View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by AFTI Scholar (Air Force Institute of Technology)

Air Force Institute of Technology
**AFIT Scholar**

Theses and Dissertations

Student Graduate Works

3-22-2018

# Analysis of a Medical Center's Cardiac Risk Screening Protocol Using Propensity Score Matching

Jake E. Johnson

Follow this and additional works at: https://scholar.afit.edu/etd

Part of the Applied Mathematics Commons

Recommended Citation

Johnson, Jake E., "Analysis of a Medical Center's Cardiac Risk Screening Protocol Using Propensity Score Matching" (2018). *Theses and Dissertations*. 1911.
https://scholar.afit.edu/etd/1911

This Thesis is brought to you for free and open access by the Student Graduate Works at AFIT Scholar. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of AFIT Scholar. For more information, please contact richard.mansfield@afit.edu.

# ANALYSIS OF A MEDICAL CENTER'S CARDIAC RISK SCREENING PROTOCOL USING PROPENSITY SCORE MATCHING

THESIS

Jake E. Johnson, 2nd Lieutenant, USAF

AFIT-ENC-MS-18-M-129

**DEPARTMENT OF THE AIR FORCE**
**AIR UNIVERSITY**

**AIR FORCE INSTITUTE OF TECHNOLOGY**

**Wright-Patterson Air Force Base, Ohio**

**DISTRIBUTION STATEMENT A.**
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

The views expressed in this thesis are those of the author and do not reflect the official policy or position of the United States Air Force, Department of Defense, or the United States Government. This material is declared a work of the U.S. Government and is not subject to copyright protection in the United States.

AFIT-ENC-MS-18-M-129

# ANALYSIS OF A MEDICAL CENTER'S CARDIAC RISK SCREENING PROTOCOL USING PROPENSITY SCORE MATCHING

THESIS

Presented to the Faculty

Department of Operational Sciences

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the

Degree of Master of Science in Operations Research

Jake E. Johnson, BS

2nd Lieutenant, USAF

March 2018

**DISTRIBUTION STATEMENT A.**
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

AFIT-ENC-MS-18-M-129

# ANALYSIS OF A MEDICAL CENTER'S CARDIAC RISK SCREENING PROTOCOL USING PROPENSITY SCORE MATCHING

Jake E. Johnson, BS

2nd Lieutenant, USAF

Committee Membership:

Lt Col Andrew J. Geyer, PhD
Chair

Dr. Raymond R. Hill
Member

## Abstract

Researchers at the University of Maryland Medical Center in Baltimore have developed a cardiac risk stratification protocol in the hopes of reducing the time-from-arrival-to-first-operation for geriatric orthopedic patients. They collected observational data for two years prior to and following the October 2014 implementation of the new screening protocol. Therefore, advanced analytical techniques are required to isolate the treatment effect of the new screening protocol. Propensity score matching (PSM) is used to handle the observational data in order to reduce the bias attributable to the confounding covariates. In addition to PSM, various regression techniques are used to help the researchers determine if their treatment has been successful in reducing the number of cardiac complications experienced by the elderly patients during and post-surgery. Recommendations are then made to the hospital's researchers.

# Acknowledgments

I would like to express my sincere appreciation to my faculty advisor, Lt Col Andrew Geyer, for his guidance and support throughout the course of this thesis effort. His insights and experience were greatly appreciated. I would also like to thank my sponsor, Dr. Bryce Haac, from the University of Maryland Medical Center, for both the support and latitude provided to me in this endeavor.

Jake E. Johnson

## Table of Contents

## List of Figures

Page

## List of Tables

Page

# ANALYSIS OF A MEDICAL CENTER’S CARDIAC RISK SCREENING PROTOCOL USING PROPENSITY SCORE MATCHING

## I. Introduction

### General Issue

The healthcare world is highly influenced by the continual advancement of science, technology, and medicine. With millions of patients’ health and billions of dollars at stake, there is a constant demand for better technology as well as increased medical research. As developers and researchers strive to meet this demand, the medical world has become more sophisticated and effective at helping patients. This means life expectancies are increasing and more patients are being saved. Throughout the world, death rates from infectious and cardiovascular diseases have fallen significantly. People are living longer than they were twenty years ago (“Life Expectancy”, 2015). While technology and research play a large role, some of this success can be attributed to the increased utilization of Operations Research techniques in the healthcare field.

Many medical studies are heavily reliant on observational data analysis. Consequently, medical researchers are turning to Operations Research analysts for assistance with producing reliable results. Due to the issues resulting from reliance on observational data, basic statistical techniques must be discarded in favor of more advanced methods. These results can be found in articles in the medical field in journals like *Operations Research for Health Care*, which features new Operations Research-based medical publications four times a year. This thesis is another application of Operations Research in the healthcare field, with a goal of improving the efficiency and effectiveness of a hospital’s patient-care and surgery process using observational data analysis.

## Problem Statement

One primary way that medical researchers improve patient care is by collecting patient treatment data and analyzing it to develop more effective patient treatment prioritizations. This thesis examines efforts at the University of Maryland Medical Center in Baltimore to streamline its orthopedic surgery screening process for elderly (geriatric) patients. Due to their age, these patients require increased attention as they are more prone to developing serious complications during and after surgery. The most severe complications are cardiac events, which can often prove fatal. Thus, the hospital developed a cardiac screener to better schedule surgical procedures with the patients at the greatest risk operated on soonest. This screener, which the hospital calls a ‘pre-operative cardiac risk stratification protocol’, is designed to assess the severity and surgical urgency faced by its geriatric patients. The hospital’s screening protocol utilizes patient data to assign geriatric surgical priority based on risk of cardiac events. The screener also serves to replace standard pre-operative testing procedures by gaining relevant patient history and risk factors. Consequently, it reduces the time normally devoted to pre-operative processes that often delay patients getting into surgery.

The subject matter expert (SME) for this problem is Dr. Bryce Haac, MD. Dr. Haac is a surgical resident and researcher at the Maryland Medical Center. Dr. Haac and her associates seek to improve their hospital’s ability to care for its geriatric patients. In this effort, her team is studying geriatric patients who require non-emergent orthopedic surgeries. As is common in the medical field, the hospital did not design an experiment with randomization and pre-planning. The researchers simply compiled geriatric orthopedic surgery patient data during the designated study time period. The study's data collection began on October 1, 2012 and lasted until October 31, 2016. The screening protocol took effect October 1, 2014. Since all the data in their study are observational in nature, analysis techniques that account for the lack of a deliberate experimental design are required to extract useful insights from the collected data.

## Research Objectives

In this thesis, propensity score matching, significance testing, and regression techniques are used to determine if the new pre-operative cardiac risk stratification protocol improved (shortened) the time from arrival to operation for geriatric orthopedic patients. Additionally, Dr. Haac's data are analyzed for indications that the new screening protocol reduced the number of patients experiencing intra- or post-operative cardiac complications, regardless of whether the protocol reduced the time needed to get into surgery. The protocol was employed in an attempt to optimize how the hospital orders and schedules its surgeries for these specific patients in order to increase efficiency and help save more lives. However, it is unclear whether the protocol has been effective so far due the large amount of complicated and highly correlated patient variables and factors, as well as many potentially confounding pieces of data. Thus, there is a need for deep, Operations Research analysis and modeling (via propensity score matching, significance testing, and regression techniques) to confidently ascertain the efficacy of the new screening protocol on the two outcomes of interest.

# II. Literature Review

## Chapter Overview

This chapter provides relevant background on the tools used during this thesis. Additionally, the literature review serves to express the importance of the problems that this thesis aims to solve. First, the general aspects of the observational data problem at hand are covered. Next, an overview of valuable and necessary data cleaning steps is supplied. Third, the primary tool for solving this problem, propensity score matching (PSM), is detailed. Fourth, a summary of similar works using PSM is given. Lastly, the techniques used for producing insights from the data are discussed. These techniques include Student's *t*-test, feature selection, and regression modeling, primarily via the negative binomial.

## Description

Dr. Haac's team is interested in questions of cause and effect. The researchers' principal goal is to test the directly attributable impact of their October 2014 risk stratification protocol. Before beginning the steps of cause and effect analysis, it is important to better understand the meaning of causal relationships, especially in observational data. In a report overviewing causal inference, Pearl (2009:97) states that most studies in health, social, and behavioral sciences are motivated by questions that are not associational, but causal in nature. He adds that questions about the effects of a drug in a population, for example, are causal since they require an understanding of the data-generating process (Pearl, 2009:97). This means they can be neither answered solely from the data nor from the distributions to which the data belong (Pearl, 2009:97). Dr. Haac's team is faced with a causal question as well, as they want to determine the effectiveness of a treatment (the screening protocol implementation) from data that is purely observational. Consequently, the methods and tools used in this thesis must be tailored towards causal problems, while accounting for the fact that the available data did not originate from a randomized, designed experiment.

While discussing the nature of causal inference, Rubin (1990:472-480) states that observational data must be structured in a specific format to distinguish cause and effect relationships. He adds that the standard approach for handling an observational dataset is to use one variable to represent the observed outcome and a separate indicator variable to show treatment assignment (Rubin 1990:476). He states that these variables are necessary despite the fact that a treatment assignment variable will be correlated with the observed outcome variable if the null hypothesis does not hold true (Rubin 1990:476). With this in mind, it is important during one's data preparation phase to ensure that this treatment variable is clearly created and defined so that proper analysis can be performed.

Thus, the problem addressed in this thesis must be framed as one of determining causal inference. But first, to address the cause and effect question posed by Dr. Haac's team, a plan is required to mold the dataset into a proper format that can be analyzed. Before any advanced analytic methods can be performed, observational data must be thoroughly sorted, transformed, and reduced. This guarantees that the final dataset is as simplified, understandable, and usable as possible. The set of steps involved in organizing an observational dataset is typically referred to as the data cleaning process. Numerous data cleaning steps can be performed on an observational dataset. A review of current academic literature as well as input from the appropriate subject matter experts provide the best avenues for determining which of these steps are required before performing analysis on any observational dataset.

Kuhn and Johnson (2013:27-48) details proper data cleaning steps that are pertinent to the issues faced in many observational datasets. In their book on predictive modeling, the authors devote a chapter to outlining key data cleaning rules and procedures. They begin by stating that data cleaning must be performed thoroughly and carefully, as these steps can make or break a model's predictive or explanatory ability (Kuhn and Johnson, 2013:27). This concurs with other research and intuition, as any observational data analytic methods will be both meaningless and powerless if the data is not first molded into a clean and usable format.

The authors cover many different steps for data cleaning. First, they discuss the issue of dealing with missing values, which often plaque observational datasets. For this common problem, Kuhn and Johnson (2013:41) provides a simple solution; often, the percentage of missing data is substantial enough to eliminate the variable (predictor) in question from the dataset. Intuitively, this makes sense, as not much can be gleaned from a vector that is significantly empty. Consequently, any columns that are found to significantly blank, say greater than 20%, can be removed. This constitutes a significant first step in reducing the variables down to only those that contain sufficient information for analysis.

Sometimes, however, the percentage of missing data in a column is not sufficient enough to warrant complete deletion. In these frequent cases, Kuhn and Johnson (2013:42) suggests imputation. Imputation is the process of filling in missing values in a dataset by relying on the meaning and information in the surrounding covariates. Often, if the amount of missing values in a column is small, say less than 5%, it is permissible to use a simple form of imputation called mean imputation (Zhang, 2015:6). Mean imputation involves using the mean of the column in question (sans the empty values) to fill in the missing data entries. This process is simple and unlikely to interfere with future modeling assumptions, provided that the amount of values that were mean imputed is small (Zhang, 2015:6). However, if the amount of missing values in a column is large (but not large enough to permit complete removal), then Kuhn and Johnson (2013:42) suggests using a linear model to predict the missing values. This approach involves generating a linear regression model that uses the surrounding covariates to estimate the values that are empty. They also discuss a more complex method, *K*-nearest neighbor imputation, which involves using a nearest neighbor model to fill in missing values with other values that are 'closest' to them, according to a set of pre-defined distance parameters (Kuhn and Johnson 2013:42-43). The process of imputation, regardless of the method applied, can help increase the meaningfulness and predictive power of a dataset. Especially in messy, observational datasets, which are prone to featuring missing values, imputation is a necessary data cleaning tool.

Next, Kuhn and Johnson (2013:27) detail specific steps like "feature engineering" and "feature extraction", which involve the transformation and combination of variables in order to make them more meaningful. They define performing feature engineering as the creation of combinations or ratios of variables to make better predictors (Kuhn and Johnson, 2013:27). These are the cases where the combination or ratio of two vectors of data is more informative than analyzing just the two columns independently (Kuhn and Johnson, 2013:27). This step is quite relevant, as there are often many messy columns in observational datasets that can be combined or transformed into better predictors for subsequent analysis.

The authors then proceed to discuss another common issue in observational data, which is called “zero variance predictors” (Kuhn and Johnson, 2103:44). They define this as columns where the data take on a single value for each row (observation) (Kuhn and Johnson, 2103:44). Hence, these columns (predictors) have zero variance and as a result, offer nothing of value to any modeling technique. Another related and even more common issue in observational data is what the authors term “near-zero variance predictors” (Kuhn and Johnson, 2103:44). These predictors are columns of data which are comprised of mostly one singular value as well as the smattering of a few other values throughout (Kuhn and Johnson, 2013:44). Consequently, these columns have near-zero variance and likely contribute little to any statistical model. Thus, the authors find that it is advantageous to remove such variables from one’s dataset (Kuhn and Johnson, 2013:45). The authors supply straightforward criteria for eliminating such columns, which can be utilized in cleaning most observational datasets (Kuhn and Johnson, 2013:44-45). They state that if the percentage of unique values over the entire column is below a low threshold, say 10%, and if the ratio of the frequencies of the most common value to the second most common value is large enough, say above 5, then the column in question can be excluded from the dataset (Kuhn and Johnson, 2013:45). This process can be conducted one-by-one, until all the variables in the dataset have been examined and removed if they meet the exclusion criteria. Once such zero and near-zero variance predictors are removed, one’s dataset will be even closer to having only meaningful and clean columns of variables to analyze.

The next data preprocessing step that Kuhn and Johnson (2013:47-48) discuss is the addition of predictors through the creation of dummy variables. The authors state that, when faced with the presence of categorical predictors, it is best to resolve them by decomposing them into sets of more specific variables (Kuhn and Johnson, 2013:47). Specifically, this process involves turning the categorical predictors into columns of dummy variables with binary (0/1) indications for each category. Hence, in this manner, a five-category predictor would be decomposed into five binary dummy columns, where each column denotes one of the categories. Kuhn and Johnson (2013:48) then add that the final column (the fifth one in example above) can be removed in certain models, such as linear regression. This is because the values in the last decomposed column can be inferred from all the other column values. However, the authors state that some models, such as tree-based models, are insensitive to the full set of dummy columns, and therefore the addition of all the decomposed columns can increase their interpretability (Kuhn and Johnson, 2013:48). In either case, the detection and transformation of categorical predictors into the necessary dummy columns is an important step in improving the clarity and meaningfulness of one's dataset.

Another common data issue that Kuhn and Johnson (2013:33-34) cover is the presence of outliers. Generally speaking, the authors consider an outlier to be an observation that is exceptionally far from the majority of the data (Kuhn and Johnson, 2013:33). The authors suggest dealing with outliers carefully, as it is often difficult to determine whether the source of an outlier was a data entry error or perhaps just a legitimate value in the data's distribution. Resolving outliers depends on what their source is believed to be. If it is likely that an outlier resulted from a data entry error, it is permissible to delete that row as such an outlier is likely to hinder later analysis. However, if it is unclear on where an outlier originated from, more significant methods are required to resolve the issue. Kuhn and Johnson (2013:34) suggest the process of selectively choosing analytic models later on based on their resistance or sensitivity to outliers. This approach guarantees that one accounts for the presence of outliers and chooses an appropriate modeling technique when performing analysis. In whole, regardless of the method used, resolving outliers is a necessary step in data cleaning and is imperative for successful analysis. By investigating outliers, the analyst can ensure the dataset is even more clean and informative.

The next step the authors suggest is correlation analysis on the all the variables (Kuhn and Johnson, 2103:45-47). In observational data, there are often many correlated and potentially redundant predictors. Consequently, not all these predictors are essential for analysis. Therefore, this step involves locating and removing these unnecessary predictors, which then allows for a less messy dataset. With regards to a redundant predictor, Kuhn and Johnson (2013:43-44) find that eliminating it should not impact the performance of the model and could lead to a more parsimonious and interpretable model. To conduct this step, software is typically required to input the dataset and calculate the correlation matrix for the covariates. It is then often beneficial to color-code the correlation matrix according to a high cut-off threshold, say ± 0.9, so as to identify the variables that are extremely correlated with one another. This step is followed by using an algorithm like the one supplied by Kuhn and Johnson (2013:47), which suggests arbitrarily removing a member of each pair of variables with an absolute correlation above the chosen threshold until no more highly correlated pairs exist. This approach safely and properly removes the strongly correlated predictors in one’s observational dataset. Once these unnecessary predictors that meet the cut-off criteria are deleted, the dataset will another step closer to being fully clean and ready for analysis.

While Kuhn and Johnson (2013:27-48) cover more generalized issues that are prevalent in observational datasets, each observational dataset is unique and presents its own share of specific challenges for the analyst. Often, one must rely on the insights of the relevant SME. As the individual who collected the data or oversaw the data collection, the SME will usually have the best guidance on how to handle the many unique types of observational data issues that an analyst may face (Loukides, 2012). This means the SME can advise the analyst on a myriad of issues to include: subjective decision-making, such as which variables can be outright deleted and which variables will be the most important for subsequent analysis. Furthermore, the SME can alleviate vagueness or the lack of descriptions in the observational data by defining variables and explaining the scale or units of certain columns. Overall, without reliance on the SME for direction, one can often become stuck and unsure of how to handle certain issues. This can then lead to false assumptions and critical mistakes such as inadvertently removing the most predictive variables or transforming variables inappropriately (Loukides, 2012). Thus, to eliminate such mistakes, one should utilize the SME as an authority frequently when conducting the data cleaning process on observational datasets.

