Assessing the effects of coronary artery bypass grafting versus complex cardiac surgery: Comparison of Methods of Adjusting for Channeling Bias

By

Yanyun Liu, B.Sc.

A Thesis

Submitted to the School of Graduate Studies in Partial Fulfillment of the Requirements

for the Degree

Master of Science

McMaster University

MASTER OF SCIENCE (2010) McMaster University

(Statistics) Hamilton, Ontario

TITLE: Assessing the effects of coronary artery bypass grafting versus complex cardiac surgery: Comparison of methods of adjusting for channeling bias

AUTHOR: Yanyun Liu, B.Sc. (Henan University of Science & Technology)

SUPERVISOR: Dr. Lehana Thabane

NUMBER OF PAGES: viii, 72

Abstract

Coronary artery bypass grafting (CABG) is the most commonly performed "open heart" operation in North America. Complex cardiac surgeries served for a large amount of the cardiac surgery population, but outcomes after these surgeries have been limited by lack of appropriate interpretation. Given the observed trend toward an increasing proportion of complex cardiac surgeries, there is a great need to understand the outcomes and patterns of resource utilization for the population who have had complex cardiac surgery.

The clinical objectives of this thesis are to compare clinical outcomes and resource usage between isolated coronary bypass grafting and complex cardiac surgery and determine the difference of outcomes for complex cardiac surgeries among cardiac surgical sites across Canada.

The statistical objective of this thesis is to compare Bayesian and classical methods of analyzing two surgeries difference in outcomes. The classical methods are multivariable logistic regression, matched propensity score method, propensity score weighted regression and stratified propensity score method. The Bayesian method is Bayesian matched propensity score.

For the primary outcome mortality, the odds ratio and 95% confidence interval for the treatment effect is 4.49 (1.92, 10.56) for propensity score matching method, 4.97 (3.62, 6.11) for propensity score weight method, 3.49 (1.91, 6.40) for propensity score strata method, 3.71 (2.10, 6.56) for multivariable regression method, and 3.82 (1.23,

13.07) for Bayesian propensity score matching method. Different methods obtained different treatment effect estimates.

We concluded that patients who are undergoing complex cardiac surgery have a greater risk for adverse postoperative events and longer ICU length of stay compared to patients who are undergoing isolated CABG. We also found that there is variability in outcomes and resource usage among Canadian cardiac centers.

Acknowledgements

I would like to express my deepest gratitude to my supervisor, Dr. Lehana Thabane for giving me the opportunity to work on this project. Without his guidance and support, this project would not have been possible.

Special thanks to my committee members, Dr. Roman Viveros-Aguilera and Dr. Eleanor Pullenayegum who offered valuable advice.

Finally, I would like to thank all of my friends and family for their love and help, especially my husband, Shaozong Li, who endured this long process with me, always offering support and understanding. I would also thank my parents, who have helped me taking care of my daughters. Without their help, it would not be having been possible for me to focus on the study.

Table of Contents

Appendix D: Tables

Appendix E: Figures

Appendix F: Code

Chapter 1

Introduction

1.1 The Multi-center Study

Coronary artery bypass grafting (CABG) is the most commonly performed "open heart" operation in North America. Outcomes of coronary artery bypass grafting surgery have been studied extensively in North America, including a recent nation-wide report on mortality rates for CABG surgery across all provinces in Canada [1]. Complex cardiac surgeries served for a large amount of the cardiac surgery population, but outcomes after these surgeries have been limited by lack of appropriate interpretation. There has been a substantial increase in the number of complex cardiac surgeries in recent years, largely due to improvements in the outcomes for patients who are undergoing these procedures [2]. This has led to a more liberal indication for complex cardiac surgery, primarily in elder patients [3-5]. Given the observed trend toward an increasing proportion of complex cardiac surgeries, there is a great need to understand the outcomes and patterns of resource utilization for the population who have had complex cardiac surgery.

Some studies have reported outcomes of combined cardiac surgical procedures in Canada, but these studies have been limited to analysis of mortality rates alone, using data from a single cardiac center [3, 6]. Although mortality in combined CABG-valve surgeries has been studied extensively, no studies involving Canadian hospitals have analyzed outcomes associated with serious morbidity following complete

cardiac surgery, such as stroke, postoperative myocardial infarction, renal failure and prolonged mechanical ventilation. Knowledge of these outcomes is becoming increasingly important, as greater focus is being placed not only on survival, but also on improvement in quality of life following cardiac surgery [7].

Compared to patients undergoing isolated CABG or isolated valve procedures, complex cardiac surgical patients have greater co-morbidities and longer surgeries [2, 5], which are associated with higher mortality rates [8], increased length of hospital stay [2, 5], and greater cost [9].

This study is based on the data from a previous multi-center cohort study [10]. It included 3500 adult (> 18 yrs) patients who had cardiac surgery at seven academic Canadian hospitals during 2004. Five hundred consecutive patients who underwent cardiac surgery from each hospital were collected, excluding infrequent procedures (heart transplantation, ventricular assist device placement, and complex congenital abnormality repair). We also excluded patients undergoing single valve procedures, isolated aortic root surgery, atrial septal defect repair and emergency procedures.

There were 22 baseline variables collected retrospectively, which are listed as follows: gender, age, weight, body surface area, urgency of surgery, smoking, left ventricle grade, angina, myocardial infarction, diabetes, hypertension, hypercholesterolemia, chronic obstructive pulmonary disorder, cerebrovascular disease, peripheral vascular disease, atrial fibrillation, congestive heart failure, shock, dialysis, heparin use, timing, and acetylsalicylic acid.

The primary outcome of interest is mortality, while secondary outcomes include stroke, renal failure (defined as initiation of dialysis post-surgery), postoperative myocardial infarction, pneumonia, re-intubation and re-exploration. ICU length of stay and the duration of ventilation were used as measures of resource usage.

1.2 Objectives of the Study

The clinical objectives of this thesis are to compare clinical outcomes and resource usage between isolated coronary bypass grafting and complex cardiac surgery and determine the difference of outcomes for complex cardiac surgeries among cardiac surgical sites across Canada. The primary outcome of interest is mortality, while secondary outcomes include stroke, renal failure (defined as initiation of dialysis post-surgery), postoperative myocardial infarction, pneumonia, re-intubations and re-exploration. ICU length of stay and the duration of ventilation were used as measures of resource usage.

The statistical objectives of this thesis include three parts. First, to compare different classical and Bayesian methods of adjusting differential propensities for patients of receiving isolated coronary bypass grafting versus complex cardiac surgery. Many of the balancing methods proposed involve the propensity score [11], used for stratifying subjects [12], matching [13], or weighting [14]. Second, to assess the sensitivity of the Bayesian results to different priors chosen. Finally, we use multiple imputations to assess the impact of missing data on the results.

1.3 Scope of the Study

This report will start by introducing the background and design of the multi–center observational study. Then we will provide an overview of the different methods for channeling bias. We will also report the results of the study for the methods we used; as well, we will compare the results from different statistical methods and find the reason for the difference in the results. Finally, conclusions will be made based on our results and discussions.

In chapter 2, we will go through all of the statistical methods used in this study. Those statistical methods include propensity score matching method, propensity score weight method, propensity score strata method, multivariable regression and Bayesian propensity score matching method. We will also discuss the sensitivity of the Bayesian models.

In chapter 3, the results for primary and secondary outcomes will be reported. We will also compare the difference in the results.

In chapter 4, we will report the key findings of the study. We will discuss reasons for why we came up with different results from different methods. Moreover, the limitation and the strength of the design will be pointed out.

Finally, in Chapter 5, conclusions will be drawn based on the results of the analysis and some suggestions for future research of this topic will be given.

Chapter 2

Statistical Methods

2.1 Overview

In this chapter, we will provide an overview of the statistical methods used in conducting this study. As well, we will discuss the sensitivity analysis regarding the Bayesian model.

The demographic and baseline diagnostic characteristics of the patients were analyzed using descriptive statistics presented as mean (standard deviation) or median (minimum, maximum) for continuous variables and count (count) for categorical variables.

For primary outcome mortality, four classical methods and one Bayesian method are suitable for this study. The available classical methods for this study are multivariable logistic regression, propensity score strata, propensity score weighted, and propensity score matching. The design of the study is summarized in Appendix E Figure 1.

The Bayesian approach used in this thesis is the Bayesian propensity score matching method. This method was chosen because combining Bayesian analysis with propensity score techniques can ease model specification and yield estimates with good frequentist properties [15, 16].

The classical analysis will be performed using SAS 9.1 (Cary, NC) and STATA 9.1, and the Bayesian analysis will be performed using WinBUGS Version 1.4.

The results of the analysis for binary outcomes are reported as odds ratio (OR), corresponding 95% confidence interval (CI) and associated p-values. The Bayesian results will be reported in accordance with the ROBUST guideline [17]. For classical analysis of continuous variables, the results are reported as the estimate of treatment effect (coefficient), corresponding 95% CI and associated p-values. The SAS code for running classical methods and WinBUGS code for Bayesian model will be presented in Appendix F.

2.2 Multivariable Logistic Regression

In studies with a dichotomous outcome, the most common adjustment method is logistic regression of the outcome on treatment and a subset of the pretreatment covariates. The model is as following:

$$Logit \quad (\pi(x)) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots\ldots + \beta_p x_p$$

Here $\pi(x)$ represents the probability of an event that depends on p covariates or independent variables. x_1 is the treatment indicator and the rest are optional confounders adjusted for in the analysis.

2.3 Propensity Score Matching Method

The idea of using propensity score approach for adjusting for baseline imbalance between groups in observed formal studies has been discussed by several authors [12-14]. The propensity score is the probability that an individual would have been treated based on that individual's observed pretreatment variables. To describe the propensity score, let

the dichotomous (0, 1) variable Z indicate treatment, and let $\mathbf{X}$ be the vector of available pretreatment covariates. The propensity score $e(\mathbf{X})$ for an individual is defined as the conditional probability of being treated given his or her covariates $\mathbf{X}$: $e(\mathbf{X}) = \Pr(Z = 1|\mathbf{X})$. The propensity score is a one-dimensional variable that summarizes the multidimensional pretreatment covariates $\mathbf{X}$. Among persons with a given propensity score, the distribution of the covariates $\mathbf{X}$ is on average the same among the treated and untreated.

In our study, we assign complex cardiac surgery as the treatment and isolated CABG as control. The estimated propensity score, $e(\mathbf{X})$, was obtained from a logistic regression model. We considered the following pretreatment variables: gender, age, weight, body surface area, urgency of surgery, smoking, left ventricle grade, angina, myocardial infarction, diabetes, hypertension, hypercholesterolemia, chronic obstructive pulmonary disorder, cerebrovascular disease, peripheral vascular disease, atrial fibrillation, congestive heart failure, shock, dialysis, heparin use, acetylsalicylic acid, and timing.

Once we had done the estimation of the propensity score, we could then start the propensity score matching procedure. We matched participants who had complex cardiac surgery (treatment) to those who had coronary artery bypass grafting (control) based on a range of ± 0.05 of the propensity score. We chose the matching range of ± 0.05 because it is commonly used, provides reasonable balance of the included covariates, and does not lose many treated individuals as unmatchable [13]. To match patients, we used an automated matching procedure in the STATA software that randomly selected a treated

individual and randomly selected an untreated individual (comparator) from the pool of potential comparators to determine whether he or she met the matching criterion. If the selected comparator was eligible, he or she was matched to the treated individual, and the pair was removed. This procedure was repeated until all treated patients were matched to one comparator or until no further comparators met the matching criteria. The approach has been used by different authors [11, 13, 31-33].

2.4 Propensity Score Weighting Method

Weighting on the propensity is not implemented as commonly as the other methods of adjustment. In propensity weighting, the treatment and control observations are re-weighted in order to make them more representative of the population [14]. The weight of a treated subject is defined as the inverse of its propensity score, $1/\hat{e}(\mathbf{X})$. The weight of a control subject is defined as the inverse of one minus its propensity, $1/ (1 - \hat{e}(\mathbf{X}))$. The approach has been applied in practice by several authors [14, 20, 21].

2.5 Propensity Score Strata Method

In stratification, the estimated propensity score is used to stratify the subjects into homogenous subclasses, with similar propensity scores. Each subclass consists of relatively the same number of subjects. According to Rosenbaum's study [12], using five strata will eliminate more than 90% of the covariate bias.

Two approaches can be used to compare the outcomes of treated and control subjects. One is comparing the outcome of two groups within each stratum and finally combining

all of the strata difference to get the overall difference of the outcome. Another is fitting logistic model includes propensity score strata as a covariate in the model [12].

2.6 Bayesian Propensity Score Matching Method

Bayesian propensity score matching method offers a natural strategy for modeling uncertainty in the propensity scores. We modeled the joint distribution of the data and parameters that we are interested in with the propensity score as a latent variable.

In our Bayesian random effect regression model [18], we assumed that the random effect follows a normal distribution with mean zero and variance unknown parameter τ^2. The uncertainty of τ^2 is counted in the model by assuming a prior distribution which will represent the researcher's pre-belief or information external to τ^2. The likelihood function represented by the observed data is used to update the researcher's pre-belief and obtain the posterior distribution. The final results are presented as the posterior distribution. For binary outcomes, we will obtain the log odds ratio of the treatment effect from the posterior distribution. However, we will easily get the odds ratio which we are interested in doing exponential transformation.

In our primary analysis using the Bayesian model, we assumed the inverse gamma conjugate prior distribution with shape parameter and scale parameter 10 and 10 respectively. We assumed that the prior distribution for all the coefficients follows a normal distribution with mean zero and variance 1.0E+6. The total number of iterations to obtain the posterior distribution is 200,000, and the burned-in number is 100, thin 5. The convergence of the Markov Chain can be evaluated from the plot of the posterior

distributions including autocorrelation plots, time series plots, dynamic trace plots and density plots. Those plots will be reported in Appendix E, Figures 2-5.

2.7 Sensitivity Analysis of Bayesian Model

In Bayesian analysis, it is necessary to assess the sensitivity of the Bayesian model by choosing different priors. First, we chose non-informative priors, which included uniform distribution with lower bound 0 and upper bound 1, 5, 10 and 50 respectively. Then we chose the conjugate priors for the variance of the random effect. The priors are Inverse Gamma (5, 5), Inverse Gamma (10, 10) and Inverse Gamma (15, 15). The Bayesian results will be reported in accordance with the ROBUST guideline [17] Table 7 in Appendix D.

Chapter 3

Results

3.1 Demographic Information and Baseline Characteristics

Of the 3500 adult patients who had cardiac surgery, the mean age of the patients in the control group is 65.19 years with standard deviation (SD) 10.02, in the treatment group the mean age of the patients was 68.66 years with standard deviation (SD) 11.52. About 80% (1820/2271) of the patients in the control group were male, and 67% (320/476) in the treatment group were male. Detailed demographic information of the study patients will be reported in Appendix D, Table 1.

3.2 Results of Primary Analysis

The primary outcome of the study is mortality. We applied four classical statistical methods and one Bayesian method. The results from the various statistical methods differed, but the estimations by the five methods all showed that there were significant differences in mortality rate between the control and treatment groups.

