS

I would like to say thank you to my thesis advisor, Dr. Jacob Turner, whose support and advice helped me along the way on this project. Thank you to the committee members for their attention and all of their hard work. Thank you to Derek Blankenship at Fresenius Medical Care for being a mentor/helping with the SCD project. Thank you to the math department at SFA for their support. Lastly, thank you to my friends and family for their continued support.

CONTENTS

LIST OF FIGURES

1 INTRODUCTION

Survival analysis is a branch of statistics that analyzes the expected duration of time until an event occurs. Often, survival analysis is used in medical research, and the event being considered is death of a patient, hence the name survival analysis. This thesis is motivated by an internship experience working with Fresenius Medical Care for the Fall Semester of 2022 and Spring Semester of 2023. Fresenius Medical Care creates equipment for and provides dialysis treatments for patients with kidney disease. As there is no known cure for kidney disease except for a kidney transplant, patients often undergo dialysis treatments regularly for the rest of their lives. Fresenius collects data from their clinics in order to analyze and improve outcomes for patients. Fresenius is with patients for long periods of time, with typical outcomes for patients being either death or transplantation. Survival analysis allows Fresenius to get an idea of how long a patient on dialysis will live and assess what clinical factors and treatment strategies impact the outcome.

The thesis is structured as follows. In this chapter, a review is given on the main methodology of survival analysis in context of using a Cox proportional hazard model to compare two populations. To help the discussion, a motivating example using a publicly available data set is provided. Many survival analyses are conducted with data generated from observational studies. While this can be handled including additional covariates in the model, we introduce an alternative method for adjustment using propensity score matching. In Chapter 2, our focus will be on the analysis of the sickle cell disease (SCD) data. A thorough review of the analysis plan, data collection, data processing, analysis, and conclusions is provided. Chapter 3 provides a discussion of current challenges and solutions when performing survival analyses in more

adaptive but realistic ways. In Chapter 4, a brief discussion of life-lessons and general experiences of working in the hospital sector are discussed along with suggestions of future work in regards to propensity score matching and survival analysis.

1.1 Survival Analysis By Example

The example dataset was collected from patients with advanced lung cancer from the North Central Cancer Treatment Group [3]. The first six rows of the data set are provided in Figure 1.1.

	inst	time	status	age	sex	ph.ecog	ph.karno	pat.karno	meal.cal	wt.loss
1	3	306	2	74	Male	1	90	100	1175	NA
2	3	455	2	68	Male	0	90	90	1225	15
3	3	1010	1	56	Male	0	90	90	NA	15
4	5	210	2	57	Male	1	90	60	1150	11
5	1	883	2	60	Male	0	100	90	NA	0
6	12	1022	1	74	Male	1	50	80	513	0

Figure 1.1: Lung Cancer Data Set

There are two main variables to focus on. The variable "time" is the survival time of the patients in the clinic until either an event occurs (in this case, death), or the data is *censored*. Censoring is a common practice in survival analysis to let us know that at the time the study stopped and the data was collected some patients had not experienced the event of interest yet. We are still aware of how long they have survived currently but it is an underestimate of what their true survival time actually will be. The censoring described is referred to as *right censored*. In practicality, the reason for the censoring can be for additional reasons other than the termination of the study. This could, for example, be a patient leaving the trial for personal reasons or a completely different event occurred that altered our ability to follow up on the patient's status. In this dataset, a 1 in the status category indicates that the patient died, and a 2 indicates that the data has been censored.

In this dataset, the status variable and takes two values 1 and 2 that indicate the

different censoring statuses. There is nothing special about the choice of values for this variable. For example, most of the time, 0 and 1 are used instead. As long as the number indicating censoring is specified, any value for censoring status is allowed.

The additional variables listed in Figure 1.1 provide additional information for the patients. Performance scores are ratings that describe how well the patient can perform usual daily activities. For example, "ph.ecog" is the physician's rating of the patient's status on a scale of 0-5 using the Eastern Cooperative Oncology Group (ECOG) system. The ECOG system is graded on a 0-5 scale, with 0 being peak condition, and 5 being death. The Karnofsky system is a similar system to the ECOG system, but on a 100 point scale. Physicians and patients both rated patient's condition on this scale. Lastly, there are other measurements such as how many calories are in a patient's meal, and how much weight the patient has lost. While all of these variables may help determine how long the person survives, we will discuss how to handle these situations a little later. For now, we will introduce key definitions and general framework of a survivor model.

With both survival time and the event status of all of the patients, we must have some sort of function to describe survival data. Three functions that are of interest to us are the survivor function, the hazard function, and the cumulative hazard function. For a more detailed discussion, see [1]. Let the actual survival time of an individual be t, and let T be the random variable associated with the survival time. Suppose that this random variable has a probability distribution with underlying probability density function $f(t)$. The distribution function of T is then given by

$$F(t) = P(T \leq t) = \int_0^t f(u)du. \tag{1.1}$$

This function represents the probability that the survival time is less than some value t. This function is called the cumulative incidence function because it summarizes the cumulative probability of an event occurring before time t. The survivor function,

$S(t)$, is defined to be the probability that the survival time is greater than t, so

$$S(t) = P(T > t) = 1 - F(t). \tag{1.2}$$

The hazard function is used to express the risk, or hazard, of an event occurring for an individual at some time t, conditional on that individual surviving (or not having an event happen yet) until that time. The hazard function can be defined as follows

$$h(t) = \lim_{\delta t \to 0} \frac{Pr(t \leq T < t + \delta t | T \geq t)}{\delta t} \tag{1.3}$$

The notion of the definition is as follows. Consider a time interval $(t, t + \delta t)$. The numerator of this expression is the conditional probability that an event will occur within this interval, given that is has not occurred before. The denominator is the width of this interval. Dividing the numerator by the denominator gives us a rate of event occurrence per unit of time and taking the limit as the interval width goes to zero gives us an instantaneous rate of occurrence.

Next, we can use the properties of conditional probabilities to show some useful relationships. Consider the conditional probability in the numerator of our hazard function. Due to the properties of conditional probabilities, it can be expressed as

$$\frac{P(t \leq T < t + \delta t)}{P(T \geq t)} \tag{1.4}$$

which is equal to

$$\frac{F(t + \delta t) - F(t)}{S(t)}. \tag{1.5}$$

Upon reexamining the hazard function definition, we can see that

$$h(t) = \lim_{\delta t \to 0} (\frac{F(t + \delta t) - F(t)}{\delta t}) \frac{1}{S(t)}. \tag{1.6}$$

Note that the quantity in parentheses,

$$\lim_{\delta t \to 0} \frac{F(t + \delta t) - F(t)}{\delta t} \tag{1.7}$$

is the definition of the derivative of $F(t)$, the density function $f(t)$. Hence, we arrive at the result

$$h(t) = \frac{f(t)}{S(t)}. \tag{1.8}$$

With real survival data, censoring always comes into play, so the survival function $S(t)$ cannot simply be estimated as a simple step function. The most commonly used estimate for survival functions is the Kaplan-Meier estimate of the survival function [1]. If we consider time intervals t_j, where each j indicates at least one event has occurred, and d_j, the number of events (or deaths) that occurred in the interval t_j, and n_j, the number of individuals known to have not have an event occur or been censored up to time t_j the Kaplan-Meier estimator is defined as

$$\hat{S}(t) = \prod_{i:t_i \leq t} \frac{n_i - d_i}{n_i} \tag{1.9}$$

and is the estimated probability that a subject survives longer than time t.

The Kaplan-Meier estimate for the lung data set is displayed in Figure 1.2.

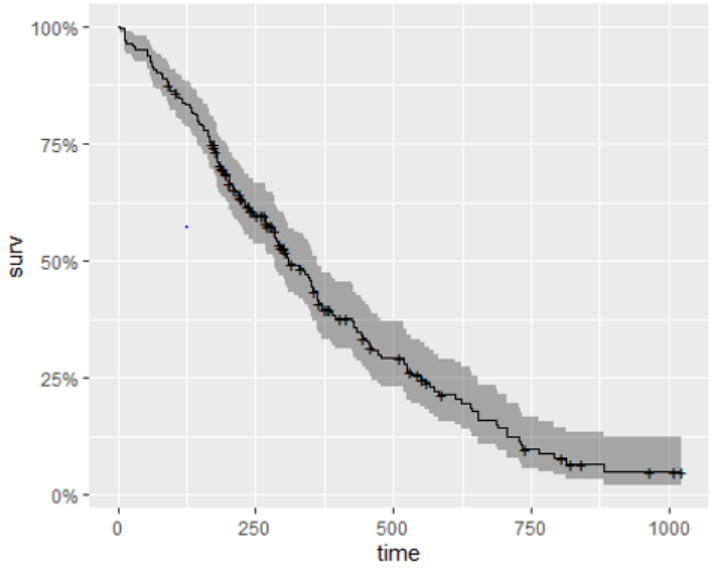

Figure 1.2: Kaplan-Meier Estimate

As shown, the Kaplan-Meier function gives an easily interpretable visualization of how long patients are expected to survive over time. Each "+" on the figure indicates that a patient's follow-up time was censored at that point. Confidence intervals can also be calculated and are included within the graph.