Now that the review of the primary data cleaning tools is complete, the next step involves reviewing the most suitable techniques for handling messy, observational data and answering causal inference questions. Before beginning this step, it is crucial to note that unfortunately, despite performing data cleaning, this problem cannot just be simply solved via a means comparison and significance testing. That is, when dealing with an observational dataset, one cannot adequately assess the effect of a treatment by simply comparing the means of the pre- and post-treatment groups. Due to the presence of many other potentially confounding and correlated variables, the user cannot just isolate the outcome of interest and compare the pre- and post-treatment means via a significance test. Since there are many other variables in play, one cannot trust the results of a significance test to be meaningful as there could likely be confounding interference due to the presence of the covariates, leading to inaccurate and biased results. Thus, questions of this nature require some understanding and accounting for the manner in which the data were collected (Pearl, 2009:97). In observational studies, researchers have no control over the treatment assignment (D'Agostino, 1998:2265). The treated and non-treated (control) groups may have noticeable differences in their observed covariates, and these differences can lead to a biased estimate of the treatment effect (D'Agostino, 1998:2265). This bias, which is inherent to most observational datasets, is called treatment selection bias and poses the primary dilemma that many researchers face when analyzing data that did not arise from designed experiments (Thavaneswaran, 2008:4). Thus, all these covariates in the data must be taken into account. Hence, one must rely on more complex methods and techniques to solve this problem.

Fortunately, this problem is well-studied in healthcare applications as well as in many other fields. This is because it is often either unethical or impossible to randomly assign patients to control and treatment groups in medical studies. Therefore, methodologies were developed to account for the treatment selection bias present in observational data. Over the past few decades, there have been several techniques created for dealing with observational datasets. One of these main techniques is propensity score matching (PSM). PSM is designed to allow the analyst to test the effect of a treatment variable by comparing the pre- and post-treatment sets of data while reducing confounding variable bias in analyses of observational data (“Evaluating Observational Data”, 2016). In an article detailing multiple uses of PSM in medical research, Austin (2008:2037) states that propensity score methods are frequently being relied on to reduce the impact of treatment selection bias when using observational data to estimate treatment effects. Therefore, PSM is usually the first technique employed to address problems of this nature.

Rosenbaum and Rubin (1983:41-42), the original source of PSM, discusses the mathematics behind propensity scores. The authors detail how using PSM can allow the analyst to properly assess how a treatment is performing even when there are many other variables in play. Propensity score techniques are based on the underlying premise that in an observational data problem, all observations fall into either the pre-treatment group or the post-treatment group, but not both, which leads to a missing data problem (Rosenbaum and Rubin, 1983:41). Holland (1986:947) refers to this dilemma, wherein it is impossible to simultaneously observe both the effects of treatment and control on an individual unit, as the *Fundamental Problem of Causal Inference*. Propensity scores remedy this issue by trying to fill in the missing values (Rosenbaum and Rubin, 1983:41). That is, propensity scores work to estimate what would have occurred had the pre-treatment observations been treated, and the post-treatment values left untreated (Rosenbaum and Rubin, 1983:41). PSM provides an avenue to take a purely observational dataset and transform it into a quasi-randomized experiment (D’Agostino, 1998:2267). From there, standard analytic methods can be applied to one's newly transformed dataset as it will now resemble data that had been collected in a designed experiment from the onset.

Rosenbaum and Rubin (1983:41) introduce propensity score methods as a technique for searching for causal effects while accounting for all the variables in the data. The authors define a propensity score as the conditional probability of assignment to a particular treatment given a vector of observed covariates (Rosenbaum and Rubin, 1983:41). They further define a propensity score as a specific type of balancing score that can be used to group treated and untreated observations. These grouped observations can then be used to make reliable direct comparisons between the treated and untreated groups (Rosenbaum and Rubin, 1983:42). The mathematical notation for the propensity score is:

$$e(x) = P(z = 1|x),$$

where this formulation indicates that a propensity score, $e(x)$, is the probability of exposure to treatment (with $z = 1$ for treated values and $z = 0$ for untreated values) given the presence of the covariates (Rosenbaum and Rubin, 1983:42). Next, they suggest methods for estimating the propensity scores. In a randomized, designed experiment, $e(x)$ is easily obtainable. In other words, $e(x)$ is a known function, such as 0.5 when the treatment is randomly applied to half the observations (Thavaneswaran, 2008:2). But in a nonrandomized, observational experiment, one must use a tool (such as a logistic regression (logit) model or discriminant scores) to estimate $e(x)$ (Rosenbaum and Rubin, 1983:42,47).

Rosenbaum and Rubin conclude by describing three propensity score techniques that can be explicitly used to account for confounding covariates in observational data. These techniques are matched sampling (or PSM), subclassification (also called stratification), and covariance adjustment (Rosenbaum and Rubin, 1983:48). These three techniques each have their own strengths and specific applications, and thus it is important to consider all three when preparing to analyze observational data (Rosenbaum and Rubin, 1983:48-54). PSM is the most widely used technique of the three as it is the most applicable to the majority of observational datasets.

The methodology of PSM is based on one primary assumption. This underlying concept is referred to as the *Strong Ignorability of Treatment Assignment* (Imai and van Dyk, 2004:855). This crucial concept assumes that when conducting PSM, there are no unobserved differences between the control and treatment groups, conditional on all the observed covariates (Stuart, 2010:5). In other words, to satisfy this PSM assumption, all variables that are known to be related to the treatment variable and/or the outcome(s) of interest must be included in the model. Consequently, it is suggested that analysts be “liberal” when adding variables to a PSM model since it can be very detrimental to a PSM model’s ability to reduce treatment selection bias if an important confounder is left out (Stuart, 2010:5). In total, this assumption helps guarantee the efficacy of PSM methods. By ensuring that all relevant variables are included in the model, one can maximize the power of PSM in balancing out the differences between the control and treated groups in observational data.

In simplest terms, PSM begins by calculating and assigning a propensity score to each observation in the dataset. The propensity score for each observation is usually estimated by researchers using a logit or probit regression model with exposure to the treatment as the dependent variable (Austin, 2008:2037). This ‘treatment dependent variable’ is indicated by a binary vector as suggested by Rubin (1990:476), where 0 denotes a pre-treatment observation and 1 denotes a post-treatment observation. Next, PSM takes each post-treatment observation and its associated propensity score, and, in accordance with the type of matching method and its parameters, matches it with a pre-treatment observation. Typically, PSM matches the treatment observation to the non-treatment one via nearest neighbor matching (Ho *et al*, 2017:6). This means that observations in one group are matched with those in the other that bear the nearest propensity score in terms of Euclidian distance or some other predetermined distance function (Ho *et al*, 2017:7).

However, there are other matching methods available for PSM that make use of varying criteria to match observations in the treated and untreated groups. The other main PSM methods are exact matching, subclassification, full matching, optimal matching, genetic matching, and coarsened exact matching (Randolph *et al*, 2014:3). Exact matching is the method of matching where pairs of control and treated observations are matched only if their values are identical for every covariate (Randolph *et al*, 2014:3). This method is often not functional for large datasets as the probability of having pre- and post-treatment observations with the exact same value for every covariate is near zero.

Subclassification is the explicit process of splitting the dataset up into multiple subclasses (usually five or six), where the distribution of the covariate data is similar in each one (Ho *et al*, 2011:10). Then, typically, the standard nearest neighbor matching proceeds inside each subclass to produce a better balance between the control and groups.

Full matching is a specific type of subclassification where the subclasses are formed in an optimal way (Randolph *et al*, 2014:3). The full matching method's goal is to minimize a weighted average of the estimated distance measures between the matched pre- and post-treatment observations in each subclass.

The next method, optimal matching, minimizes the average absolute distance across all the matched pairs (Randolph *et al*, 2014:3). Genetic matching is more complex than the rest of the methods in that it uses a "computationally intensive genetic search algorithm" in order to create matches between the pre- and post-treatment sets (Randolph *et al*, 2014:3). The final primary PSM method, coarsened exact matching, is mostly intended for datasets with multi-level treatments, as opposed to the standard 0-1 binary treatment found in most observational datasets. This method creates balance in the groups by matching on one covariate while maintaining the balance of the other covariates (Randolph *et al*, 2014:3). Each of these matching methods could result in similar, but disparate results from the other matching methods. Therefore, it is important to apply all the available methods to one's observational dataset when conducting PSM to determine the best matching methodology (Randolph *et al*, 2014:3).

Once the matching method is specified, PSM then repeats the process of matching pre- and post-treatment values, typically without replacement, until all observations are matched up. If the size of the pre- and post-treatment groups are unequal, the two groups are said to be imbalanced. In an imbalanced dataset, some values cannot be matched one-to-one with values in the other group. If this occurs during the matching process, one can employ weights as a way of matching up all the observations (Ho *et al*, 2017:7). In PSM, this technique is found to be most effective for imbalanced datasets for one-to-two up to one-to-five matching (Randolph *et al*, 2014:3). This weighted matching technique results in more balanced control and treated sets, which allows for more reliable analysis. Other matching methods, like genetic matching, discard some of the pre-treatment observations in order to more efficiently match the two groups and achieve the best balance between them (Ho *et al*, 2017:7).

A matching method is deemed successful if it greatly improves the balance between the pre- and post-treatment groups. After applying any of the various PSM methods, this balance is typically measured by determining the reduction in the mean differences between all the pre- and post-treatment covariates (Randolph *et al*, 2014:3). In other words, the success of a particular matching method is gauged by comparing the pre- and post-matching balances. The balances are assessed by calculating how the mean differences of the pre- and post-treatment covariates changed before and after the certain matching method is applied. This assessment is referred to as the percent balance improvement and is calculated according to:

$$100\left(\frac{|a|-|b|}{|a|}\right),$$

where $a$ is the pre-matching balance and $b$ is the post-matching balance (Ho *et al*, 2011:13). Once the matching is concluded via the most appropriate method and parameter settings, the dataset accounts for the inherent treatment selection bias resulting from the use of observational data. The post-PSM dataset now features matched pairs (or sets) of pre- and post-treatment observations thereby permitting the analyst to operate under the assumption that these dataset values were randomly assigned to each group.

This key assumption then allows for analysis of the treatment's effect as if it was collected in a randomized, controlled test.

In a controlled test, the only difference between the pre- and post-treatment groups of data is treatment assignment. Thus, any differences between the two groups can be attributed to the effect of the treatment. Hence, post-PSM, it is possible to look at the isolated treatment effect separate from the effects of the other covariates and the inherent treatment selection bias. When examining the isolated treatment effect, the *Fundamental Problem of Causal Inference* described in Holland (1986:947) must be kept in mind. Since it is impossible, even after applying PSM, to ever determine the effect of a treatment on a single unit, the treatment effect must be looked at more broadly. Hence, researchers using PSM assess treatments by looking at the mean effect across all the observations as well as just across the treated values. The mean effect on all the units in the dataset is called the average treatment effect (ATE) and is calculated according to the following equation:

$$ATE = \frac{1}{n}\sum_{i=1}^{n} E[Y_i(1) - Y_i(0) \mid X_i].$$

Here $Y_i(1) - Y_i(0)$ represents a quantity that can never be known for any specific observation due to the *Fundamental Problem of Causal Inference* (Ho *et al*, 2007:204). But over the span of the dataset, the ATE can be an informative metric in aiding researchers to assess the value of a treatment. The ATE can be interpreted as the effect of the treatment on all the observations had all observations experienced the treatment. In most cases, the ATE is the primary measurement desired by researchers. However, in addition to examining the ATE, researchers also use the average treatment effect on the treated (ATT) (Ho *et al*, 2007:204). This value is obtained similarly to the ATE, but is limited only to the treatment population:

$$ATT = \frac{1}{\sum_{i=1}^{n} T_i} \sum_{i=1}^{n} T_i E[Y_i(1) - Y_i(0) \mid X_i].$$

Here, the only distinguishing difference between the ATE and the ATT is that the ATT represents the effect of the treatment on only those who underwent the treatment. This metric is also useful to researchers, especially in cases where they are only interested in determining the specific effect of the treatment on those that received the treatment and/or those who *would* receive the treatment (Ho *et al*, 2007:204).

To explain the difference between using the two metrics, Ho *et al* (2007:204) uses the example of a study involving a medical drug treatment, where the researchers were only interested in determining the effect of the drug on those who would or do receive the drug, and hence, they only focused on the ATT. In total, these two averages comprise the primary two metrics that analysts use to study treatment effects in observational data. Both the ATE and ATT are worth calculating and interpreting in order to provide the best measure of assessing one's treatment variable.

In practice, to conduct PSM and analyze the results, one must rely on a statistical software program. There are many capable programs, but the most appropriate one for problems of this nature is R (R Core Team, 2017). As a large scale, free, and open-source programming language, R enables the creation of user-built 'packages'. These are bundles of code designed to help other users solve specific problems. Fortunately, there are already a packages in existence that allow one to perform PSM in R.

One of the most notable and utilized PSM packages is MatchIt, which was created by Ho, Imai, King, and Stuart (2007:2-7). This package is comprehensive and user friendly. It allows one to tailor a large set of PSM parameters to fit the specific nature of one's observational data problem. In their package documentation, the authors list the parameters that can be altered when using MatchIt, like *formula*, *method*, and *distance*, among others (Ho *et al*, 2007:6). These parameters will vary some of the aforementioned considerations involved in PSM, such as the independent and dependent variables of interest, the type of matching mechanism used, and the distance function used to match data points (Ho *et al*, 2007:3).

To gain a better understanding of how to practically apply PSM to real observational data problems, previous research efforts where PSM was applied to observational data problems, including those where MatchIt was used, is reviewed.

## Relevant Research

PSM and other related methods have been used prevalently in medical research due to the frequent observational data issues faced in healthcare and pharmaceutical studies. Perkins *et al* (2000:93) used PSM in a pharmacoepidemiologic research study on the effect of two pharmaceutical drugs. To conduct their research, they relied on propensity scores to gauge the effect of two drugs, Ibuprofen and Sulindac, from an observational dataset (Perkins *et al*, 2000:93). Specifically, they employed a logistic regression model to calculate the propensity scores to combat possible confounding effects from imbalanced covariates (Perkins *et al*, 2000:93). In their case, the propensity scores equated to the probabilities of each patient being prescribed the treatment, Sulindac. Next, to determine the drug’s effect, they used the subclassification technique to stratify the data according to sample quintiles of the propensity score distribution in order to produce an average treatment effect (Perkins *et al*, 2000:93). In the end, the authors found PSM to be a successful technique, as they were able to balance out the few dozen covariates in their data and isolate the treatment effect, which was found to be statistically insignificant (Perkins *et al*, 2000:93). They concluded that propensity score analysis provided a simple but effective way of controlling the effects of covariates in order to obtain a less biased estimate of the treatment effect (Perkins *et al*, 2000:93).

Despite being frequently applied in medical studies, PSM use is not limited to healthcare research. Ho *et al* (2007:199-236) applied PSM (via the MatchIt R package they created) to a political science problem, where the goal was to estimate the effect of electoral advantage of incumbency for Democrats in the United States House of Representatives (Ho *et al*, 2007:202). The authors relied on PSM because they were dealing with causal effects, ‘missing’ data, and an observational dataset. They addressed this issue by first understanding that their problem revolved around determining the effect of the Democratic Party nominating an incumbent or not. This is because the Party can only do one or the other. Thus, Ho *et al* (2007:230) employed PSM via one-to-one, nearest neighbor matching to solve their problem in order to produce a reliable estimate of the ‘treatment’ in question. Ultimately, they found that PSM improved the balance of the covariates substantially and that it was a successful tool for the causal and observational data issues they faced (Ho *et al*, 2007:231).

Randolph *et al* (2014:1-6) presents another case of using PSM via MatchIt in the field of education. Specifically, they were interested in assessing the effect of designating certain schools as high performing ‘Schools to Watch’. The outcomes of interest were the students’ reading and mathematics achievement scores (Randolph *et al*, 2014:1). Like Ho *et al*, these authors utilized the nearest neighbor method, with one-to-one matching (Randolph *et al,* 2014:3). They examined all possible matching methods and combinations in MatchIt until they found the lowest mean differences between the control and treatment groups (Randolph *et al*, 2014:3). They advise that other analysts do so as well in order to find the most suitable set of parameters for their specific problem.

After running PSM with their optimal set of parameters, they showed that they were able to greatly fix their imbalanced dataset and produce two groups of quasi-randomized data, which could then be analyzed informatively. The authors further demonstrated the efficacy and importance of PSM. In their follow-up analysis, they found that without PSM, there was a significant difference in the mathematics performance of treated (‘Schools to Watch’) and untreated schools. However, after they employed PSM via MatchIt, the difference was no longer significant (Randolph *et al*, 2014:5). This highlights the need for PSM when analyzing observational data, as one can draw false conclusions about the effect of a treatment if the data is not properly transformed and balanced through PSM beforehand. Without PSM, the data are subject to treatment selection bias.

The researchers’ conclusions from the aforementioned studies demonstrate the value of PSM. Their work makes it clear that once the data cleaning steps are complete, PSM an appropriate technique to use to adequately prepare observational data for analysis. Now, the final part of preparing for analysis is to review which analytic tools are necessary to infer causal relationships from a newly clean and propensity score-matched dataset. Thus, it is paramount that the analysis techniques used post-PSM are capable of properly determining a treatment effect. That is, the techniques must be able to ascertain whether or not changes in pre- and post-treatment outcomes are statistically significant enough to warrant a conclusion on the positive or negative efficacy of the treatment.

Once the original messy, observational dataset is cleaned and the data are matched via PSM into a quasi-randomized dataset, statistical regression analysis is performed as most models require the assumption of randomized, experimental data. This underlines the importance of first thoroughly cleaning and then applying PSM to observational datasets. One of these tools is the calculation and assessment of mean differences via significance testing. Earlier, this tool was mentioned as inadequate for observational data, but now that the data has been matched, it is a sufficient method of analyzing a treatment's effect by the comparing pre- and post-treatment groups. As Randolph *et al* (2014:5) showed, significance testing of a PSM-matched dataset is a suitable way of computing a treatment effect and gaining insights about the original, unmatched observational dataset. Typically, the specific type of *t*-test that should be applied is a two-sample *t*-test using non-pooled variances (Kim, 2015:540). Since the pre- and post-treatment sets resulting from PSM can be treated as independent groups and can often vary in size, it is best to utilize this type of *t*-test (Kim, 2015:540). With the assumption that the two groups are now randomized, this *t*-test shows whether or not there is a significant difference in the means of the two groups, for a specific dependent variable outcome, according to a predetermined $\alpha$-level of significance. If there is a significant difference in the pre- and post-treatment outcome means, the analyst can be statistically confident that there was a positive (or negative) effect directly attributable to the treatment variable.