For the propensity score matching method, first, we estimated the propensity score for each patient and then matched patients who had complex cardiac surgery (treatment) to those who had coronary artery bypass grafting (control) based on a range of ± 0.05 of

the propensity score. It is a common method to balance groups on covariates. After doing the match, the mean age of the patients in the control group was 69.19 years with standard deviation (SD) 8.47 years; in treatment group the mean age of the patients was 68.66 years with standard deviation (SD) 11.66 years. About 75% (325/432) of the patients were males in the control group and 69% (296/432) were males in the treatment group. The detailed demographic information of the patients will be reported in Appendix D Table 2 which shows how difference is eliminated after matching.

For the propensity score strata method, once we estimated propensity score for each patient, we used rank to place the patients into five stratums in which they are similar in the characteristics since they have a close propensity score. Adjusting the quintile as a covariate, the odds ratio and 95% confidence interval were 3.49 (1.91, 6.40).

For the propensity score weighted method, the odds ratio and 95% confidence interval of the treatment effect were 4.97 (3.62, 6.11) by using the inverse of its propensity score as the weight to a treated subject, and the inverse of one minus its propensity as the weight to a control subject.

For the multivariable logistic regression method, the treatment effect of the primary outcome obtained by adjusting the covariates age, gender, body surface area, hypertension, smoke, congestive heart failure, diabetes, site, redo number and timing (urgency of surgery). The odds ratio and 95% confidence interval for treatment effect of the primary outcome were 3.71 (2.10, 6.56).

For the Bayesian propensity score matching method, we modeled each patient's propensity score as a latent variable in the matched sample. Without adjusting any

covariate, the odds ratio and 95% credible interval of the treatment effect were 3.82(1.23, 13.07).

Comparing the results from the different statistical methods, we found that the estimates for the treatment effect were different. We also found that the Bayesian propensity score matching method gave the widest credible interval for the odds ratio of treatment effect. The reason is because the fact that the Bayesian model counted all of the uncertainty of the parameters. As well, we found that the propensity score weighted model gave the narrowest confidence interval for the odds ratio of the treatment effect.

We will discuss the outcome difference of various statistical methods in Chapter 4. The detailed results for the primary outcome are presented in Appendix D, Table 3 and Figure 6 in Appendix E.

3.3 Results of Secondary Analysis

The secondary outcomes of the study included binary outcomes and continuous outcomes. Binary outcomes include stroke, renal failure (defined as initiation of dialysis post-surgery), postoperative myocardial infarction, pneumonia, re-intubations, sepsis, tracheotomy and re-exploration. ICU length of stay and the duration of ventilation are continuous outcomes.

The secondary outcomes were analyzed by using four classical statistical methods. For the secondary outcome 'stroke', the odds ratio of the treatment effect and its corresponding 95% confidence interval were 7.15 (2.01, 25.74) for the propensity score matching method, 7.51 (2.18, 15.10) for the propensity score weighted method, 6.48 (2.45, 12.49) for the propensity score strata method and 3.94 (1.85, 8.41) for the

multivariable logistic regression with adjusted covariates age, gender, body surface area, hypertension, smoke, congestive heart failure, diabetes, site, redo number and timing (urgency of surgery).

For renal failure, the odds ratio of the treatment effect and its corresponding 95% confidence interval were 0.93 (0.54, 1.62) for the propensity score matching method, 3.49 (2.63, 4.60) for the propensity score weighted method, 1.72 (0.97, 3.01) for the propensity score strata method and 2.21(1.24, 3.95) for the multivariable logistic regression with adjusted covariates age, gender, body surface area, hypertension, smoke , congestive heart failure, diabetes, site, redo number and timing (urgency of surgery). We found that there was no significant difference of renal failure between the two surgeries when we matched patients by their propensity scores.

In terms of ICU length of stay, we found that there were differences of ICU length of stay between Isolated CABG surgery and Complex surgery. For example, the odds ratio and its corresponding 95% confidence interval were 1.24 (0.96, 1.59) for the propensity score matching method, 2.23 (2.01, 2.49) for the propensity score weighted method, 2.13 (1.70, 2.65) for the propensity score strata method and 1.81 (1.47, 2.23) for the multivariable logistic regression with adjusted covariates age, gender, body surface area, hypertension, smoke, congestive heart failure, diabetes, site, redo number and timing (urgency of surgery). The detailed results of the secondary outcomes will be reported in Appendix D, Table 4 and figure 6-9 in Appendix E.

3.4 Results of Bayesian Analysis

To do the sensitivity analysis of the Bayesian random effect logistic model, we evaluated the impact of different prior distributions of the variance parameter of interest. For the primary outcome of mortality, first we chose non-informative priors including uniform (0, 1), uniform (0, 5), uniform (0, 10) and uniform (0, 50). We found that the standard deviation never had a convergence. Then we chose Inverse Gamma (5, 5), Inverse Gamma (10, 10) and Inverse Gamma (15, 15). We found that the result were quite consistent when we chose conjugate prior Inverse Gamma with shape parameter and scale parameter greater than 5. The odds ratios and their corresponding 95% credible interval are 3.82 (1.12, 13.07), 3.79 (1.12, 13.07) and 3.80 (1.12, 13.07). This result will be reported in Table 6 in Appendix D.

3.5 Missing Data in Our Study

There are missing values on covariates in our study. There are three missing values for covariate gender, sixteen missing values for covariate smoke status, five missing values for covariate angina, eighteen missing values for covariate diabetes and eight missing values for covariate hypertension. There are no missing values for our primary outcome. To estimate each patient's propensity score, we need to deal with the missing values on covariates. We used the MI (multiple imputations) procedure in SAS to impute the missing data and estimated each patient's propensity score. Interestingly, for the propensity score weighted, we obtained the same treatment effect estimate as in the analysis with no MI procedure. So we found the same with the propensity score strata

method. The reason might be that it is not a problem for estimating propensity score if each patient missed only one or two covariate values. Thus, we assume that, in this study, the missing values on covariates are ignorable.

Chapter 4

Discussions

4.1 Key Findings

The effect estimates resulting from the five different statistical methods were quite different and are summarized in Appendix D, Table 3 and Table 4. For the primary outcome of mortality, the propensity score stratification analysis yielded the smallest odds ratio of 3.49 with its 95% confidence interval (1.91, 6.40), followed by the multiple variable logistic regression odds at 3.71 with its 95% confidence interval (2.10, 6.56). Adjusting propensity score as a latent variable in the Bayesian model yielded the estimated odds ratio of 3.82 with its 95% credible interval (1.23, 13.07), which is the widest confidence interval among the methods we used for conducting the study. The propensity score matching method yielded an odds ratio of 4.49 with its 95% confidence interval (1.92, 10.56). The propensity score weighted method yielded extreme odds ratio estimate of 4.97 with its 95% confidence interval (3.32, 6.11).

For mortality, we found that the five different methods to control for confounding yielded extremely different treatment effect estimates. Now we argue that the variation we observed in effect estimates cannot be ascribed to the small numbers of subjects in the low-propensity strata and the variability of the associated estimated odds ratios. Furthermore, we argue that this variation does not prove or even suggest that any one of the five methods is superior for controlling confounding.

If we do not consider unmeasured confounding, the propensity score strata method estimates the average treatment effect in a population whose distribution of risk factors is close. In stratum one, most of the patients were in the lower propensity score with a low associated risk of death and lower odds ratio, also; in the first three stratums, there were 65% of Isolated CABG patients, and they have a lower odds ratio. In stratum five, most of the patients were in the group with a higher associated risk of death. When counting the average of the odds ratio, it is not surprising that the propensity score method odds ratio was 3.49 with its 95% confidence interval (1.91, 6.40). In contrast, the propensity score weighted method estimates the average effect of treatment in the entire study population; that is, for patients who had Isolated CABG and Complex cardiac surgery. Each patient had an estimated propensity score, and when with re-weighted to the treatment subject and the control group, more uncertainty was counted in; therefore it is no surprise that the propensity score weighted method yielded extreme odds ratio of 4.97.

Similarly, it is no surprise that the propensity score matching estimate is 4.49. When the number of untreated subjects is many times larger than the number of treated subjects, as in the present example, the propensity score matching will typically result in all or nearly all treated patients being successfully matched, while many untreated patients will remain unmatched and be excluded from the analysis. As a result, the distribution of covariates in the (successfully) matched subpopulation will be close to that in the treated study population. In this study, the propensity-matched estimate is very close to the propensity score-weighted estimate. Although the propensity score-weighted and matched propensity analyses gave similar results in this particular data set, the propensity score-

weighted analysis has the theoretical advantages that 1) data from all patients are used, and 2) it is not affected by further uncontrolled confounding attributable to the inability to find an exact match for each treated subject [19].

However, the similarity of the results obtained with the propensity score-weighted and propensity score–matched analyses to the results of the observational study should not be taken as evidence that compared with other multivariable outcome models, these two methods are a better tool to adjust for covariates in observational research. In addition, in most studies in the literature, the effect estimates from multivariable regression models were quite close to the effect estimates derived from various implementations of the propensity score, as long as the number of outcome events was much larger than the number of potential confounders [20-22]. An apparent advantage in using the propensity score, however, may be that the strong effect modification in this clinical example is very obvious across propensity score matching. This effect modification may be difficult to unveil when evaluating individual risk factors.

For the Bayesian propensity score matching method, it is no surprise that the odds ratio has the widest 95% credible interval, since the Bayesian model included more uncertainty than the other models.

In this study, the estimated treatment effect from a multivariable logistic model that includes only the treatment indicator with adjustment for covariates, age, gender, body surface area, hypertension, smoke , congestive heart failure, diabetes, site, redo number and timing (urgency of surgery). It considered the whole sample but not deleted some sample with moderate odds ratio. Thus, it is no surprise that the odds ratio yielded from

multivariable logistic regression is somewhat lower than that of the propensity score matching method.

4.2 Comparison with Similar Studies

A literature search of relevant sources found some similar studies about comparing the outcome differences between isolated CABG and complex cardiac surgery.

Our clinical results of primary outcomes are comparable to similar studies. Our study reported the crude mortality rates of 1.6% for isolated CABG and 7.1% for complex cardiac surgeries. This is close to recent North American studies on mortality rates in isolated CABG and combined cardiac surgical procedures, which reported mortality rates of 4.3% to 7.1% in combined cardiac surgeries [23-26]. For the secondary outcome 'stroke', we reported the rate of 3.8% for complex cardiac surgery, which is lower than that observed in large analyses of postoperative stroke in combined cardiac procedures [27].

Moreover, about the clinical results, our study measured usage among the complex cardiac surgical population and reported an average of 5.2 days in the ICU following the complex cardiac surgery. This result is similar to recent findings by Gulbins et al, who reported an average LCU length of stay of 5.7 days in their cohort of 124 patients, aged 70-80, and undergoing combined CABG-AVR procedures [28].

In our study, we reported a treatment effect estimated by Bayesian method of widest credible interval, which agreed with previous studies [29]. The reason for this is because the Bayesian model accounted for more uncertainty than any of the other models.

Prior studies have compared the estimation of treatment effects using different sets of propensity score methods.

Ralph's study [30] compared five methods for evaluating the effect of tissue plasminogen activator on death among 6,269 ischemic stroke patients registered in a German stroke registry: multivariable logistic regression, propensity score–matched analysis, regression adjustment with the propensity score, and two propensity score–based weighted methods—one estimating the treatment effect in the entire study population (inverse probability-of-treatment weights), and another in the treated population (standardized-mortality-ratio weights). This study showed five different methods to control for confounding yielded different treatment effect estimates. Our study agreed with this result. The reason is that the different methods are effectively estimating the effect in different populations, with different distributions of covariates.

In the system review by Peter C. Austin [31] on propensity-score matching in the cardiovascular showed that analysis of propensity score–matched samples tended to be poor in the cardiovascular surgery literature. Most statistical analyses ignored the matched nature of the sample. Propensity score matching may require more analytic steps than competing propensity score methods. We do have the above limitation in our study since after propensity score matching process, the left over unmatched sample were ignored in the analysis.

In the study by Peter C. Austin [32] on comparing four propensity score methods for estimating the reduction in all-cause mortality due to statin therapy for patients hospitalized with acute myocardial infarction. The four propensity score methods are:

propensity-score matching, stratification using the propensity score, covariate adjustment using the propensity score, and weighting using the propensity score. The study demonstrated the breadth of propensity score methods and that these methods allow the estimation of adjusted as well as absolute and relative treatment effects.

Another study by Peter C. Austin [33] on the performance of different propensity score methods for estimating marginal odds ratios performed a series of Monte Carlo simulations to assess the performance of propensity score matching, stratifying on the propensity score, and covariate adjustment using the propensity score to estimate marginal odds ratios. They showed that matching on the propensity score resulted in the least biased estimates of marginal odds ratios, whereas stratifying on the quintiles of the propensity score resulted in the greatest degree of bias amongst the three different propensity score methods examined. See Table 7 for a summary of the comparisons of the results with other studies discussed above.

4.3 Some Limitations of the Study

Our study does have some limitations. The first is the retrospective design of the study. This makes our results susceptible to errors associated with selection biases, since data collection for the complete cohort of cardiac surgery patients at the seven study centers was unfeasible.

The other limitation is that this study is observational study, not randomized. As we know that in a randomized experiment, the randomization of patients to different treatments minimizes the chance of differences on observed or unobserved covariates [30]. However, in nonrandomized studies, systematic differences can exist between the

treatment group and the control group. This will yield a potentially biased estimate of treatment effect. A caution with regard to the use of weighted methods is that they can perform poorly when the weights for a few subjects are very large. Although some partial approximate fixes have been described [34, 35], there is no perfect solution to this problem. For multivariable regression analysis, we have limited ability to adjust for potential confounding variables in the subgroup analysis of complex patients alone, due to a limited number of events in this group. This makes it difficult to interpret the results of analysis done on this group, such as our finding that variability in mortality and ICU length stay among the study sites for these patients was limited, since this could be affected by confounding factors that were not adjusted for in the univariate analysis.

Furthermore, the different propensity score methods had different sample size. For the propensity score matching method in our study, patients of Isolated CABG have a score that is almost five times larger than that of complex cardiac subjects. The propensity score matching will typically result in all or nearly all treated patients being successfully matched, while many untreated patients will remain unmatched and be excluded from the analysis, which will generate a biased treatment effect estimate. In addition, the propensity score method only balances the observed confounders and could not balance the unobserved confounders, this will generate optional bias.

4.4 Implications of Clinical Results

Our study reported that the complex cardiac patients in our cohort of seven Canadian cardiac centers had a significantly longer length of stay in ICU, as well as a longer duration of mechanical ventilation compared to isolated CABG patients. We

reported an average stay of 5.2±10.8 days in the ICU following the complex cardiac surgery and an average stay of 2.5±6.8 days for isolated CABG. These findings provide the first analysis of resource usage specifically in the complex cardiac population, which has not been reported in studies of Canadian cardiac centers to date.

In addition, our study provides the first analysis of variability in primary outcome among Canadian cardiac centers (Table 5 in Appendix 5). Our data showed that there was limited variability in mortality for cardiac surgery between sites, when isolated CABG and complex procedures were analyzed collectively, and by complex procedures alone. These findings indicate a fairly consistent quality of care among the study sites, which has been the goal of recent efforts to improve outcomes in key health care services across Canada, particularly in cardiac surgery [36].