For this thesis, we are primarily concerned with the scenario where we wish to compare survival times between two groups. First, consider the Kaplan-Meier function for the lung cancer data, but stratified by sex.

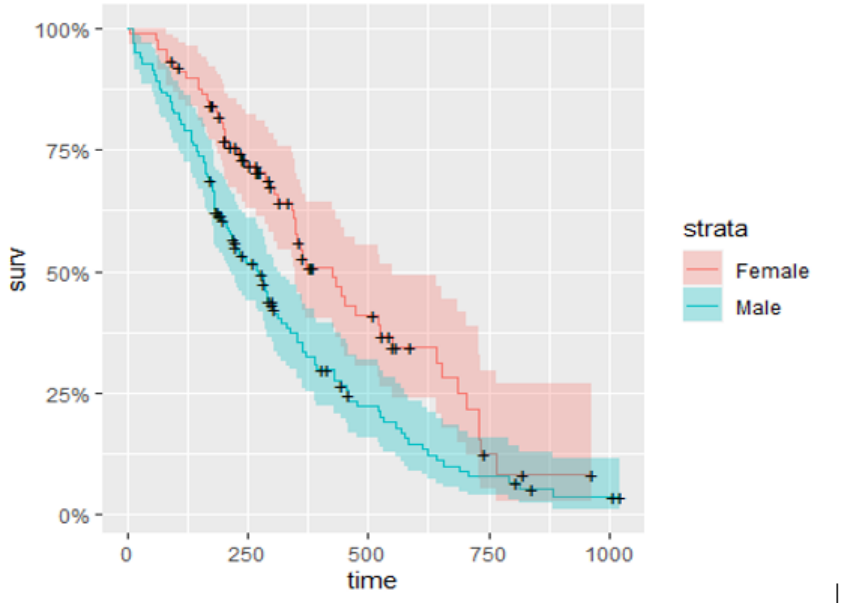

Figure 1.3: Kaplan-Meier Graph Separated by Sex

It seems possible that there is a difference in the distribution of survival times between the two sexes based on the Kaplan-Meier estimates. We need a method that allows us to answer the following questions: does this difference exist, and if so, how large is it? There are multiple methods that exist to tackle these questions, but the method we are focused on is the Cox Proportional-Hazards Model. The term "proportional-hazards" here is important, as it is an underlying assumption we must make in order to use this method. Let $h_1(t)$ be the hazard function for group 1 and let $h_2(t)$ be the hazard function for group 2. If the two hazards are proportional, then we can say that $h_1(t) = \phi h_2(t)$, where ϕ is a constant that does not depend on time. This assumption can be assessed informally by looking at a graph of the estimated survival functions. If the two functions do not cross, then the assumption is probably met. For example, the estimates in Figure 1.3 show that the lines for male and female do not cross, and therefore we see that our assumption has been met. This assumption can be tested more rigorously with the "cox.zph" function in

R's survival package.

Under the proportional hazard model assumption $h_1(t) = \phi h_2(t)$, ϕ is known as the relative hazard or hazard ratio. If $\phi > 1$, then we could say that the hazard of death at time t is greater for group 1. If $\phi < 1$, we would say the opposite.

Next, we can re-parameterize ϕ. Let $\phi = e^{\beta}$. Then $\beta = ln(\phi)$ and now a value greater than 1 for ϕ will lead to a positive value for β and a value between 0 and 1 will lead to a negative value for β (ϕ is a ratio, so $\phi > 0$).

The purpose of this parameterization is a common strategy among the generalized linear model framework [7]. The parameter ϕ is strictly positive and when trying to include covariates as one would attempt to do with multiple linear regression, the linear regression equation to model the hazard ratio $\phi = \beta_0 + \beta_1 x_1 + ... + \beta_p x_p$, is not guaranteed to keep ϕ strictly positive. However, $e^{\phi} = \beta_0 + \beta_1 x_1 + ... + \beta_p x_p$ will be strictly positive.

If we assume a scenario with a control and treatment group, let x_1 be an indicator variable such that $x_1 = 0$ designates the control group, and $x_1 = 1$ designates the treatment group. Then our hazard function can now be expressed as

$$h(t) = e^{\beta x_i} h_0(t) \tag{1.10}$$

This is the Cox Proportional Hazards model with one predictor. In this function, $h_0(t)$ essentially serves as our intercept and represents the hazard function for the reference group $x_1 = 0$. Under the null hypothesis, it is the common hazard function characterizing both groups.

If we have more predictor variables, then the model can be extended similar to that of multiple linear regression. The full Cox Proportional Hazards model is expressed as

$$h(t) = (e^{\beta_1 x_1 + \beta_2 x_2 + ... + \beta_n x_n}) h_0(t) \tag{1.11}$$

which can be represented as

$$\frac{h(t)}{h_0(t)} = e^{\beta_1 x_1 + \beta_2 x_2 + ... + \beta_n x_n} \tag{1.12}$$

and finally

$$ln\frac{h(t)}{h_0(t)} = \beta_1 x_1 + \beta_2 x_2 + ... + \beta_n x_n \tag{1.13}$$

Since the predictor variables are linked to ϕ, the general model is in a form that is easily interpretable. The regression coefficients, $\beta_1, ..., \beta_n$, represent the expected change in the natural log of the hazard ratio for one unit change in X, holding all other predictors constant. Using software, the hazard ratio across two groups can easily be estimated after fitting the proportional hazards model, as well as a confidence interval for this hazard ratio. Parameter estimation is conducted by maximum likelihood (MLE) and hypothesis testing and confidence intervals are conducted using the asymptotic properties of MLE's [1].

For the lung cancer data, if we are interested estimating the effect that sex has on survival, we can fit the following Cox proportional model:

$$ln\frac{h(t)}{h_0(t)} = \beta_1 Sex \tag{1.14}$$

where Sex is a dummy variable coded as 1 for females and 0 for males.

Figure 1.4 provides the fit and summary of results using the R package **gtsummary**. The estimate of the hazard ratio, or HR for short, is $e^{\hat{\beta} x_1} = 0.59$. We see that 1 is not within our confidence interval and our p-value, testing $H_0 : \beta_1 = 0$ versus $H_a : \beta_1 \neq 0$ is significant (.001) concluding that the hazard ratio is different from 1. Based on the interval, we are 95% confident that the hazard for female patients is 0.42 to 0.82 times less than the male patients. This is reflected in Figure 1.3, as the survival curve for females is consistently higher than the curve for males.

This example analysis and discussion highlights that the Cox proportional hazard model can be utilized for a simple two group analysis but has the flexibility to include

Characteristic	HR	95% CI	p-value
sex	0.59	0.42, 0.82	0.001

Figure 1.4: Hazard Ratio

additional explanatory variables in the model to allow for estimates of the HR to be adjusted for other potential confounding variables in the model. For our purposes, we are mostly interested in the one predictor case, as we will be using propensity score matching to control for confounding factors across the two groups. This is a standard practice for many researchers and will be discussed in the next section.

1.2 Propensity Score Matching

Because clinical studies are often observational studies and are thus not randomized experiments, often times potential bias from confounding variables, or covariates, must be considered. Propensity Score Matching is a method commonly used in clinical settings that attempts to reduce this bias when comparing outcomes across two groups [8]. The process of Propensity Score Matching is typically done by estimating propensity scores for each patient, which are generally created by building a logistic regression model using the treatment group status as the response and any potential confounding variables as the explanatory variables. Propensity scores in this case are the predicted probabilities of an individual patient being in the treatment group based on the confounding variables in the logistic model. As data sets get increasingly large, one could theoretically use non-parametric classification models, such as random forests and classification trees, to generate the propensity scores.

After propensity scores are calculated, the main goal is to down-sample the original data set so that the smaller data set exhibits properties of a randomized exper-

iment. Primarily, the confounding variables are evenly distributed across the treatment groups. To do this, we wish to pair up patients, one patient in the treatment group and one patient in the control group, that have the same propensity score. There are various methods that could be used for matching, but the method that is be used for our purposes is generally be nearest neighbor matching [9]. With this method, patients in the treatment group are be matched with patients from the control group based on their propensity scores being as close as possible. In many clinical settings, the number of observations in the treatment group is smaller than that of the control group. Because of this, each observation from the treatment group is matched to its "closest" observation among the control group. This process yields only the removal of observations from the treatment group generally, as most of the time the control group will be much larger than the treatment group. Oftentimes, if the control pool is large enough, matching will be done in a ratio. For example, 3:1 matching would mean that 3 patients from the control pool are be matched with 1 patient from the treatment pool.

In theory, after matching is done, the covariates are much more balanced between the treatment and control groups in the matched patients compared to the covariates across both groups when considering all patients. Typically, this is measured by comparing the difference of the mean between the control and treatment groups before and after matching for every independent variable included in the matching process. This difference is then standardized by dividing by the standard deviation of the treatment group before matching.

Below is an example from a 2018 study on patients with Chron's disease [5]. Various covariates were used is the propensity score calculation and the standardized mean differences between groups were compared before and after matching.