To conduct this *t*-test, an analyst first computes the means of the two groups' values for the outcome of interest. The analyst also calculates the standard deviations of each group and records the sizes (*n*) of each group. These values are then imported into the appropriate *t*-test calculator or statistical software. The results will show, according to the significance level chosen (typically $\alpha$ = 0.05), whether or not the groups have a significantly different means, and hence, the presence of an effectual treatment (Kim, 2015:541). This process is used to compare the change due to the treatment's implementation in any other outcome variable that one wishes to analyze. In the end, the *t*-test is a simple, but powerful way of analyzing randomized or quasi-randomized datasets to search for significant treatment effects. But while this tool is informative, it does not provide the analyst with much more than a comparison between two groups. Consequently, if one wants to gain more insights into a dataset, *e.g.* what variables, if any, best predict or explain cause and effect in the dataset, other, more complex tools need to be utilized.

As stated, significance testing does not afford one the ability to conduct predictive modeling or determine which variables, if any, explain the outcome. To gain these insights, one can use regression-based methods. The simplest form of regression is linear regression. This basic type assumes a linear relationship between the predictor(s) and the outcome, or response, variable. Observational data, particularly medical data of this nature, can feature linear relationships among the data, but it can also feature more complex, nonlinear relationships (Higgins, 2002:247). Thus, other more complex regression methods should also be explored. For datasets of an observational nature, where variables are often nonnegative and follow a Poisson process, the most appropriate regression tools are negative binomial and Poisson models (Cameron and Trivedi, 1999:1-2). These regression techniques allow for the modeling of count or arrival data as well as data that is purely positive (Cameron and Trivedi, 1999:1-2). If an observational dataset has primarily count data or nonnegative variables, it precludes applying any standard linear regression modeling technique. Thus, the negative binomial and the Poisson, which is a special case of the negative binomial, are appropriate models for observational datasets that fit these characteristics (Cameron and Trivedi, 1999:1-2).

The negative binomial model, while intended for count data, also works well for continuous data (Cameron and Trivedi, 1999:10). This is important since observational datasets are just as likely to feature nonnegative continuous variables as count data. In practice, the negative binomial regression takes the data and produces a model that predicts the counts of the desired outcome based off all the predictors chosen in the model (Cameron and Trivedi, 1999:7). Additionally, in some cases, a more particular model is better when the modeling circumstances allow.

The Poisson model is a specific instance of the negative binomial. The Poisson model carries the critical assumption that the means and variances of the data’s variables are equivalent (Cameron and Trivedi, 1999:2). This is a highly useful characteristic and allows for a simple and interpretable model. Thus, the Poisson model is preferred whenever possible. However, there are occurrences, especially in observational data, where the Poisson model’s assumptions are violated.

In the case where the variance of the data exceeds the mean, the Poisson model fails. This common situation is called overdispersion (Cameron and Trivedi, 1999:6). When this occurs, the negative binomial model is preferred because it features an extra parameter that allows the variance to remain independent of the mean. While more complex than the Poisson regression model, the negative binomial is less restrictive and most appropriate for cases when the Poisson's assumptions do not hold true (Cameron and Trivedi, 1999:6). In these cases, one can trust the negative binomial to yield the most reliable regression results.

In order to fully understand the Poisson and negative binomial regressions, it is important next to discuss the nature of the outputs given by these models. Whereas a simple linear regression model reports coefficients that can be linearly summed to predict the response variable, the Poisson and negative binomial rely on a log-link internally, which means their output coefficients are not in linear terms (Popovic, 2016). Consequently, these coefficients must be transformed in order to derive meaning and make predictions about an outcome variable. Typically, this conversion occurs by exponentiating the Poisson or negative binomial coefficients into a metric called odds ratios (Szumilas, 2010:227). Generally, an odds ratio is defined as a measure of association between a predictor and an outcome variable. More specifically, in the case of transforming Poisson or negative binomial regression coefficients, an odds ratio is given by $e^{\beta}$, where β is the coefficient in question (Szumilas, 2010:227). Once all the coefficients have been transformed into odds ratios, the odds ratios are used to generate predictions on the response variable. Together, the odds ratios (raised to the power of the predictor value), and the intercept, can be multiplied together to give the predictive equation for the outcome variable. Because the odds ratios are multiplicative, their general effects can be divided into three possible levels. If an odds ratio is less than 1, the predictor is associated with lower odds of the outcome occurring. If an odds ratio equals 1, the predictor has no effect on the outcome. And if an odds ratio is greater than 1, the predictor is associated with higher odds of the outcome occurring (Szumilas, 2010:227). In total, odds ratios provide an effective and understandable way of interpreting and applying the coefficients of Poisson and negative binomial models.

The next step is to discuss how to build proper models. Once the most suitable regression model for the dataset is chosen, the next task in the analysis is to select which variables, or features, to include in the model. The first priority of this process is that a more parsimonious model is always desired. Thus, the model chosen should include the minimal set of features required to explain as much variation in the response variable as possible (Guyon and Elisseeff, 2003:1158).

The process of narrowing down the list of features in the dataset to best set of predictors is referred to as feature selection. To conduct feature selection, there are a few tactics available. The most straightforward types of feature selection techniques are stepwise regression and best subsets regression (Zhang, 2016:2). These regression methods are designed to help select the best, or most significant, variables in the model. Stepwise regression works by either selecting the most significant variables in order of decreasing value, eliminating the least significant variables in order of increasing value, or a combination of both, depending on the method (Zhang, 2016:2-4). Specifically, these three methods are called forward selection, backward selection, and ‘both’ selection, respectively. Once stepwise regression is concluded, only the variables that are most important and significant to the model will remain. Best subsets regression operates slightly differently by choosing the best set of predictors for each possible model size according to specified criteria in the model (Zhang, 2016:4). Once the best subsets model is complete, the user has a list of the optimal set of predictors that are significant in the model that may or may not match the results of stepwise regression. Thus, it is important to use both methods, with comparisons made between them based on the values of their $R^2$. A model with higher $R^2$ is preferred, as a larger $R^2$ means more variance in the response is explained by the model's predictors.

Once this process is complete, the analyst has a better understanding of which predictors matter, which ones can be ignored, and can then proceed to the process of generating informative regression models that are able to reliably answer causal effect inquiries. The next chapter describes how these techniques are applied to the data generated by Dr. Haac's research effort in order to determine if the data supports the conclusion that there is a statistically significant causal effect on either time-from-arrival-to-first-operation or the number of cardiac events experienced by geriatric patients due to the implementation of the October 2014 risk screening protocol.

# III. Methodology

## Chapter Overview

This chapter describes the techniques that were applied to Dr. Haac's research data to statistically infer a significant causal relationship between the implementation of the October 2014 screening protocol and either the time-from-arrival-to-first-operation for geriatric patients, the number of cardiac events experienced by geriatric patients, or both. The analysis began with data cleaning. After that, PSM was performed to create a matched dataset. Then, the actual analysis began with significance testing of the factors present in the matched dataset. Next, features were selected based on techniques discussed in Chapter II. Once the best set of predictors were identified, predictive models were generated via the negative binomial regression model.

## Data Cleaning

As is usual in medical observational research studies, the original dataset provided by Dr. Haac's research team required substantial effort to put the data in a format that could be analyzed via the statistical methods discussed in Chapter II.

Dr. Haac supplied the dataset in the form of a Microsoft Excel spreadsheet with multiple sheets of data. The first sheet and primary set of data, titled "Demographics", is sorted by the study's 745 patients' unique identification numbers (ID #'s). For each patient, ninety columns of data are recorded. The first few columns contain simple demographic, or background, data like patient sex, race, and age. The data then progresses into patient medical factors and measurements. These values include hospital admittance date, discharge date, injury severity scores, systolic blood pressure, diastolic blood pressure, body temperature, blood alcohol content, as well as a few dozen other very specific measurements of patient motor, verbal, cognitive, and verbal skills. For each patient entry, there are also multiple columns of categorical, ‘string’ data that describe the patient’s category of injury, full injury description, date of injury, and injury scene. Finally, rounding out the “Demographics” dataset is a group of columns of numerical data measuring the level of certain drugs/narcotics, like marijuana or opiates, in the patients’ bodies.

The second and third sheets are titled, “Injuries AIS ’90” and “Injuries AIS ’05”, respectively. These two sheets each contain around two thousand rows, where the patients’ injuries are broken up individually into separate rows of data. Each of the injuries references the same unique patient ID # used on the “Demographics” sheet. Additionally, the same admit and discharge dates are referenced. Next, the sheets contain specific injury text descriptions. Also, for each injury, further columns of numerical data are recorded, with names such as “AIS”, “ICD9”, “ISS BR”, and “AIS PREDOT”. These columns contain specific scores and information that detail the severity and nature of the patient injuries.

The next sheet, “OR Procedures”, contains all the data referencing each of the 795 patients’ specific operating room procedures. During the time period of this dataset, there were 1,833 individual surgical procedures. As with the previous two sheets, each procedure references the patient by his or her unique ID #. Each procedure is tabulated by row, with each row featuring many pieces of descriptive data. The data includes columns that record the type of surgery, the minutes each patient had to wait to get into the operating room (the main outcome of interest for Dr. Haac’s team), the patient disposition (the plan for continuing healthcare after discharge from the medical center), the surgery start and end times, and a textual, ‘string’ description of each surgery. This sheet also contains some more columns of data like “ICD9”, “CPT”, and “SERVICEID” that contain bits and pieces of medical terminology that list further information about each surgery that occurred in the study.

The fifth sheet, titled “Co-morbidities”, tabulates a list of all the non-surgical medical risk factors possessed by each patient. The sheet contains close to 2,500 entries, and for each entry, the specific patient is again referenced by his or her unique ID #. These co-morbidities are extenuating medical problems held by some of the patients. They include issues like heart attacks, depression, hypertension, tobacco use, and hypothyroidism. On this sheet, each of these specific types of co-morbidities is designated by its own unique ID #. The purpose of this sheet is clear; it is intended to record any underlying medical factors for each patient, as these issues can play a serious role in predicting outcomes such as the patient’s required number of surgeries, likelihood of complications, length of stay (LOS) in the hospital, and time-from-arrival-to-first-operation.

The sixth and final sheet is titled “Complications”. This sheet records any complications that arose during and/or after the patients’ procedures and surgeries. Like earlier sheets, each complication is broken up into an individual row with the unique patient ID # referenced. In each row, a description of the medical complication is given as well as a unique complication ID #. The age of the patient is referenced again, as well as the admit and discharge dates. As with the “Co-morbidities” sheet, this data sheet exists to record the complications that result from the patients’ surgical operations. These complications can correlate strongly with outcomes like time-from-arrival-to-first-operation or LOS. This sheet also contains the relevant data that record the number of cardiac events experienced by the patients throughout the study, which is the other outcome of interest in the researchers' study.

Significant data cleaning steps were necessary to transform these six sheets of observational medical data into a clean and informative dataset. Beginning with the "Demographics" sheet, it was imperative to reduce the ninety columns of variables down to a smaller set of solely useful and understandable variables. Using the steps and criteria provided by Kuhn and Johnson (2013:27-48) as well as advice from the SME, Dr. Haac, it was a straightforward process to transform, combine, and eliminate some variables in this sheet, until the final set was as meaningful and clean as possible. Furthermore, since the specific goal of this analysis was to assess the effect of a treatment on certain outcomes, the data cleaning process became even more straightforward than initially expected as many irrelevant variables that were unrelated to answering the researchers' questions could be removed immediately.

Starting with the "Demographics" sheet, the first step suggested in Kuhn and Johnson (2013:41), eliminating variables that are over 20% empty, was conducted simply by using Excel to ascertain which columns were 'full' enough to satisfy this requirement. Using this criterion, it was possible to remove thirty-eight variables, all of which featured more than 20% missing values. These removed variables were primarily comprised of various patient medical measurements and skills data that the hospital did not record for every patient in the study. It was assumed that these few dozen variables did not provide much useful information given their lack of completeness in the dataset. Therefore, their removal should not have damaged future analysis. Next, another eighteen variables were then eliminated by following the guidance in Kuhn and Johnson (2013:44-45) regarding zero and near-zero variance predictors.

The dataset was then further cleaned by taking into account the scope of this problem. Intuitively, this meant that certain variables could be removed simply due to the fact that they offer no relevance to the questions this analysis aims to solve. Thus, variables on the “Demographics” sheet that contained only textual information regarding the category, location, code, and description of the specific patient injuries were simply deleted from the spreadsheet. This step reduced the variable count by another six variables.

The next step in cleaning this dataset involved appealing to the SME, Dr. Haac, for guidance on how best to simplify the remaining data as well as how to handle certain unfamiliar medical variables. Dr. Haac (2017) stated that the injury severity score variables, the “ISS 90” and “ISS 05” columns could be combined, with any overlap resolved by giving precedence to the “ISS 05” scores. The same was true for the similar “TRISS 90” and “TRISS 05” columns. These combinations allowed for the reduction of two more columns from the dataset. However, it was suggested a binary indicator variable be generated in order to express which injury severity score was used (05 or 90). Thus, another variable was added to the dataset to denote this.

The next step involved transforming some of the categorical, patient background data on the “Demographics” sheet into a more suitable format for analysis. Specifically, this step meant turning the categorical sex and race variables into dummy (binary) variables via the manner discussed in Kuhn and Johnson (2013:47-48). For the sex variable, the column was turned into a binary column where a 1 indicates a male patient and a 0 indicates a female patient. For the race variable, more columns of dummy variables were added to the dataset where each race was indicated with a 1 if the patient belongs to that race or 0 if he or she does not. Since the patients in this study are predominantly of European or African descent, these separate dummy variables were generated first. Patients that were listed as being of Asian descent, two or more races, “other”, or “unknown”, were combined into one “Other” column due to the low number of patients in those categories present in the data. Thus, the dataset was expanded by two columns in order to provide clear dummy variables that indicate the race of each patient.

A further search of the dataset and some verification in Excel revealed that the variables “U-Amp”, “U-Coc”, and “U-PCP” (levels of amphetamines, cocaine, and PCP, respectively, in the patients’ bloodstreams) are all identical columns, making two of these columns redundant and unnecessary for modeling purposes. Consequently, two of them were removed. Furthermore, the dataset featured both an admittance date variable as well as a length of stay variable. Therefore, discharge date column was unnecessary, as it could be inferred from the other two, and consequently was removed.

The next cleaning process involved filling in the relatively small amount of missing values still present in the remaining data. As seen in Zhang (2015:6), mean imputation was an appropriate process for filling in the few empty values in the dataset. Specifically, the mean values of the columns (disregarding the empty cells) were inserted into each missing entry until all the variables had complete columns. In total, this filled sixty missing entries across five variables’ columns. This process ensured that the dataset was fully complete, thereby guaranteeing that no columns were excluded in the later analysis due to missing data.

With the original “Demographics” spreadsheet finalized, the next step was to combine all the spreadsheets into one comprehensive dataset. After discussion with Dr. Haac (2017), the sheets entitled “Injuries AIS ’90” and “Injuries AIS ’05” were completely deleted as they were deemed to be medically irrelevant to the analysis. However, the next three sheets were all significant to the researchers’ questions and had to be handled appropriately in order to transfer meaningful data over to the “Demographics” sheet.

By relying on the SME’s guidance and the tools outlined in the previous chapter, the “OR Procedures” sheet was handled similarly to the “Demographics” sheet. In this manner, the sheet was reduced to just three informative columns: patient ID #, total time in the operating room, and time-from-arrival-to-first-operation. However, these three columns spanned 1,843 rows instead of the expected 745 rows (for 745 patients). This is because many patients in Dr. Haac’s team’s study underwent more than one operation. While this would normally heavily complicate the process of adding data to the “Demographics” sheet, Dr. Haac (2017) stated that her team is solely interested in the first operation underwent by each patient. Thus, any data on subsequent operations for each patient was removed. Some sorting and maneuvering in Excel allowed for a straightforward reduction of this sheet down to only the first-operation data for each patient. This condensed sheet seemed to confirm that each patient reflected in the “Demographics” sheet did undergo at least one surgical operation and consequently every patient in the cleaned dataset had a corresponding time-from-arrival-to-first-operation value. With this reduction complete, the “OR Procedures” sheet was merged with the “Demographics” sheet, which expanded it by three columns.

The next sheet, “Co-morbidities”, presented its own share of challenges. As previously described, this sheet featured three columns: the patient ID #, the specific co-morbidity each patient has, and the ID # that is unique to each co-morbidity on the sheet. A quick glance showed that this sheet, like the last, spans more than the 745 rows of “Demographics”. Just as many patients underwent multiple surgeries, many patients entered the hospital with multiple co-morbidities. These co-morbidity data capture the risk factors and medical issues associated with each patient. Therefore, all of the data on this sheet were relevant to PSM and the subsequent statistical modeling in this analysis. Hence, binary dummy indicator variables were created for the sixty-one co-morbidities listed on the sheet. Visual Basic for Applications (VBA) was used to perform these steps in Microsoft Excel. With this process complete, the “Demographics” sheet now featured sixty-one more (binary dummy variable) columns.

The final sheet in the dataset provided by Dr. Haac was entitled “Complications”. This sheet contained data on all the complications experienced by the 745 patients both during and after surgery. The age and admittance date columns were redundant with those in the “Demographics” sheet and were deleted. Next, using Dr. Haac’s guidance (2017), the date of complication (calendar date) as well as the day of complication (integer value for day after arrival) columns were removed. This is because Dr. Haac’s research team was only concerned with whether or not the patients experienced any complications, and were not interested in the timing of the complications. However, the rest of the data on the “Complications” sheet were deemed valuable to Dr. Haac’s team’s questions.

The remaining variables on the “Complications” sheet were the patient ID #, the complication name, the complication’s specific ID #, and the category to which the complication belongs. As requested by Dr. Haac (2017), only the cardiac events were considered in the analysis. Consequently, the rest of the non-cardiac data was removed. This step revealed that only a few dozen patients actually experienced at least one of the eight complications that fall under the cardiac category. Two more columns were then added to the clean dataset in order to capture the relevant data in the original “Complications” sheet. Specifically, a binary dummy column was created that shows a 1 for every patient that experienced at least one of the eight cardiac complications and a 0 for every patient who did not. Another column was generated that counts the number of cardiac events experienced by each patient. Once complete, these variables were merged with the rest of the data in the “Demographics” sheet.