Chapter 5

Conclusions

For the primary outcome of the study, we provided retrospective evidence that patients undergoing complex cardiac surgery have a higher mortality rate than patients undergoing isolated CABG. As well, for the secondary outcome, our results showed that patients undergoing complex cardiac surgery had longer ICU length of stay in comparison with patients undergoing isolated CABG. In addition, this study provided a quantitative assessment of the risk of mortality, morbidities, and ICU length of stay in the complex cardiac surgery population in Canada, as well as the variability in outcomes and resource usage that exists among Canadian cardiac centers.

Among all the statistical methods we used in our study, Bayesian propensity score matching provided the widest 95% credible interval of treatment effect estimate, since the Bayesian analysis counted all kinds of variability and therefore yielded conservative evidences.

For the primary outcome 'mortality', all of the estimates have the valid interpretation that patients undergoing complex cardiac surgery had higher mortality rates than patients undergoing isolated CABG to the given population. However, they are valid only if the assumptions of the propensity score methodology are satisfied; that is, if there are no unmeasured confounders, the propensity score model is correctly specified, the study size

is large enough to make the asymptotically unbiased estimator in fact unbiased, and the standard error is reliable.

If the effect of treatment varies between individuals, different propensity based methods of balancing covariates may give different answers in a given population. Each estimate may only reflect a parameter of interest in that population. However, none of the estimates will reflect the effect of treatment in a different population. It is therefore necessary by using propensity-based methods to test whether the treatment effect varies between individuals.

We found that five different methods to control for confounding yielded extremely different treatment effect estimates. However, there is no evidence to show that any one of the five methods is superior for controlling confounding. There are few methodological studies to count for determining which method is better. Simulation study including all of the statistical methods under all kinds of situations is required to determine which method is the best.

References

1. Ghali WA, Quan H, Shrive FM, Hirsch GM. Canadian Cardiovasculare Outcomes Research Team. Outcomes after coronary artery bypass surgery in Canada: 1992/93 to 2000/01. *Can J Cardiol* 2003; 19 (7): 774-81.

2. Gulbins H, Malkoc A, Ennker J. Combined cardiac surgical procedures in octogenarians: operative outcome. *Clin Res Cardiol* 2008; 97 (3): 176-80.

3. Aggarwal SK, Fox SA, Stitt L, Kiaii B, McKenzie FN, Menkis AH, et al. The new cardiac surgery patient: defying previous expectations. *Can J Surg* 2006; 49 (2): 117-22.

4. Hassan A, Newman AM, Gong Y, Kennedy C, Humphries KH, Ghali WA, et al. Use of valve surgery in Canada. *Can J Cardiol* 2004; 20 (2): 149-54.

5. Rao V, Christakis GT, Weise RD, Buth KJ, Ikonomidis JS, Shirai T, et al. Changing pattern of valve surgery. *Circulation* 1996; 94 (9 suppl): II 113-20.

6. Novick RJ, Fox SA, Stitt LW, Butter R, Kroh M, Hurlock-Chorostecki C, Cheng DC. Impact of the opening of a specialized cardiac surgery recovery unit on postoperative outcomes in an academic health sciences center. *Can J Anesth* 2007; 54 (9): 737-43.

7. Eddleton JM, White P, Guthrie E. Survival, morbidity, and quality of life after discharge from intensive care. *Crit Care Med*; 28 (7): 2293-9.

8. Hassan A, Quan H, Newman A, Ghali WA, Hirsh GM. Canadian cardiovascular outcomes Research Team. Outcomes after aortic and mitral valve replacement surgery in Canada: 1994/95 to 1999/2000. *Can J Cardio* 2004; 20 (2): 155-63.

9. Ferraris VA, Ferraris SP, Singh A. Operative outcome and hospital cost. *J Thorac Cardiovasc Surg* 1998; 115(3): 593-602.

10 Karkouti K, Wijeysundera DN, Beattie WS, Callum JL, Cheng D, Dupuis JY, et al. Variability and predictability of large-volume red blood cell transfusion in cardiac surgery: a multicenter study. *Transfusion* 2007; 47 (11): 2081-8.

11. Rosenbaum PR, Ribin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983 70 (1): 41-55.

12. Rosenbaum PR, Ribin DB. Reducing bias in observational studies using subclassification on the propensity score. *J Am Stat Assoc* 1984; 79 (387): 516-524.

13. Rosenbaum PR, Ribin DB. Constuucting a control group using multivariate matched sampling methods that incoporate the propensity score. *Am Stat* 1985; 39 (1): 33-38.

14. Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology.* 2000; 11 (5): 550-560.

15. Ribin DB. The use of propensity scores in applied Bayesian inference. In *Bayesian Statistics 2,* Bernardo JM, De Groot MH, Lindley DV, Smith AFM (eds). Valecia University Press, North- Holland: Amsterdam, 1985: 63-72.

16. Little RJA. To model or not to model? Competing modes of inference for finite population sampling. *Journal of the American Statistical Association* 2004; 291: 1864-1870.

17. Sung L, Hayden J, Greenberg ML, Koren G, Feldman BM, Tomlinson GA. Seven items were identified for inclusion when reporting a Bayesian analysis of a clinical study. *Journal of Clinical Epidemiology* 2005; 58: 261-268.

18. Gelman, Andrew, John B. Carlin, Hal s. Stern, and Donald B. Rubin. 2004. *Bayesian Data Analysis.* Chapman & Hall/CRC, second edition.

19. Rosenbaum PR, Rubin DB. The bias due to incomplete matching. *Biometrics* 1985; 41:103–16.

20. Cepeda MS, Boston R, Farrar JT, et al. Comparison of logistic regression versus propensity score when the number of events is low and there are multiple confounders. *Am J Epidemiol* 2003;158:280–7.

21. Braitman LE, Rosenbaum PR. Rare outcomes, common treatments: analytic strategies using propensity scores. *Ann Intern Med* 2002;137:693–5.

22. Shah BR, Laupacis A, Hux JE, et al. Propensity score methods gave similar results to traditional regression modeling in observational studies: a systematic review. *J Clin Epidemiol* 2005;58:550–9.

23. Hassan A, Quan H, Newman A, Ghali WA, Hirsch GM; Canadian Cardiovascular Outcomes Research Team. Outcomes after aortic and mitral valve replacement surgery in Canada: 1994/95 to 1999/2000. *Can J Cadiol* 2004; 20 (2): 155-63.

24. Hannan EL, Racz MJ, Jones RH, Gold JP, Ryan TJ, Hafner JP, *et al.* Predictors of mortality for patients undergoing cardiac valve replacements in New York State. *Ann Thorac surg* 2000; 70: 1212-8.

25. Birkmeyer NJ, Marrin CA, Morton JR, Leavitt BJ, Lahey SJ, Charlesworth DC, et al. Decreasing mortality for aortic and mitral valve surgery in Northern New England. *Ann Thorac Surg* 2000; 70: 432-7.

26. Edwards FH, Peterson ED, Coombs LP, Delong ER, Jamieson WR, Shroyer ALW, et al. Prediction of operation mortality after valve replacement surgery. *J Am Coll Cardiol* 2001; 37: 885-92.

27. Bucerius J, Gummert JF, Borger MA, Walter T, Doll N, Onnasch JF, Metz S, Falk V, Mohr FW. Stroke after cardiac surgery: a risk factor analysis of 16,184 consecutive adult patients. *Ann Thorac Surg*. 2003 Feb;75 (2):472-8.

28. Gulbins H, Malkoc A, Ennker J. Combined cardiac surgical procedures in octogenarians: operative outcome. *Clins Res Cardiol* 2008; 97 (3): 176-80.

29. J. Ma. Comparing Bayesian and Classical Methods in the Analysis of a Cluster Randomized Trial.

30. Ralph B. Propensity Scores in Cardiovascular Research American Heart association 2007; 115:2340-2343.

31. Austin PC. Propensity-score matching in the cardiovascularsurgery literature from 2004 to 2006: a systematic review and suggestions for improvement. *J Thorac Cardiovasc Surg*. 2007; 134: 1128–35.

32. Austin PC, Mamdani MM. A comparison of propensity score methods: a case-study estimating the effectiveness of post-AMI statin use. *Stat Med*. 2006;25:2084–106.

33. Austin PC. The performance of different propensity score methods for estimating marginal odds ratios. *Stat Med*. 2007; 26: 3078–94.

34. Scharfstein DO, Rotnitzky A, Robins JM. Adjusting for nonignorable drop-out using semiparametric non-response models. *J Am Stat Assoc* 1999; 94:1096–120.

35. Potter FJ. The effect of weight trimming on nonlinear survey estimates. In: Proceedings of the Section on Survey Research Methods of the American Statistical Association, San Francisco, California, 1993. *J Am Stat Assoc* 1993:758–63.

36. Mathew J, Fox-Robichaud A, Cheng J, Thabane L, Whitlock R, Sawchuk C. Comparison of outcomes between coronary artery bypass grafting and complex cardiac surgery: a multicenter study.

Appendix D

Tables

Table1: Groups Comparison Demographic and Preoperative Variables before Matching

Variables		**Isolated CABG (n=2271)**	**Complex (n=476)**	**P - value**
Age (years)	Mean	65.19	68.66	<0.0001
	SD	10.02	11.52	
Weight (kg)	Mean	82.20	79.95	<0.0001
	SD	16.48	18.04	
Height (cm)	Mean	169.09	167.09	<0.0001
	SD	11.77	12.92	
Body Surface Area (m^2)	Mean	1.94	1.89	<0.0001
	SD	0.23	0.25	
Gender (%)	Male	1820 (80.25)	320 (67.23)	<0.0001
	female	448 (19.75)	156 (32.77)	
Left Ventricle Grade (%)	>=60%	50 (2.35)	19 (4.18)	<0.0001
	40-59%	306 (14.39)	60 (13.19)	
	20-39%	755 (35.50)	112 (24.62)	
	<20%	1016 (47.77)	264 (58.02)	
Smoke (%)	Non	828 (36.65)	175 (36.84)	<0.0001
	Stopped	420 (18.59)	79 (16.63)	
	Smoker	1011 (44.75)	221 (46.53)	
Myocardial infarction (%)	Yes	1155 (51.38)	132 (27.85)	0.0046
	No	1093 (48.62)	342 (72.15)	
Angina (%)	None	244 (10.77)	242 (50.84)	<0.0001
	Stable	869 (38.35)	142 (29.83)	
	Unstable	1153 (50.88)	92 (19.33)	
Diabetes (%)	Type I	153 (6.78)	15 (3.18)	<0.0001
	Type II	627 (27.78)	109 (23.09)	
Hypertension (%)	Yes	1630 (71.96)	296 (62.45)	<0.0001
	No	635 (28.04)	178 (37.75)	
Hypercholesterolemia (%)	Yes	1810 (79.98)	274 (58.17)	<0.0001
	No	453 (20.02)	197 (41.83)	
Chronic Obstructive Pulmonary Disorder (%)	Yes	227 (10.07)	63 (13.32)	<0.0001
	No	2028 (89.93)	410 (86.68)	
Cerebrovascular disease (%)	Yes	213 (9.45)	66 (13.89)	<0.0001
	No	2041 (90.55)	409 (86.11)	
Peripheral Vascular Disease (%)	Yes	273 (12.06)	58 (12.18)	<0.0001
	No	1990 (87.94)	418 (87.82)	
Atrial Fibrillation (%)	Yes	111 (4.90)	110 (23.11)	<0.0001

	No	2152 (95.10)	399 (76.89)	
Congestive Heart Failure (%)	Yes	289 (12.79)	212 (44.54)	<0.0001
	No	1971 (87.21)	264 (55.46)	
Shock (%)	Yes	8 (0.35)	5 (1.05)	<0.0001
	No	2251 (99.65)	471 (98.95)	
Dialysis (%)	Yes	21 (0.93)	14 (2.95)	<0.0001
	No	2242 (99.07)	461 (97.05)	
Heparin Use (%)	Yes	405 (17.88)	60 (12.63)	<0.0001
	No	1860 (82.12)	415 (87.37)	
Acetylsalicylic Acid (%)	Yes	1193 (52.83)	157 (33.12)	<0.0001
	No	1065 (47.17)	317 (66.88)	

CABG= Coronary artery bypass grafting
SD= Standard deviation

Table 2: Groups Comparison of Preoperative Variables after Matching

Variables		**Isolated CABG (n=432)**	**Complex (n=432)**	**P - value**
Age (years)	Mean	69.19	68.66	0.4445
	SD	8.47	11.65	
Weight (kg)	Mean	79.47	80.64	0.3059
	SD	15.29	18.06	
Height (cm)	Mean	167.13	167.34	0.7926
	SD	9.87	13.22	
Body Surface Area (m^2)	Mean	1.89	1.90	0.6735
	SD	0.22	0.25	
Gender (%)	Male	325 (75.23)	296 (68.52)	0.0282
	female	107 (24.77)	136 (31.48)	
Left Ventricle Grade (%)	>=60%	14 (3.24)	18 (4.17)	0.2183
	40-59%	76 (25.69)	56 (12.96)	
	20-39%	111 (17.599)	107 (24.77)	
	<20%	231 (53.47)	251 (158.10)	
Smoke (%)	Non	115 (26.62)	128 (29.63)	0.5442
	Stopped	105 (24.31)	95 (21.99)	
	Smoker	212 (49.07)	209 (48.3)	
Myocardial infarction (%)	Yes	139 (32.18)	120 (27.78)	0.1583
	No	293 (67.82)	312 (72.22)	
Angina (%)	None	212 (49.07)	226 (52.31)	0.3026
	Stable	146 (33.80)	125 (28.94)	
	Unstable	74 (17.13)	81 (18.75)	
Diabetes (%)	Type I	27 (6.25)	15 (3.47)	0.0316
	Type II	124 (28.70)	104 (24.07)	
Hypertension (%)	Yes	304 (70.37)	273 (63.19)	0.0251
	No	128 (29.63)	159 (36.81)	
Hypercholesterolemia (%)	Yes	248 (57.41)	251 (58.10)	0.8363
	No	184 (42.59)	181 (41.90)	
Chronic Obstructive Pulmonary Disorder (%)	Yes	52 (12.04)	57 (13.19)	0.6084
	No	380 (87.96)	375 (86.81)	
Cerebrovascular disease (%)	Yes	122 (28.24)	63 (14.58)	<0.0001
	No	310 (71.76)	369 (85.42)	
Peripheral Vascular Disease (%)	Yes	60 (13.89)	53 (12.27)	0.4800
	No	372 (86.11)	379 (87.73)	
Atrial Fibrillation (%)	Yes	95 (21.99)	103 (23.84)	0.5173
	No	337 (78.01)	329 (76.16)	

Congestive Heart Failure (%)	Yes	172 (39.81)	197 (45.60)	0.0855
	No	260 (60.19)	235(54.40)	
Shock (%)	Yes	1 (0.23)	5 (1.16)	0.1013
	No	431 (99.77)	427 (98.84)	
Dialysis (%)	Yes	13 (3.01)	12 (2.78)	0.8392
	No	419 (96.99)	420 (97.22)	
Heparin Use (%)	Yes	56 (12.96)	54 (12.50)	0.8383
	No	376 (87.04)	378 (87.50)	
Acetylsalicylic Acid (%)	Yes	149 (34.49)	143 (33.10)	0.6661
	No	283 (65.51)	289 (66.90)	

CABG= Coronary artery bypass grafting
SD= Standard deviation

Table 3: Comparison of Different Methods of Mortality.

