Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Christopher Locandro Date

Red Blood Cell Transfusion as a Predictor of Outcome after Cardiac Surgery in Neonates and Young Infants

By

Christopher Locandro

Master of Science in Public Health

Biostatistics and Bioinformatics

Traci Leong, PhD

(Thesis Advisor)

Robert Lyles, PhD

(Reader)

Red Blood Cell Transfusion as a Predictor of Outcome after Cardiac Surgery in Neonates and Young Infants

By

Christopher Locandro

B.S.

Georgia State University

2015

Thesis Committee Chair: Traci Leong, PhD

Reader: Robert Lyles, PhD

An abstract of

A thesis submitted to the Faculty of the

Rollins School of Public Health of Emory University

in partial fulfillment of the requirements for the degree of

Master of Science in Public Health

in Biostatistics

2017

Abstract

Red Blood Cell Transfusion as a Predictor of Outcome after Cardiac Surgery in Neonates and Young Infants

By: Christopher Locandro

Introduction: Little is known about the adverse outcomes associated with red blood cell (RBC) transfusion in neonates and young infants undergoing cardiac surgery. The goal of this paper was to examine associations between RBC transfusion, both intra-operatively and post-operatively, and adverse outcomes in this patient population. We sought to adjust for surgical risk scores and implement a standardized measure of RBC volume per kilogram for each patient. We then extend these results to develop a clinical tool that predicts patient complication risk post-surgery.

Methods: We retrospectively analyzed a cohort of 605 patients (666 surgeries in total) aged 6 months or less who underwent cardiac bypass surgery. Clinical parameters included age, weight, risk adjustment for congenital heart surgery (RACHS1 and STAT) scores, intensive care unit length of stay (ICU LOS), RBC volume transfused intra-operatively, RBC volume transfused 24 hours post-operatively, cross-clamp time, and cardiopulmonary bypass (CPB) time. Risk-adjusted logistic regression and lognormal regression were used to assess the influence of RBC transfusion on complication risk and LOS, respectively. Finally, we trained and tested 3 models (random forest, decision tree, and logistic regression) for predicting patient complication risk post-surgery. We generated the receiver operating characteristics (ROC) curve for each model and calculated area under the curve (AUC) as a performance metric.

Results: In our cohort, we observed 137 (20.6%) complications. Patients that were transfused post-operatively had a significantly higher risk of post-surgery complications (95% C.I. for OR, [3.06,7]). Both intra-operative and post-operative standardized RBC transfusion volumes were associated with higher complication risk, after controlling for patient age and surgery risk scores ($p<0.001$ for both). We found similar results when considering our secondary outcome, ICU LOS. Both intra-operative ($p<0.001$) and post-operative ($p=0.004$) standardized RBC transfusion volumes were associated with increased ICU LOS. The random forest model had the highest predictive accuracy with an AUC of 0.718.

Discussion: Our findings suggest that this younger population of pediatric cardiac surgical patients is particularly volume-sensitive to RBC transfusion, even after controlling for variables such as surgery risk. Our predictive model may assist in identifying patients that are high risk for complication immediately following surgery.

Red Blood Cell Transfusion as a Predictor of Outcome after Cardiac Surgery in Neonates and Young Infants

By

Christopher Locandro

B.S.

Georgia State University

2015

Thesis Committee Chair: Traci Leong, PhD

Reader: Robert Lyles, PhD

A thesis submitted to the Faculty of the

Rollins School of Public Health of Emory University

in partial fulfillment of the requirements for the degree of

Master of Science in Public Health

in Biostatistics

2017

Acknowledgements

I would like to thank the department of Biostatistics and Bioinformatics at Emory University for their support and guidance over the past two years. I would especially like to thank Traci Leong for her willingness to be my thesis advisor and all of the help and feedback that she has provided on my thesis. I would also like to thank Robert Lyles for his feedback on my work. Lastly, I would like to thank my family and friends for their love, support, and encouragement throughout this project.

Table of Contents

1: INTRODUCTION

1.1 Overview

Transfusion of red blood cell (RBC) products has been regarded as one of the most significant contributions to modern medicine and surgical practice. The primary purpose of RBC transfusion is to correct for insufficient tissue oxygenation, particularly when the patient's physiological mechanisms of compensation are failing to do so (Liumbruno, Bennardello, Lattanzio, Piccoli, & Rossetti, 2009). Factors that influence tissue oxygenation include hemoglobin (Hb) concentration, tissue oxygen (O_2) demand, and Hb saturation (Blackwood et al., 2010). Therefore, part of the decision to transfuse is based on whether Hb levels fall below a predefined threshold. Other factors determining the need for RBC transfusion include blood loss (as a percentage of total blood volume) and clinical features of the patient (Hill et al., 2000). However, the demand for RBC transfusion, as well as its associated benefits and risks, varies across patient populations and surgical procedures (Spiess, 2013).

1.2 Problem Statement

We believe the following issues regarding RBC transfusion have not been appropriately studied as of yet. There are no studies that focus exclusively on RBC transfusion associated with cardiac surgery in neonates and young infants less than 6 months of age. While these patients are more volume-sensitive than older children and adults, none of the previous studies specifically took into account the volume of RBCs transfused intra-operatively versus post-operatively when examining clinical endpoints. Furthermore, most previous studies have been small (fewer than 300 patients) and of short duration.