The next step for cleaning the “Demographics” sheet was to insert a binary indicator variable to denote treatment assignment. In this case, this meant the inclusion of a column where a 0 denotes a pre-treatment patient and 1 denotes a post-treatment patient. As the treatment took effect October 1, 2014, but likely took some time to fully function, a buffer range was created. All of the observations during October 2014 were removed to create a clear divide between the pre-treatment group of observations and the post-treatment group. This step reduced the number of patients by eighteen (745 rows to 727 rows). With this step complete, it was simple in Excel to generate a treatment column where pre-October 2014 patients were assigned a 0, and the post-October 2014 patients were assigned a 1. This step, in total, added the highly necessary and important treatment variable to the dataset.

With all the data combined into a single dataset, the process of outlier investigation began. To keep all the original meaning in Dr. Haac's dataset, only values that could be assumed to be entry errors were removed. Furthermore, the focus of the outlier investigation was solely on the primary outcome of interest, time-from-arrival-to-first-operation. In this column, a purely visual inspection yielded the presence of four clear outliers. First, it was immediately apparent that, in fact, one patient in the dataset did not undergo any surgical procedures. Patient #518 was the only patient of the 727 remaining in the dataset who did require any surgery. Thus, since this patient registered a 0 for time-from-arrival-to-first-operation, her record in the dataset was a threat to later analysis and was removed. Next, it could be seen that there were three clear outliers on the high side of time-from-arrival-to-first-operation. These patients, #679, #467, and #102, all required extremely large amounts of time (18,832, 18,291, and 13,324 minutes, respectively) to get into their first surgery. These values roughly equate to ten to fourteen days required to go from arrival to surgery, which were time levels that were significantly higher than the rest of the patients. Since these values were so disparate from the remaining majority of the data, it could be assumed that these values were data entry errors, and consequently were removed. Thus, with patient #518's removal, as well as the deletion of the three huge time-from-arrival-to-first-operation patients, the outlier investigation reduced the dataset by four rows. The dataset now featured complete information for 723 patients.

The final step in cleaning the combined dataset was to compute the correlation matrix to see which variables, if any, were strongly correlated, and thus removable. Using R to generate the correlation matrix showed a handful of strongly correlated variable pairs. In order to reduce the collinearity of the dataset, a member of each of these pairs was removed arbitrarily according to a threshold of ±0.9 in the manner described in Kuhn and Johnson (2013:47). This removed four variables, which brought the dataset to a final size of eighty-eight variables.

## Preliminary Analysis

The next task involved importing the Excel data into R for further analysis. Appendix A.3 contains the R code used to upload the Excel data in its exact form into an R data frame. Storing the dataset in a singular data frame allowed for straightforward analysis and modeling.

Before diving into advanced analytic methods like PSM and regression modeling, it usually advisable to first become familiarized with the dataset at hand. Elementary methods such as data visualization and summary/descriptive statistics are a simple but effective way of capturing the essence of the data before conducting deeper analysis. This first visualization and descriptive statistics steps should involve focusing on the primary outcome in Dr. Haac’s study. As stated, this is the dependent variable, time-from-arrival-to-first-operation, which serves as a metric of capturing the efficacy of the researchers’ screening protocol. This variable is measured in minutes and records exactly how long it took each patient after arriving at the medical center to enter his or her first surgical procedure. While advanced analytic techniques were required to confidently determine a significant change in this outcome variable after the screener's implementation, it was still worth using elementary statistics and visualization to gauge a potential treatment effect. Consequently, simple descriptive statistics were applied to the data.

Due to the cleaning steps and data reduction, there were now only 723 patients' data recorded in the finalized dataset. Of these 723 patients, 338 of them were pre-treatment (control) observations, which meant 385 of them were post-treatment observations. This split was nearly even, which gave this dataset a decent balance (46.7% control vs. 53.3% treated). Next, the means of the two groups were compared to assess the effect of treatment across the two groups. Hence, the means were calculated in R (as well as the standard deviations) to provide initial insights into the treatment's effect. Table 1 gives the results of this elementary means comparison. Note that, the primary outcome, time-from-arrival-to-first-operation, is referred to as "Min_to_1st_OR", as it is titled in the main R data frame.

Table 1. Mean Time-from-Arrival-to-First-Operation

| | | n | Mean (minutes) |
|---|---|---|---|
| Treatment | No | 338 | 1727.2 |
| | Yes | 385 | 1591.6 |
| Overall | | 723 | 1655.0 |

Based on Table 1, there is an apparent difference in the two means and therefore, a potential treatment effect. While it remains to be seen if the treatment had a statistically significant effect on "Min_to_1st_OR", there is certainly a noticeable difference in the pre- and post-treatment sample means. On average, those patients who received the treatment required almost 136 minutes less time to get into surgery than their untreated counterparts. Table 2 compares the standard deviations. Table 2's results show that the outcome variable, "Min_to_1st_OR", is characterized by large variation, with standard deviations that are almost as large as the means. This implies that the time-from-arrival-to-first-operation for the patients in Dr. Haac's study significantly varied from patient to patient. Due to this large spread amongst the outcome variable, appropriate steps were taken in order to account for the large standard deviations in the two groups.

Table 2. Standard Deviation for Mean Time-from-Arrival-to-First-Operation

| | | n | Standard Deviation (minutes) |
|---|---|---|---|
| Treatment | No | 338 | 1456.8 |
| | Yes | 385 | 1251.8 |
| Overall | | 723 | 1352.2 |

To further elucidate the outcome variable, "Min_to_1st_OR", as well as the treatment effect, data visualization was performed. R, via the package ggplot2 (Wickham, 2009), provides simple but elegant avenues of displaying and assessing one's data. Figure 1 provides a snapshot of the variable "Min_to_1st_OR"'s underlying distribution. The three large outliers that were removed earlier were re-added to the dataset in order to provide a visual depiction of their separation from the mainstream outcome data. The R code used to generate the preliminary data analysis graphics can be referenced in Appendix A.5.

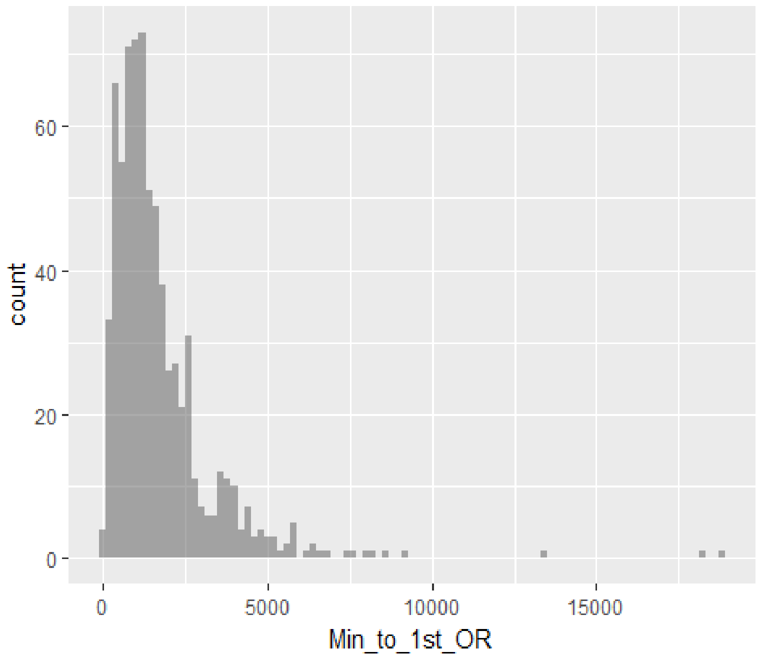

Figure 1. Histogram of Patient Counts for Time-from-Arrival-to-First-Operation

Figure 1 indicates that “Min_to_1st_OR” does not follow the Gaussian, or normal, distribution. The data do not fall under a bell curve and are clearly skewed right. Additionally, the histogram confirms that there were a few major outliers in the outcome data, specifically the three observations that fall above 12,500 minutes. Despite the heavy majority of the data falling within the 0 - 5,500-minute range, there were a few patients who required over 12,500 minutes (eight and a half days) before they were admitted into surgery. These values were extremely large comparatively, and were likely due to entry errors. Thus, their removal was warranted. Removing these three points now afforded a more in-depth look at the outcome variable’s distribution, which can be seen in Figure 2.

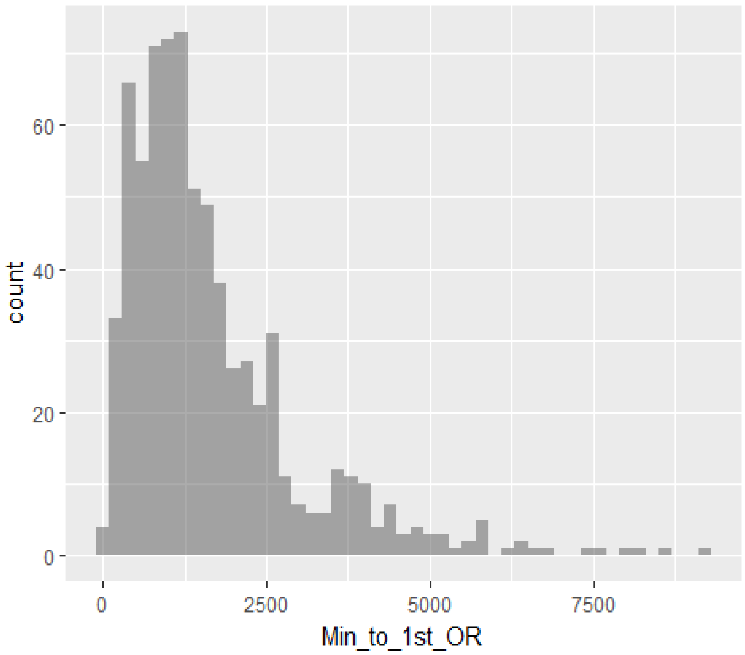

Figure 2. Histogram of Patient Counts for Time-from-Arrival-to-First-Operation (Outliers Removed)

The outlier removal reveals a clearer picture of the outcome data's distribution. Clearly, the vast majority of patients in Dr. Haac's study underwent surgery within roughly three thousand minutes (fifty hours). Because this data is so skewed, it is worth attempting a data transformation in order to make the distribution appear more normal. The most common transformation that can be applied is a logarithmic base 10 (log) transformation. Figure 3 depicts a standard log transformation of the outcome variable, "Min_to_1st_OR".

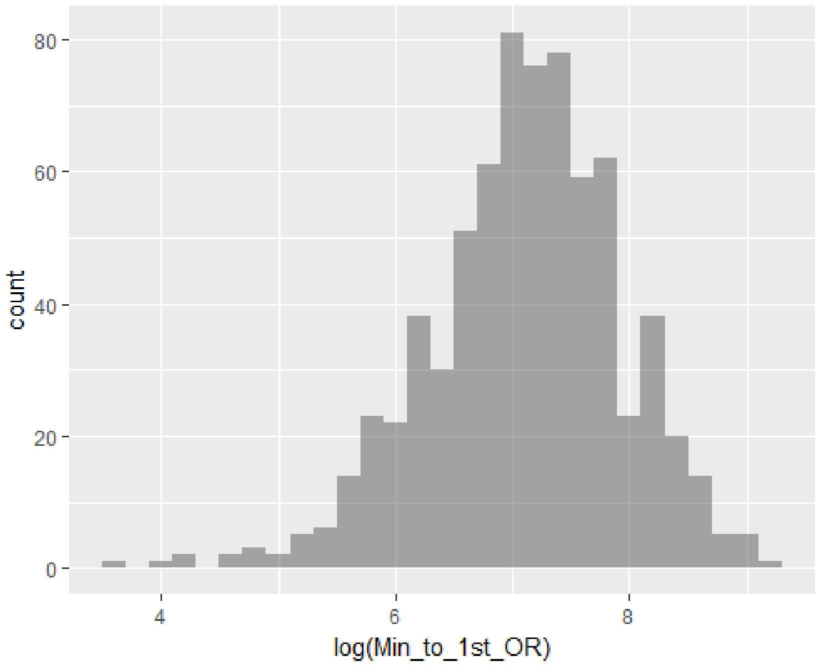

Figure 3. Log Transformation of Time-from-Arrival-to-First-Operation

While it is not perfectly normal, this log transformation was effective at removing most of the skew from the previous histograms. In later analysis, such a transformation could be considered in order to improve the effectiveness of particular models or to satisfy modeling assumptions. The next step was to visualize the differences in the outcome according to the control and treatment split. This was done simply by adding a few more parameters to the ggplot2 histogram code in R. Figure 4 is an overlaid histogram that is color-coded to show the differences between the control and treatment groups.

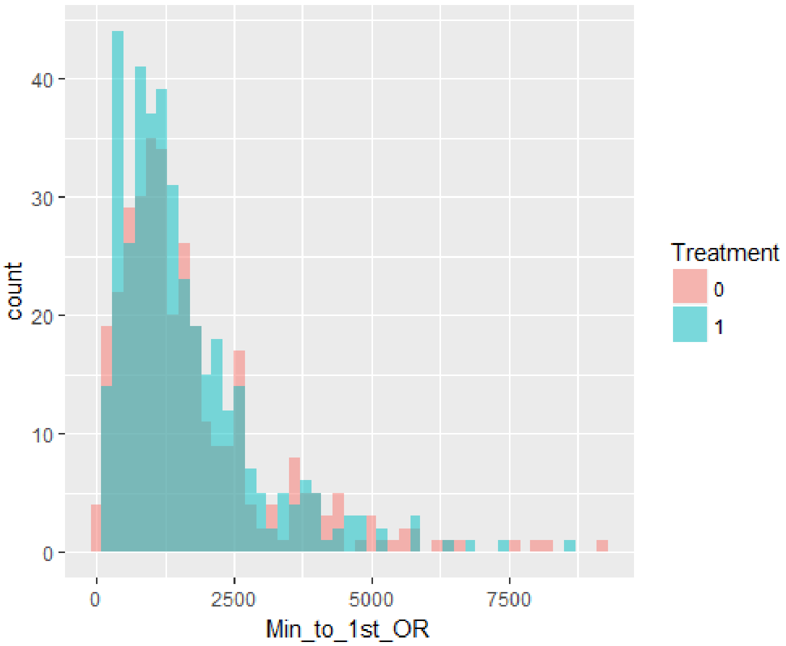

Figure 4. Histogram of Pre- & Post-Treatment Times-from-Arrival-to-First-Operation

Figure 4 provides visual insights into the treatment’s effect on the primary outcome, “Min_to_1st_OR”. This overlaid histogram shows that, despite the differences in the groups’ means calculated earlier, there is a large amount of overlap and no easily discernable difference between the control and treatment groups. That is not to say that more advanced analytic techniques would not detect a significant difference, but from this plot, the treatment effect difference is undetectable. However, different plotting techniques provide a different view of the two groups’ differences. Another visualization that compares distributions is a density plot, which is shown in Figure 5.

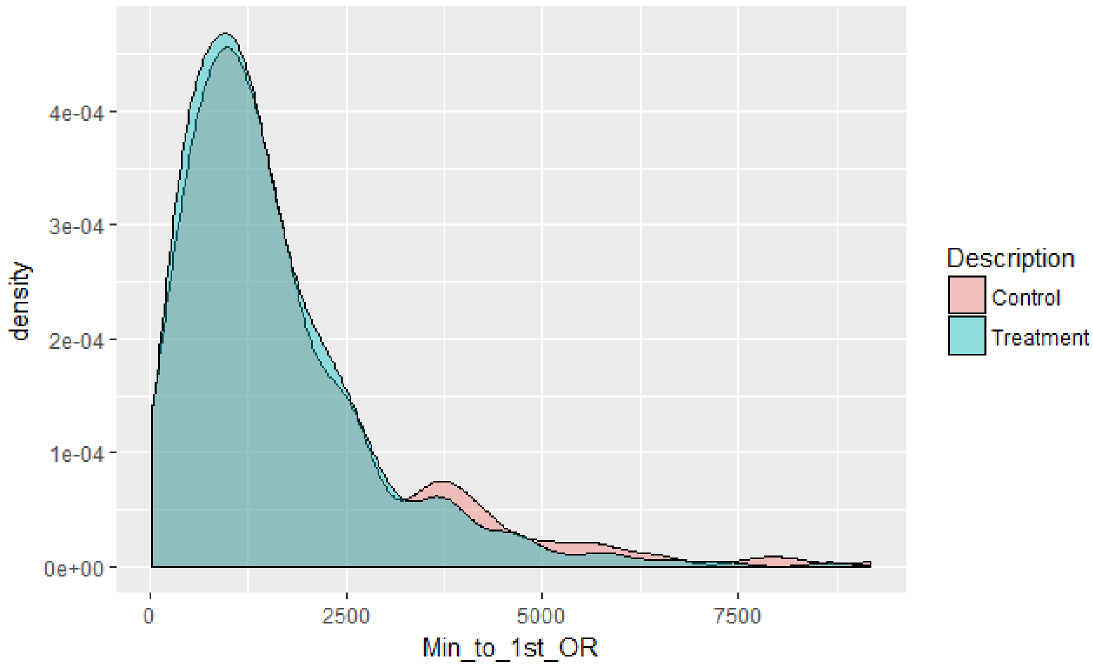

Figure 5.  Density Plot of Pre- & Post-Treatment Times-from-Arrival-to-First-Operation

Figure 5 shows more distinguishable differences in the two groups' data than the histogram in Figure 4 did. From this plot, at around the 3,000-minute mark, noticeably fewer people from the treatment group start requiring large amounts of time to get into surgery. Patients who received the protocol treatment, on average, were admitted into surgery prior to three thousand minutes (fifty hours) 88.8% of the time, whereas the control patients went into surgery in under fifty hours only 85.5% of the time. In other words, this plot suggests that the researchers' treatment may have had an influence on reducing the number of patients who required longer than fifty hours, or around two days, to receive their first operation. This could potentially mean that the treatment was in part effective at lowering the "Min_to_1st_OR" outcome and producing a positive result in the study.