Outcomes	Methods	Odds Ratio	95% Confidence or Credible Limits	P-value
Death	Propensity match	4.49	(1.92 10.56)	0.0006
	Propensity weight	4.97	(3.32 6.11)	<0.0001
	Propensity strata	3.49	(1.91 6.40)	0.0001
	Multiple regression	3.71	(2.10 6.56)	<0.0001
	Bayesian propensity score	3.82	(1.23 13.07)	

Table 4: Summaries of Second Outcomes Differences by Using Different Methods.

Outcomes	Methods	Odds Ratio	95% Wald Confidence Limits	P-value
Stroke	Propensity match	7.15	(2.01 25.74)	0.0014
	Propensity weight	7.51	(2.78 15.10)	<0.0001
	Propensity strata	6.48	(2.45 12.49)	0.0001
	Multiple regression	3.94	(1.85 8.41)	<0.0001
	Bayesian propensity score			
Renal failure	Propensity match	0.93	(0.54 1.62)	0.2314
	Propensity weight	3.48	(2.63 4.60)	<0.0001
	Propensity strata	1.72	(0.97 3.07)	0.0701
	Multiple regression	2.21	(1.24 3.95)	<0.0001
	Bayesian propensity score			
Pneumonia	Propensity match	1.61	(0.91 2.86)	0.0601
	Propensity weight	1.04	(0.79 1.36)	0.0525
	Propensity strata	2.22	(1.31 3.75)	0.0700
	Multiple regression	2.22	(1.35 3.65)	0.0005
	Bayesian propensity			
Pulmonary Embolism	Propensity match	3.28	(0.64 16.89)	0.1398
	Propensity weight	3.34	(1.66 6.72)	<0.0001
	Propensity strata	3.55	(0.99 12.60)	0.2310
	Multiple regression	2.26	(0.69 7.36)	0.0706
	Bayesian propensity			
Sepsis	Propensity match	2.41	(0.42 13.88)	0.6027
	Propensity weight	2.96	(0.98 6.10)	0.0505
	Propensity strata	2.74	(0.48 15.71)	0.1670
	Multiple regression	2.71	(0.44 16.70)	0.0740
	Bayesian propensity	2.23	(0.36 17.94)	
SternalDebrid	Propensity match	1.49	(0.68 3.30)	0.7902
	Propensity weight	2.19	(1.57 3.04)	<0.0001

	Propensity strata	1.87	(0.92 3.78)	0.0700
	Multiple regression	2.48	(1.29 4.72)	0.0007
	Bayesian propensity score			
Re-intubation	Propensity match	1.45	(0.89 2.38)	0.0616
	Propensity weight	3.30	(2.68 4.07)	<0.0001
	Propensity strata	2.47	(1.53 3.99)	0.0020
	Multiple regression	2.88	(1.85 4.50)	<0.0001
	Bayesian propensity score	1.26	(0.73 2.82)	
Tracheotomy	Propensity match	3.84	(1.51 9.77)	0.0024
	Propensity weight	3.92	(1.97 5.41)	<0.0001
	Propensity strata	3.74	(1.61 8.71)	0.0020
	Multiple regression	3.71	(1.86 7.39)	<0.0001
	Bayesian propensity score	3.78	(0.94 11.24)	
Re-Exploration	Propensity match	2.44	(1.48 4.02)	0.0036
	Propensity weight	3.11	(2.50 3.87)	<0.0001
	Propensity strata	2.86	(1.87 4.37)	0.0001
	Multiple regression	3.54	(2.38 5.28)	<0.0001
	Bayesian propensity			
ICU Length of stay	Propensity match	1.24	(0.96 1.59)	0.2034
	Propensity weight	2.23	(2.01 2.49)	<0.0001
	Propensity strata	2.13	(1.70 2.65)	<0.0001
	Multiple regression	1.81	(1.47 2.23)	<0.0001
	Bayesian propensity			
Duration of Ventilation	Propensity match	2.02	(1.59 2.56)	<0.0001
	Propensity weight	2.80	(2.54 3.09)	<0.0001
	Propensity strata	2.46	(2.01 3.03)	<0.0001
	Multiple regression	2.12	(1.74 2.57)	<0.0001
	Bayesian propensity			

Table 5: Mortality Comparison between Different Centers. Site 7 is used as reference.

Mortality	Odds Ratio	95% Wald Confidence Limits	P-value
SiteID 1 vs 7	2.41	(0.82 7.09)	0.4111
SiteID 2 vs 7	2.36	(0.83 6.75)	0.4268
SiteID 3 vs 7	0.96	(0.25 3.72)	0.1356
SiteID 4 vs 7	2.17	(0.75 6.28)	0.6327
SiteID 5 vs 7	2.64	(0.92 7.56)	0.2299
SiteID 6 vs 7	2.79	(0.97 8.02)	0.1631

- OR is based on the scale of one unit of change

Table 6: Bayesian Result of Different Priors

Outcome	Prior Dist	Odds ratio	95% CI
Mortality	IGamma (5, 5)	3.82	(1.12 13.07)
	IGamma (10, 10)	3.79	(1.12 13.07)
	IGamma (15, 15)	3.82	(1.12 13.07)
Pulmonary Embolism	IGamma (5, 5)	3.11	(0.59 18.89)
	IGamma (10, 10)	3.11	(0.59 18.89)
	IGamma (15, 15)	3.11	(0.59 18.89)
Sepsis	IGamma (5, 5)	2.23	(0.36 17.94)
	IGamma (10, 10)	2.21	(0.36 17.69)
	IGamma (15, 15)	2.24	(0.36 17.67)
Tracheotomy	IGamma (5, 5)	3.78	(0.94 11.24)
	IGamma (10, 10)	3.78	(0.94 11.24)
	IGamma (15, 15)	3.78	(0.94 11.24)

Table 7: Summary of the Comparison of the Results with Other Studies

Methods	My study	Ralph's study *	Austin's study **	Austin's study ***
Propensity match (I)	√	√	√	√
Propensity weight (II)	√		√	
Propensity strata (III)	√		√	√
Multivariable Regression (IV)	√	√	√	
Bayesian propensity(V)	√			
Type of study	Empirical	Empirical	Empirical	Simulation
Conclusions	-II has narrowest interval -V has widest interval -Effect differ but the conclusions remain robust	-Estimate vary -Conclusions remain robust	-Effect differ but the conclusions remain robust -II has narrowest interval	-I yields least bias estimate -III yields greatest degree of bias

*= Reference number 30

**= Reference number 32

***= Reference number 33

Appendix E

Figures

Figure 1. Scheme of Study Analysis

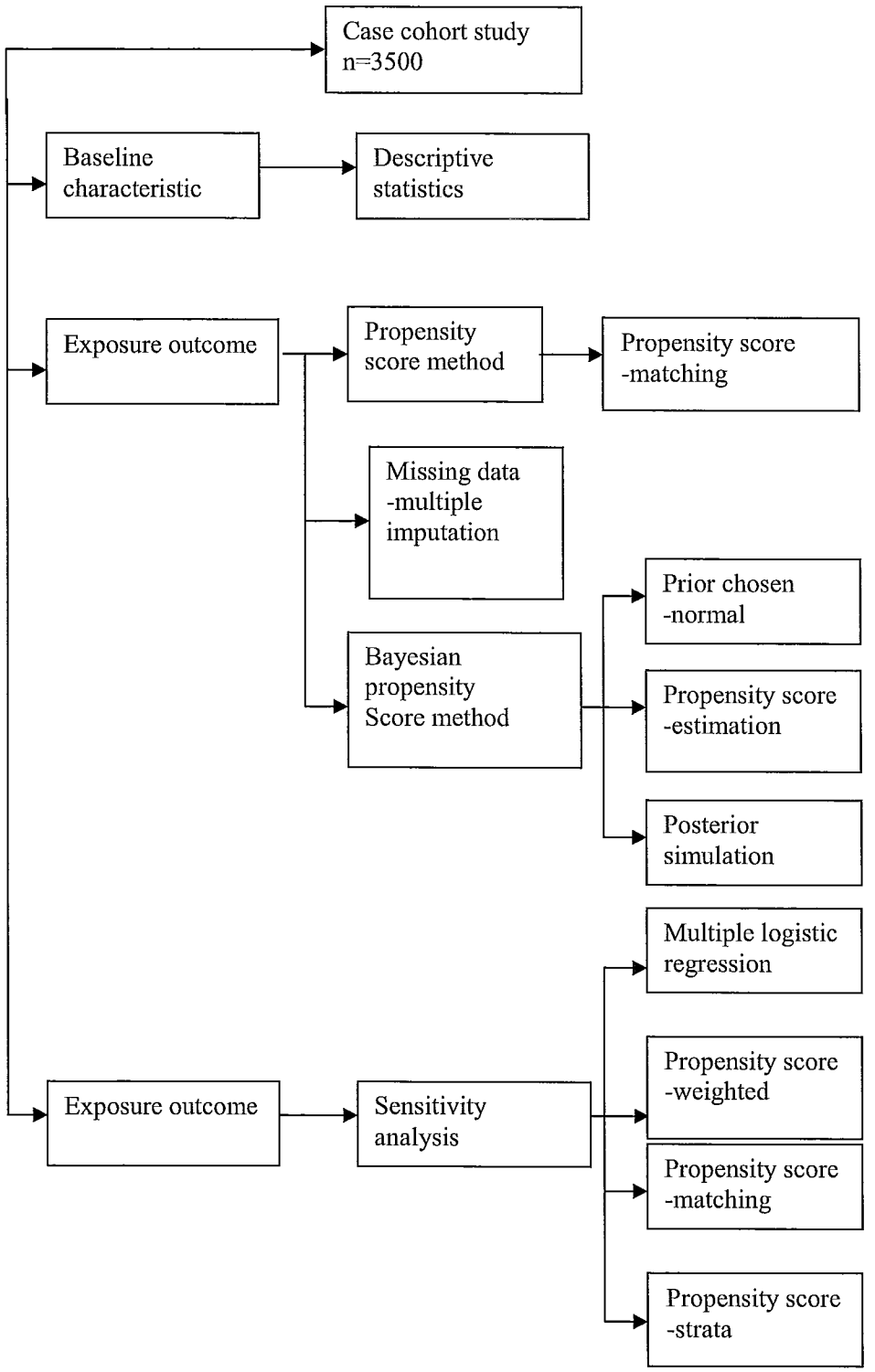
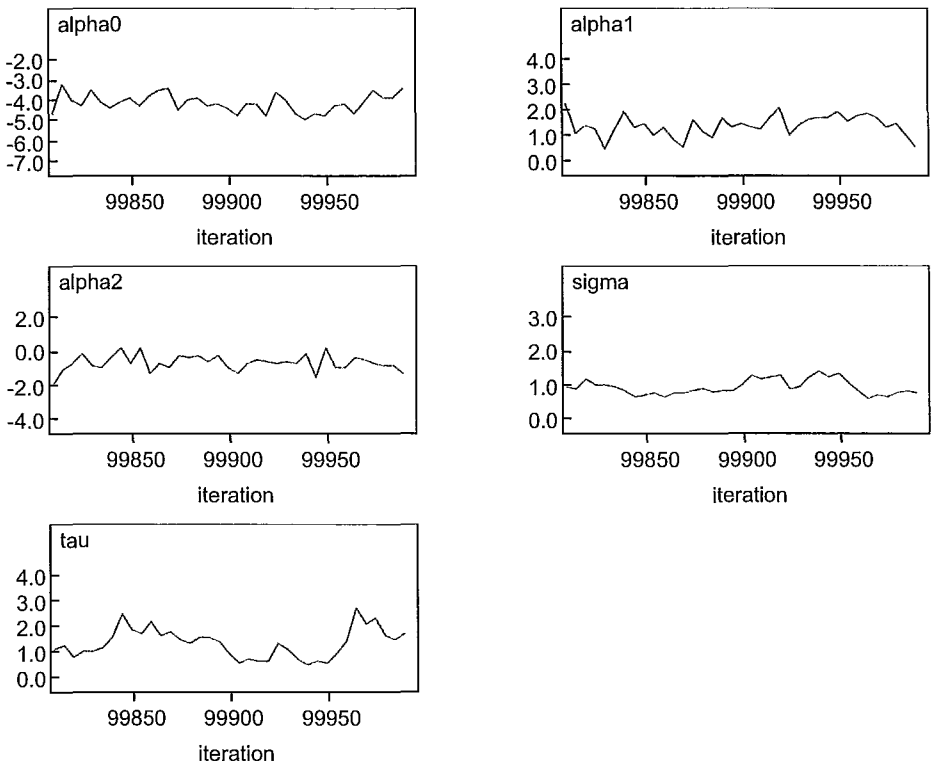
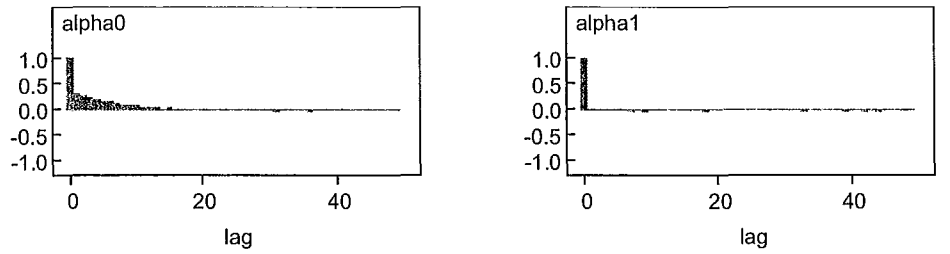

Case cohort study
n=3500
Baseline
characteristic
Descriptive
statistics
Exposure outcome
Propensity
score method
Propensity score
-matching
Missing data
-multiple
imputation
Bayesian
propensity
Score method
Prior chosen
-normal
Propensity score
-estimation
Posterior
simulation
Exposure outcome
Sensitivity
analysis
Multiple logistic
regression
Propensity score
-weighted
Propensity score
-matching
Propensity score
-strata

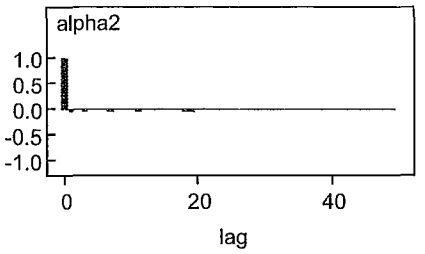
Figure 2 Diagnosis Plot for Bayesian Analysis -- Mortality