1.3 Purpose Statement

We propose to correct what we believe are drawbacks in prior studies in order to accurately correlate RBC transfusion volume with outcome measures in newborns and neonates less than 6 months of age undergoing cardiac surgery, and to develop a model to predict which of these patients are more likely to be at risk for any deleterious effects of transfusion. The primary goal of this study is to examine the impact of intra-operative and post-operative RBC transfusion, with standardized quantification of transfusion volume in each patient, upon specific post-operative outcomes in a cohort of neonates and infants less than six months of age undergoing cardiac surgery at a large, urban medical center.

1.4 Significance Statement

This paper will examine the effect of both intraoperative and postoperative RBC transfusion on multiple outcome measures in a rarely studied pediatric cardiac surgery patient population. This paper will provide evidence for the effects of RBC transfusion on outcome in neonates as compared to the older pediatric population. Additionally, a predictive model will be incorporated into a web application that can readily be used as a clinical tool to anticipate patient complication risks following cardiac surgery.

2: BACKGROUND/ LITERATURE REVIEW

2.1 RBC Transfusion Overview

Transfusion of blood products, primarily RBCs, is common in adult and pediatric patients undergoing surgery. The benefits of RBC transfusion have been well established in the literature. In patients with severe anemia, RBC transfusion may help prevent organ ischemia due to decreased oxygen delivery (Iyengar et al., 2013; Marik & Corwin, 2008). In patients with less severe anemia, transfusion of RBCs may also help prevent hemodynamic instability due to intra-operative or post-operative blood loss. However, RBC transfusions are not without risk. Transfusion has been associated with several adverse outcomes in diverse patient populations, including increased rates of infection, ischemia, kidney failure, circulatory overload, and mortality (Marik & Corwin, 2008). In adults, RBC transfusion has also been associated with increased 30-day mortality and increased length of stay (Corwin et al., 2004). More recently, RBC transfusion has been linked with adverse outcomes particularly relevant to cardiac surgery, including an increased risk of new-onset cardiac arrhythmias and conduction abnormalities in patients with acute myocardial infarction (Athar et al., 2011). Numerous studies involving adult cardiac surgery patients have also identified associations between RBC transfusion and increased morbidity and mortality (Marik & Corwin, 2008; Murphy et al., 2008; Ranucci et al., 2011; Szekely et al., 2009).

2.2 RBC Transfusion in Pediatric Cardiac Surgery

Because the majority of adult cardiac surgery patients included in these studies underwent coronary artery bypass grafting, it is difficult to extrapolate from their results to the effects of transfusion on pediatric cardiac surgery patients, most of whom have congenital heart

disease. Furthermore, RBC transfusion reaction rates and types differ between these two populations due to physiological differences (e.g. metabolic rate, oxygen consumption, cardiac output) (Bharadwaj, Khandelwal, & Bhargava, 2014). For instance, in a large cohort study of adult and pediatric RBC transfusion reactions, Oakley et al. (2015) observed a significantly higher incidence of allergic, hypotensive, and febrile nonhemolytic transfusion reactions in pediatric patients compared to adults.

In addition, the available evidence is too sparse to either confirm or refute a relationship between RBC transfusion and postoperative morbidity and mortality in critically ill pediatric patients, particularly neonates and infants. For example, Willems et al. (2010) examined 125 pediatric patients after cardiac surgery and found a mean ICU length of stay of 7 days and a 3.2% 28-day mortality. However, this study excluded patients less than 28 days of age, a substantial segment of the pediatric cardiac surgery population. Salvin et al. (2011) found that postoperative RBC transfusion was associated with prolonged hospitalization in children after cardiac surgery. However, in this study greater than 47% of children were over one year of age. Other studies that focused on neonates primarily enrolled non-cardiac populations. For example, for patients in non-cardiac pediatric ICUs, RBC transfusion was found to be independently associated with increased inotropic requirements, prolonged duration of mechanical ventilation and greater mortality (Keung et al., 2009). Furthermore, the study determined that only 41% of released blood products were transfused but did not determine the volume transfused per patient. Schmotzer et al. (2010) examined RBC usage at Children's Healthcare of Atlanta at Scottish Rite, but their analysis lacked cardiac surgical patients and only examined six months of data. Agarwal et al. (2014) discovered associations between cardiopulmonary bypass time and complications in a cohort of pediatric cardiac surgeries, however, they did not consider RBC

transfusion volumes. Similarly, Costello et al. (2010) identified the number of RBC transfusions as a risk factor for surgical site infection in pediatric cardiac surgery, but they ignored transfusion volume in their analysis. There are few studies examining the effects of RBC transfusion on clinical outcomes after cardiac surgery.