For the points labeled "unadjusted", these values are the difference in means (standardized) between the treatment and control groups before matching, while the

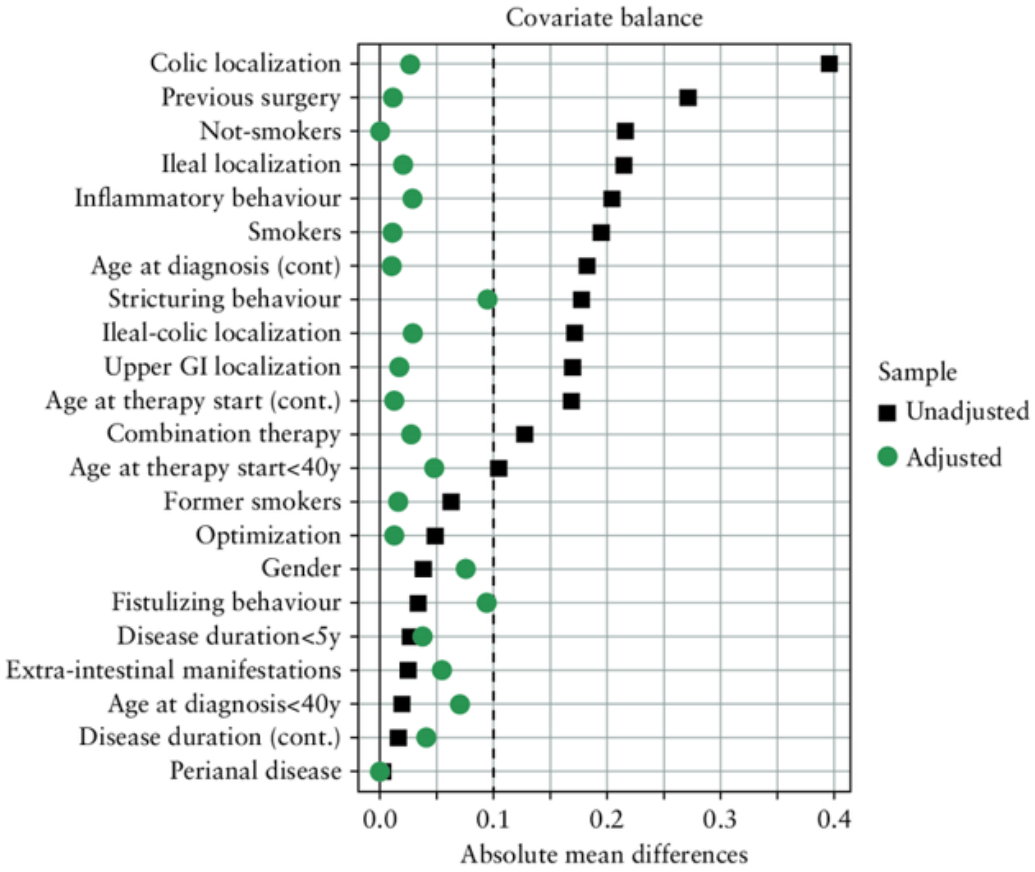

Figure 1.5: Example of Covariate Balance before and after Propensity Score Matching

points labeled "adjusted" are the differences after matching. For variables that are categorical, dummy variables were assigned to each potential level. In that case, the absolute mean difference shown is the difference between the proportions of the treatment and control group that have that trait. For example, dummy variables were assigned for both "smokers" and "non-smokers", and the shown differences are the difference of the proportions of the control and treatment groups that have each trait.

Overall, after matching, the absolute standardized mean difference for each covariate was kept below .1 in this case. Thus, it is evident that the covariates are much

more balanced after propensity score matching. After propensity score matching, the researchers would be able to proceed with the analysis between treatment and control groups without worrying about potential bias caused by covariates such as colic localization, smoker vs. non-smoker status, etc.

A common question asked among analysts who have not been exposed to propensity score matching is "why bother when you can just include the confounding variables in your Cox proportional hazard model?". There are a few logistic advantages. The first is in regard to the process of building a model with multiple explanatory variables. If the number of potential confounding variables is large, the modeler must make numerous decisions such as feature selection, transformations, and adding model complexity such as interaction terms. These time-consuming processes are what allow us to obtain an estimate of the hazard ratio between two groups holding the confounding variables fixed. The matching process effectively does this and only a single predictor is needed to estimate the hazard ratio. The second advantage is the simplicity of the final model and the ease of communication of the entire analysis process. For any given analysis, a traditional approach will be a unique process that must effectively be explained each time. Under the propensity score matching framework, the focus is just building a predictive model that can produce accurate predictive probabilities, comparing matching results such as the output displayed in Figure 1.5, and a simple Cox model with one explanatory variable is applied. Another concern may be that a potentially large amount of data in the control pool could be lost in the process of propensity score matching. With this in mind, propensity score matching should only be performed if it is feasible that subjects in the treatment group could have reasonably come from the control group.

2 EFFECTS OF SICKLE CELL DISEASE ON DIALYSIS

In this chapter a report is provided on one of the primary analyses that was completed during an internship experience with Fresenius Medical Care on patients with sickle cell disease (SCD) who received in-center care from Fresenius. Sickle cell disease is an inherited blood disorder that causes a problem with hemoglobin in the body. Normally these red blood cells are disc-shaped and flexible enough to move through the blood vessels, but a patient with SCD has sickle shaped red blood cells. This causes complications and makes it difficult for the cells to bend or move easily, and can block blood flow to the rest of the body or cause other medical complications [6]. SCD is a genetic mutation that evolved to combat malaria, so it is primarily found in patients from parts of the world where malaria is an issue. For that reason, SCD is primarily found in people of African descent, however a person of any race or ethnicity can have SCD. The goal of this report was to quantify the effect of sickle cell disease on mortality for dialysis patients. In doing so, Fresenius Medical Care and other dialysis providers may be able to better understand how SCD impacts a patient health, and ultimately provide better care. The report is summarized in 5 sections that provide the study design, data collection and cleaning processes, propensity score matching, interpretation of results from a cox proportional hazard model, and additional discussions.

2.1 Study Design and Data Collection

In-center dialysis patients with and without SCD treated by Fresenius Medical Care from 1/1/2017 to 12/31/2021 were observed and regular measurements were collected including basic demographic data, what method of access was used for dialysis,

hospitalizations, lab measurements, quality of life data, etc. The data is deidentified so that it is not possible to identify any patients based on their information. Basic steps for deidentification include each patient being assigned a random ID, as well as all dates recorded being relative to the patient's first date of dialysis (FDD).

Because not all patients started dialysis with Fresenius Medical Care, it is possible that not all medical data for a patient is recorded. However, a patient's first date of dialysis (FDD) is always well recorded, regardless of the dialysis provider. All recorded dates for a patient are relative to this FDD, meaning it is possible to determine whether a patient started dialysis with Fresenius, or if they started dialysis somewhere else and then became a patient with Fresenius. With this in mind, there are two dates that are important for each patient: the patient's first date of dialysis (FDD) and the patient's first Fresenius date (FFD). For some patients these two dates may align, and for others they will not.

For each patient, there is a baseline period and a performance period. A patient's baseline period is considered 6 months from First Fresenius Date (FFD) and follow-up/performance period is thereafter. The day a patient's performance period starts is considered the index date for that patient. Performance period ends with last day of follow-up or first occurring event among death or transplantation.

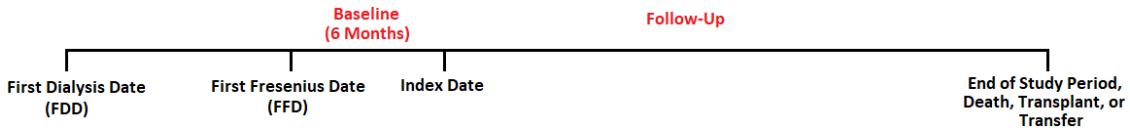

Figure 2.1: Study Timeline

Patients that started dialysis with Fresenius and patients that started dialysis elsewhere and then came to Fresenius make up two different populations. These two populations are referred to as the "incident" and "prevalent" groups respectively, and a separate analysis was done for each group. In this study, a patient's incidence/prevalence status was determined based on the patient's first fresenius date. If

the patient's first Fresenius date was within 120 days of the patient's first date of dialysis, then the patient is considered an incident patient. Otherwise, the patient falls into the prevalent group. With this distinction, there will be two separate analysis done: one for the incident population, and one for the prevalent population.

2.2 Data Processing and Summary Statistics

At the start of the project, data was pulled from all Fresenius Medical Care patients from 1/1/2017 to 12/31/2021. From the various sources of data, the first Fresenius date for each patient was found, along with statistics for each patient's baseline period. The total number of patients to begin with was 541,978. From here, patients who did not have a 180 day baseline period due to death, transplantation, or any other reason were not considered. Follow-up time for patients for which an event did not occur was calculated by subtracting the index date for a patient from the last time a patient showed up in any data. Lastly, after many discussions it was decided that the study would apply to patients who only received in-center care during the baseline period. This left the study with 365,748 patients. Figure shows the breakdown of these patients by SCD and incident status.