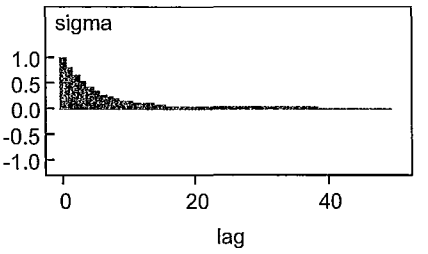
Dynamic Trace

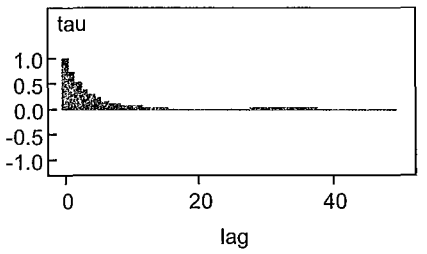


Autocorrelation function

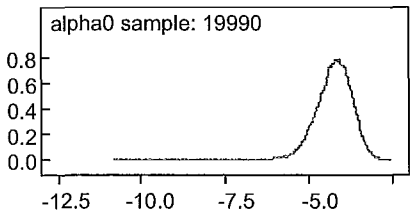


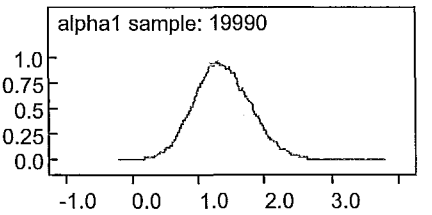


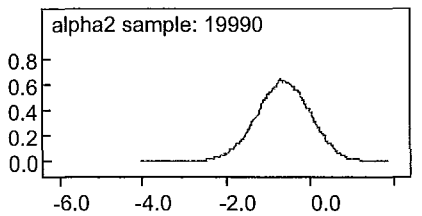


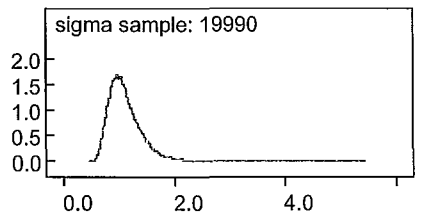


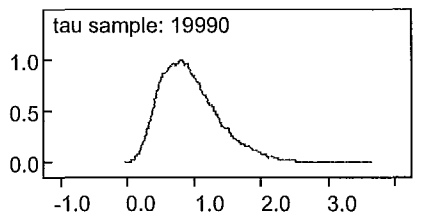
Kernel Denisty



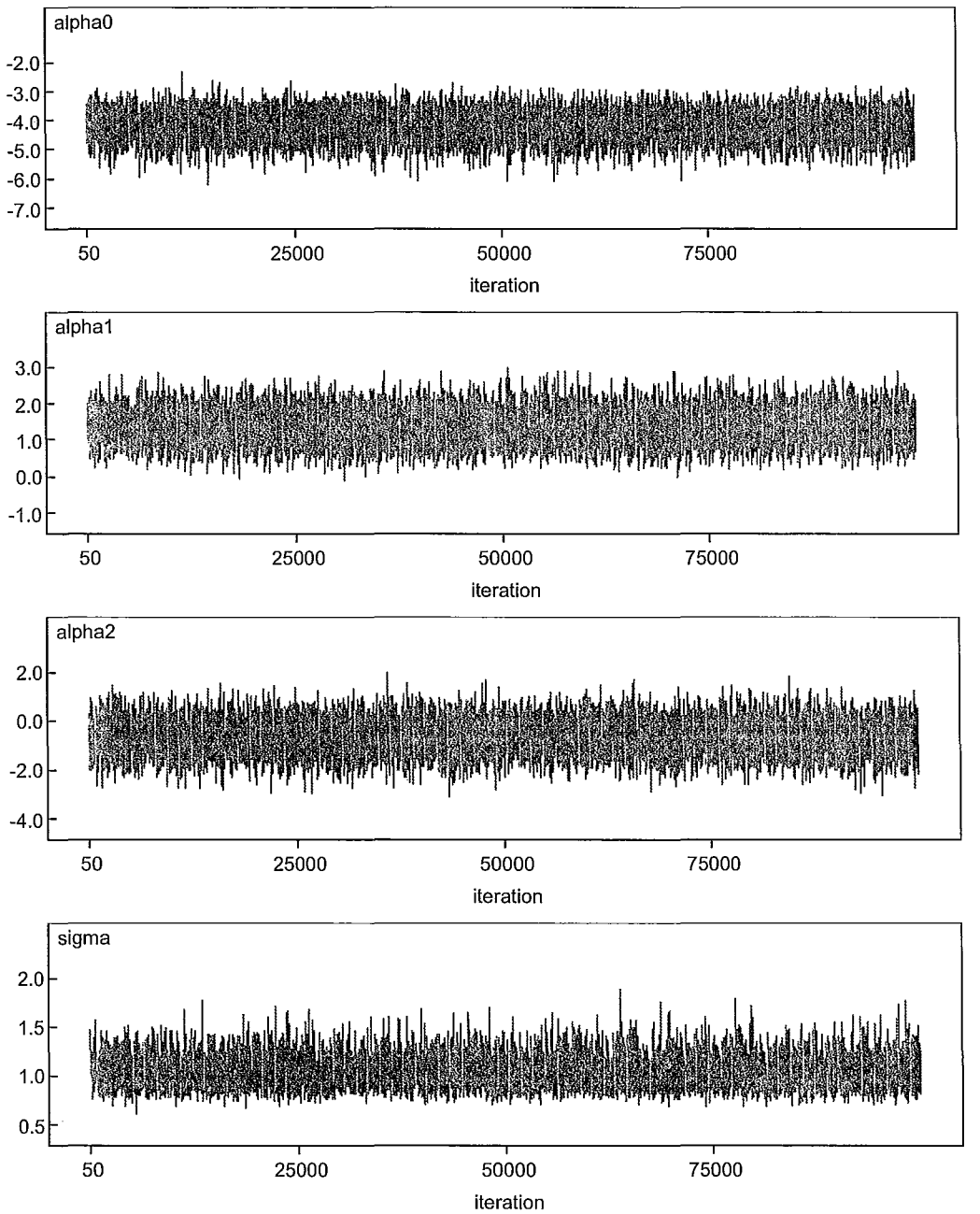








Time series


alpha0
-2.0
-3.0
-4.0
-5.0
-6.0
-7.0
50
25000
50000
75000
iteration
alpha1
3.0
2.0
1.0
0.0
-1.0
50
25000
50000
75000
iteration
alpha2
2.0
0.0
-2.0
-4.0
50
25000
50000
75000
iteration
sigma
2.0
1.5
1.0
0.5
50
25000
50000
75000
iteration

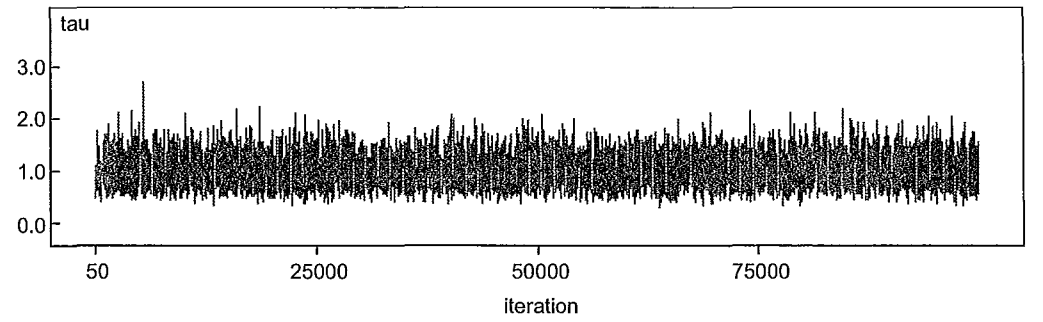
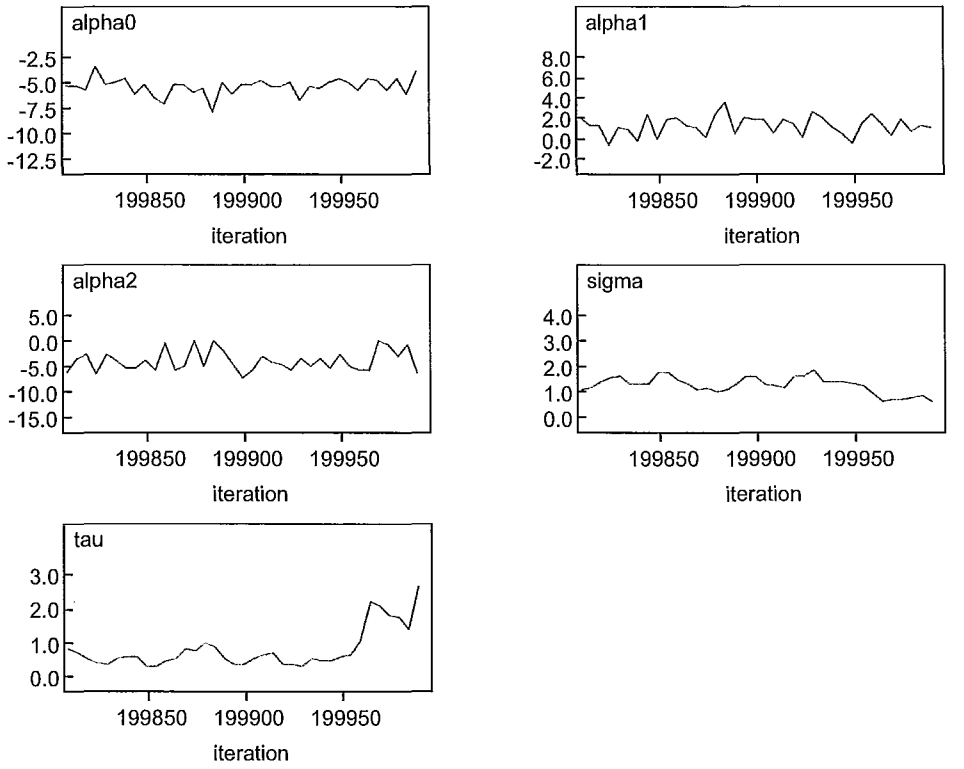
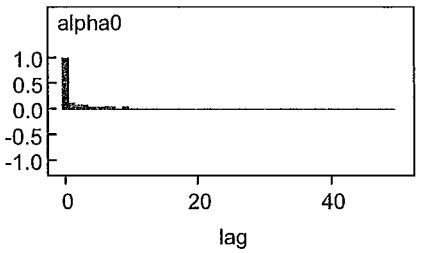


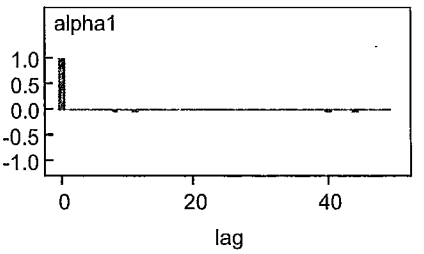
Figure 3. Diagnostic Plot for Bayesian Analysis-- Pulmonary Embolism

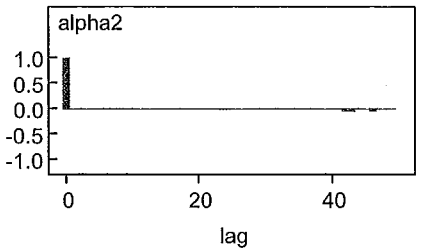
Dynamic Trace

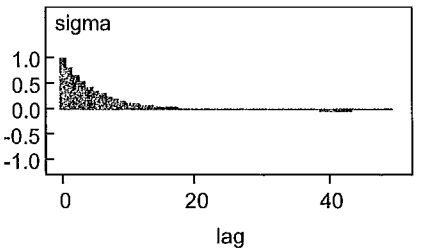


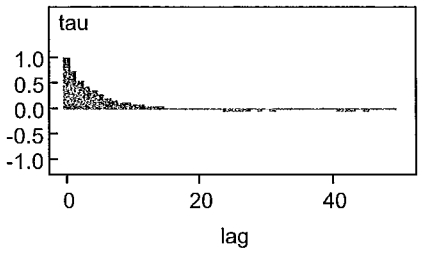
Autocorrelation function



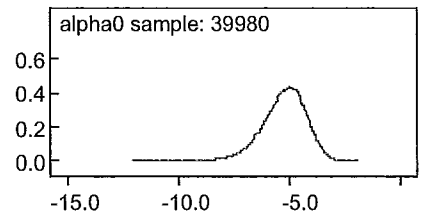


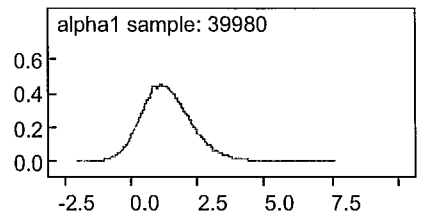


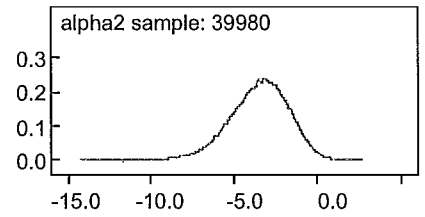


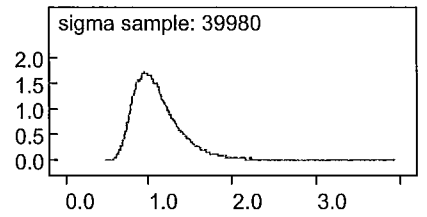


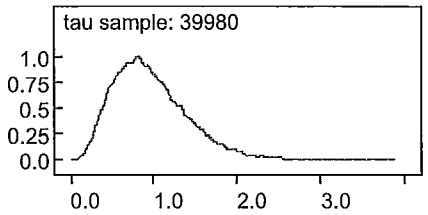
Kernel Denisty



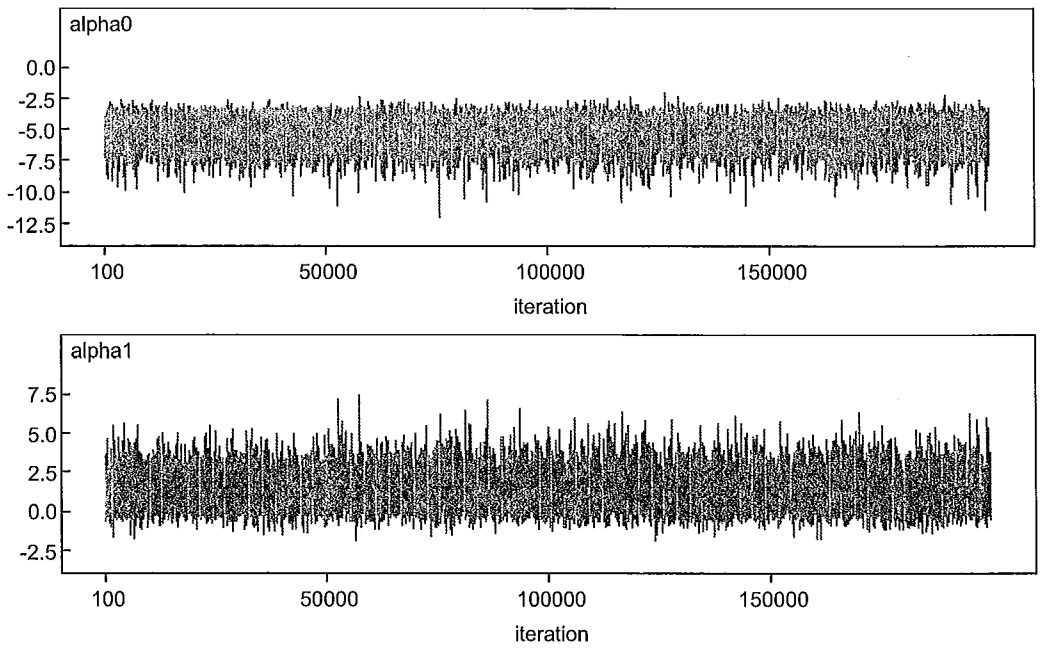







tau sample: 39980
1.0
0.75
0.5
0.25
0.0
0.0
1.0
2.0
3.0

Time series


alpha0
0.0
-2.5
-5.0
-7.5
-10.0
-12.5
100
50000
100000
150000
iteration
alpha1
7.5
5.0
2.5
0.0
-2.5
100
50000
100000
150000
iteration