2.3 Statistical Considerations

In our analysis, we consider intensive care unit (ICU) length of stay as one of our primary endpoints. ICU LOS is known to have a right-skewed, nonnegative distribution, which renders ordinary least squares (OLS) regression inappropriate. Numerous alternative approaches have been proposed, including zero-truncated Poisson, zero-truncated negative binomial, Cox proportional hazards, and lognormal models (Austin, Rothwell, & Tu, 2002; Faddy, Graves, & Pettitt, 2009). Each approach has its associated benefits and drawbacks. For instance, if one considers LOS as a measure of time-to-event (i.e. time to discharge from hospital) data, then the Cox proportional hazards model becomes a natural choice that can also account for censoring. However, the proportional hazards assumption may or may not be justified (Verburg, Keizer, Jonge, & Peek, 2014). The log-transformed linear model approach offers interpretability after appropriate retransformation of the regression coefficients. However, Manning (1998) shows that if the error term for the log-transformed dependent variable is heteroscedastic, this can lead to biased coefficient estimates on the untransformed scale. Ultimately, model fit dictates the optimal approach, however, model interpretability is also an attractive feature in the clinical setting. Knowing the associated increase in LOS corresponding to unit increases in covariate values is often of interest to clinical investigators. We factor these considerations into our model selection process.

2.4 Predictive Modeling

In the final part of our analysis, we develop a predictive model for complication risk as a function of patient and operative characteristics. With complication as a dichotomous variable (yes/no), several modeling options are feasible. The traditional approach, logistic regression, considers the log-odds (logit) of complication as outcome. With logistic regression, we can explicitly model the probability of complication (risk) and then generate predictions for new covariate values. However, supervised machine learning methods, such as classification and regression trees (CART), may offer improved predictive accuracy (Austin, Lee, Steyerberg, & Tu, 2012). CART works by establishing a sequence of splitting rules on each of the predictors in the model, such that each series of splits corresponds to a constant outcome value (or class, in the case of categorical outcomes). The CART algorithm then chooses the optimal set of splitting criteria by minimizing an impurity measure (e.g. misclassification error, Gini index, or cross-entropy) (Loh, 2014).

This approach offers many advantages over logistic regression, particularly in the context of clinical prediction models (Verplancke et al., 2009). First, CART is a nonparametric approach that does not require any assumptions about the distributional form of the outcome. Second, the interpretation of CART closely resembles the clinical decision making process. Finally, CART is generally robust to outliers and missing data (Hastie, Tibshirani, & Friedman, 2009). However, decision trees are subject to instability and high variance. Additionally, they may fail to capture additive structures within data (Loh 2014).

Another feasible modeling option for binary classification is the random forest approach. Random forests extend CART by generating many optimized trees, each on a bootstrap-sampled

version of the training set. For new observations, the predicted class is generated by taking the majority vote from this set of trees (Hastie et al., 2009). By taking an averaged prediction over many decision trees, the random forest method reduces the inherent instability of the traditional CART approach. In classification problems involving patient mortality, for instance, random forests have been shown to offer superior predictive accuracy when compared to other approaches such as CART and logistic regression (Austin, 2011; Peng, Chuang, Kang, & Tseng, 2010).

3. METHODOLOGY

3.1 Study Design

In our study, we retrospectively examined a cohort of 605 patients that underwent pediatric cardiac surgery with cardiopulmonary bypass (CPB) from January 2012 through July 2013. Blood transfusion and outcomes data were pulled from the electronic medical record EPIC database at Egleston Hospital with authors N. Guzzetta and S. Niazi reviewing each case for accuracy (Niazi, Leong, Meyer, & Guzzetta, 2017). In patients with congenital defects requiring multiple surgeries, each surgery was considered independently in the analysis (666 operations in total). We required patients to be fully recovered from a prior surgery for the subsequent surgery to qualify.

3.2 Data Collection

Demographic variables, including patient weight (kg) and age (days), were collected from the electronic medical record. Operative variables included cross clamp time (min), cardiopulmonary bypass time (min), perfusion time (min), time on ventilation (hrs), RBC volume transfused intra-operatively (mL), RBC volume transfused post-operatively (mL), whether or not the patient received RBC transfusion prior to surgery, and whether or not the patient required Extracorporeal Membrane Oxygenation (ECMO). Prior to surgery, patients are classified into risk categories defined by two scoring systems: the Society of Thoracic Surgeons-European Association for Cardio-Thoracic Surgery Score (STAT) and the Risk Adjustment for Congenital Heart Surgery Score (RACHS1) (Jenkins, 2014). High-risk surgeries correspond to higher STAT and RACHS1 scores. We use these two metrics to control for the inherent variability in surgery risks across the cohort. In the original dataset, STAT scores ranged from 1-5 with some surgeries classified further into subgroups (e.g. "1a"). However, due to the sparsity of these sub-grouped classifications, we ultimately combined them into their numerical group (e.g. "1a" to"1"). We also consider cardiopulmonary bypass time as an indicator of surgery risk. Outcomes considered were discharge mortality, ICU length of stay (hrs), total length of stay (days), renal failure, infection, and thrombosis.