Another way to visualize the main outcome variable, “Min_to_1st_OR”, is to graph it over the entirety of the study. As the study lasted from October 1, 2012 to October 31, 2016 (with the buffer month of October 2014 removed), there were exactly four years of outcome data. Plotting the outcome over time can show any positive or negative trend in the results. Adding a best fit (or trend) line to the plot depicts if the outcome variable decreased over the span of the study. Figure 6 is a scatterplot (with the three large outliers re-included to show their effect) that depicts the outcome over time.

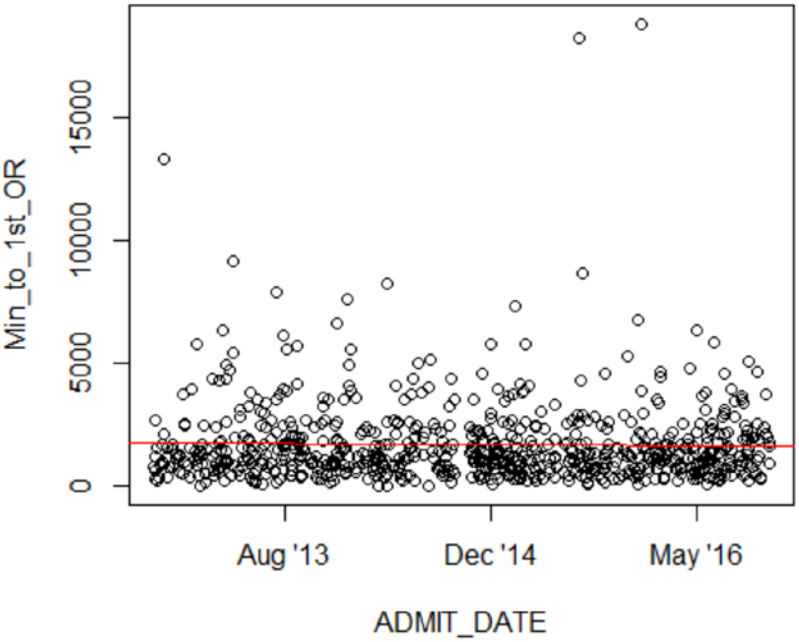

Figure 6. Scatterplot of Date vs. Time-from-Arrival-to-First-Operation

From Figure 6, it is apparent that there is no easily detectable change in “Min_to_1st_OR”. Upon very close examination, though, there does appear to be a very slight decrease in the best fit line over time (slope = -0.066). However, the high p-value of 0.647 means that “ADMIT_DATE” is not a significant predictor of “Min_to_1st_OR”. Figure 7 shows the results of the scatterplot sans the three outliers included in Figure 6.

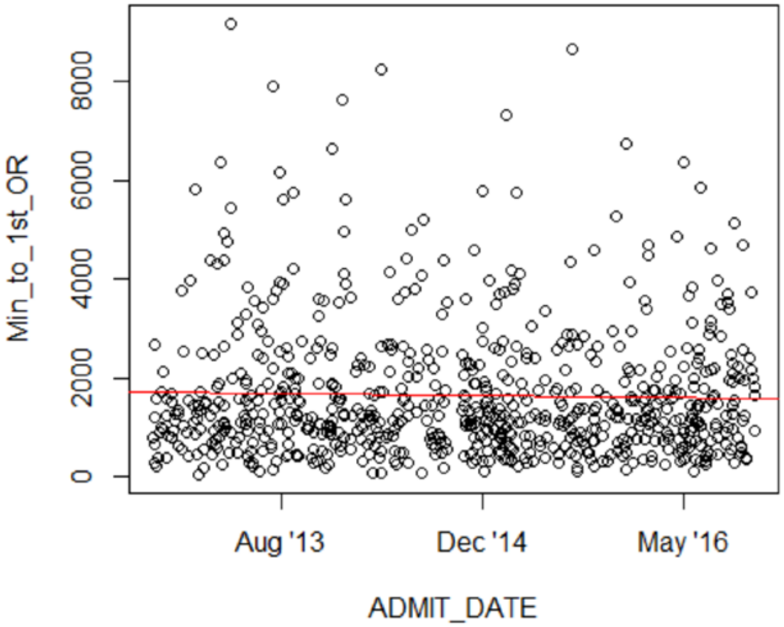

Figure 7. Scatterplot of Date vs. Time-from-Arrival-to-First-Operation (Outliers Removed)

Figure 7 seems to indicate that the outlier removal did indeed result in a more noticeable negative slope. The influence of “ADMIT_DATE” is still insignificant (p-value = 0.458), but the slope is slightly more negative (slope = -0.086). This slope can be interpreted as an indication that for every day that passes over the four-year study, the patients required five seconds fewer on average to go from arrival to their first surgery. This is a minute change, and close to zero. Therefore, more advanced methods will be required to statistically prove whether or not the current dataset supports the hypothesis that the screening protocol treatment had any effect on the primary outcome, time-from-arrival-to-first-operation.

## Propensity Score Matching

PSM is the next step. All the available matching methods were applied to the dataset in order to search for the optimal parameters and method that best balanced the data to remove any inherent treatment selection biases. The simplest form of PSM is nearest neighbor, 1-to-1 matching. Thus, it was the first PSM model run and served as the benchmark upon which to judge other PSM methods. All PSM models were generated using the MatchIt package (Ho *et al*, 2017,1-8) in R.

Before running the initial PSM model, it was necessary to decide which variables should be included. Despite inclusion into the final dataset that was imported into R, there were a few variables that could be ignored for PSM purposes. These variables were patient ID, admittance date, length of stay, total time in surgery, days spent in the trauma unit (TRU), and days spent in the intensive care unit (ICU). As the first set of PSM models were designed with the intent of isolating the treatment effect on “Min_to_1st_OR”, it was unnecessary and inappropriate to include other response variables that were irrelevant with regards to time-from-arrival-to-first-operation or that occurred after the outcome was recorded by the researchers. For example, patient ID is not needed for PSM as it is just a unique identifier, and neither is total time in surgery, as that can be considered a dependent variable that was measured after time-from-arrival-to-first-operation. Therefore, these two (and other similar variables) were not considered predictors that needed to be balanced in a model that is focused on “Min_to_1st_OR”.

Once these variables were removed, the initial model was run. For an initial, simplistic look at the nearest neighbor method, all the co-morbidity binary variables were ignored for the time being. Tables 3, 4, 5, and 6 provide a glimpse (via the first six variables in the model) of the results provided by the initial nearest neighbor method with 1-to-1 matching. Appendix A.8 contains all the code used to generate the PSM models.

Table 3. Nearest Neighbor (1-to-1) Matching (Partial Model) Sample Sizes

| | Control | Treated |
|---|---|---|
| All | 338 | 385 |
| Matched | 338 | 338 |
| Unmatched | 0 | 47 |
| Discarded | 0 | 0 |

Table 4. Nearest Neighbor (1-to-1) Matching (Partial Model) Balance for All Data

| | Mean Treated | Mean Control | SD Control | Mean Difference |
|---|---|---|---|---|
| Distance | 0.772 | 0.260 | 0.347 | 0.512 |
| Sex | 0.351 | 0.296 | 0.457 | 0.055 |
| White | 0.816 | 0.879 | 0.327 | -0.063 |
| Black | 0.130 | 0.053 | 0.225 | 0.077 |
| Other | 0.055 | 0.068 | 0.252 | -0.014 |
| Age | 77.875 | 77.299 | 8.889 | 0.577 |
| ISS_90.05 | 7.797 | 8.518 | 2.760 | -0.720 |

Table 5. Nearest Neighbor (1-to-1) Matching (Partial Model) Balance for Matched Data

| | Mean Treated | Mean Control | SD Control | Mean Difference |
|---|---|---|---|---|
| Distance | 0.799 | 0.260 | 0.347 | 0.540 |
| Sex | 0.385 | 0.296 | 0.457 | 0.089 |
| White | 0.802 | 0.879 | 0.327 | -0.077 |
| Black | 0.148 | 0.053 | 0.225 | 0.095 |
| Other | 0.050 | 0.068 | 0.252 | -0.018 |
| Age | 78.420 | 77.299 | 8.889 | 1.121 |
| ISS_90.05 | 7.698 | 8.518 | 2.760 | -0.820 |

Table 6. Nearest Neighbor (1-to-1) Matching (Partial Model) Balance Improvement

| | Mean Difference Improvement (%) |
|---|---|
| Distance | -5.383 |
| Sex | -61.992 |
| White | -21.880 |
| Black | -23.571 |
| Other | -31.474 |
| Age | -94.499 |
| ISS_90.05 | -13.768 |

Table 3 confirms that 1-to-1 matching occurred as the newly matched dataset featured 338 treated and 338 control observations, which means forty-seven treated patients were discarded. Table 4 tabulates the means of the variables for the entire dataset separated into the treated and control groups prior to matching. Table 5 shows these same variables' means and differences, but for the post-matching dataset. If the matching was successful, the means would be much closer, leading to reduced mean differences. Table 6 sums up this comparison by providing the percentage change in the mean differences between the groups for all the variables included in the model. It can clearly be seen that this initial PSM model is ineffective at improving the balance between the pre- and post-treatment groups in the dataset. For all the variables shown, including the propensity scores (listed as "Distance"), this PSM model actually created worse balance. It appears that the standard PSM method, nearest neighbor with 1-to-1 matching, is not suitable for this data. To confirm this result, all the co-morbidities were then added to see if any positive change occurred.

The next iteration of PSM operated the same as the previous model, just with the addition of the co-morbidity binary variables. Twelve co-morbidities were completely imbalanced. These twelve co-morbidities did not appear in either the pre-treatment or post-treatment sets. This meant matching would be impossible across these co-morbidity covariates. PSM will not operate effectively if it is forced to match only 1's with only 0's. Thus, these twelve co-morbidities were excluded from subsequent PSM models, which meant forty-nine co-morbidities were added to the next rendition of PSM. Tables 7, 8, and 9 below provide a look at the results of adding the forty-nine co-morbidity variables to the original propensity score model.

Table 7. Nearest Neighbor (1-to-1) Matching (Full Model) Balance for All Data

| | Mean Treated | Mean Control | SD Control | Mean Difference |
|---|---|---|---|---|
| Distance | 0.803 | 0.225 | 0.321 | 0.578 |
| Sex | 0.351 | 0.296 | 0.457 | 0.055 |
| White | 0.816 | 0.879 | 0.327 | -0.063 |
| Black | 0.130 | 0.053 | 0.225 | 0.077 |
| Other | 0.055 | 0.068 | 0.252 | -0.014 |
| Age | 77.875 | 77.299 | 8.889 | 0.577 |
| ISS_90.05 | 7.797 | 8.518 | 2.760 | -0.720 |

Table 8. Nearest Neighbor (1-to-1) Matching (Full Model) Balance for Matched Data

| | Mean Treated | Mean Control | SD Control | Mean Difference |
|---|---|---|---|---|
| Distance | 0.843 | 0.225 | 0.321 | 0.619 |
| Sex | 0.373 | 0.296 | 0.457 | 0.077 |
| White | 0.799 | 0.879 | 0.327 | -0.080 |
| Black | 0.148 | 0.053 | 0.225 | 0.095 |
| Other | 0.053 | 0.068 | 0.252 | -0.015 |
| Age | 78.080 | 77.299 | 8.889 | 0.781 |
| ISS_90.05 | 7.663 | 8.518 | 2.760 | -0.855 |

Table 9. Nearest Neighbor (1-to-1) Matching (Full Model) Balance Improvement

| | Mean Difference Improvement (%) |
|---|---|
| Distance | -7.020 |
| Sex | -40.393 |
| White | -26.568 |
| Black | -23.571 |
| Other | -9.562 |
| Age | -35.482 |
| ISS_90.05 | -18.697 |

Tables 7-9 show that this model did not perform well either after the addition of all the co-morbidity variables. Table 9 shows that the balance was degraded across most of the variables, including “Distance”, after the matching occurred. Together, the results of the partial and full nearest neighbor models, with 1-to-1 matching, show that this method is not effective on this dataset.

At this point, the only parameter adjustment that could potentially yield positive results was to turn off 1-to-1 matching. Increasing the matching ratio to 2-to-1 failed to improve the balance and was just as ineffective as the 1-to-1 models. It is clear that the approach of nearest neighbor matching was unsuccessful on this dataset. Thus, other methods were attempted to try to improve the balance between the pre- and post-treatment groups.

The other methods available in MatchIt (Ho *et al*, 2017:1-8) include exact matching, subclassification, full matching, optimal matching, genetic matching, and coarsened exact matching. The exact, subclassification, full, optimal, and coarsened exact matching models all failed to run successfully or produce any informative matching results. As expected, exact matching was impossible for a dataset of this magnitude since there were no perfectly exact records between the two groups. The subclassification and optimal matching models failed to run as they likely could not resolve the large number of binary covariates in the dataset. However, full matching did execute, and was able to match the 338 control observations to the 385 treated observations, but the balance across the covariates was not improved over that of the nearest neighbor method. Lastly, coarsened exact matching ran properly, but was only able to match one control observation to one treated observation. Thus, it was also unsuccessful at producing informative matching results.

As stated, the fact that a few of the methods were ineffective is not an alarming issue. The dataset features many binary covariates, which are likely to impede the ability of some of the matching methods. Genetic matching did execute successfully and appears to be capable of matching this medical data at least somewhat effectively. The first run of genetic matching was conducted without the co-morbidity variables to test its efficacy on just the background variables. Tables 10, 11, 12, and 13 capture the results of the partial model. It is key to note that genetic matching also requires the R package Matching to run (Sekhon, 2011)

Table 10. Genetic Matching (Partial Model) Sample Sizes

| | Control | Treated |
|---|---|---|
| All | 338 | 385 |
| Matched | 100 | 385 |
| Unmatched | 238 | 0 |
| Discarded | 0 | 0 |

Table 11. Genetic Matching (Partial Model) Balance for All Data

| | Mean Treated | Mean Control | SD Control | Mean Difference |
|---|---|---|---|---|
| Distance | 0.772 | 0.260 | 0.347 | 0.512 |
| Sex | 0.351 | 0.278 | 0.450 | 0.073 |
| White | 0.816 | 0.816 | 0.390 | 0.000 |
| Black | 0.130 | 0.130 | 0.338 | 0.000 |
| Other | 0.055 | 0.055 | 0.228 | 0.000 |
| Age | 77.875 | 77.790 | 8.601 | 0.086 |
| ISS_90.05 | 7.797 | 7.751 | 2.736 | 0.047 |

Table 12. Genetic Matching (Partial Model) Balance for Matched Data

| | Mean Treated | Mean Control | SD Control | Mean Difference |
|---|---|---|---|---|
| Distance | 0.772 | 0.759 | 0.115 | 0.013 |
| Sex | 0.351 | 0.278 | 0.450 | 0.073 |
| White | 0.816 | 0.816 | 0.390 | 0.000 |
| Black | 0.130 | 0.130 | 0.338 | 0.000 |
| Other | 0.055 | 0.055 | 0.228 | 0.000 |
| Age | 77.875 | 77.790 | 8.601 | 0.086 |
| ISS_90.05 | 7.797 | 7.751 | 2.736 | 0.047 |

Table 13. Genetic Matching (Partial Model) Balance Improvement

| | Mean Difference Improvement (%) |
|---|---|
| Distance | 97.416 |
| Sex | -32.735 |
| White | 100.000 |
| Black | 100.000 |
| Other | 100.000 |
| Age | 85.132 |
| ISS_90.05 | 93.510 |

Tables 10-13 above show that while the genetic matching algorithm discarded 238 control units, the balance improved between the groups. Across the majority of the covariates, including “Distance”, the mean differences are much smaller, which demonstrates that the genetic matching method was effective on this dataset. Some of the variables, including “White”, “Black”, and “Other”, were even matched 100% effectively by the genetic model as seen by the equivalent means in Table 10.  This means that the genetic matching process was able to create perfect balance between the groups across these select variables.

Tables 14, 15, 16, and 17 below capture the results of the full genetic PSM model with the forty-nine co-morbidity variables included.

Table 14.  Genetic Matching (Full Model) Sample Sizes

| | Control | Treated |
|---|---|---|
| All | 338 | 385 |
| Matched | 87 | 385 |
| Unmatched | 251 | 0 |
| Discarded | 0 | 0 |

Table 15.  Genetic Matching (Full Model) Balance for All Data

| | Mean Treated | Mean Control | SD Control | Mean Difference |
|---|---|---|---|---|
| Distance | 0.803 | 0.225 | 0.321 | 0.578 |
| Sex | 0.351 | 0.296 | 0.457 | 0.055 |
| White | 0.816 | 0.879 | 0.327 | -0.063 |
| Black | 0.130 | 0.053 | 0.225 | 0.077 |
| Other | 0.055 | 0.068 | 0.252 | -0.014 |
| Age | 77.875 | 77.299 | 8.889 | 0.577 |
| ISS_90.05 | 7.797 | 8.518 | 2.760 | -0.720 |

Table 16. Genetic Matching (Full Model) Balance for Matched Data

| | Mean Treated | Mean Control | SD Control | Mean Difference |
|---|---|---|---|---|
| Distance | 0.803 | 0.765 | 0.169 | 0.038 |
| Sex | 0.351 | 0.281 | 0.452 | 0.070 |
| White | 0.816 | 0.839 | 0.370 | -0.023 |
| Black | 0.130 | 0.127 | 0.335 | 0.003 |
| Other | 0.055 | 0.034 | 0.182 | 0.021 |
| Age | 77.875 | 77.148 | 9.225 | 0.727 |
| ISS_90.05 | 7.797 | 8.307 | 2.568 | -0.509 |

Table 17. Genetic Matching (Full Model) Balance Improvement

| | Mean Difference Improvement (%) |
|---|---|
| Distance | 93.489 |
| Sex | -27.994 |
| White | 62.961 |
| Black | 96.610 |
| Other | -53.899 |
| Age | -26.151 |
| ISS_90.05 | 29.327 |

Based on the Tables 14-17, the full genetic model discarded even more control units than before (251) to best match up all the covariates. Despite the addition of all the covariates, this model had moderate success reducing the mean differences between the variables in the pre- and post-treatment groups. The mean differences are noticeably smaller across the majority of the covariates included in the model, although some did worsen. Most importantly, the balance between the "Distance" means was significantly improved, which demonstrates that the model ran as intended. In whole, across the entire dataset, the genetic matching outperformed the other methods. This indicates that the genetic matching algorithm was appropriate for this dataset as it was capable of handling the multiple types of variables in the data while producing positive results.