	Non-SCD (N=364605)	SCD (N=1143)	Total (N=365748)
incident			
Incident Patient	194804 (53.4%)	602 (52.7%)	195406 (53.4%)
Non-incident patient	169801 (46.6%)	541 (47.3%)	170342 (46.6%)

Figure 2.2: Patient Population by Status

Many summary statistics were collected for these patients, Figure 2.3 shows important summary statistics that were discussed.

	Non-SCD		SCD	
	Incident Patient (N=194804)	Non-incident patient (N=169801)	Incident Patient (N=602)	Non-incident patient (N=541)
Age				
Mean (SD)	63.3 (14.3)	62.3 (14.3)	54.8 (15.6)	53.3 (15.5)
Median [Min, Max]	64.8 [4.50, 104]	63.3 [2.70, 103]	54.8 [16.9, 92.7]	53.5 [15.1, 94.3]
Sex				
F	82321 (42.3%)	72783 (42.9%)	321 (53.3%)	302 (55.8%)
M	112483 (57.7%)	97018 (57.1%)	281 (46.7%)	239 (44.2%)
Race				
BLACK	52722 (27.1%)	59763 (35.2%)	480 (79.7%)	445 (82.3%)
OTHER	9986 (5.1%)	8847 (5.2%)	4 (0.7%)	4 (0.7%)
UNKNOWN	19497 (10.0%)	11416 (6.7%)	43 (7.1%)	27 (5.0%)
WHITE	112599 (57.8%)	89775 (52.9%)	75 (12.5%)	65 (12.0%)
Ethnicity				
	21452 (11.0%)	13001 (7.7%)	54 (9.0%)	36 (6.7%)
Hispanic or Latino	28465 (14.6%)	25442 (15.0%)	21 (3.5%)	15 (2.8%)
Not Hispanic and Not Latino	144887 (74.4%)	131358 (77.4%)	527 (87.5%)	490 (90.6%)
First Fresenius Date				
Mean (SD)	10.9 (19.5)	1550 (1440)	10.6 (18.4)	1850 (1680)
Median [Min, Max]	4.00 [1.00, 119]	1110 [120, 16700]	4.00 [1.00, 118]	1380 [121, 11200]
Average Albumin				
Mean (SD)	3.63 (0.392)	3.79 (0.343)	3.65 (0.392)	3.84 (0.338)
Median [Min, Max]	3.67 [1.50, 4.98]	3.83 [1.54, 5.13]	3.70 [1.92, 4.63]	3.87 [2.27, 4.63]
Missing	1736 (0.9%)	8761 (5.2%)	5 (0.8%)	27 (5.0%)
base_HGB_avg				
Mean (SD)	10.5 (0.837)	10.8 (0.949)	9.26 (1.55)	9.71 (1.74)
Median [Min, Max]	10.5 [5.16, 17.3]	10.8 [4.84, 19.2]	9.70 [3.96, 13.8]	10.3 [3.88, 14.4]
Missing	1230 (0.6%)	6480 (3.8%)	2 (0.3%)	20 (3.7%)
Average NLR				
Mean (SD)	4.42 (3.14)	4.19 (2.86)	3.38 (2.03)	3.17 (1.81)
Median [Min, Max]	3.72 [0, 182]	3.59 [0.0429, 153]	2.85 [0.420, 19.5]	2.73 [0.517, 16.4]
Missing	32580 (16.7%)	45097 (26.6%)	81 (13.5%)	130 (24.0%)
Baseline Access Method				
avf	86554 (44.4%)	115615 (68.1%)	210 (34.9%)	296 (54.7%)
avg	23019 (11.8%)	33245 (19.6%)	102 (16.9%)	157 (29.0%)
cath	85231 (43.8%)	20941 (12.3%)	290 (48.2%)	88 (16.3%)

Figure 2.3: Summary Statistics

Of note in Figure 2.3 is the "Missing" line beneath several variables. If a patient was missing data for HGB or albumin lab readings, the missing data was replaced by an average value for the SCD and incidence/prevalence status of the patient. Variables were then made to notate which patients had data that was filled in by averages.

2.3 Propensity Score Matching

Propensity score matching of non-SCD patients to SCD patients was employed with data as of 6 months post FFD. Matching was done with a 1:1 ratio in order to get as much accuracy as possible on the match.

Variables matched on were decided based upon a previously written analysis plan for the study, before any work was done. Matching was done with the following variables included in a logistic regression model with the probability of the patient being in the SCD group as the response variable: sex, race, ethnicity, age, vintage, albumin, hemoglobin, access type, the indicator variables for missing data being replaced, and all two-way interaction terms for the mentioned variables.

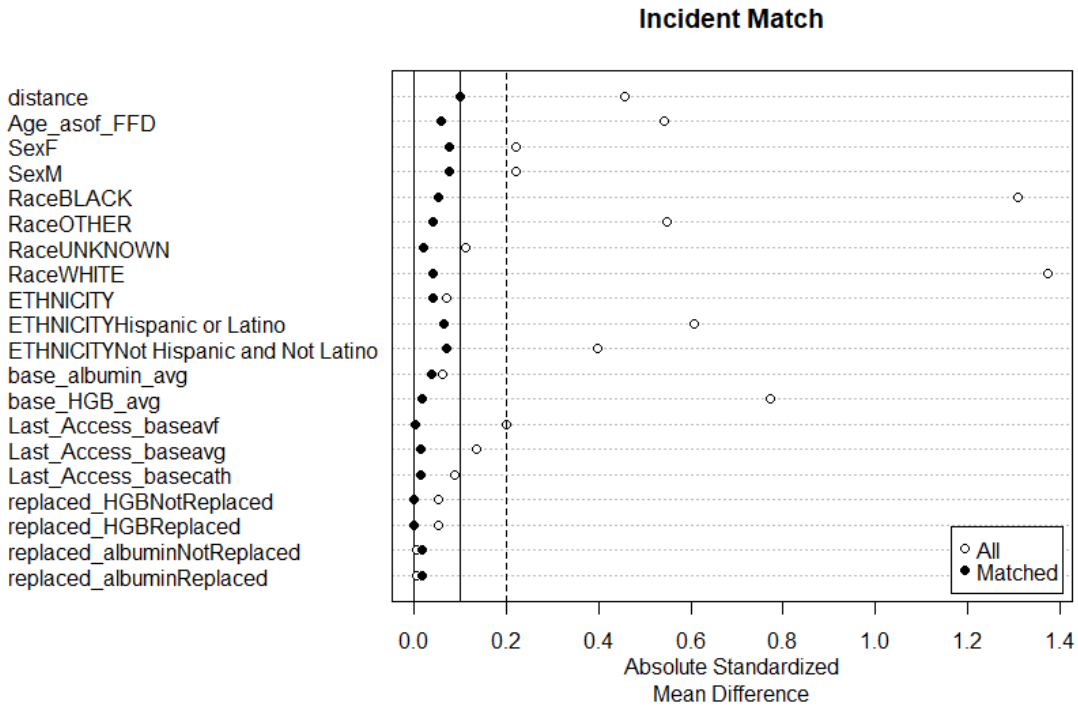

Figure 2.4: Incident Match Balance

Figure 2.4 shows absolute standardized mean differences before and after matching

for the incident population. A similar graph for the prevalent population would show similar results. The "distance" variable shown on Figure 2.4 is simply the standardized difference between the mean of the propensity scores between groups. The decision to include two-way interaction terms in the model was based on the idea that interaction terms should provide better predictive capabilities for the model, as interpretation is not as much of an issue for the model. Summary statistics before and after matching have been provided for the incident population.

	Non-SCD (N=194804)	SCD (N=602)
Age		
Mean (SD)	63.3 (14.3)	54.8 (15.6)
Median [Min, Max]	64.8 [4.50, 104]	54.8 [16.9, 92.7]
Sex		
F	82321 (42.3%)	321 (53.3%)
M	112483 (57.7%)	281 (46.7%)
Race		
BLACK	52722 (27.1%)	480 (79.7%)
OTHER	9986 (5.1%)	4 (0.7%)
UNKNOWN	19497 (10.0%)	43 (7.1%)
WHITE	112599 (57.8%)	75 (12.5%)
Ethnicity		
Hispanic or Latino	28465 (14.6%)	21 (3.5%)
Not Hispanic and Not Latino	144887 (74.4%)	527 (87.5%)
Unknown	21452 (11.0%)	54 (9.0%)
Average Albumin		
Mean (SD)	3.63 (0.390)	3.65 (0.390)
Median [Min, Max]	3.67 [1.50, 4.98]	3.70 [1.92, 4.63]
Average Hemoglobin		
Mean (SD)	10.5 (0.834)	9.26 (1.55)
Median [Min, Max]	10.5 [5.16, 17.3]	9.70 [3.96, 13.8]
Baseline Access Method		
avf	86554 (44.4%)	210 (34.9%)
avg	23019 (11.8%)	102 (16.9%)
cath	85231 (43.8%)	290 (48.2%)
Replaced Hemoglobin Value		
NotReplaced	193574 (99.4%)	600 (99.7%)
Replaced	1230 (0.6%)	2 (0.3%)
Replaced Albumin Value		
NotReplaced	193068 (99.1%)	597 (99.2%)
Replaced	1736 (0.9%)	5 (0.8%)