In order to relate RBC transfusion parameters to outcomes, we account for patient weight by calculating a standardized ratio of RBC volume (mL) transfused per kg for each patient and incorporate this quantity into our analysis. Additionally, we define complication as the occurrence of any of the following: mortality, infection, thrombosis, renal failure, or ECMO. This leads to more stable maximum likelihood estimates in our logistic regression model, compared to a model with a low mortality rate as the outcome.

3.3 Statistical Analysis

Depending on skewness, continuous variables were summarized using median (IQR) or mean (SD). Categorical variables were summarized using frequencies and percentages. For univariate analyses, independent t-tests were used for pairwise comparisons of normally distributed continuous variables. Chi-square tests were used for pairwise comparison of groups of categorical variables.

When considering complication as the outcome, logistic regression was employed with stepwise variable selection. We report the odds ratios (OR) and associated 95% confidence intervals. For LOS as the outcome, zero-truncated negative binomial, zero-truncated Poisson, lognormal, and Cox proportional hazards models were considered with stepwise variable selection. The Akaike Information Criterion (AIC) was then employed for model selection of fully-parametric methods (all except Cox proportional hazards). However, AIC-based model comparison is only valid for models with the same outcome variable. Therefore, we use statistical software to calculate the corrected lognormal model AIC for the untransformed version of LOS, based on the likelihood function (as in Burnham & Anderson, 2010). For the semi-parametric Cox proportional hazards model, model fit was assessed by Cox-Snell residuals. For predictive modeling with complication as the outcome, CART, logistic regression, and random forests were considered. Classification trees were developed using Gini impurity as the splitting criterion. Each model was developed and tested using a 70/30 train-test split on the data. We then compare the performance of each binary classifier by examining receiver operating characteristic (ROC) curves and computing area under the curve (AUC) as a performance metric.

4: RESULTS

4.1 Univariate Analysis

The median (IQR) patient age at time of surgery was 78 days (131). The mean (SD) patient weight was 4.39 kg (1.45). Table 1 shows selected patient characteristics by post-operative transfusion status (yes/no). Of the 666 operations, death as discharge status occurred 34 (5.1%) times, infection occurred 64 (9.6%) times, thrombosis occurred 55 (8.3%) times, renal failure occurred 12 (1.8%) times, and 40 (6%) operations required extracorporeal membrane oxygenation (ECMO). The complication rate for the cohort was 20.6%. Median (IQR) ICU length of stay was 101.5 hours (172.5) and median hospital LOS was 11 days (16). Two patients died during surgery, therefore no ICU LOS was reported.

Age and weight at time of surgery were found to be negatively associated with risk of complication ($p<0.001$ for both). A 60-day reduction in age was associated with a 2.12 increase in the risk of complication (95% C.I. for OR [1.34,2.90]). Both cross clamp time and cardiopulmonary bypass time were associated with higher risk of complication ($p<0.001$ for each). Furthermore, patients that were transfused within 7 days prior to surgery had a significantly higher risk of complication (95% C.I. for OR [1.8,4.31], $p<.001$). Patients transfused within 7 days post-operation had an even higher risk of complication (95% C.I. for OR [3.06,7], $p<.001$). Post-operative standardized RBC volume ($p<0.001$) and intra-operative standardized RBC volume ($p<0.001$) were associated with higher risk of complication in the cohort. We also assessed complication rate by post-operative transfusion status (yes/no) (Table 2). Within each group, we observed the same effects of patient weight, age, and

cardiopulmonary bypass time as we did without subgrouping by transfusion status, as in Table 1. However, in patients that did not receive post-operative RBC transfusion, cross clamp time was not significantly associated with risk of complication ($p=0.08$). On the other hand, cross clamp time did have a significant effect on complication risk in patients that were transfused post-operatively ($p<0.001$).

We considered ICU LOS as a secondary outcome in our analysis. Patient age was negatively associated with ICU LOS ($p<.001$). A 60-day reduction in age corresponded to an increase in ICU length of stay of approximately 49.6 hours. Similarly, patient weight was negatively associated with ICU length of stay ($p<.001$). A 6-kg reduction in weight, for example, corresponded to an 8-hour increase in length of stay. Both intra-operative standardized RBC volume ($p<0.001$) and post-operative standardized RBC volume ($p=0.002$) were associated with increased LOS. A 20-unit (mL/kg) increase in standardized RBC volume was associated with a 20.1-hour increase in LOS for intra-operative transfusion and a 20.0-hour increase in LOS for post-operative transfusion.

4.2 Multivariate Analysis

In considering associations between patient characteristics and outcome, we then accounted for the surgery risk as measured by both the STAT/RACHS1 scores and CPB time. As expected, Figure 1 illustrates that the distribution of STAT/RACHS1 scores for our cohort is higher in operations that resulted in a complication compared to operations that did not. Therefore, in modeling complication as the outcome, we forced these variables into our logistic regression model. We found that models including STAT vs. RACHS1 performed similarly, since these scores are based on similar clinical criteria. Thus, we only kept RACHS1 score in

our final model to reduce the number of parameters. Additionally, since the standardized ratios of RBC transfusion volume already account for patient weight, we excluded weight from the model. Stepwise variable selection was then used to choose the most important of the remaining predictors. The final logistic regression model included RACHS1 score, CPB time, standardized post-operative RBC transfusion volume, standardized intra-operative RBC transfusion volume, and patient age. After accounting for RACHS1 score, CPB time, and standardized transfusion volumes, age was found to remain significantly associated with complication risk ($p=0.010$). Similarly, after controlling for the other predictors in the model, the effect of standardized postoperative transfusion remained significant ($p=0.006$).