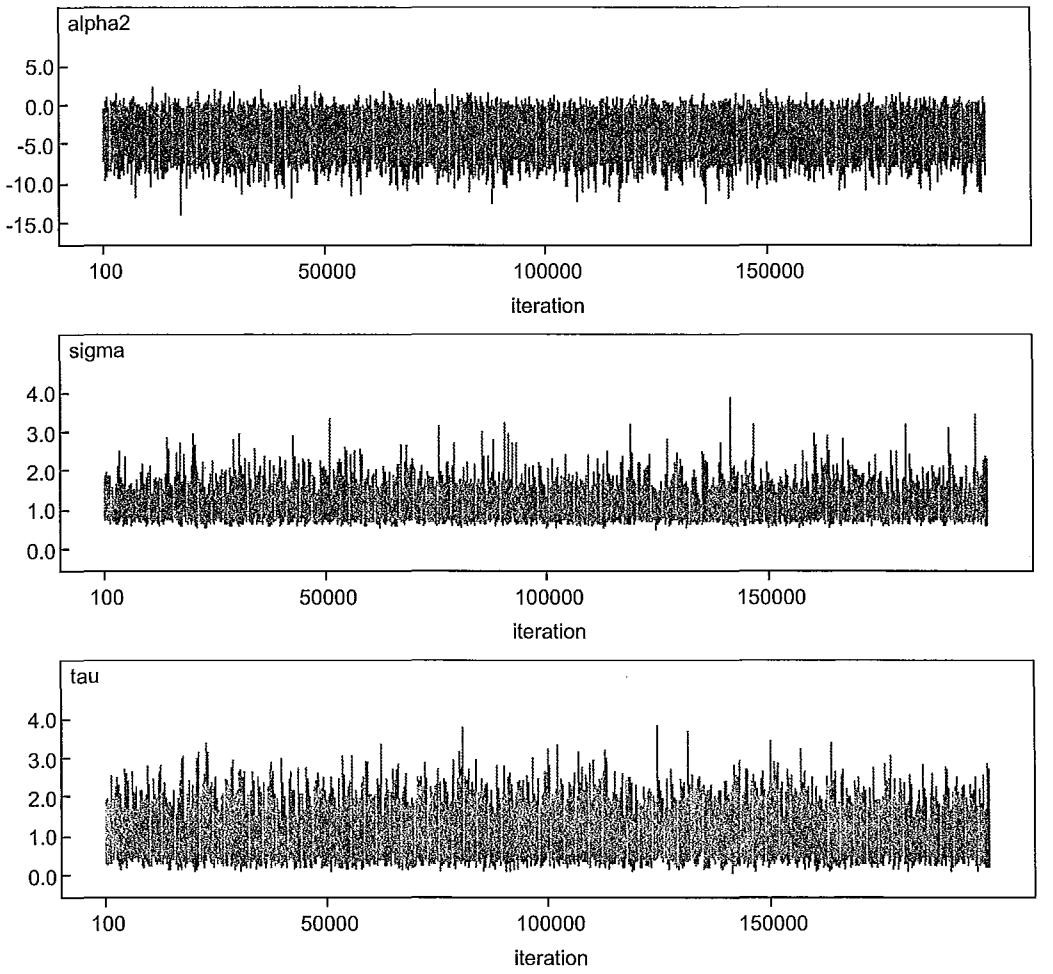
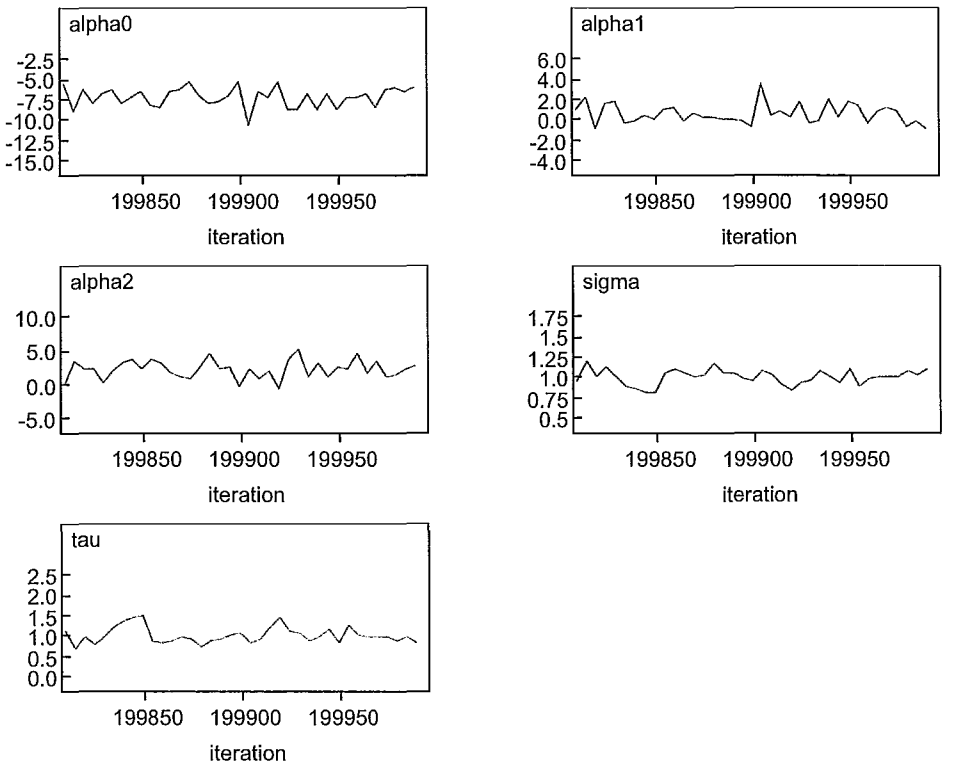
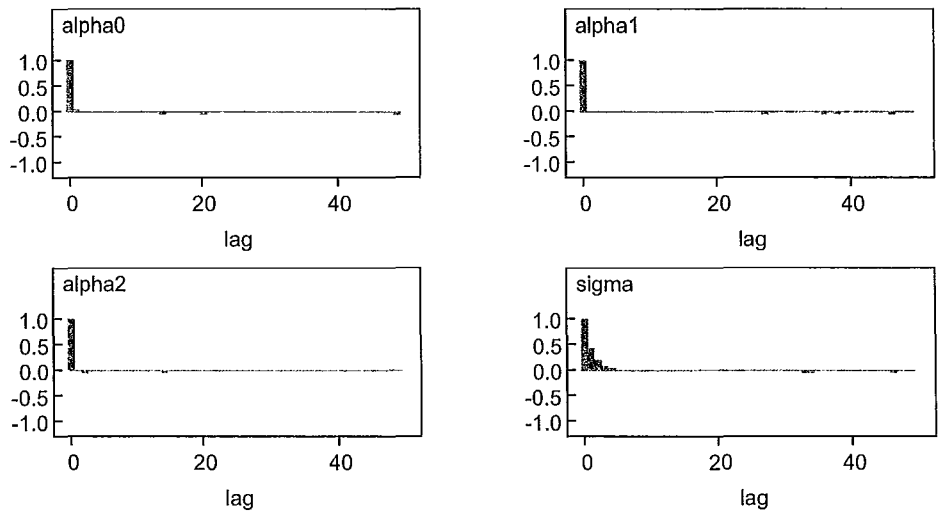

alpha2
5.0
0.0
-5.0
-10.0
-15.0
100
50000
100000
150000
iteration
sigma
4.0
3.0
2.0
1.0
0.0
100
50000
100000
150000
iteration
tau
4.0
3.0
2.0
1.0
0.0
100
50000
100000
150000
iteration

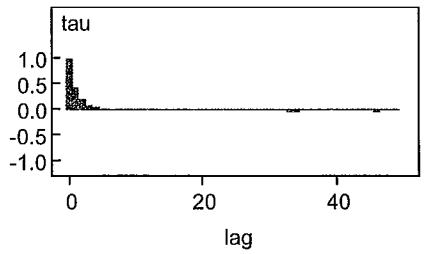
Figure 4. Diagnostic Plot for Bayesian Analysis-- Sepsis

Dynamic Trace

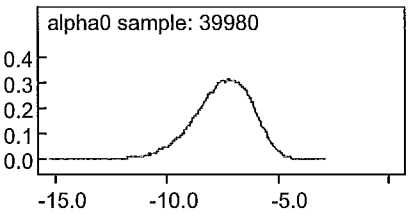


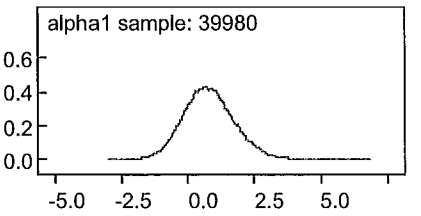
Autocorrelation function

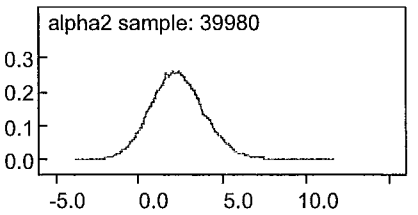


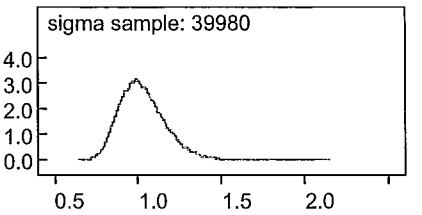


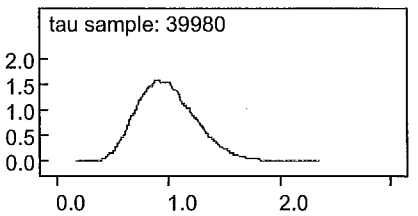
Kernel Denisty



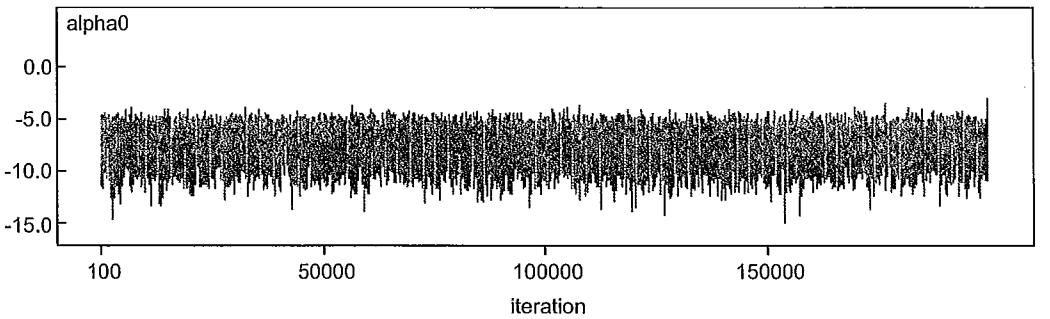








Time series



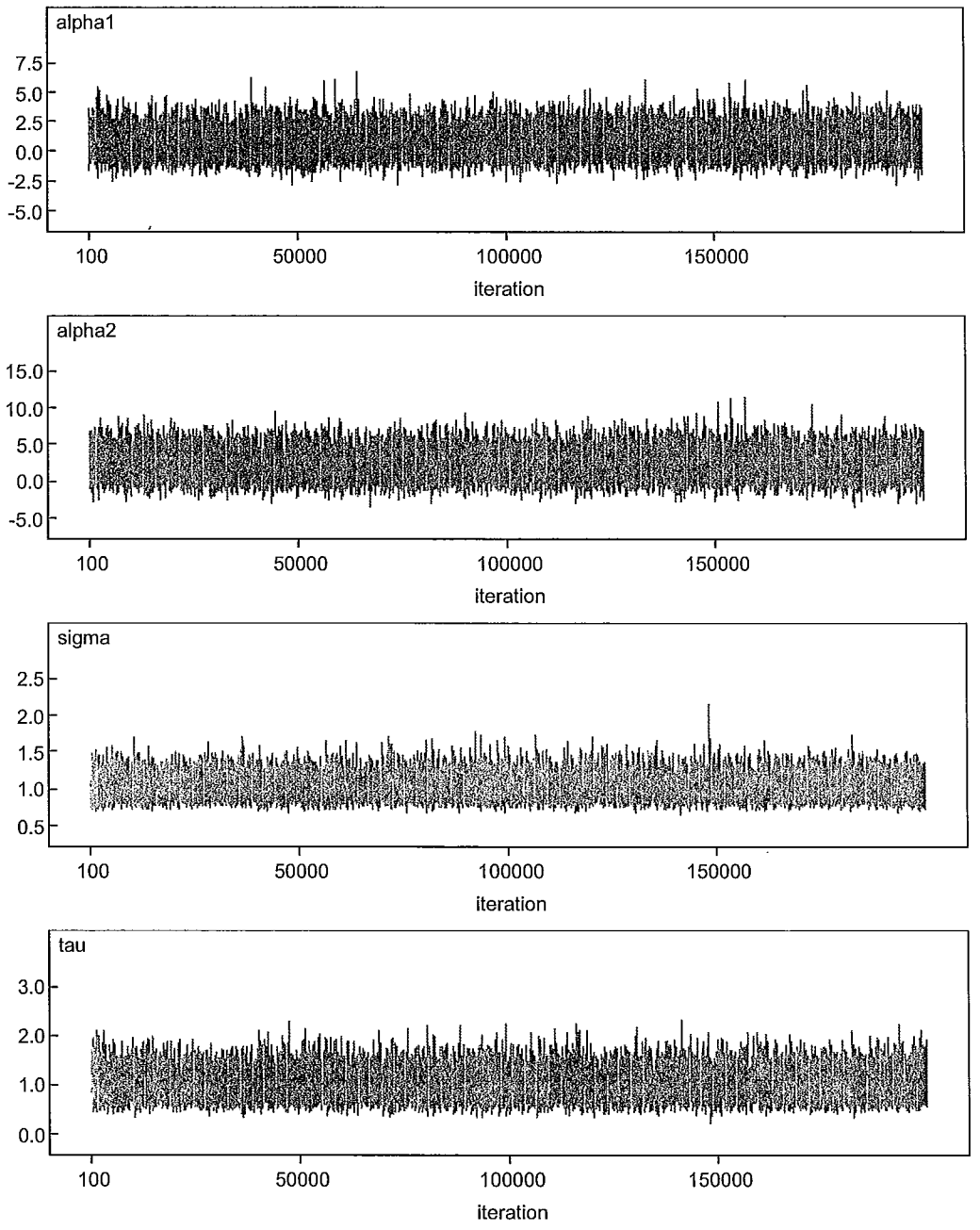
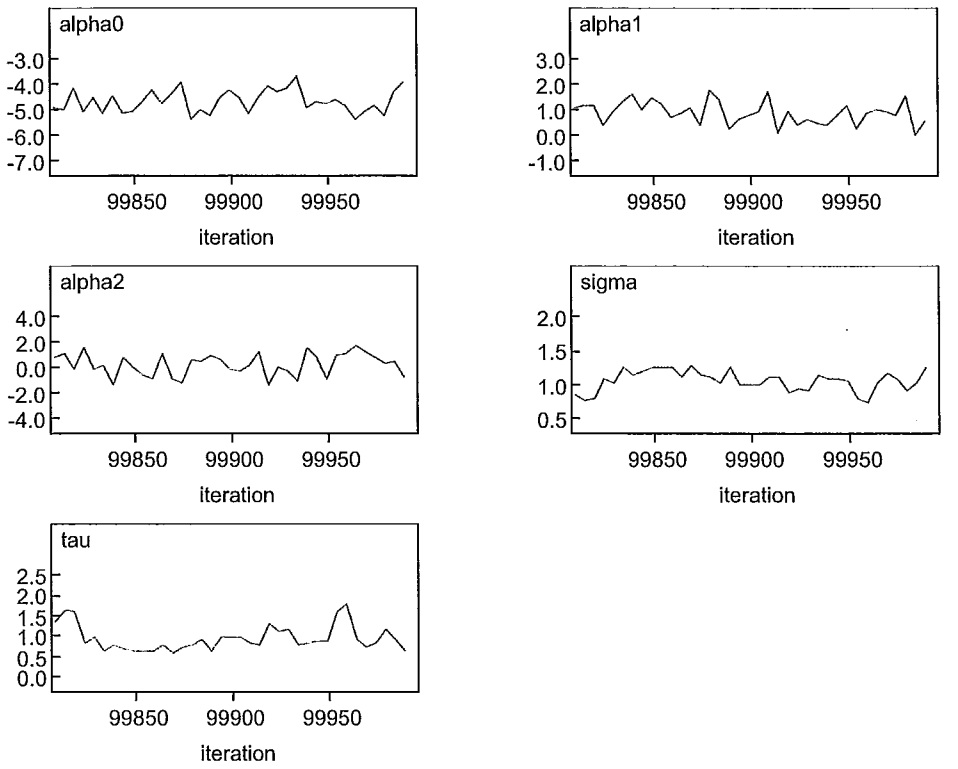
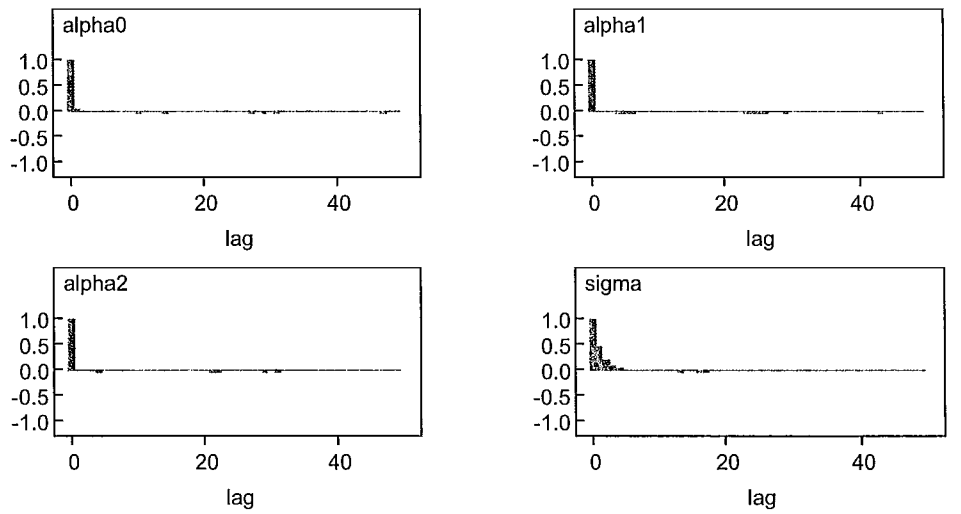