Figure 2.5: Summary Stats Before Matching in Incident Population

	Non-SCD (N=602)	SCD (N=602)
Age		
Mean (SD)	55.7 (15.9)	54.8 (15.6)
Median [Min, Max]	57.5 [16.8, 90.0]	54.8 [16.9, 92.7]
Sex		
F	298 (49.5%)	321 (53.3%)
M	304 (50.5%)	281 (46.7%)
Race		
BLACK	467 (77.6%)	480 (79.7%)
OTHER	6 (1.0%)	4 (0.7%)
UNKNOWN	46 (7.6%)	43 (7.1%)
WHITE	83 (13.8%)	75 (12.5%)
Ethnicity		
Hispanic or Latino	28 (4.7%)	21 (3.5%)
Not Hispanic and Not Latino	513 (85.2%)	527 (87.5%)
Unknown	61 (10.1%)	54 (9.0%)
Average Albumin		
Mean (SD)	3.64 (0.407)	3.65 (0.390)
Median [Min, Max]	3.68 [2.17, 4.62]	3.70 [1.92, 4.63]
Average Hemoglobin		
Mean (SD)	9.24 (1.31)	9.26 (1.55)
Median [Min, Max]	9.52 [5.16, 13.4]	9.70 [3.96, 13.8]
Baseline Access Method		
avf	209 (34.7%)	210 (34.9%)
avg	99 (16.4%)	102 (16.9%)
cath	294 (48.8%)	290 (48.2%)
Replaced Hemoglobin Value		
NotReplaced	600 (99.7%)	600 (99.7%)
Replaced	2 (0.3%)	2 (0.3%)
Replaced Albumin Value		
NotReplaced	598 (99.3%)	597 (99.2%)
Replaced	4 (0.7%)	5 (0.8%)

Figure 2.6: Summary Stats After Matching in Incident Population

As is evident in Figure 2.5 and Figure 2.6, after matching, the covariates are balanced between groups after matching. The prevalent population showed similar results after matching.

2.4 Interpretation of Results from Cox Proportional Hazards Model

After performing both matches, the incident group of 602 SCD patients was matched with 602 non-SCD patients and the prevalent group of 541 SCD patients

was matched with 541 non-SCD patients. Both of these matched groups have been fit to a Cox proportional hazards model. Kaplan Meier curves stratified by SCD status for both populations have been provided.

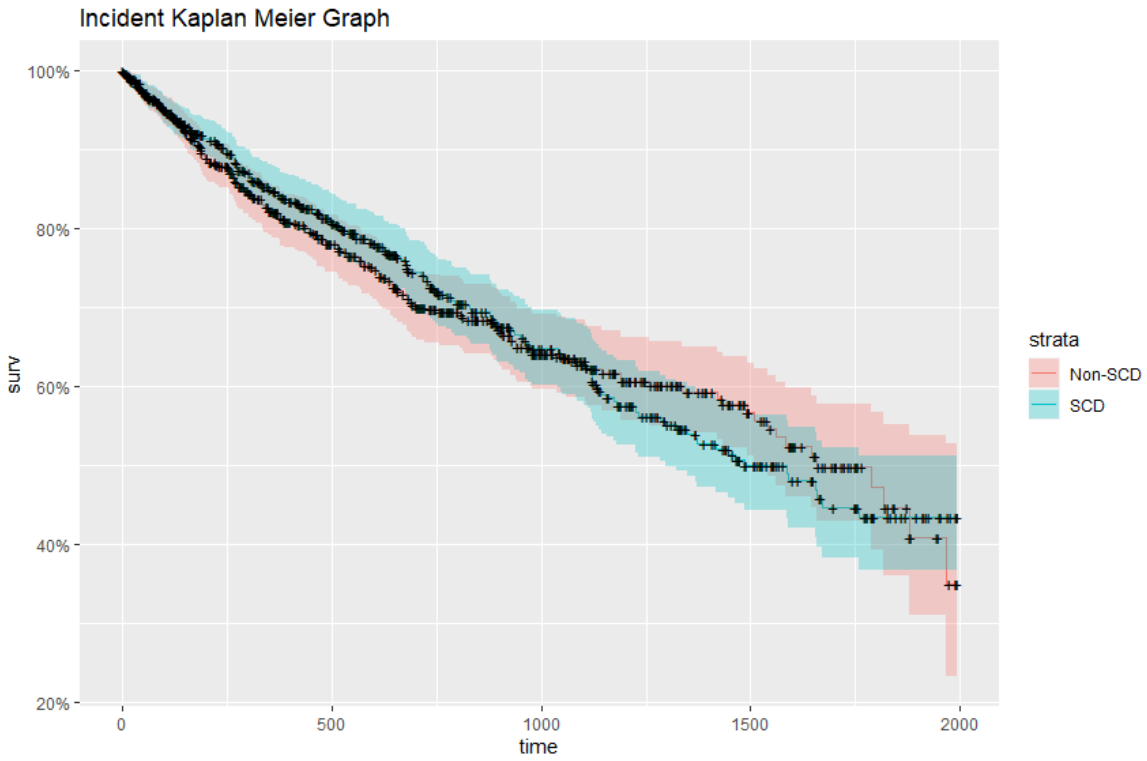

Figure 2.7: Kaplan Meier Curve for Incident Patients

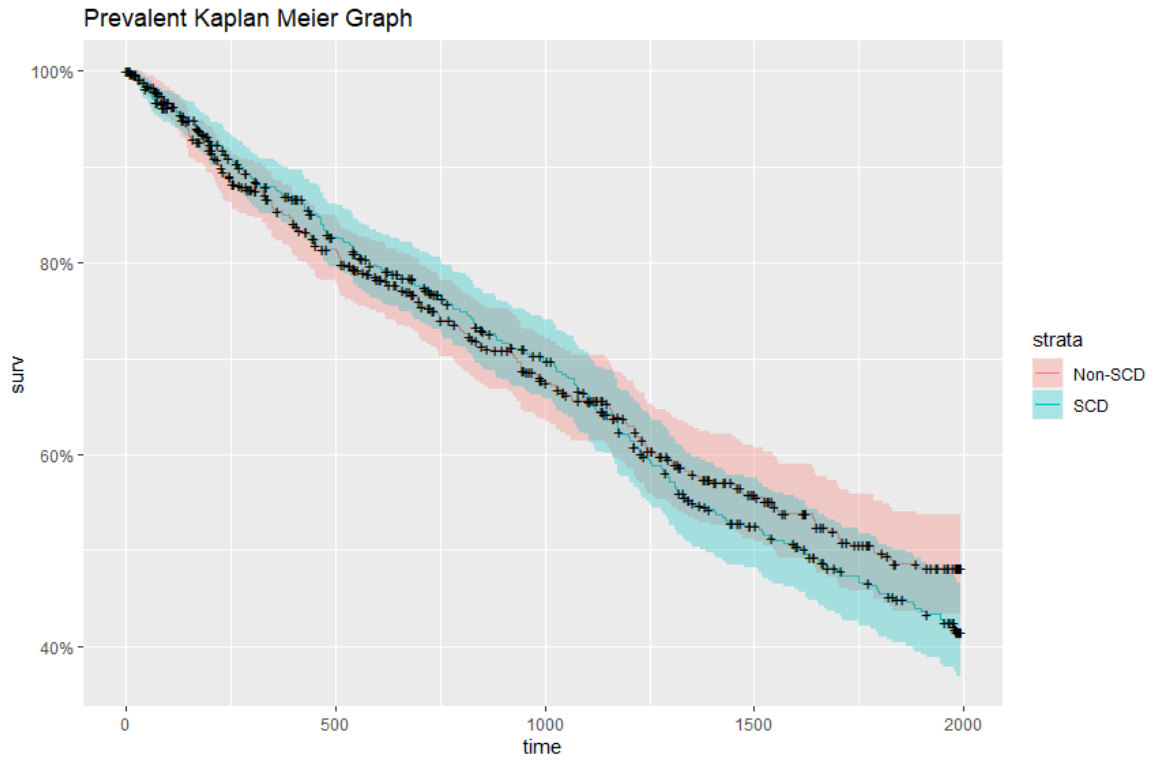

Figure 2.8: Kaplan Meier Curve for Prevalent Patients

On Figures 2.4 and 2.7, we see that the curves for the SCD and non-SCD groups look very similar for both incident and prevalent patients.

Characteristic	HR[1]	95% CI[1]	p-value
SCD	0.99	0.81, 1.21	>0.9

[1] HR = Hazard Ratio, CI = Confidence Interval

Figure 2.9: HR for Incident Population

Characteristic	HR[1]	95% CI[1]	p-value
SCD	1.09	0.91, 1.31	0.3

[1] HR = Hazard Ratio, CI = Confidence Interval

Figure 2.10: HR for Prevalent Population

Our hazard ratios include 1 in their confidence intervals and our p-value is greater than .05, so we cannot say that there is a statistically significant difference between the SCD and non-SCD groups in either population.