We then performed multivariate analysis with ICU LOS as the outcome. The following candidate models were considered: zero-truncated negative binomial, zero-truncated Poisson, lognormal, and Cox proportional hazards. Patients that died while in the ICU were considered censored in the Cox proportional hazards model. Surgery risk variables (RACHS1 score and CPB Time) were forced into each of these candidate models and then stepwise variable selection was used to choose the remaining predictors in the model. Table 4 shows the AIC for each of the 4 parametric models, with the lognormal model having the lowest. The Cox proportional hazards model had significant upper tail deviation in the Cox-Snell residual plot (Figure 2), indicating poor fit. For these reasons, we ultimately chose the lognormal model in our analysis. Regression estimates for this final multivariate model are reported in Table 5. After controlling for CPB time, RACHS1 score, and standardized RBC transfusion, patient age was found to remain negatively associated with ICU LOS ($p<0.001$). Similarly, standardized post-operative transfusion remained a significant predictor of ICU LOS even after accounting for risk measures

and patient age (p=0.004). Intra-operative transfusion was also positively associated with ICU LOS, after accounting for risk measures and patient age (p<0.001).

4.2 Predictive Modeling

Predictive models with complication as outcome were then constructed using the logistic regression model and two supervised machine learning techniques (CART and random forest). A random sample of 70% of the original dataset was used to train each model. In all three cases, the predictions measured the probability of having a complication, given specific patient and operative characteristics (patient weight, patient age, RACHS1 score, STAT score, standardized post-operative RBC volume, standardized intra-operative RBC volume, cross clamp time, and CPB time). However, since the observed data takes on a binary form (complication vs. no complication), it was necessary to define some threshold, k, to dichotomize the predicted probability into similar classes. Plotting true positive rate against the false positive rate for different values of k then produced the ROC curve for each classifier. The area under the ROC curve (AUC) was used to compare the predictive performance of each model. The random forest model had the highest AUC at 0.718 (Figure 3). AUC's for the decision tree and logistic regression model were 0.683 and 0.659, respectively.

The random forest model was then developed into an online web application (Figure 4). The application allows the user to input the following variables: patient weight, patient age, RACHS1 score, STAT score, RBC volume transfused intra-operatively, RBC volume transfused post-operatively, cross clamp time, and CPB time. The application then converts patient weight and transfusion volumes into standardized RBC transfusions volumes. It then inputs the

predictor values into the random forest model. Finally, the application outputs the probability of complication for a given patient, as well as the probability of mortality, infection, renal failure, and thrombosis. These predictions update in real-time, enabling the user to examine the effect of changing certain predictor values while keeping others fixed.

5: DISCUSSION

5.1 Introduction and Implications

The purpose of this analysis was to determine the effects of RBC transfusion (both intra-operatively and post-operatively) on outcome measures in a cohort of neonates and infants undergoing pediatric cardiac surgery. The risk-adjusted impact of RBC transfusion on complication rate and length of stay, with careful quantification of standardized transfusion volume, has not been previously studied in this subset of the pediatric patient population. Another goal of the analysis was to generate a predictive model for patient complication risk as a function of operative parameters and patient characteristics.

Our findings that age and weight were negatively associated with complication risk reinforce the vulnerability of this younger subset of the pediatric patient population. After adjusting for risk scores, we also found that increased standardized post-operative transfusion volume was associated with increased complication. Iyengar et al. (2011) noticed a similar association, though they treated pulmonary complications as the outcome. Salvin et al. (2011) also observed a higher risk of morbidity among patients with increased post-operative transfusion volumes. Willems et al. (2010) observed no effect of post-operative transfusion volume on complication risk, but their study population excluded neonates and young infants.

We also observed an association between both intra-operative and post-operative RBC transfusion and ICU LOS, after controlling for surgery risk. This finding is well supported by other evidence from the literature in other patient populations (Crown et al., 2004, Ranucci et al., 2011, Keung et al., 2009).

Our findings suggest that the population of neonate and young infant cardiac surgical patients are particularly volume-sensitive to RBC transfusion. Even after controlling for surgical risk and standardizing RBC volume by patient weight, both intra-operative and post-operative transfusion were found to have deleterious effects as measured by ICU LOS and complication risk. Our results suggest that caution should be exercised in determining operative RBC transfusion volumes for this vulnerable subset of the pediatric cardiac surgical population. Our predictive model may assist in this process by identifying high-risk patients immediately following cardiac surgery. Patients identified as having a high risk of complication could then be followed more closely after surgery.