alpha1
7.5
5.0
2.5
0.0
-2.5
-5.0
100
50000
100000
150000
iteration
alpha2
15.0
10.0
5.0
0.0
-5.0
100
50000
100000
150000
iteration
sigma
2.5
2.0
1.5
1.0
0.5
100
50000
100000
150000
iteration
tau
3.0
2.0
1.0
0.0
100
50000
100000
150000
iteration

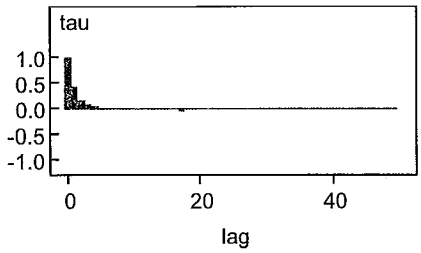
Figure 5. Diagnostic Plot for Bayesian Analysis-- Tracheotomy

Dynamic Trace

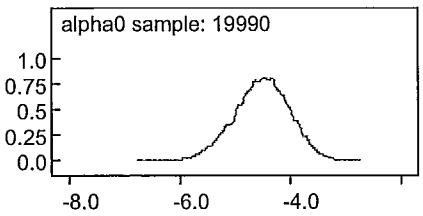


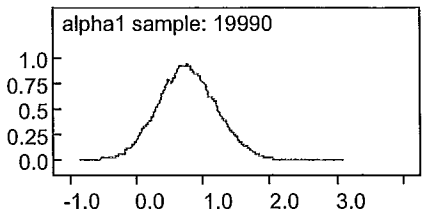
Autocorrelation function

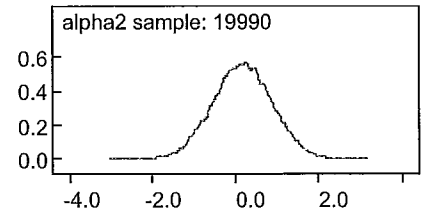


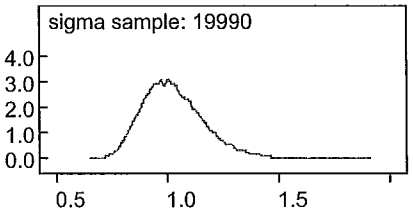


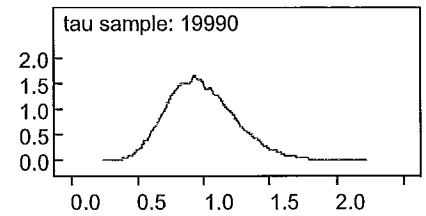
Kernel Denisty



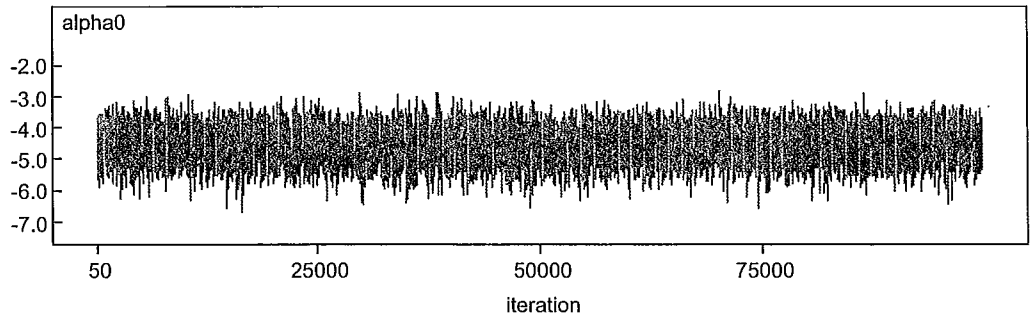








Time series



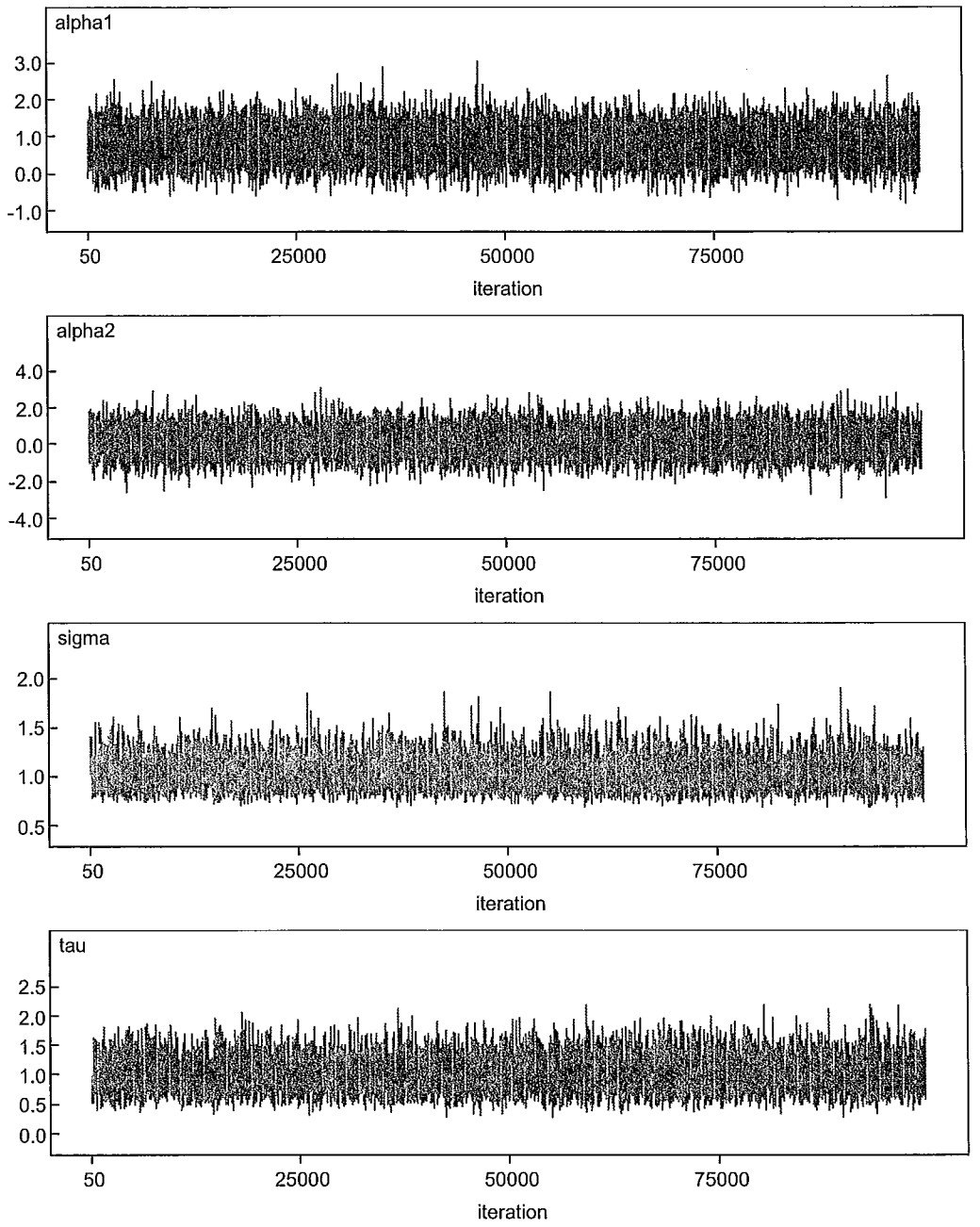

alpha1
3.0
2.0
1.0
0.0
-1.0
50
25000
50000
75000
iteration
alpha2
4.0
2.0
0.0
-2.0
-4.0
50
25000
50000
75000
iteration
sigma
2.0
1.5
1.0
0.5
50
25000
50000
75000
iteration
tau
2.5
2.0
1.5
1.0
0.5
0.0
50
25000
50000
75000
iteration