2.5 Discussions and Further Investigative Work

The fact that no difference was found between the SCD and non-SCD groups in both populations was surprising, given that similar studies in the past have yielded results that contradict this [10]. A further review of methods between our analysis and similar studies could provide some insight. Two potential limitations in our analysis are the method used to calculate follow-up time for a patient, and the variables included in the propensity score. Follow-up time for a patient for which an event did not occur was calculated as the time from the patient's index date to the last time the patient appeared in any data. While this should be a relatively accurate estimate, an exact amount of follow-up time is not possible to know due to the way the data was gathered. In regards to the variables included in the propensity score, it is possible that some of the variables related to lab results for a patient could be part of the *causal pathway*. The notion of causal pathway is that certain variables may actually be part of what causes or is a fundamental characteristic of the primary conditions we wish to compare. For example, if one were interested in comparing patients with an autoimmune disease such as lupus versus patients without and included in their matching routine genetic measurements related to inflammation, the resulting

matching could potentially remove the effect one would expect to see between the two groups. The matching would directly force patients with similar immune responses in each group, but it is the very unique immune response is what we wish to compare. It is for this reason, matching on variables that are part of the causal pathway is not advised. In the SCD analysis, lab values such as hemoglobin appear to have a difference between the SCD and non-SCD groups and could potentially be a part of the causal pathway for SCD. Further discussions with a medical professional would help to determine if matching with these variables in our model is appropriate. From a statistical standpoint, including these variables is also problematic based on the distance between propensity scores being so high. The larger this distance is, the idea that our treatment group could have feasibly come from the control pool comes into question. If matching only includes non-SCD patients with the comparable levels of hemoglobin to SCD patients, then it is possible that the patients from the non-SCD group included in the match are not representative of the entire non-SCD group, and thus our analysis would definitely be biased.

2.6 Analysis with Different Matched Variables

As discussed in the previous section, it is possible that some of the variables previously used in the matching process could be a part of the causal pathway. It is suspected that hemoglobin specifically may be a variable that should not be included in the matching process due to this. In this section we will perform the same analysis as the previous section, but without the lab values (hemoglobin and albumin) in our matching process. Keeping the matching process down to only essential variables that have plenty of representation accross both groups should ensure that our results are not biased.

So, the matching process now only includes age, sex, race, ethnicity, access type,

and two-way interaction terms in the logistic regression model. The standardized mean differences before/after matching have been included.

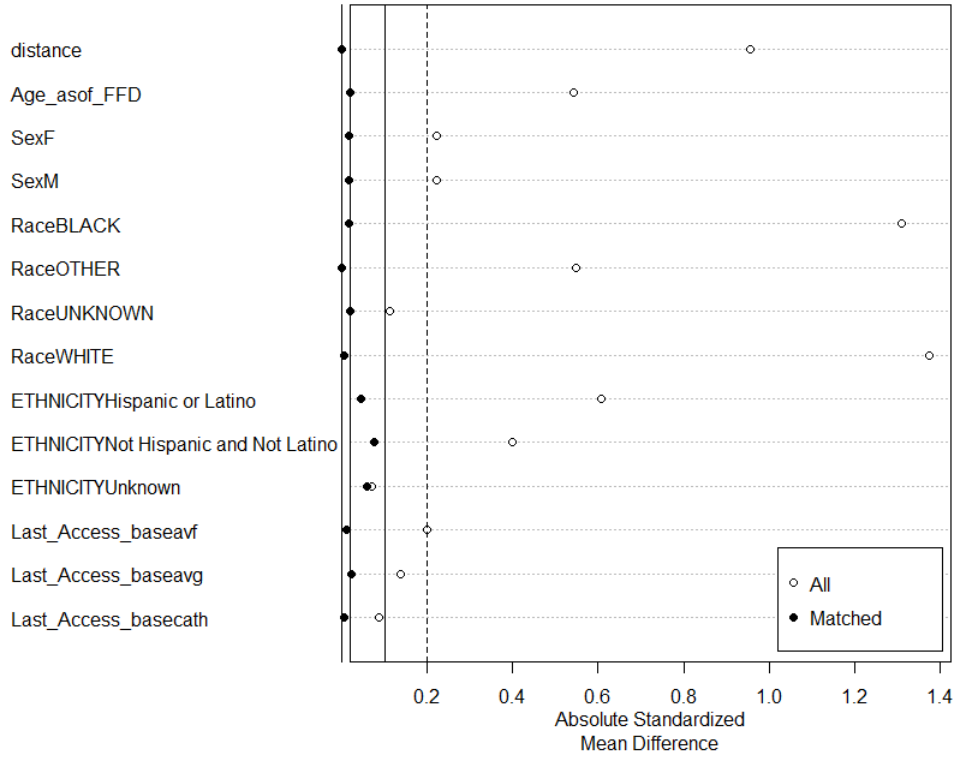

Figure 2.11: Matching Balance Without Lab Values

A similar figure for the prevalent population would look similar to Figure 2.11. Of note in 2.11 is the fact that "distance" is now very small after matching. With this in mind, matching with a ratio higher than 1:1 could be possible, however, in this analysis this was not done for the sake of consistency. Now, fitting a Cox proportional hazards model with the patients included in this match yields different results than seen previously.

Figure 2.12: Kaplan Meier Curve for Incident Patients Without Lab Values

Figure 2.13: Kaplan Meier Curve for Prevalent Patients Without Lab Values

We now see a difference between the two curves visually in both the incident and prevalent populations (Figures 2.12 and 2.13). Hazard ratios have been provided for each population.

Characteristic	HR[1]	95% CI[1]	p-value
SCD	1.45	1.18, 1.79	<0.001

[1] HR = Hazard Ratio, CI = Confidence Interval

Figure 2.14: Hazard Ratio for Incident Patients Without Lab Values

Characteristic	HR[1]	95% CI[1]	p-value
SCD	1.23	1.03, 1.48	0.023

[1] HR = Hazard Ratio, CI = Confidence Interval

Figure 2.15: Hazard Ratio for Prevalent Patients Without Lab Values

Under this alternative matching strategy, the test suggests there is enough evidence to conclude that a difference in mortality between the SCD and non-SCD groups exists for both populations based on our p-values being significant for each population and the hazard ratios not containing 1. With a HR of 1.45, a patient with SCD in the incident group is 45% more likely to pass away than a patient without SCD at any given time. Similarly, a patient in the prevalent group with SCD is 23% more likely to pass away than a patient without SCD at any given time.

The difference in our analysis based on the variables in the matching process emphasizes the importance of speaking to an expert in the subject matter before doing an analysis. While it is important not to "fish" for results, it seems plausible that hemoglobin is a part of the causal pathway in this case. If that is the case, then the first analysis is flawed and therefore the analysis in this section is more accurate. Alternatively, it is possible that the previous study referenced is flawed and should have considered lab values for patients in their matching process. It is speculated that the truth is that hemoglobin is part of the causal pathway for SCD, however, that is a conclusion that would be better decided after discussions with a medical expert.

3 SEQUENTIAL ANALYSIS CONSIDERATIONS

Projects involving survival analysis, like the one presented in Chapter 2, tend to be conducted over numerous years, and follow-ups are conducted throughout the studies and not just at the final endpoint. Similar to clinical trials, it is natural to want to perform preliminary analysis (looks) at an earlier time point or at multiple time points such as every 3 months. This creates a multiple testing issue. One solution is to take the approach that clinical trials take and adjust the significance level at each look. this is referred to as alpha spending and is a well-developed and studied procedure [4]. In this chapter, we will offer some helpful equations to derive effective sample sizes and power calculations and offer some insight as to how these calculations are impacted if an alpha spending approach is applied. Some additional discussion on potential future work are also discussed.

3.1 Alpha Spending

A brief summary of the alpha spending approach is provided, while a full review of technical details and summary of alpha spending in clinical trials can be found in [4] and [2]. The main objective in taking multiple looks during a study is to maintain the control of the type-I error rate. To do this for K looks, assuming a two-sided test statistic $Z(k)$ at the k^{th} look, we wish to find critical values for each interim analysis $(Z_c(k), k = 1, 2, ..., K)$ such that the type-I error rate over the sequential run of tests maintains a specified level. The procedure starts at the first look and decides to continue data collection if $|Z(1)| < Z_c(1)$ and otherwise the analysis is stopped since H_a is concluded. If data collection continues, the next interim analysis is conduced with the same decision process except using a different critical value $Z_c(2)$. The

process continues until a rejected test is observed or the final analysis is conducted.