5.2 Recommendations

One of the issues we encountered in our study was a low mortality rate. Because of this, we grouped other outcomes together to create a complication variable. Future studies with larger sample sizes would enable meaningful analysis of those specific outcomes (i.e. mortality, renal failure, infection, and thrombosis). Our study also did not include patient lab measurements (e.g. hemoglobin and creatinine concentrations), because the data were unavailable. Future studies may examine, for instance, the relationship between hemoglobin concentration and outcome. Furthermore, these measurements may improve the accuracy of our random forest model.

Additionally, some of the patients in our cohort underwent multiple cardiac surgeries. Although we only include such cases where patients made a full recovery between surgeries, future analyses may consider the possibility that these observations could be correlated. Another consideration is the possibility of "hidden" confounders that we were unable to control for in the analysis. For instance, surgeon skill may confound the relationship between RBC transfusion volumes and outcome. Lastly, the predictive modeling application could be modified to allow continual updating of the training set. The random forest model could then be retrained on this new dataset and achieve increased predictive accuracy over time.

REFERENCES

Agarwal, H. S., Wolfram, K. B., Saville, B. R., Donahue, B. S., & Bichell, D. P. (2014). Postoperative complications and association with outcomes in pediatric cardiac surgery. *The Journal of Thoracic and Cardiovascular Surgery,148*(2).

Athar, M. K., Bagga, S., Nair, N., Punjabi, V., Vito, K., Schorr, C., & Gerber, D. R. (2011). Risk of cardiac arrhythmias and conduction abnormalities in patients with acute myocardial infarction receiving packed red blood cell transfusions. *Journal of Critical Care,26*(4), 335-341.

Austin, P. C., Lee, D. S., Steyerberg, E. W., & Tu, J. V. (2012). Regression trees for predicting mortality in patients with cardiovascular disease: What improvement is achieved by using ensemble-based methods? *Biometrical Journal,54*(5), 657-673.

Bharadwaj, A., Khandelwal, M., & Bhargava, S. (2014). Perioperative neonatal and paediatric blood transfusion. *Indian Journal of Anaesthesia,58*(5), 652.

Blackwood, J., Joffe, A. R., Robertson, C. M., Dinu, I. A., Alton, G., Penner, K., . . . Rebeyka, I. M. (2010). Association of Hemoglobin and Transfusion With Outcome After Operations for Hypoplastic Left Heart. *The Annals of Thoracic Surgery,89*(5).

Burnham, K. P., & Anderson, D. R. (2010). *Model selection and multimodel inference: a practical information-theoretic approach*. New York: Springer.

Corwin, H. L., Gettinger, A., Pearl, R. G., Fink, M. P., Levy, M. M., Abraham, E., . . . Shapiro, M. J. (2004). The CRIT Study: Anemia and blood transfusion in the critically ill—Current clinical practice in the United States*. *Critical Care Medicine,32*(1), 39-52.

Costello, J. M., Graham, D. A., Morrow, D. F., Morrow, J., Potter-Bynoe, G., Sandora, T. J., . . . Laussen, P. C. (2010). Risk Factors for Surgical Site Infection After Cardiac Surgery in Children. *The Annals of Thoracic Surgery,89*(6), 1833-1842.

Faddy, M., Graves, N., & Pettitt, A. (2009). Modeling Length of Stay in Hospital and Other Right Skewed Data: Comparison of Phase-Type, Gamma and Log-Normal Distributions. *Value in Health,12*(2), 309-314.

Hastie, T., Tibshirani, R., & Friedman, J. H. (2016). *The elements of statistical learning: data mining, inference, and prediction*. New York: Springer.

Hill, S., Carless, P. A., Henry, D. A., Carson, J. L., Hebert, P. P., Henderson, K. M., & Mcclelland, B. (2000). Transfusion thresholds and other strategies for guiding allogeneic red blood cell transfusion. *Cochrane Database of Systematic Reviews.*

Iyengar, A., Scipione, C. N., Sheth, P., Ohye, R. G., Riegger, L., Bove, E. L., . . . Hirsch-Romano, J. C. (2013). Association of Complications With Blood Transfusions in Pediatric Cardiac Surgery Patients. *The Annals of Thoracic Surgery,96*(3), 910-916.

Jenkins, KJ (2004). Risk adjustment for congenital heart surgery: the RACHS-1 method. *Semin Thorac Cardiovasc Surg Pediatr Card Surg Annu,7*, 180-184.

Keung, C. Y., Smith, K. R., Savoia, H. F., & Davidson, A. J. (2009). An audit of transfusion of red blood cell units in pediatric anesthesia. *Pediatric Anesthesia,19*(4), 320-328.

Liumbruno, G., Bennardello, F., Lattanzio, A., Piccoli, P., & Rossetti, G. (2009) Recommendations for the transfusion of red blood cells. *Blood Transfusion*, *7*(1), 49–64.

Loh, W. (2014). Fifty Years of Classification and Regression Trees. *International Statistical Review,82*(3), 329-348.

Manning, W. G. (1998). The logged dependent variable, heteroscedasticity, and the retransformation problem. *Journal of Health Economics,17*(3), 283-295.

Marik, P. E., & Corwin, H. L. (2008). Efficacy of red blood cell transfusion in the critically ill: A systematic review of the literature*. *Critical Care Medicine,36*(9), 2667-2674.