Although the results of the genetic matching model are visible in the tables, it is often best to assess the effectiveness of PSM through visualization. The propensity score jitter plot in Figure 8 provides a visual depiction of the matching performed by the genetic model.

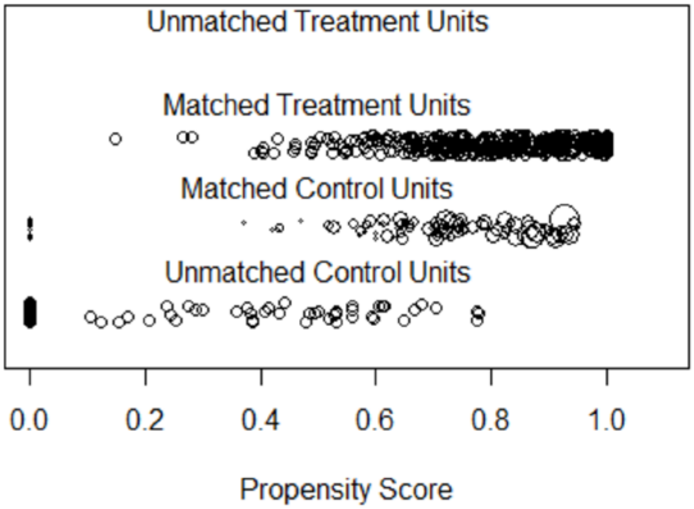

Figure 8.  Distribution of Genetic Matching (Full Model) Propensity Scores

The jitter plot in Figure 8 gives a visual representation of the matching that occurred in the genetic PSM model. This plot confirms the moderate success detected in the tables above as the matched treatment units are mostly similar to the matched control units, according to their propensity scores. This plot also confirms that many control observations were discarded by the genetic algorithm. However, the plot shows that the genetic matching was effective at matching the available units. In other words, the genetic model worked well since it discarded and did not match on the large cluster of control units that are centered on the propensity score = 0 mark. As there are no treatment units with a near-zero propensity score, this decision by the genetic model to discard these control units ensured meaningful results. This depiction of the cluster of control units towards the zero propensity score mark might possibly explain the lack of success in the other matching methods like 1-to-1 nearest neighbor matching. Due to the requirement of having to match up equivalent amounts of units, these extremely low propensity scores in the control group had to be matched to dissimilar treatment units, which likely prevented positive results.

In addition to the jitter plot, there are other forms of visualization that can display the results of PSM. The set of histograms of the propensity scores in Figure 9 further depict the genetic matching results.

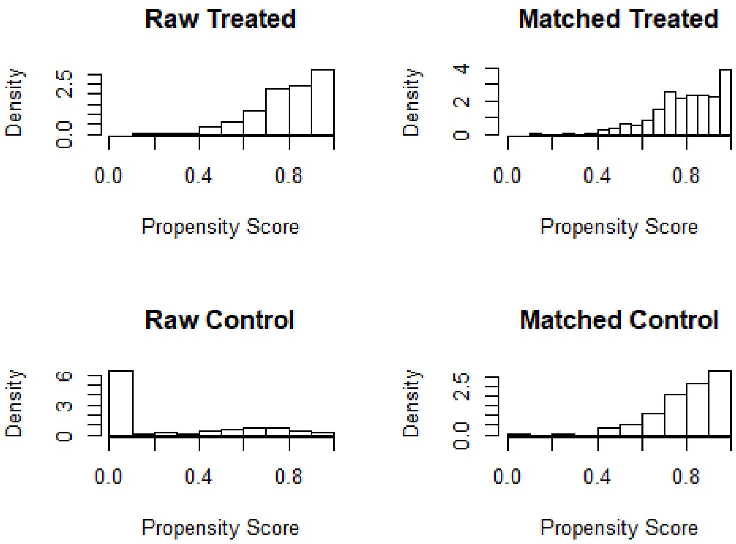

Figure 9. Genetic Matching (Full Model) Propensity Score Histograms

These histograms in Figure 9 portray the changes in balance between the treated and control groups that occurred due to matching. From the raw histograms, it is apparent that two groups were originally noticeably imbalanced due to their dissimilar distributions.

However, post-matching, the balance appears to have significantly improved due to the similar shapes depicted in the matched histograms. These plots confirm the results seen in the genetic matching output tables. While the results are not perfect and the balance did not improve across all the covariates, the genetic matching method certainly outperformed the other matching techniques. Consequently, the genetic model's matching results were used for subsequent analysis, including treatment assessment.

## Significance Testing

The first step used in analyzing the isolated treatment effect was examining the Student's. This process involved taking the genetic-matched data and selecting the outcomes of interest for significance testing. The first and primary outcome analyzed was the time-from-arrival-to-first-operation for the geriatric orthopedic patients who entered the hospital. Due to the success of the genetic matching and the elimination of the treatment selection bias, this outcome was isolated and analyzed by comparing the means of the pre- and post-treatment groups. Note these two means were reported in Table 1, but due to the genetic matching discarding a group of control units, the control mean has changed. Table 18 provides the post-matched summary of means.

Table 18. Mean Time-from-Arrival-to-First-Operation (Post-Matching)

| | | n | Mean (minutes) |
|---|---|---|---|
| Treatment | No | 87 | 1907.3 |
| | Yes | 385 | 1591.6 |
| Overall | | 472 | 1649.8 |

Table 18 shows that the mean difference in the two groups widened noticeably due to the genetic matching. A *t*-test confirms whether or not this difference is significantly large enough to show if the treatment was effective at reducing time-from-arrival-to-first-operation. The results of the significance are shown below in Table 19.

Table 19. Post-Matching Summary of *t*-test

| Alternative hypothesis: true difference in means is not equal to 0 | | |
|---|---|---|
| t = 1.733 | df = 111.24 | p-value = 0.086 |
| 95% confidence interval: (-45.232, 676.710) | | |
| Sample estimates: | mean of control: 1907.332 | mean of treated: 1591.593 |

Table 19 provides all the necessary information to form an initial analysis on the treatment's effect on the primary outcome, "Min_to_1st_OR". According to the standard level of significance (α = 0.05), this mean difference, though large, is not below the alpha threshold. It is close, however, as 0.086 falls nearly within the range of significance. The important takeaway from this *t*-test is not that the treatment was insignificant, but rather that more methodology was needed to fully make a conclusion about the treatment's efficacy on "Min_to_1st_OR". With a mean difference this large, it was very possible that regression techniques could provide convincing evidence that the treatment was in fact significant in explaining the decrease in time-from-arrival-to-first-operation over the course of Dr. Haac's team's study.

The next task involved analyzing the secondary outcome of interest in the medical researchers' study. Dr. Haac and her team are also interested in their screening protocol's effect on the number of patients who experienced cardiac events intra- and post-operation. Table 20 provides a look at the cardiac event-specific data after the genetic matching was completed.

Table 20. Pre- and Post-Matching Summary of Cardiac Events

| | Pre-matching | Post-matching |
|---|---|---|
| Control | 15 (4.4%) | 1 (1.1%) |
| Treated | 10 (2.6%) | 10 (2.6%) |
| Total | 25 | 11 |

Table 20 shows that unfortunately, the genetic matching discarded fourteen of the fifteen control patients who experienced a cardiac event. Therefore, not much can be gleaned from the post-matching results. It can be seen that pre-matching, there was a higher percentage of patients who experienced cardiac events (4.4% > 2.6%), but the matched results have failed to show much more than that. Consequently, no serious statistical analyses can be conducted on this outcome due to a lack of data. Table 20 provides simple inferences about the decrease in patients experiencing cardiac events, but that is the limit of insights possible. The next analytic steps focused solely on the primary outcome, time-from-arrival-to-first-operation.

**Feature Selection**

Feature selection was performed next. Using all the variables in the dataset and the process outlined in Zhang (2016:2-5), the most basic method, stepwise regression, was applied first, using a blend of forward and backward stepwise regression. Table 21 shows the list of variables that were selected as the most important in explaining “Min_to_1st_OR” as well as their corresponding significance levels.

Table 21. Initial Stepwise Regression Results

| Variable | Estimate | Std. Error | t | p-value |
|---|---|---|---|---|
| (Intercept) | 5793.501 | 2380.878 | 2.433 | 0.015 |
| Age | 15.642 | 6.823 | 2.292 | 0.022 |
| I.T_05 | -1095.774 | 457.816 | -2.393 | 0.017 |
| Treatment | -233.705 | 158.476 | -1.475 | 0.141 |
| Adm_Sys_BP | -2.917 | 1.986 | -1.469 | 0.143 |
| Adm_SaO2 | -38.907 | 22.985 | -1.696 | 0.091 |
| Lower_Ext_Sev | 107.580 | 57.134 | 1.883 | 0.060 |
| CM_None | -687.212 | 348.973 | -1.969 | 0.050 |
| CM_6 | 1098.072 | 639.661 | 1.717 | 0.087 |
| CM_16 | -423.982 | 186.280 | -2.276 | 0.023 |
| CM_18 | 1144.577 | 651.795 | 1.756 | 0.080 |
| CM_22 | -207.104 | 136.197 | -1.521 | 0.129 |
| CM_28 | 294.140 | 207.960 | 1.414 | 0.158 |
| CM_69 | 318.099 | 161.558 | 1.969 | 0.050 |

Table 21 shows that stepwise regression selected thirteen out of the seventy-nine possible variables. According to the criteria of stepwise feature selection, these thirteen variables are deemed the most valuable in predicting the primary response. Note that the treatment variable was chosen as one of these top thirteen variables. This is by no means a conclusive evaluation of the treatment's effect, but it does suggest that the treatment had at least a small influence on the "Min_to_1st_OR" outcome variable. However, of the thirteen variables selected, only five (as well as the intercept) had p-values below the standard $\alpha = 0.05$ cut-off. This means that only five variables, "Age", "I.T_05", "CM_None", "CM_16", and "CM_69", were technically significant predictors according to the methodology of stepwise regression. Specifically, these variables correspond to patient age, the binary variable indicating which injury score scale (05 or 90) the patients received, the absence of any co-morbidities, tobacco use, and coronary artery disease, respectively. The treatment assignment variable, with a p-value of 0.141, hence, was not a significant predictor. However, it was important to recognize that the most significant variable, “I.T_05” (p-value = 0.017), was strongly correlated with “Treatment”. Earlier correlation analysis calculated a value of 0.691 between the two variables. While this value was not high enough to warrant removal of “I.T_05”, it was likely affecting the significance of “Treatment” reported by the stepwise regression. Given that there is a large amount of equivalent overlap between the two binary variables due to “I.T_05” taking effect at the start of 2014, it was deemed vital to remove “I.T_05” in order to more thoroughly isolate and assess the effect of the treatment variable.

This second iteration of stepwise regression was run with the “I.T_05” indicator variable removed. The results from this updated model are captured in Table 22.

Table 22. Stepwise Regression Results (“I.T_05” Removed)

| Variable | Estimate | Std. Error | t | p-value |
|---|---|---|---|---|
| (Intercept) | 4534.475 | 2362.523 | 1.919 | 0.056 |
| Sex | 195.139 | 129.010 | 1.513 | 0.131 |
| Age | 17.518 | 6.989 | 2.506 | 0.013 |
| Treatment | -352.139 | 152.393 | -2.311 | 0.021 |
| Adm_Sys_BP | -3.083 | 1.991 | -1.548 | 0.122 |
| Adm_SaO2 | -37.855 | 23.073 | -1.641 | 0.102 |
| Lower_Ext_Sev | 110.73 | 57.438 | 1.928 | 0.055 |
| CM_None | -637.135 | 349.790 | -1.821 | 0.069 |
| CM_6 | 1194.642 | 644.439 | 1.854 | 0.064 |
| CM_16 | -475.070 | 188.960 | -2.514 | 0.012 |
| CM_18 | 1420.390 | 645.018 | 2.202 | 0.028 |
| CM_22 | -230.265 | 136.649 | -1.685 | 0.093 |
| CM_28 | 367.918 | 206.715 | 1.780 | 0.076 |
| CM_69 | 284.682 | 162.188 | 1.755 | 0.080 |

First, Table 22 shows that again, thirteen variables were selected as the most important in explaining “Min_to_1st_OR”. In addition, this table shows that “Sex” replaced “I.T_05” in the list of thirteen most valuable variables according to the stepwise regression model. Next, Table 22 shows that the removal of “I.T_05” did in fact make “Treatment” significant (p-value = 0.021). This intuitively makes sense as “I.T_05” was acting as a redundant predictor and therefore was masking the predictive power of “Treatment”. Further, this updated stepwise model only shows four variables, “Age”, “Treatment”, “CM_16”, and “CM_18” as statistically significant, according to α = 0.05 (where “CM_18” corresponds to Alzheimer’s Disease/Dementia). To verify whether stepwise regression selected the best variables for subsequent modeling and analysis, best subsets was then used to compare selection results.

In the first best subsets regression model generated, the “I.T_05” variable was re-included in order to provide a thorough comparison between the two feature selection methods. The results are provided in Table 23.

Table 23. Best Subsets Results

| 1 var. | 2 var. | 3 var. | 4 var. | 5 var. | 6 var. | 7 var. | 8 var. | 9 var. |
|---|---|---|---|---|---|---|---|---|
| I.T_05 | I.T_05 | I.T_05 | I.T_05 | I.T_05 | I.T_05 | I.T_05 | I.T_05 | I.T_05 |
| | Age | Age | Age | Age | Age | Age | Age | Age |
| | | Low_Ext | Low_Ext | Low_Ext | Low_Ext | Low_Ext | Low_Ext | Low_Ext |
| | | | CM_69 | CM_69 | CM_69 | CM_69 | CM_69 | CM_69 |
| | | | | Sys_BP | Sys_BP | Sys_BP | Sys_BP | Sys_BP |
| | | | | | CM_16 | CM_16 | CM_16 | CM_16 |
| | | | | | | CM_None | CM_None | CM_None |
| | | | | | | | CM_18 | CM_18 |
| | | | | | | | | SaO2 |

From the feature selection output in Table 23, the best subsets regression determined the nine most important variables in the model. The first column reports that the most important variable in predicting “Min_to_1st_OR” is the “I.T_05” binary indication variable, which concurs with the results of the first stepwise model generated. If only one predictor could be used in a model explaining “Min_to_1st_OR”, best subsets reports that “I.T_05” is the best one to use. Following along the table, best subsets shows in order of decreasing value the next eight most important variables. By examining the final column, best subsets reports nine of the thirteen variables that were selected by the first stepwise regression model. This means that both techniques overlap and do agree on which variables warrant consideration in future modeling.

Note the set of nine best variables in Table 23 does not include “Treatment”. To confirm the results of the second stepwise model regarding the statistical significance of the treatment variable, “I.T_05” was removed again and an updated model was generated. Table 24 lists the results of the next best subsets regression model.

Table 24. Best Subsets Results (“I.T_05” Removed)

| 1 var. | 2 var. | 3 var. | 4 var. | 5 var. | 6 var. | 7 var. | 8 var. | 9 var. |
|---|---|---|---|---|---|---|---|---|
| Age | Age | Age | Age | Age | Age | Age | Age | Age |
|  | Low_Ext | CM_18 | Treatment | Treatment | Treatment | Treatment | Treatment | Sex |
|  |  | CM_28 | CM_18 | AdmDBP | Low_Ext | AdmDBP | AdmDBP | Treatment |
|  |  |  | CM_28 | CM_18 | CM_16 | Low_Ext | Low_Ext | AdmDBP |
|  |  |  |  | CM_28 | CM_18 | CM_16 | CM_None | Low_Ext |
|  |  |  |  |  | CM_28 | CM_18 | CM_16 | CM_6 |
|  |  |  |  |  |  | CM_28 | CM_18 | CM_16 |
|  |  |  |  |  |  |  | CM_28 | CM_18 |
|  |  |  |  |  |  |  |  | CM_28 |

Table 24 shows that this updated best subsets model ran quite differently than the first model. Here, it can be seen that depending on the model size, some variables were selected, excluded, and then re-selected across the nine different model sizes. Overall, ten different variables were chosen at least once across the nine models, with “Age” being the clearly most significant variable according to this best subsets methodology. Next, as expected, the removal of “I.T_05” enabled the inclusion of “Treatment” into the best subsets models of four to nine variables. Most importantly, Table 24 shows that this best subsets model agrees mostly with the second stepwise model with regards to the variables selected across the varying model sizes. The second stepwise model reported that only four variables, “Age”, “Treatment”, “CM_16”, and “CM_18”, were statistically significant at the $\alpha = 0.05$ level. This result is supported by the best subsets output above, where the six-variable model includes these four variables. In order to merge the results of these two methods, it made sense then to select the six variables mentioned that contain the four significant ones from the stepwise model. Thus, the combined results of the two feature selection methods showed that, given “I.T_05”’s exclusion, there were six top variables, including the treatment assignment, that best explain the main outcome, “Min_to_1st_OR”. While this number of variables was small compared to the original dataset size, it did allow for a parsimonious model that was easily interpretable.

Overall, the feature selection process was successful and the next steps of analysis can reliably use these best selected predictors. Thus, for the rest of the modeling steps, the predictors in the matched dataset were narrowed down to the six significant ones (“Age”, “Treatment”, “Lower_Ext_Sev” (severity of injuries to patient’s lower half), “CM_16”, “CM_18”, and “CM_28” (congestive heart failure)) from feature selection.

**Regression**

The nature of this data leads one to believe that the negative binomial or the Poisson regression models were the most appropriate. While the negative binomial can operate on a wider range of data, the Poisson is built on the critical assumption that the mean and the variance of the data are equivalent. Table 25 below captures the results of the primary outcome variable's post-matching mean and variance comparison.

Table 25. Post-Matching Summary of Time-from-Arrival-to-First-Operation

| | Mean | Variance |
|---|---|---|
| Control | 1907.3 | 2,332,593 |
| Treated | 1591.6 | 1,566,881 |
| Total | 1649.8 | 1,754,988 |

Table 25 shows that the mean and variance of "Min_to_1st_OR" are markedly different and not even close to being equivalent. Thus, the Poisson model's primary assumption is violated. This means that, although the Poisson model is preferred, it could not work effectively on these data; the negative binomial regression was utilized instead. By taking the four best variables reported by the two feature selection methods, the negative binomial efficiently provides a reliable analysis on the treatment and other predictors. The results of the negative binomial model on the variables chosen are shown in Table 26 below.