The critical values must be chosen so that, under the null hypothesis,

$$P(|Z(1)| \geq Z_c(1), or|Z(2)| \geq Z_c(2), or..., or|Z(K)| \geq Z_c(K) = \alpha \tag{3.1}$$

If the distribution of the test statistic is Normally distributed, the joint distribution of $Z(1), Z(2), ..., Z(K)$ can be determined and their covariances are simply functions of the sample sizes observed at each look . The critical values can be determined in a sequential way, obtaining a critical value for $Z(1)$, then $Z(2)||Z(1)| \leq Z_c(1)$, and so on. Under this conditioning, a significance level α_k must be specified and there is no unique approach to its selection. However, since the sum of the conditional probabilities must equal the overall significance level α, each α_k represents the amount of α being spent at each interim analysis, and the cumulative proportion of α spent at the k^{th} test is $\frac{1}{\alpha}\Sigma_{j=1}^{k}\alpha_k$. For example, if it were decided to conduct 3 looks, and we choose $\alpha_1 = .001, \alpha_2 = .011, \alpha_3 = 0.038$, the evidence reflected in the earlier test statistics must be extremely strong in order for a rejection to be made. This is often a reasonable choice given that the preliminary looks have smaller sample sizes and thus stronger effect sizes should be observed in order to reject.

Demett and Lang proposed that the choice of how much α is getting spent at each sequential look can be specified using a spending function $\alpha(t_k)$, which returns the total significance level spent at the k^{th} look which can be expressed as $\Sigma_{j=1}^{k}\alpha_k$ [4, 2]. The input to the function t_k is referred to as the information fraction at the k^{th} look. For comparing two means, t_k is the ratio of the current sample size at look k divided by the total sample size that will be observed at the final look. For survival analysis, t_k is the total number of deaths at look k divided by the total number of expected deaths at the final look.

Two common spending functions noted in [2] are the O'Brien-Fleming and Pocock

functions, denoted $\alpha_1(t_k)$ and $\alpha_2(t_k)$ respectively, are as follows:

$$\begin{aligned} \alpha_1(t_k) &= 2 - 2\Phi(z_{\alpha_2}/\sqrt{t_k}) \\ \alpha_2(t_k) &= \alpha ln(1 + (e-1)t_k) \end{aligned} \tag{3.2}$$

Note that both functions return 0 at $t_k = 0$ and α at $t_k = 1$ and are strictly increasing. As long as the spending function is a non-decreasing function with the property that $\alpha(0) = 0$ and $\alpha(1) = \alpha$, critical values can be obtained in a sequential fashion using the previously described method and setting $\alpha_k = \alpha(t_k) - \alpha(t_{k-1})$.

To illustrate the spending functions' utility, suppose that we wish to take 5 looks during a survival analysis study. If we assume that each of the looks will be conducted once one-fifth of the total number of expected deaths are observed, the information fractions are thus 0.2, 0.4, 0.6, 0.8, and 1. Figure 3.1 provides a plot of the two-sided critical values at each look using the O'Brien-Fleming spending function with $\alpha = 0.05$. This particular spending function spends very little of α for low information fractions. This can be seen by observing the large critical regions at the early look

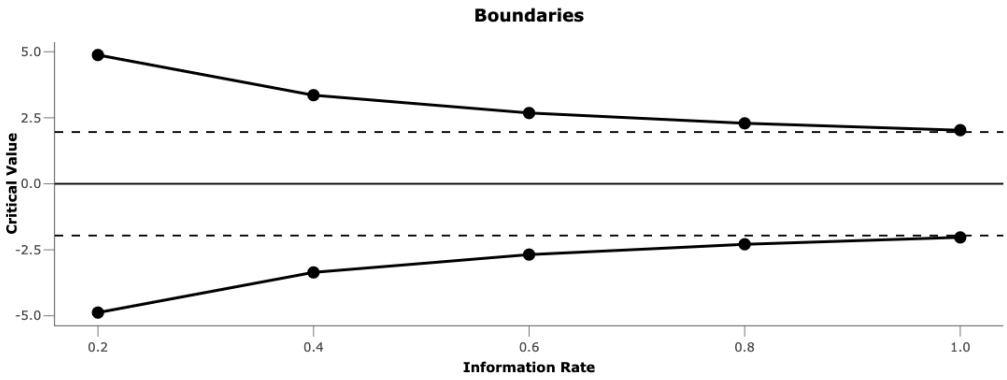

Figure 3.1: Critical Values For 5 Looks Using the O'Brien Fleming Spending Function

The advantage of the spending function approach is sequential testing. Original testing approaches had restrictions such as specifying the number of tests in advance as well as restrictions on what situations the preliminary looks should occur. Due to

its sequential nature and its ability to create critical values based on information fraction, critical values at later looks can be easily updated to reflect real-world changes and challenges during data collection.

Since the critical values of the alpha spending approach are different for each look, in situations like the SCD data analysis where data was collected over a 5-year period, the number of preliminary looks could be quite large. The larger number of tests and choice of spending function could yield situations in which the power to detect a meaningful hazard ratio could be astronomically low. In these cases, it would be of interest to know what effect sizes still maintain good power so a decision could be made if the additional look is even warranted.

Sample size planning for survival analysis involved not just the number of patients but the number of deaths observed over the time period of the study. While it is difficult to determine power calculations for Cox proportional hazard models in general, it is recommended that power analysis derived from the log Rank test is an appropriate surrogate when comparing two groups [1]. Suppose one wishes to detect a hazard ratio $H_a : \phi \neq 1$, such that the power of the log Rank test is $1 - \beta$ and significance level α, the total number of deaths, d that should be observed by the end of the study is:

$$d = \frac{(z_{\alpha/2} - z_{\beta})^2}{\pi(1-\pi)\theta^2} \tag{3.3}$$

where π is the proportion of observations in one of the two groups, z_l is the upper $l^{th}-$percentile of the standard normal distribution, and $\theta = log(\phi)$.

Recall that when using spending functions in survival analysis, the information rate t_k is the fraction of observed deaths at look k divided by the total expected number of deaths. If the design is planned such that the expected number of deaths is d, determining the value of ϕ at look k such that the power remains at $1-\beta$ can be easily obtained by multiplying both sides of Equation (3.4) by the information rate

yields

$$t_k d = \frac{(z_{\alpha/2} - z_\beta)^2}{\pi(1-\pi)\frac{\theta^2}{t_k}}. \tag{3.4}$$

Since $t_k d$ is the number of deaths at look k, the power of the test holding the remaining values fixed remains at $1 - \beta$ when the log hazard ratio is $\frac{\theta^2}{t_k}$. Using the thresholds produced by Figure 3.1, Figure 3.2 plots the hazard ratio effect size needed to maintain a power of 0.8 and the intended effect size at the end of the study was $\phi = 1.5$.

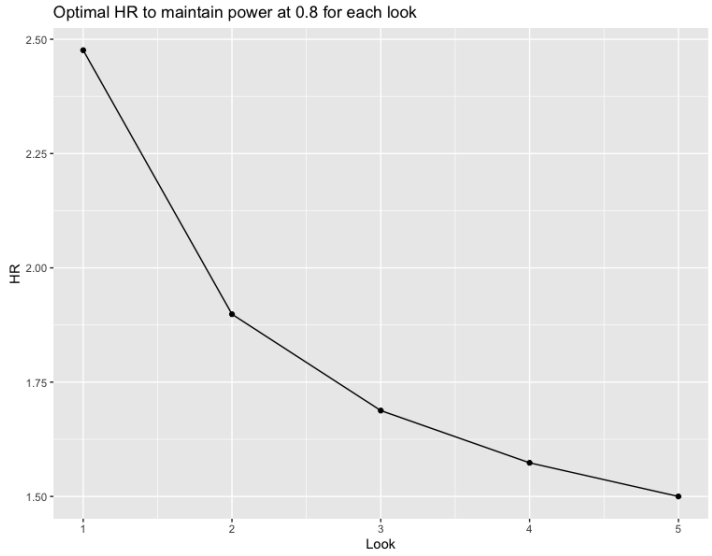

Figure 3.2: HR Effect Sizes Versus Information Fraction

At the first look, $t_k = 0.2$, that hazard ratio must be almost 2.5 in order to have statistical power of 0.8. This could very well not make any practical sense to expect or implore. Since the information rates do not have to be equally spaced, it may be more beneficial to taking multiple looks for values of t_k that are larger.

Rather than consider different hazard ratios, it may also be beneficial to examine the power for a fixed hazard ratio that is of key interest and was used to determine

the needed number of deaths d at the end of the study period. The power at each look can be derived by first solving Equation (3.4) for z_β which yields

$$z_\beta = \sqrt{t_k d\pi(1-\pi)\theta^2} - z_{\alpha/2}. \tag{3.5}$$

For a given look k, the values of t_k and α in Equation (3.5) are updated where α is determined by the corresponding significance level using the critical value $Z_c(k)$. Once z_β is computed, the power is $\Phi(z_\beta)$. Using our current example which produced Figures 3.1 and Figure 3.2, Figure 3.3 provides the statistical power at each look when ϕ is held fixed at 1.5. Since the number of deaths observed is just a small fraction, the statistical power is essentially 0 at the first look and gradually improves as the information fraction increases to one.