Murphy, G., Reeves, B., & Rogers, C. (2008). Increased mortality, postoperative morbidity, and cost after red blood cell transfusion in patients having cardiac surgery. *Journal of Vascular Surgery,47*(4), 894.

Niazi, S., Leong, T., Meyer, E., & Guzzetta, N.A. (2017). Red Blood Cell Transfusion as a Predictor of Outcome after Cardiac Surgery in Neonates and Young Infants. *Congenital Cardiac Anesthesia Surgery Meeting*.

Oakley, F. D., Woods, M., Arnold, S., & Young, P. P. (2014). Transfusion reactions in pediatric compared with adult patients: a look at rate, reaction type, and associated products. *Transfusion,55*(3), 563-570.

Peng, S., Chuang, Y., Kang, T., & Tseng, K. (2010). Random forest can predict 30-day mortality of spontaneous intracerebral hemorrhage with remarkable discrimination. *European Journal of Neurology,17*(7), 945-950.

Ranucci, M., Aronson, S., Dietrich, W., Dyke, C. M., Hofmann, A., Karkouti, K., . . . Heymann, C. V. (2011). Patient blood management during cardiac surgery: Do we have enough evidence for clinical practice? *The Journal of Thoracic and Cardiovascular Surgery,142*(2).

Salvin, J. W., Scheurer, M. A., Laussen, P. C., Wypij, D., Polito, A., Bacha, E. A., . . . Thiagarajan, R. R. (2011). Blood Transfusion After Pediatric Cardiac Surgery Is

Associated With Prolonged Hospital Stay. *The Annals of Thoracic Surgery,91*(1), 204-210.

Schmotzer, C. L., Brown, A. E., Roth, S., Johnson, J., Ines-Castillejo, M., Reisner, A., . . . Josephson, C. D. (2010). Procedure-specific preoperative red blood cell preparation and utilization management in pediatric surgical patients. *Transfusion,50*(4), 861-867.

Spiess, B. D. (2013). Blood Transfusion and Infection After Cardiac Surgery. *The Annals of Thoracic Surgery,95*(6), 1855-1858.

Székely, A., Cserép, Z., Sápi, E., Breuer, T., Nagy, C. A., Vargha, P., . . . Treszl, A. (2009). Risks and Predictors of Blood Transfusion in Pediatric Patients Undergoing Open Heart Operations. *The Annals of Thoracic Surgery,87*(1), 187-197.

Verburg, I. W., Keizer, N. F., Jonge, E. D., & Peek, N. (2014). Comparison of Regression Methods for Modeling Intensive Care Length of Stay. *PLoS ONE,9*(10).

Verplancke, T., Vansteelandt, S., Benoit, D., Depuydt, P., Turck, F. D., & Decruyenaere, J. (2009). Prediction of hospital mortality by classification trees and smoothing splines in patients with haematological malignancies admitted to the ICU. *Critical Care,13*.

Willems, A., Harrington, K., Lacroix, J., Biarent, D., Joffe, A. R., Wensley, D., . . . Tucci, M. (2010). Comparison of two red-cell transfusion strategies after pediatric cardiac surgery: A subgroup analysis. *Critical Care Medicine,38*(2), 649-656.

APPENDIX: Tables & Figures

Table 1

Patient Characteristics by Post-Operative Transfusion Status (Yes/No)

Variable	Overall	Post-Operative Transfusion	No Post-Operative Transfusion
Age (days)	78 (131)	31 (125.5)	94 (131.5)
Weight (kg)	4.39 (1.45)	4.05 (1.40)	4.55 (1.45)
CC Time (min)	56 (45)	59 (55)	55 (42.5)
CPB Time (min)	101.5 (76.75)	131 (88)	91 (67)

Note. Weight is reported as mean with standard deviation in parentheses. All other variables are reported as median with interquartile range in parentheses. CC = Cross Clamp, CPB = Cardiopulmonary Bypass.

Table 2

Patient Characteristics by Complication and Transfusion Status (Yes/No)

	Post-Operative Transfusion			No Post-Operative Transfusion		
Variable	**Comp (n=82)**	**NoComp (n=129)**	**P Value**	**Comp (n=55)**	**NoComp (n=400)**	**P Value**
Age (days)	37.26	80.9	<0.0001	46.4	88.5	<0.0001
Weight (kg)	3.46	4.42	<0.0001	3.62	4.68	<0.0001
ICU LOS (hrs)	722.4	257.4	<0.0001	581.24	107.64	<0.0001
Total LOS (days)	49.05	20.66	<0.0001	43.95	12.62	<0.0001
CPB Time (min)	173.34	116.95	<0.0001	129.56	99.35	0.001
CC Time (min)	74.91	52.47	<0.0001	74.91	55.28	0.083

Note. Weight is reported as mean. All other variables are reported as median. P values correspond to pairwise univariate tests between subgroups. Comp=Complication, NoComp =No Complication, CC = Cross Clamp, CPB = Cardiopulmonary Bypass.