Figure 6. Forest Plot: Mortality without Adjustment for Covariates

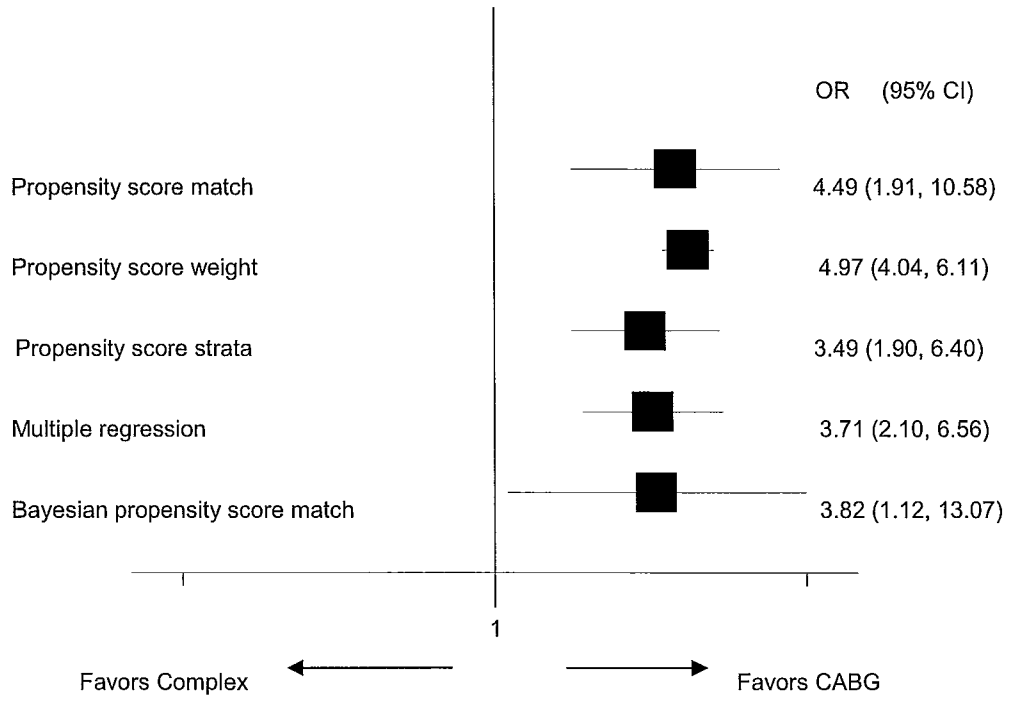


Figure 7. Forest Plot: Pulmonary Embolism without Adjustment for Covariates

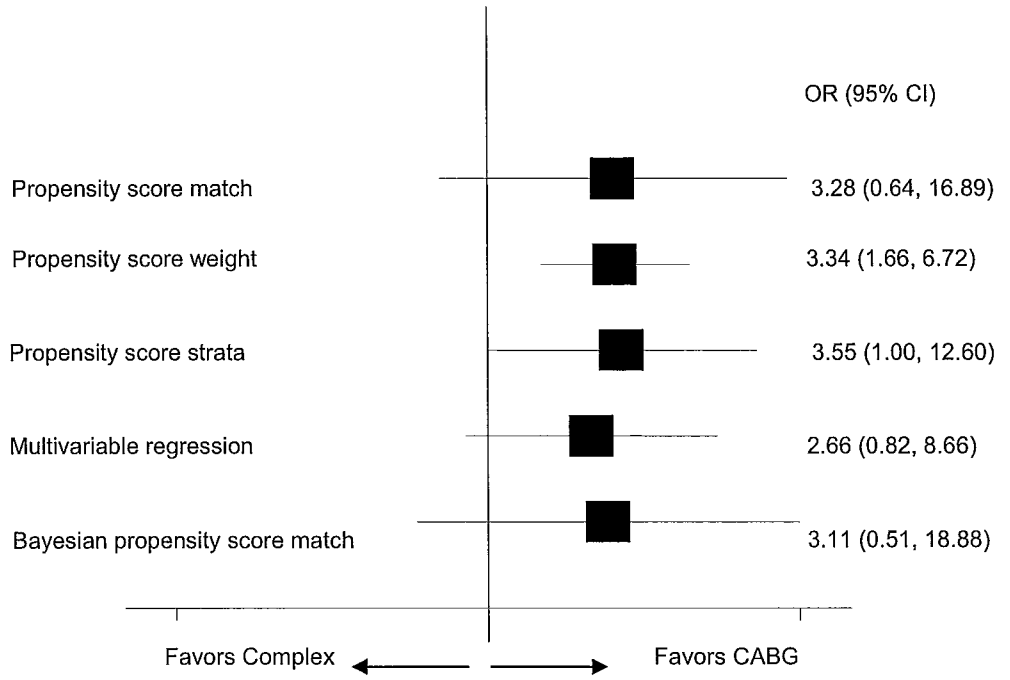


Figure 8. Forest Plot: Sepsis without Adjustment for Covariates

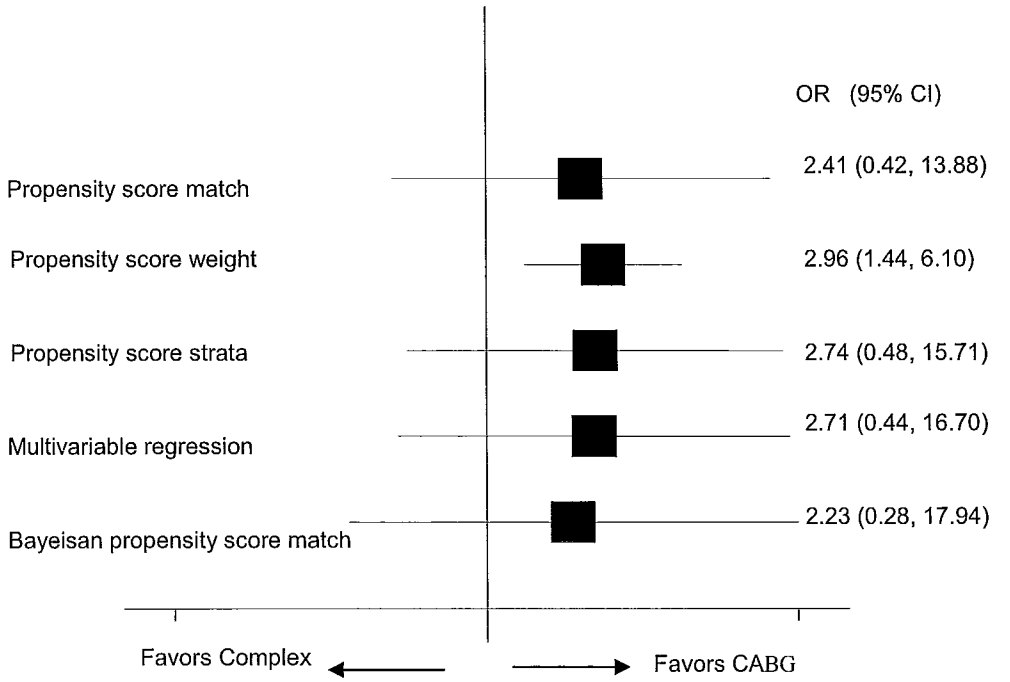
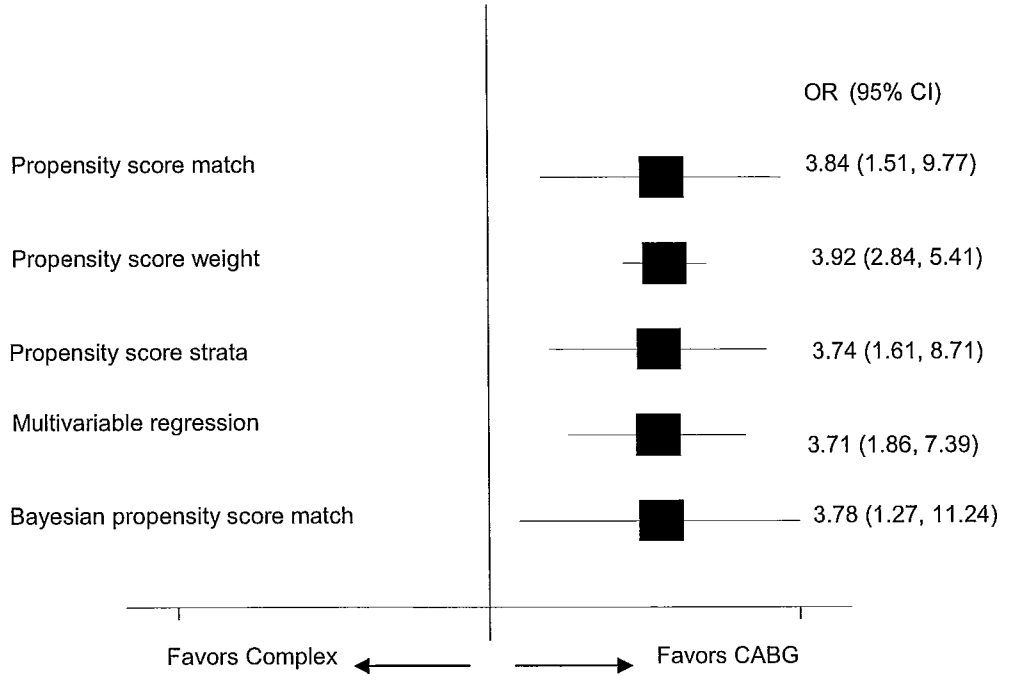


Figure 9. Forest Plot: Tracheotomy without Adjustment for Covariates



Appendix F

Code

F1: WinBugs Codes for Bayesian Analysis

```
*****************************************
*      Model for mortality
*****************************************

model
  {
    for (i in 1:864 )
{
      r[i]~dbern(p[i])
      b[i] ~ dnorm(0, tau)
      logit(p[i]) <- alpha0 + alpha1 * x1[i] + alpha2 * x2[i]
                + b[i]
    }
    alpha0 ~ dnorm(0, 1.0E-6)
    alpha1 ~ dnorm(0, 1.0E-6)
    alpha2 ~ dnorm(0, 1.0E-6)

 tau~dgamma(5,5)

   sigma<-1/sqrt(tau)
  }

*****************************************
*      Model for Pulmonary Embolism
*****************************************

model
  {
    for (i in 1:864 )
{
      r[i]~dbern(p[i])
      b[i] ~ dnorm(0, tau)
      logit(p[i]) <- alpha0 + alpha1 * x1[i] + alpha2 * x2[i]
                + b[i]
    }
    alpha0 ~ dnorm(0, 1.0E-6)
    alpha1 ~ dnorm(0, 1.0E-6)
```

```
    alpha2 ~ dnorm(0, 1.0E-6)

tau~dgamma(5,5)

   sigma<-1/sqrt(tau)
  }
```

****

*** Model for Sepsis**

****

Model for Sepsis

```
model
  {
    for (i in 1:864 )
{
      r[i]~dbern(p[i])
      b[i] ~ dnorm(0, tau)
      logit(p[i]) <- alpha0 + alpha1 * x1[i] + alpha2 * x2[i]
              + b[i]
    }
    alpha0 ~ dnorm(0, 1.0E-6)
    alpha1 ~ dnorm(0, 1.0E-6)
    alpha2 ~ dnorm(0, 1.0E-6)

 tau~dgamma(5,5)

   sigma<-1/sqrt(tau)
  }
```

```
**************************************
*      Model for Tracheotomy
**************************************
```

```
model
  {
    for (i in 1:864 )
{
      r[i]~dbern(p[i])
      b[i] ~ dnorm(0, tau)
      logit(p[i]) <- alpha0 + alpha1 * x1[i] + alpha2 * x2[i]
              + b[i]
    }
    alpha0 ~ dnorm(0, 1.0E-6)
    alpha1 ~ dnorm(0, 1.0E-6)
    alpha2 ~ dnorm(0, 1.0E-6)

 tau~dgamma(5,5)

   sigma<-1/sqrt(tau)
  }
```

F2: SAS Codes for Descriptive Statistics of Baseline Diagnostic Characteristics

```
proc import datafile="C:\Users\yanyun\desktop\prodata.xls"  out=datatemp
        dbms=excel
        replace;
        getnames=yes;
run;
ods listing close;
 ods rtf file="C:\users\yanyun\yy.rtf";

proc ttest data=datatemp;
class ProcID;
var  Age Weight Height BSA ICUDur DurVent ;
run;
```

```
quit;
ods rtf close;
ods listing;
run;


proc import datafile="C:\Users\yanyun\desktop\prodata.xls"  out=datatemp
        dbms=excel
        replace;
        getnames=yes;
run;

ods listing close;
 ods rtf file="C:\users\yanyun\yy.rtf";
proc sort data=datatemp; by ProcID;
run;
proc freq data=datatemp;
tables (Gender LV Smoke Angina PrevMI Diabetes Hyperten Hyperchol Copd CVD
PVD Prefib CHF Shock Predialysis Preheparin PreASA) *ProcID/chisq;

run;
quit;
ods rtf close;
ods listing;
run;
```

F3: Stata Codes for Propensity Score Match

```
. insheet using "C:\Users\yanyun\Desktop\propensity.CSV"
(39 vars, 2427 obs)

. set seed 1000

. generate x=uniform()

. sort x


. psmatch2 procid,pscore(pr) caliper(0.1) noreplacement descending
```

F4: SAS Codes for propensity Score Strata

```
proc import datafile="C:\Users\yanyun\desktop\prodata.xls" out=datatemp
        dbms=excel
        replace;
        getnames=yes;
run;

ods listing close;
 ods rtf file="C:\users\yanyun\yy.rtf";
proc logistic data=datatemp descending;
model ProcID = Gender Age Weight Height BSA Smoke LV Angina PrevMI Diabetes
Hyperten Hyperchol Copd
CVD PVD Prefib CHF Shock Predialysis Preheparin PreASA/link=logit rsquare;
output out = psdataset pred = ps;
run;
proc rank data= psdataset groups=5 out= r;
ranks rnks;
var ps;
run;
data quintile; set r;
quintile = rnks + 1;
run;

%MACRO logitout (vdata);
proc logistic data = quintile desc;
model &vdata = ProcID quintile / cl;
run;

%mend logitout;
%logitout(vdata= Death);
%logitout(vdata= Stroke);

%logitout(vdata=Renal);
%logitout(vdata= pneumonia);
%logitout(vdata= PMI);
%logitout(vdata= Pulmonary);

%logitout(vdata= Sepsis);
%logitout(vdata= SternalDebrid);
%logitout(vdata= relntub);
%logitout(vdata= Trach);
```

```
%logitout(vdata= reexp);
%logitout(vdata= ICUDur);
%logitout(vdata= DurVent);

quit;
ods rtf close;
ods listing;
run;

5: different strata outcome difference: (q1,q2,q3,q4,q5)
proc import datafile="C:\Users\yanyun\desktop\q1.xls" out=datatemp
        dbms=excel
        replace;
        getnames=yes;
run;


ods listing close;
 ods rtf file="C:\users\yanyun\yy.rtf";

%MACRO logitout (vdata);
proc logistic data = datatemp desc;
model &vdata = ProcID / cl;
run;

%mend logitout;
%logitout(vdata= Death);
%logitout(vdata= Stroke);

%logitout(vdata=Renal);
%logitout(vdata= pneumonia);
%logitout(vdata= PMI);
%logitout(vdata= Pulmonary);

%logitout(vdata= Sepsis);
%logitout(vdata= SternalDebrid);
%logitout(vdata= relntub);
%logitout(vdata= Trach);
%logitout(vdata= reexp);
%logitout(vdata= ICUDur);
%logitout(vdata= DurVent);
quit;
ods rtf close;
ods listing;
```

```
run;
```

F5: SAS Codes for Propensity Score Weighted

```
proc import datafile="C:\Users\yanyun\desktop\prodata.xls" out=datatemp
        dbms=excel
        replace;
        getnames=yes;
run;
proc logistic data=datatemp descending;
class ProcID;
model ProcID = Gender Age Weight Height BSA Smoke LV Angina PrevMI Diabetes
Hyperten Hyperchol Copd
CVD PVD Prefib CHF Shock Predialysis Preheparin PreASA SiteID Redonum
Timing/selection=stepwise risklimits lackfit rsquare parmlabel;
output out=preds pred=pr;
run;
proc print data=preds (obs=10);
run;

data preds;
set preds;
if pr=. then delete;
run;
PROC EXPORT
  DATA=preds
  OUTFILE="c:\Users\yanyun\desktop\pen.xls"
  DBMS=EXCEL2000 REPLACE;
RUN;



proc import datafile="C:\Users\yanyun\desktop\propensity1.xls" out=datatemp
        dbms=excel
        replace;
        getnames=yes;
run;
data ps_weight;
set datatemp;
if ProcID=1 then weight=1/pr;
else weight=1/(1-pr);
```

```
run;
ods listing close;
 ods rtf file="C:\users\yanyun\yy.rtf";

%MACRO logitout (vdata);
proc logistic data = ps_weight desc;
model  &vdata = ProcID  ;
weight weight;
 run;

%mend  logitout;
%logitout(vdata= Death);
%logitout(vdata= Stroke);

%logitout(vdata=Renal);
%logitout(vdata= pneumonia);
%logitout(vdata= PMI);
%logitout(vdata= Pulmonary);

%logitout(vdata= Sepsis);
%logitout(vdata= SternalDebrid);
%logitout(vdata= relntub);
%logitout(vdata= Trach);
%logitout(vdata= reexp);
%logitout(vdata= ICUDur);
%logitout(vdata= DurVent);
quit;
ods rtf close;
ods listing;
run;
```

F6: SAS Codes for Multivariable Logistic Regression

```
proc import datafile="C:\Users\yanyun\desktop\prodata.xls"  out=datatemp
        dbms=excel
        replace;
        getnames=yes;
run;
```

```
ods listing close;
 ods rtf file="C:\users\yanyun\yy.rtf";

%MACRO logitout (vdata);
proc logistic data = datatemp desc;
model  &vdata = ProcID  Age Gender BSA  Hyperten Smoke CHF Diabetes SiteID
PrevMI Redonum Timing  /cl;
run;

%mend  logitout;
%logitout(vdata= Death);
%logitout(vdata= Stroke);

%logitout(vdata=Renal);
%logitout(vdata= pneumonia);
%logitout(vdata= PMI);
%logitout(vdata= Pulmonary);

%logitout(vdata= Sepsis);
%logitout(vdata= SternalDebrid);
%logitout(vdata= relntub);
%logitout(vdata= Trach);
%logitout(vdata= reexp);
%logitout(vdata= ICUDur);
%logitout(vdata= DurVent);
quit;
ods rtf close;
ods listing;
run;


## site different mortality

proc import datafile="C:\Users\yanyun\desktop\prodata.xls"  out=datatemp
        dbms=excel
        replace;
        getnames=yes;
run;
ods listing close;
 ods rtf file="C:\users\yanyun\yy.rtf";


proc logistic data=datatemp descending;
class SiteID;
```

```
model Death=ProcID SiteID/cl;
run;
quit;
ods rtf close;
ods listing;
run;

Site different ICU length of stay
proc import datafile="C:\Users\yanyun\desktop\prodata.xls"  out=datatemp
        dbms=excel
        replace;
        getnames=yes;
run;
ods listing close;
 ods rtf file="C:\users\yanyun\yy.rtf";


proc logistic data=datatemp descending;
class SiteID;
model ICUDur=ProcID SiteID/cl;
run;
quit;
ods rtf close;
ods listing;
run;
```