Table 26.  Negative Binomial Model Summary (6 Variables)

| Variable | Estimate | Std. Error | Z | p-value |
|---|---|---|---|---|
| (Intercept) | 6.648 | 0.303 | 21.934 | 0.000 |
| Age | 0.009 | 0.004 | 2.465 | 0.014 |
| Treatment | -0.174 | 0.085 | -2.048 | 0.041 |
| Lower_Ext_Sev | 0.066 | 0.032 | 2.061 | 0.039 |
| CM_16 | -0.249 | 0.103 | -2.417 | 0.016 |
| CM_18 | 0.603 | 0.359 | 1.678 | 0.093 |
| CM_28 | 0.229 | 0.110 | 2.073 | 0.038 |

Table 26’s output shows the level of significance for each of the four variables in the negative binomial model. Of these six, the negative binomial reports that “Age”, “Treatment”, “Lower_Ext_Sev”, “CM_16”, and “CM_28” are significant according to α = 0.05. “CM_18” is not considered significant with an associated p-value of 0.093.

To achieve the best negative binomial model possible, “CM_18” was dropped and the model was re-run with only the five significant variables. The results of this reduced negative binomial model are tabulated in Table 27.

Table 27.  Negative Binomial Model Summary (5 Variables)

| Variable | Estimate | Std. Error | Z | p-value |
|---|---|---|---|---|
| (Intercept) | 6.605 | 0.304 | 21.756 | 0.000 |
| Age | 0.010 | 0.004 | 2.621 | 0.009 |
| Treatment | -0.179 | 0.085 | -2.103 | 0.035 |
| Lower_Ext_Sev | 0.069 | 0.032 | 2.160 | 0.031 |
| CM_16 | -0.247 | 0.103 | -2.391 | 0.017 |
| CM_28 | 0.218 | 0.111 | 1.971 | 0.049 |

Table 27 shows that all the variables included in the model were now reported as significant. The next step is to interpret the results after model diagnostics are confirmed.

The most appropriate way to judge a model's assumptions and fit is by calculating the residuals and generating a normal quantile quantile (Q-Q) plot. The results of this process are shown in Figure 10.

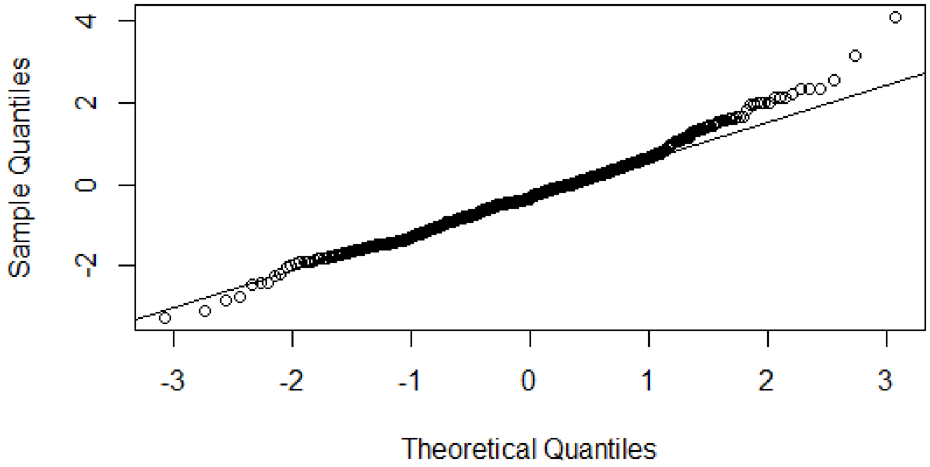

Figure 10. Negative Binomial Normal Q-Q Plot (5 Variables)

Figure 10 demonstrates that the negative binomial model's assumptions for the data were mostly correct. The residuals for the most part fall nicely along the Q-Q line shown as desired. The results of this plot imply that the negative binomial with five variables is appropriate for the dataset and could be relied on for making conclusions about the variables.

The values to be interpreted are the coefficients reported by the five-variable negative binomial model. The negative binomial coefficients are exponentiated into the form of odds ratios. These odds ratios are used to determine the multiplicative effect of one-unit increases in the predictors on the response variable. The results of transforming the negative binomial coefficients from Table 27 into odds ratios as well as 95% confidence intervals for the odds ratios are listed in Table 28.

Table 28. Negative Binomial Odds Ratios Summary

| Variable | Odds Ratio | 95% CI |
|---|---|---|
| (Intercept) | 738.735 | (401.018, 1360.308) |
| Age | 1.010 | (1.002, 1.018) |
| Treatment | 0.836 | (0.705, 0.986) |
| Lower_Ext_Sev | 1.071 | (1.005, 1.140) |
| CM_16 | 0.781 | (0.640, 0.962) |
| CM_28 | 1.244 | (1.006, 1.556) |

Table 28 provides critical information on assessing the five significant variables reported by the feature selection and negative binomial steps. First, the odds ratio for the intercept is 738.735, which means that the predictive equation for the outcome includes the constant multiplier value of 738.735 minutes. Next, the close-to-1 odds ratio for "Age" (1.010) initially appears to show that age of the patient has near-zero effect on the outcome. However, this odds ratio, when raised to the power of a patient's age, for example, eighty, can have a large effect on the outcome variable ($1.010^{80}$ = 2.217).

Finally, it can be seen that both "Treatment" and "CM_16" have odds ratios less than 1, which means that these variables are associated with lower odds of the outcome occurring. These binary variables' odds ratios indicate that when they are each set to 1, the predicted "Min_to_1st_OR" value is decreased. This means that when the treatment is applied to patients, on average, they require less time to go from arrival to their first surgery (and similarly for possessing "CM_16"). Conversely, "Lower_Ext_Sev" and "CM_28" are associated with higher odds of the outcome. To provide clarity on the effects of the intercept and variables, the following example predicts "Min_to_$1^{st}$_OR" for a hypothetical patient who is seventy-two years old, received the treatment, has an lower extremity injury level 2, has “CM_28” (congestive heart failure), but does not possess “CM_16” (tobacco use):

$$Min\_to\_1st\_OR = 738.735 * 1.010^{72} * 0.836^{1} * 1.071^{2} * 0.781^{0} * 1.244^{1} = 1803\ min.$$

This example serves to depict how the odds ratios work to predict the primary outcome variable. Specifically, when the treatment is applied to patients, they will, on average, require 1 – 0.836 = 16.4% less time to get into surgery compared to patients with the same age, lower extremity injury level, and status on suffering from tobacco use and congestive heart failure who did not receive the treatment. Overall, this odds ratio for “Treatment” and the example above support the hypothesis that the treatment variable is significant in predicting time-from-arrival-to-first-operation, and that the effect is as desired by the researchers. According to the results above, treatment assignment appears to lead to patients getting into surgery faster, which means the primary goal of the study appears to have been accomplished.

The confidence intervals portion of Table 28 shows the ranges that the odds ratios fall within, according to a 95% confidence level. While the intervals above do not report a significance level for each variable, they do offer a look at how assured the negative binomial is that the variables’ coefficients fall within a specific window. From Table 28, it can be seen that the intercept has a large confidence interval, which means that the true intercept likely falls within the wide window of 401.018 to 1360.308 minutes. This large spread is due to the variance present in the outcome variable (reference Table 25). However, the intervals for the five significant predictors are all smaller, and most importantly, do not include the value of 1. Table 28 shows that one can be 95% confident that the true odds ratios for the five variables are not equal to 1, and therefore have an effect on time-from-arrival-to-first-operation. Overall, these intervals support the findings above and provide more evidence that these five variables, including treatment, are significant predictors.

# IV. Conclusions and Recommendations

With the necessary analytic steps complete, insights to Dr. Haac's team on the effect of their risk stratification protocol treatment can be offered. Beginning with the primary outcome, all the analysis results can be pooled together in order to form one solid conclusion about the treatment's effect on time-from-arrival-to-first-operation. Overall, the treatment's effect on the main outcome was shown to be significant. The mean difference between the pre- and post-treatments groups was large pre-matching, and was even larger post-matching. While the *t*-test used was unable to report that the treatment was completely significant according to the standard α-level of 0.05, the p-value was quite small, thereby suggesting a significant mean difference. Moreover, the feature selection process picked the treatment variable as an important and highly significant predictor. The negative binomial model then supported this result by showing its effect on "Min_to_1st_OR" to be significant again. Finally, the converted negative binomial output, in the form of odds ratios, showed via a less than 1 odds ratio and a confidence interval that did not include 1, that the treatment variable had a significant effect on the main outcome variable. In total, these results point to the conclusion that there is certainly a noticeable decrease in time-from-arrival-to-first-operation for the patients in the study after the treatment took effect, that the change in means is directly attributable to the treatment's implementation, and that the treatment assignment on average gets the geriatric patients into surgery faster.

The secondary outcome is the number of cardiac events experienced. Unfortunately, due to the small percentage of patients (around 3%) who experienced a cardiac event intra- or post-operation, there was not enough data for the majority of the analysis. The genetic matching, while successful at improving the balance between the sets and reducing the treatment selection bias within, did eliminate many control patients who had experienced a cardiac event. While there was a noticeable drop in the percentage of patients prior to matching, there is just not enough data to be able to form a concrete conclusion about the treatment’s effect on cardiac events experienced. The pre-matching dataset does suggest that there was an effect as seen by the visible decrease in patients experiencing cardiac events, but there just needs to be more data collected in order to ensure this decrease is significant. Thus, nothing can be stated significantly about the number cardiac events experienced by the geriatric patients in the study.

Overall, the small indication that the treatment decreased the number of cardiac events as well as the major lack of data implies that the addition of more evidence could show this effect to be definitively significant. That is, if the study were extended another year or two in order to collect more data, it is certainly possible that the analytic methods applied could find the treatment effect on cardiac events to be significant. However, despite the successful matching process and the suggestive results in the analysis, there is just not enough clear data to be able to state that the treatment lowered the number of patients who experienced cardiac events.

Due to the overwhelming conclusion that the treatment decreased time-from-arrival-to-operation as well as the evidence that significantly more data is required to demonstrate an effect on cardiac events, two concrete recommendations can be made to Dr. Haac’s team. Provided there is deliberate accounting for the strong effect of “I.T_05”, the treatment variable can be shown to be highly significant in predicting time-from-arrival-to-first-operation according to a large swath of techniques conducted in Chapter

III. Overall, this means that the researchers risk stratification protocol was an effective implementation that statistically lowered the wait times for orthopedic geriatrics patients requiring non-emergent surgeries. Thus, it is recommended that the protocol remain in place as it is definitively increasing the hospital's efficiency as well as improving patient care.

The second recommendation that can be made is that the study be extended further in order to collect more data. Given the low percentage of patients who enter the medical center with cardiac risks, the study needs to multiply in size in order to provide the requisite data in order to confidently ascertain the treatment's effect on cardiac events in the same manner performed for time-from-arrival-to-first-operation. Overall, due to the high variation in the "Min_to_1st_OR" variable as well as the tiny percentage of patients experiencing cardiac events, the dataset needs to be expanded in order for more concrete insights to be extracted. It can be stated that the small amount of data at hand suggests that the treatment effect was at least partially responsible for the declines in the cardiac events outcome, but it is not a sure conclusion.

The final takeaway from this analysis is that there are a few other variables that were significant in explaining the primary outcome. In addition to treatment assignment, "Age" and "CM_16" explain the decrease in the primary outcome. Consequently, it is suggested that the researchers explore further why this is the case. Specifically this means studying why patient age and the tobacco use risk factor have such a significant impact on time-from-arrival-to-first-operation. Moreover, perhaps the results of the modeling analysis can provide insightful information to the researchers about their treatment in such a way that they can devise a more effective treatment process for an additional study. Either collecting more data in the current study and/or building a new one are both recommendable ways for potentially achieving and statistically demonstrating a larger treatment effect on the outcomes of interest.

## V. Future Work

If allotted more time to study and analyze the current dataset, there are few more methods and tools that could be applied. First, the various matching methods could be more significantly explored. It is possible that adjusting certain parameters in some of the ineffective methods could actually lead to successful results. While the genetic matching method was effective on Dr. Haac's dataset, there was certainly room for improvement as far as achieving optimal balance across the entirety of the dataset. Consequently, the bulk of future work would be invested in varying all the parameters in all the PSM methods to the maximum extent possible until every combination had been tried. This exhaustive process might uncover a better matching model for this dataset, which could provide better analyses of the treatment effect on the two outcomes of interest.

Another focus in future work would involve exploring other regression/treatment assessment techniques. Negative binomial regression is by no means the only available tool for assessing a treatment in observational data. Consequently, additional techniques would be researched and applied in order to determine if any other variables were significant or if there were differing results regarding the treatment effect.

Finally, time-permitting, the crux of this analysis, PSM, might be discarded in favor of an alternative treatment selection bias reduction method if available. While PSM was clearly the most appropriate technique for this dataset and the questions asked, there are likely other methods that could have resolved the issues within and reported differing conclusions about the treatment assignment's effect on the outcome variables.

Overall, these few ideas comprise the steps that would be taken if more time was provided for future work. While the results and conclusions reached in this thesis are solid and reliable, it never hurts to conduct deeper, additional analysis. Consequently, these are the main steps that would be taken in order to further support the findings of this thesis.

# V. Appendix

## A.1 Overview

All of the R code used to import the data, conduct data summaries and visualizations, perform PSM, conduct analysis, and form conclusions is included in this Appendix for the sake of reproducibility. Note that the Excel steps used to perform the majority of the data cleaning are not included.

## A.2 Required Packages

```
# These packages must be installed & loaded in order to run the R code

library(xlsx)
library(rJava)
library(xlsxjars)
library(ggplot2)
library(magrittr)
library(dplyr)
library(MatchIt)
library(Matching)
library(MASS)
library(leaps)
```

## A.3 Importing Dataset from Excel

```
# Full dataset (3 outliers included)
my_DATA <- read.xlsx("thesis_data_16.xlsx", 1)

# Partial dataset (outliers removed)
my_DATA2 <- my_DATA[c(1:12,14:485,487:555,557:726),]
```

## A.4 Data Familiarization

```
#Primary outcome

#means
attach(my_DATA2)
mean(Min_to_1st_OR)
mean(Min_to_1st_OR[1:338])
mean(Min_to_1st_OR[339:723])

#standard deviation
sd(Min_to_1st_OR)
sd(Min_to_1st_OR[1:338])
sd(Min_to_1st_OR[339:723])
```

## A.5 Outcome Visualization

```
#Figure 1
ggplot(my_DATA, aes(x = Min_to_1st_OR)) +
   geom_histogram(binwidth = 200, alpha = .5, position = "identity")

#Figure 2
ggplot(my_DATA2, aes(x = Min_to_1st_OR)) +
   geom_histogram(binwidth = 200, alpha = .5, position = "identity")

#Figure 3
ggplot(my_DATA, aes(x = log(Min_to_1st_OR))) +
   geom_histogram(binwidth = 0.2, alpha = 0.5, position = "identity")

#Figure 4
ggplot(my_DATA2, aes(x = Min_to_1st_OR, fill = Treatment)) +
   geom_histogram(binwidth = 200, alpha = .5, position = "identity")

#Figure 5
my_DATA2 %>%
  mutate(
    Description = ifelse(Treatment == "0", "Control",
                         "Treatment")) %>%
  ggplot(aes(Min_to_1st_OR, fill = Description)) +
  geom_density(alpha = 0.4)

#Figure 6
attach(my_DATA)
dates <- c("Aug '13", "Dec '14", "May '16")
plot(ADMIT_DATE, Min_to_1st_OR, xaxt = 'n')
fit <- lm(Min_to_1st_OR ~ ADMIT_DATE)
abline(fit, col = "red")
axis(1,at=c(41500,42000,42500),labels=dates[1:3])
summary(fit)

#Figure 7
attach(my_DATA2)
plot(ADMIT_DATE, Min_to_1st_OR, xaxt = 'n')
fit2 <- lm(Min_to_1st_OR ~ ADMIT_DATE)
abline(fit2, col = "red")
axis(1, at=c(41500, 42000, 42500),labels = dates[1:3])
summary(fit2)
```