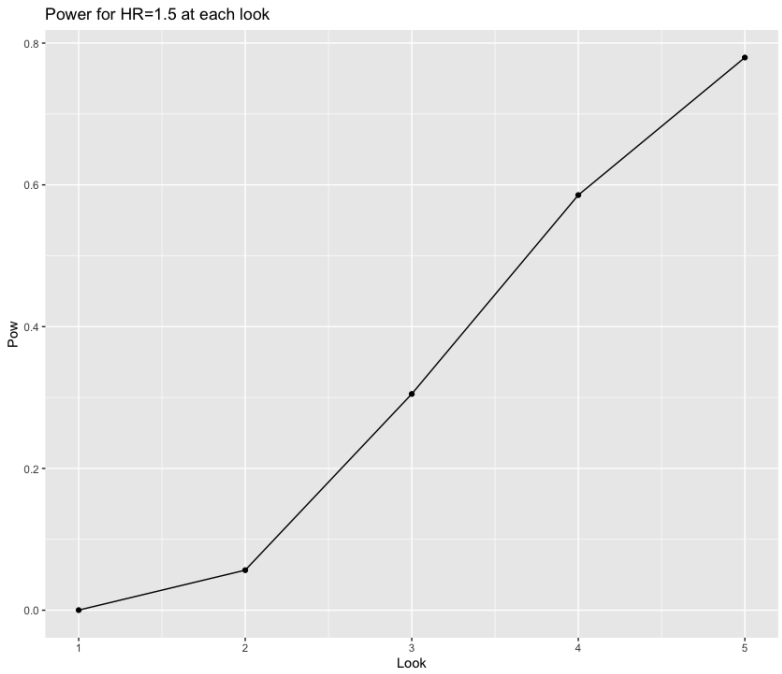

Figure 3.3: Power for HR of 1.5 at Each Look

While the example may not be realistic, the formulas provided in this section will allow for further exploration of how to optimally incorporate multiple looks in

survival analysis and how it can potentially help healthcare providers make better decisions. For large studies taken over a long period of time, an investigation of best practices in terms of picking an appropriate spending function and selecting appropriate information fractions while handling a large number of looks is needed. There are also some additional questions on how the incorporation of propensity score matching would play a role in sequential testing. Is it appropriate to keep the groups balanced (1:1 matching), or are there any benefits to matching in unique or unbalanced ways that could help studies in terms of statistical properties or logistical issues? If the tests are quite large, another consideration is if type-I error is the ultimate metric we wish to control or perhaps if the number of looks is large, controlling a different error rate would be meaningful. We leave these questions for further work.

4 DISCUSSIONS AND FUTURE WORK

4.1 The Internship Experience

This section will provide a brief review of my internship experience at Fresenius Medical Care and provide some final discussions on future work. Key advice to future students ho will experience an internship in a healthcare setting similar to mine is discussed.

Firstly, managing an internship and coursework at the same time is not easy. With regular meetings and work to complete for an internship, it can be hard to stay on top of your coursework as a student. In my experience, most mornings before classes were spent working on internship work, and afternoons were spent for coursework generally. "Real work" and "school work" require slightly different skill sets, but the statistical knowledge gained from school is definitely necessary. To expand on this idea, often times for my internship work I'd have a problem that needed to be solved, but lacked the programming knowledge to solve whatever problem I had. A lot of time can be spent searching for programming solutions to problems, and basic tasks can lead to other "sub-problems" that consume more time. Over time I learned some basic data processing strategies which may or may not be useful in the future, but the overall process of having a programming problem and having to search for or come up with a solution is a process that I grew accustomed to and expect to encounter in any future work experience.

Another major difference between school and work experience is the way data is prepared. In my school experience, often times I would simply be presented with a dataset and told to do an analysis. In my internship experience, I learned that it is usually never that simple. Companies can have massive databases of data, often

times with many different entries per patient that need to be aggregated to one entry per variable per patient. The data processing steps from the raw data to the eventual "final dataset" are usually heavily discussed and different decisions can be made. In my experience, doing actual analysis on data is not the time consuming part of any project. The time consuming part of a project is the data processing, discussions, and eventual decisions that will lead to a dataset being finalized for analysis.

One expectation from school that definitely held true in this experience is the fact that clinical/real-world data is not always clean. Missing data or data not recorded properly can be encountered regularly, and identifying these problems and fixing them is a process in itself. Sometimes missing data can be indicative of a larger problem or be a hint to something else. For example, in the SCD project, at one point it was found that approximately four percent of the patients did not have any data for what access method was used in the baseline period. After multiple discussions and explorations of this, the decision was made to pull data about where patients received their treatment and to explore. It was found that the patients missing access data were patients not receiving in-center treatments. The decision was then made to exclude them from the study and have the study's population be patients who only received in-center treatments. A problem like this can be very difficult to trace down to the root of the problem, and in this instance, due to the fact that new data needed to be pulled, finding the solution took time due to the team that pulls data having other responsibilities and workloads to handle before they can immediately help with any requests.

Ultimately, the internship experience was a great way to learn and gain experience in an environment outside of academia. I personally learned a lot, and would highly recommend a similar experience to any student interested.

4.2 Final Remarks

In addition to the previous problems listed in Chapter 3 involving sequential testing in settings outside of clinical trials. A closer investigation of propensity scores could be another potential route for future research. Propensity scores can be utilized to adjust for confounding factors outside of matching. Two such examples of this are inverse probability weighting or including the propensity score as a confounding variable in a model. While these methods make the modeling process a little bit more difficult to explain, exploring whether they are equivalent approaches that yield similar results would be worthwhile. Additionally, exploring whether there any situations that are problematic for propensity score matching such that another approach would be more appropriate may be worthwhile.

A confounding variable is a variable that is correlated to independent variable(s), as well as the response variable. Since the confounder should be associated with both the group status variable and the outcome, including key confounding variables in the Cox model could be beneficial even after matching. A better cox model fit would theoretically results in smaller confidence intervals and more powerful tests (helpful when studies are underpowered).

In summary, this thesis set out to provide a general foundation in survival analysis and provide an overview of the propensity score matching strategy to help adjust the effects of confounding variables. The analysis of the SCD data in Chapter 2 demonstrated propensity score matching strategies, provided an example of survival analysis, and provided an important discussion on the choice of what information to perform PSM. Additionally, the SCD data set is curated in such a way that it could be effectively used as a case study to investigate the behavior of different sequential strategies and PSM strategies by future students at SFASU.

BIBLIOGRAPHY

[1] David Collett, *Modelling survival data in medical research, 3rd ed.*, Chapman Hall/CRC, 2014.

[2] David L. Demets and K. K. Gordon Lan, *Interim analysis: The alpha spending function approach*, Statistics in Medicine **13** (1994), no. 13-14, 1341–1352.

[3] Loprinzi CL et. al., *Prospective evaluation of prognostic variables from patient-completed questionnaires*, North Central Cancer Treatment Group. J Clin Oncol. (1994).

[4] K. K. Gordon Lan and David L. DeMets, *Discrete sequential boundaries for clinical trials*, Biometrika **70** (1983), no. 3, 659–663.

[5] Fabio Salvatore Macaluso, Walter Fries, Antonio Carlo Privitera, Maria Cappello, Sebastiano Siringo, Gaetano Inserra, Antonio Magnano, Roberto Di Mitri, Filippo Mocciaro, Nunzio Belluardo, and et al., *A propensity score-matched comparison of infliximab and adalimumab in tumour necrosis factor- inhibitor-naïve and non-naïve patients with crohn's disease: Real-life data from the sicilian network for inflammatory bowel disease*, Journal of Crohn's and Colitis **13** (2018), no. 2, 209–217.

[6] Ankit Mangla, Moavia Ehsan, Nikki Agarwal, and Smita Maruvada, *Sickle cell anemia*, StatPearls Publishing, Treasure Island (FL), 2022.

[7] P. McCullagh and J.A. Nelder, *Generalized linear models, 2nd ed*, Chapman Hall/CRC, 1989.

[8] Paul Rosenbaum and Donald Rubin, *The central role of the propensity score in observational studies for causal effects*, Biometrilka **70** (183), 41–55.

[9] Felix J. Thoemmes and Eun Sook Kim, *A systematic review of propensity score methods in the social sciences*, Multivariate Behavioral Research **46** (2011), no. 1, 90–118, PMID: 26771582.

[10] Rima Zahr, Marianne McPherson Yee, Kenneth I. Ataga, Jeffrey Winer, and Jeffrey D. Lebensburger, *Outcomes of patients with sickle cell disease and eskd in the usrds registry*, Blood **138** (2021), 4056.

VITA

Adam Garrison was born 8/9/97 in Longview, TX. He graduated from Hallsville High School in 2015. After high school, he attended the University of Texas at Tyler and received a Bachelor's degree in Mathematics in 2019. In 2021, he attended Stephen F. Austin to pursue a Master's of Science in Mathematics with an emphasis on statistics with plans to graduate in May 2023.

Permanent Address: 327 W College St.
Nacogdoches, TX 75965

The style manual used in this thesis is A Manual For Authors of Mathematical Papers published by the American Mathematical Society.

This thesis was prepared by Adam Garrison using LaTeX.