## A.6 PSM

```
# Nearest neighbor, 1-to-1 (partial)
Match_Out1 <- matchit(Treatment ~ Sex + White + Black + Other + Age +
                        ISS_90.05 + TRISS_90.05 + I.T_05 + Adm_Sys_BP +
                        ADM_D_BP + ADM_HR + ADM_RR + ADM_SaO2 + AdmTemp
                        + Upper.Ext.sev + Lower_Ext_sev + U.Op,
                        data = my_DATA2, method = "nearest",
                        ratio = 1)

summary(Match_Out1)

# Nearest neighbor, 1-to-1 (full)
Match_Out2 <- matchit(Treatment ~ Sex + White + Black + Other + Age +
                          ISS_90.05 + TRISS_90.05 + I.T_05 + Adm_Sys_BP
                          + ADM_D_BP + ADM_HR + ADM_RR + ADM_SaO2 +
                          + AdmTemp + Upper.Ext.sev +

                          Lower_Ext_sev + U.Op +
                          CM_NONE + CM_6 + CM_7 + CM_8 + CM_9 + CM_10 +
                          CM_12 + CM_13 + CM_14 + CM_16 + CM_17 + CM_18
                          + CM_19 + CM_20 + CM_22 + CM_24 + CM_26 +
                          CM_27 + CM_28 + CM_29 + CM_30 + CM_31 +
                          CM_32 + CM_34 + CM_35 + CM_36 + CM_39 +
                          CM_41 + CM_43 + CM_49 + CM_54 + CM_55 +
                          CM_56 + CM_58 + CM_59 + CM_60 + CM_61 +
                          CM_63 + CM_65 + CM_69 + CM_74 + CM_76 + CM_78
                          + CM_85 + CM_88 + CM_95 + CM_103 + CM_114 +
                          CM_115, data = my_DATA2, method =
                          "nearest", ratio = 1)

#ratio can be set to 2 or higher as desired

#method can be adjusted to exact, subclass, optimal, full, and cem to
#test the additional PSM matching methods
#note: additional packages will be required for some methods including
#optmatch and cem


# Genetic matching (partial)
Match_Out3 <- matchit(Treatment ~ Sex + White + Black + Other + Age +
                         ISS_90.05 + TRISS_90.05 + I.T_05 + Adm_Sys_BP
                         + ADM_D_BP + ADM_HR + ADM_RR + ADM_SaO2 +

                         AdmTemp +
                         Upper.Ext.sev + Lower_Ext_sev + U.Op,
                         data = my_DATA2, method = "genetic",
                         pop.size = 1000)
summary(Match_Out3)

# Genetic Matching (full model)
Match_Out4 <- matchit(Treatment ~ Sex + White + Black + Other + Age +
                         ISS_90.05 + TRISS_90.05 + I.T_05 + Adm_Sys_BP
                         + ADM_D_BP + ADM_HR + ADM_RR + ADM_SaO2 +
                         AdmTemp +
                         Upper.Ext.sev + Lower_Ext_sev + U.Op +
                         CM_NONE + CM_6 + CM_7 + CM_8 + CM_9 + CM_10 +
                         CM_12 + CM_13 + CM_14 + CM_16 + CM_17 + CM_18
                         + CM_19 + CM_20 + CM_22 + CM_24 + CM_26 +
                         CM_27 + CM_28 + CM_29 + CM_30 + CM_31 + CM_32
                         + CM_34 +  CM_35 + CM_36 + CM_39 + CM_41 +
                         CM_43 + CM_49 + CM_54 + CM_55 + CM_56 +
                         CM_58 + CM_59 + CM_60 + CM_61 +
                         CM_63 + CM_65 + CM_69 + CM_74 + CM_76 +
                         CM_78 + CM_85 + CM_88 + CM_95 +
                         CM_103 + CM_114 + CM_115,
                         data = my_DATA2, method = "genetic",
                         pop.size = 1000)

#note: the pop.size command ensures that the run-time is limited, as
#otherwise it takes a lengthy amount of time due to the high number of
#variables

summary(Match_Out4)
```

### A.7 PSM Visualization

```
#Figure 8
plot(Match_Out4, type = "jitter", interactive = F)

#Figure 9
plot(Match_Out4, type = "hist")
```

### A.8 PSM Output / Results

```
#retrieves matched (reduced) dataset from genetic PSM
M_gen_out <- match.data(Match_Out4)[1:ncol(my_DATA2)]
#note: 1:ncol removes the additional two columns that get appended to
#the dataset after matching (distance, weights)

mean(M_gen_out$Min_to_1st_OR[1:87])
mean(M_gen_out$Min_to_1st_OR[88:472])
mean(M_gen_out$Min_to_1st_OR)
```

### A.9 Significance Test

```
t.test(M_gen_out$Min_to_1st_OR[M_gen_out$Treatment == 0],
       M_gen_out$Min_to_1st_OR[M_gen_out$Treatment == 1],

       paired = FALSE)
```

## A.10 Secondary outcome analysis

```
sum(my_DATA2$CARDIAC[1:388])
sum(my_DATA2$CARDIAC[389:723])

sum(M_gen_out$CARDIAC[1:87])
sum(M_gen_out$CARDIAC[88:472])
```

## A.11 Feature Selection

```
# Stepwise regression process
fullmod <- lm(Min_to_1st_OR ~ Sex + White + Black + Other + Age +
                ISS_90.05 + TRISS_90.05 + I.T_05 + Treatment +

                Adm_Sys_BP +
                ADM_D_BP + ADM_HR + ADM_RR + ADM_SaO2 + AdmTemp +
                Upper.Ext.sev + Lower_Ext_sev + U.Op +
                CM_NONE + CM_6 + CM_7 + CM_8 + CM_9 + CM_10 +
                CM_12 + CM_13 + CM_14 + CM_16 + CM_17 + CM_18 +
                CM_19 + CM_20 + CM_22 + CM_23 + CM_24 + CM_26 +
                CM_27 + CM_28 + CM_29 + CM_30 + CM_31 + CM_32 + CM_34 +
                CM_35 + CM_36 + CM_37 + CM_38+  CM_39 + CM_41 + CM_43 +
                CM_47 + CM_49 + CM_50 + CM_54 + CM_55 + CM_56 +
                CM_57 + CM_58 + CM_59 + CM_60 + CM_61 + CM_62 +
                CM_63 + CM_64 + CM_65 + CM_69 + CM_74 + CM_76 +
                CM_78 + CM_85 + CM_88 + CM_95 + CM_96 + CM_98 +
                CM_100 + CM_103 + CM_109 + CM_114 +
                CM_115, data = M_gen_out)

step <- stepAIC(fullmod, trace = F)
step$anova
summary(step)

# I.T_05 removed
fullmod_2 <- lm(Min_to_1st_OR ~ Sex + White + Black + Other + Age +
                ISS_90.05 + TRISS_90.05 + I.T_05 + Treatment +

                Adm_Sys_BP +
                ADM_D_BP + ADM_HR + ADM_RR + ADM_SaO2 + AdmTemp +
                Upper.Ext.sev + Lower_Ext_sev + U.Op +
                CM_NONE + CM_6 + CM_7 + CM_8 + CM_9 + CM_10 +
                CM_12 + CM_13 + CM_14 + CM_16 + CM_17 + CM_18 +
                CM_19 + CM_20 + CM_22 + CM_23 + CM_24 + CM_26 +
                CM_27 + CM_28 + CM_29 + CM_30 + CM_31 + CM_32 + CM_34 +
                CM_35 + CM_36 + CM_37 + CM_38+  CM_39 + CM_41 + CM_43 +
                CM_47 + CM_49 + CM_50 + CM_54 + CM_55 + CM_56 +
                CM_57 + CM_58 + CM_59 + CM_60 + CM_61 + CM_62 +
                CM_63 + CM_64 + CM_65 + CM_69 + CM_74 + CM_76 +
                CM_78 + CM_85 + CM_88 + CM_95 + CM_96 + CM_98 +
                CM_100 + CM_103 + CM_109 + CM_114 +
                CM_115, data = M_gen_out)
#note: the default setting is "both"

step_2 <- stepAIC(fullmod_2, trace = F)
step_2$anova
summary(step_2)

# Best subsets
sub <- regsubsets(Min_to_1st_OR ~ Sex + White + Black + Other + Age +
             ISS_90.05 + TRISS_90.05 + I.T_05 + Treatment + Adm_Sys_BP
             + ADM_D_BP + ADM_HR + ADM_RR + ADM_SaO2 + AdmTemp +
             Upper.Ext.sev + Lower_Ext_sev + U.Op +
             CM_NONE + CM_6 + CM_7 + CM_8 + CM_9 + CM_10 +
             CM_12 + CM_13 + CM_14 + CM_16 + CM_17 + CM_18 +
             CM_19 + CM_20 + CM_22 + CM_23 + CM_24 + CM_26 +
             CM_27 + CM_28 + CM_29 + CM_30 + CM_31 + CM_32 + CM_34 +
             CM_35 + CM_36 + CM_37 + CM_38+  CM_39 + CM_41 + CM_43 +
             CM_47 + CM_49 + CM_50 + CM_54 + CM_55 + CM_56 +
             CM_57 + CM_58 + CM_59 + CM_60 + CM_61 + CM_62 +
             CM_63 + CM_64 + CM_65 + CM_69 + CM_74 + CM_76 +
             CM_78 + CM_85 + CM_88 + CM_95 + CM_96 + CM_98 +
             CM_100 + CM_103 + CM_109 + CM_114 +
             CM_115, data = M_gen_out, really.big=T)
#note: really.big must be turned on to enable a satisfactory

#finish time

summary(sub)

sub2 <- regsubsets(Min_to_1st_OR ~ Sex + White + Black + Other + Age +
             ISS_90.05 + TRISS_90.05 + Treatment + Adm_Sys_BP +
             ADM_D_BP + ADM_HR + ADM_RR + ADM_SaO2 + AdmTemp +
             Upper.Ext.sev + Lower_Ext_sev + U.Op +
             CM_NONE + CM_6 + CM_7 + CM_8 + CM_9 + CM_10 +
             CM_12 + CM_13 + CM_14 + CM_16 + CM_17 + CM_18 +
             CM_19 + CM_20 + CM_22 + CM_23 + CM_24 + CM_26 +
             CM_27 + CM_28 + CM_29 + CM_30 + CM_31 + CM_32 + CM_34 +
             CM_35 + CM_36 + CM_37 + CM_38+  CM_39 + CM_41 + CM_43 +
             CM_47 + CM_49 + CM_50 + CM_54 + CM_55 + CM_56 +
             CM_57 + CM_58 + CM_59 + CM_60 + CM_61 + CM_62 +
             CM_63 + CM_64 + CM_65 + CM_69 + CM_74 + CM_76 +
             CM_78 + CM_85 + CM_88 + CM_95 + CM_96 + CM_98 +
             CM_100 + CM_103 + CM_109 + CM_114 +
             CM_115, data = M_gen_out, really.big = T)

summary(sub2)
```

### A.11 Regression

```r
# mean vs variance check
mean(M_gen_out$Min_to_1st_OR[1:87])
mean(M_gen_out$Min_to_1st_OR[88:472])
var(M_gen_out$Min_to_1st_OR)

var(M_gen_out$Min_to_1st_OR[1:87])
var(M_gen_out$Min_to_1st_OR[88:472])
var(M_gen_out$Min_to_1st_OR)

# Negative binomial on six variables chosen
mod_final <- glm.nb(round(Min_to_1st_OR) ~ Age +
                      Treatment + Lower_Ext_sev + CM_16 +
                      CM_18 + CM_28, data = m11_2_out2)
summary(mod_final)

# negative binomial on five significant variables
mod_final2 <- glm.nb(round(Min_to_1st_OR) ~ Age +
                      Treatment + Lower_Ext_sev + CM_16 +
                      CM_28, data = m11_2_out2)
summary(mod_final2)
#Note: round() must be applied in order for glm.nb to function.

#Q-Q plot, Figure 10
Resid <- rstandard(mod_final2)
qqnorm(Resid)
qqline(Resid)
```

### A12. Odds Ratios

```r
#exporting odds ratios and 95 % C.I.'s
capture.output(exp(cbind(OR = coef(mod_final2),
                         confint(mod_final2))),file="odds.txt",

                         append=TRUE)


#note: this exports a .txt file with the odds ratios and conf ints
```

# Bibliography

Austin, Peter C. "A Critical Appraisal of Propensity-Score Matching in the Medical Literature Between 1996 and 2003," *Statistics in Medicine*, 2037-2049 (2008).

Cameron, A. Colin and Pravin K. Trivedi. "Essentials of Count Data Regression." *A Companion to Theoretical Economics*. Ed. Badi Baltagi. Malden, MA: Blackwell Publishing Ltd, 1999.

D'Agostino, Jr., Ralph B. "Propensity Score Methods for Bias Reduction in the Comparison of a Treatment to a Non-Randomized Control Group," *Statistics in Medicine*, 2265-2281 (1998).

"Evaluating Observational Data Analyses: Confounding Control and Treatment Effect Heterogeneity." *Brown.edu*. Brown School of Public Health. 15 May 2016. Retrieved on 10 October 2017.

Guyon, Isabelle and André Elisseeff. "An Introduction to Variable and Feature Selection," *Journal of Machine Learning Research*, 1157-1182 (March 2003).

Haac, Bryce. Surgical Resident/Researcher, University of Maryland Medical Center, Baltimore, MD. Personal Correspondence. 5 July 2017.

Higgins, John P. "Nonlinear Systems in Medicine," *Yale Journal of Biology and Medicine*, 247-260 (2002).

Ho, Daniel E., Kosuke Imai, Gary King, Elizabeth A. Stuart. "Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference," *Political Analysis*, 199-236 (January 2007).

Ho, Daniel E., Kosuke Imai, Gary King, Elizabeth A. Stuart, and Alex Whitworth.

"MatchIt: Nonparametric Preprocessing for Parametric Casual Inference," *Journal of Statistical Software*, 1-38 (June 2011).

Ho, Daniel E., Kosuke Imai, Gary King, Elizabeth A. Stuart, and Alex Whitworth. "MatchIt: Nonparametric Preprocessing for Parametric Casual Inference," *Journal of Statistical Software*, 1-8 (April 2017).

Holland, Paul W. "Statistics and Causal Inference," *Journal of the American Statistical Association*, 945-960 (December 1986).

Imai, Kosuke and David A. van Dyk. "Causal Inference with General Treatment Regimes: Generalizing the Propensity Score," *Journal of the American Statistical Association*, 854-866 (September 2004).

Kim, Tae Kyun. "T Test as a Parametric Statistic," *Korean Journal of Anesthesiology,* 540-546 (November 2015).

Kuhn, Max and Kjell Johnson. *Applied Predictive Modeling*. New York: Springer, 2013.

"Life Expectancy Increases Globally as Death Toll Falls from Major Diseases." *Healthdata.org*. Institute for Health Metrics and Evaluation. 30 January 2015. Retrieved on 13 October 2017.

Loukides, Mike. "The Unreasonable Necessity of Subject Experts," *O'Reilly*. 20 March 2012. RadarOReilly.com. Retrieved on 31 January 2018.

Pearl, Judea. "Causal Inference in Statistics: An Overview," *Statistics Surveys*, 3: 96-146 (September 2009).

Perkins, Susan M., Wanzhu Tu, Michael G. Underhill, Xiao-Hua Zhou, and Michael D. Murray. "The Use of Propensity Scores in Pharmacoepidemiologic Research," *Pharmacoepidemiology and Drug Safety*, 93-101 (March 2000).

Popovic, Gordana. "Interpreting Coefficients in glms," *EnvironmentalComputing.net*. 11 November 2016. Retrieved on 4 February 2018.

R Core Team (2017). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Randolph, Justus J., Kristina Falbe, Austin K. Manuel, and Joseph L. Balloun. "A Step-by-Step Guide to Propensity Score Matching in R," *Practical Assessment, Research & Evaluation*, 1-6 (November 2014).

Rosenbaum, Paul R. and Donald B. Rubin. "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, 41-55 (April 1983).

Rubin, Donald B. "On the Application of Probability Theory to Agricultural Experiments," *Statistical Science*, 472-480 (November 1990).

Sekhon, Jasjeet S. "Multivariate and Propensity Score Matching Software with Automated Balance Optimization: The Matching Package for R," *Journal of Statistical Software*, 1-51 (May 2011).

Stuart, Elizabeth A. "Matching Methods for Causal Inference: A Review and a Look Forward," *Statistical Science*, 1-21 (2010).

Szumilas, Magdalena. "Explaining Odds Ratios," *Journal of the Canadian Academy of Child and Adolescent Psychiatry*, 227-229 (2010).

Thavaneswaran, Arane. "Propensity Score Matching in Observational Studies," *Manitoba Centre for Health Policy*, 1-28 (April 2008).

Wickham, Hadley. *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag, 2009.

Zhang, Zhongheng. "Missing Data Imputation: Focusing on Single Imputation," *Annals* *of Translational Medicine*, 2-7 (December 2015).

Zhang, Zhongheng. "Variable Selection with Stepwise and Best Subsets Approaches," *Annals of Translational Medicine*, 1-6 (January 2016).

# REPORT DOCUMENTATION PAGE

*Form Approved*
*OMB No. 0704-0188*

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

**PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

| | | |
|---|---|---|
| **1. REPORT DATE** *(DD-MM-YYYY)*<br>22-03-2018 | **2. REPORT TYPE**<br>Master's Thesis | **3. DATES COVERED** *(From - To)*<br>Aug 2016 - Mar 2018 |

**4. TITLE AND SUBTITLE**

Analysis of a Medical Center’s Cardiac Risk Screening Protocol Using Propensity Score Matching

**5a. CONTRACT NUMBER**

**5b. GRANT NUMBER**

**5c. PROGRAM ELEMENT NUMBER**

**6. AUTHOR(S)**

Johnson, Jake E., 2d Lt, USAF

**5d. PROJECT NUMBER**

**5e. TASK NUMBER**

**5f. WORK UNIT NUMBER**

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**

Air Force Institute of Technology
Graduate School of Engineering and Management (AFIT/EN)
2950 Hobson Way
Wright-Patterson AFB OH 45433-7765

**8. PERFORMING ORGANIZATION REPORT NUMBER**

AFIT-ENC-MS-18-M-129

**9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)**

University of Maryland Medical Center
Dr. Bryce Haac, MD
22 S Greene St,
Baltimore, MD 21201
brycehaac@gmail.com

**10. SPONSOR/MONITOR'S ACRONYM(S)**

**11. SPONSOR/MONITOR'S REPORT NUMBER(S)**

**12. DISTRIBUTION/AVAILABILITY STATEMENT**

Distribution Statement A. Approved for public release; distribution unlimited.

**13. SUPPLEMENTARY NOTES**

This material is declared a work of the U.S. Government and is not subject to copyright protection in the United States.

**14. ABSTRACT**

Researchers at the University of Maryland Medical Center in Baltimore have developed a cardiac risk stratification protocol in the hopes of reducing the time-from-arrival-to-first-operation for geriatric orthopedic patients. They collected observational data for two years prior to and following the October 2014 implementation of the new screening protocol. Therefore, advanced analytical techniques are required to isolate the treatment effect of the new screening protocol. Propensity score matching (PSM) is used to handle the observational data in order to reduce the bias attributable to the confounding covariates. In addition to PSM, various regression techniques are used to help the researchers determine if their treatment has been successful in reducing the number of cardiac complications experienced by the elderly patients during and post-surgery. Recommendations are then made to the hospital’s researchers.

**15. SUBJECT TERMS**

Propensity score matching, genetic, treatment, observational data, medical

| **16. SECURITY CLASSIFICATION OF:** | | | **17. LIMITATION OF ABSTRACT** | **18. NUMBER OF PAGES** | **19a. NAME OF RESPONSIBLE PERSON** |
|---|---|---|---|---|---|
| **a. REPORT** | **b. ABSTRACT** | **c. THIS PAGE** | | | Lt Col Andrew J. Geyer, AFIT/ENC |
| U | U | U | UU | 102 | **19b. TELEPHONE NUMBER** *(Include area code)*<br>(937) 255-6565 x4584 Andrew.Geyer@afit.edu |