Table 3

Multivariate Logistic Regression Model Estimates

Parameter	β Estimate (SE)	P Value
Intercept	1.43 (3.56)	0.656
PostBloodPerKG	0.0032 (0.0011)	0.006
IntraBloodPerKG	0.0071 (0.0023)	0.0016
RACHS1 Score*		
1	-0.32 (0.54)	0.277
2	-0.12 (0.23)	0.301
3	0.05 (0.39)	0.551
4	-0.002 (0.07)	0.489
5	0.21 (0.11)	0.700
6	0.43 (0.76)	0.714
CPB Time	0.0041 (0.0024)	0.093
Age (Days)	-0.0061 (0.0025)	0.014

Note. PostBloodPerKG= Post-operative RBC volume per kg. IntraBloodPerKG= Intra-operative RBC volume per kg. CPB= Cardiopulmonary Bypass.. Patients that did not receive any transfusion were assigned a value of 0 for PostBloodPerKG and IntraBloodPerKG and included in this analysis. * RACHS1 Score 0 is treated as the reference group.

Table 4

AIC Comparison of Candidate Models for ICU LOS

Model Type	AIC
Naïve (Normal)	9813.039
Lognormal	6412.99*
Zero-Truncated Negative Binomial	8091.488
Zero-Truncated Poisson	11342.12

Note. * Because AIC is only a valid comparison technique when models have the same outcome measure (untransformed ICU LOS in our case), we use statistical software to obtain the corrected AIC for the lognormal model based on a modified version of the likelihood (as in Burnham & Anderson, 2010).

Table 5

Multivariate Lognormal Regression Estimates for ICU LOS

Parameter	β Estimate (SE)	P Value
Intercept	5.47 (0.87)	<0.001
PostBloodPerKG	0.0009 (0.00032)	0.004
IntraBloodPerKG	0.0024 (0.00064)	<0.001
RACHS1 Score*		
1	-1.62 (1.05)	0.123
2	-0.99 (0.86)	0.255
3	-0.89 (0.86)	0.303
4	-0.87 (0.86)	0.314
5	-0.21 (1.21)	0.865
6	-0.28 (0.86)	0.744
CPB Time	0.002 (0.0008)	0.005
Age (Days)	-0.004 (0.0007)	<0.001

Note. PostBloodPerKG= Post-operative RBC volume per kg. IntraBloodPerKG= Intra-operative RBC volume per kg. CPB= Cardiopulmonary Bypass. This model treats log(ICU LOS) as outcome. Patients that did not receive any transfusion were assigned a value of 0 for PostBloodPerKG and IntraBloodPerKG and included in this analysis.
*RACHS1 Score 0 is treated as the reference group.

Figure 1: STAT and RACHS1 Scores by Complication (Yes/No)

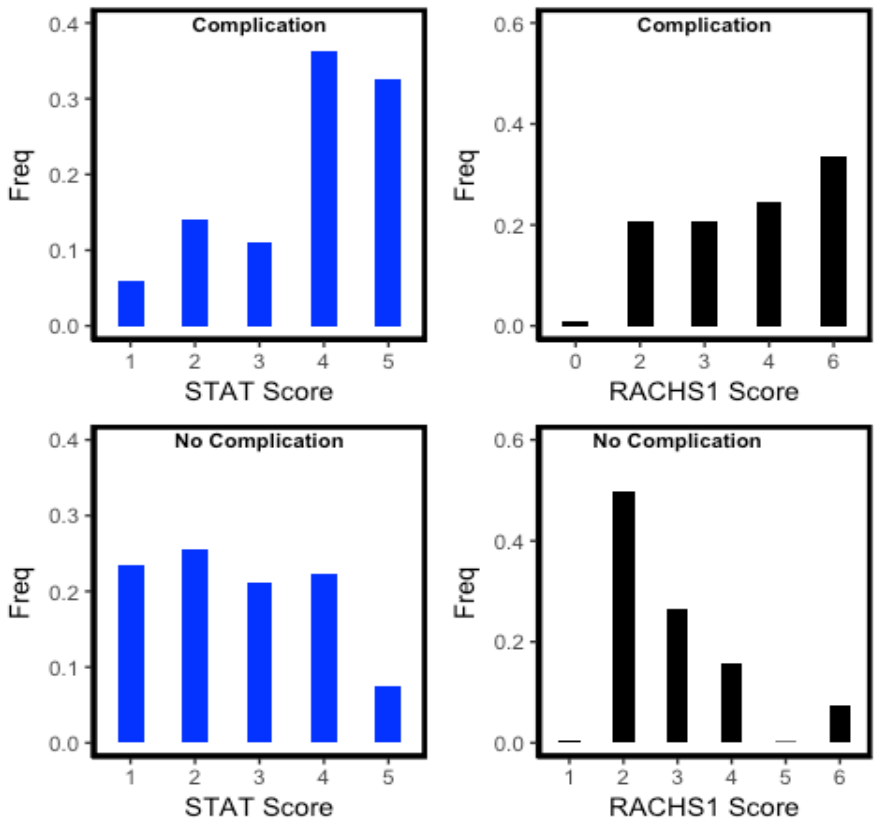

Figure 2: Cox-Snell Residuals for Cox PH Model

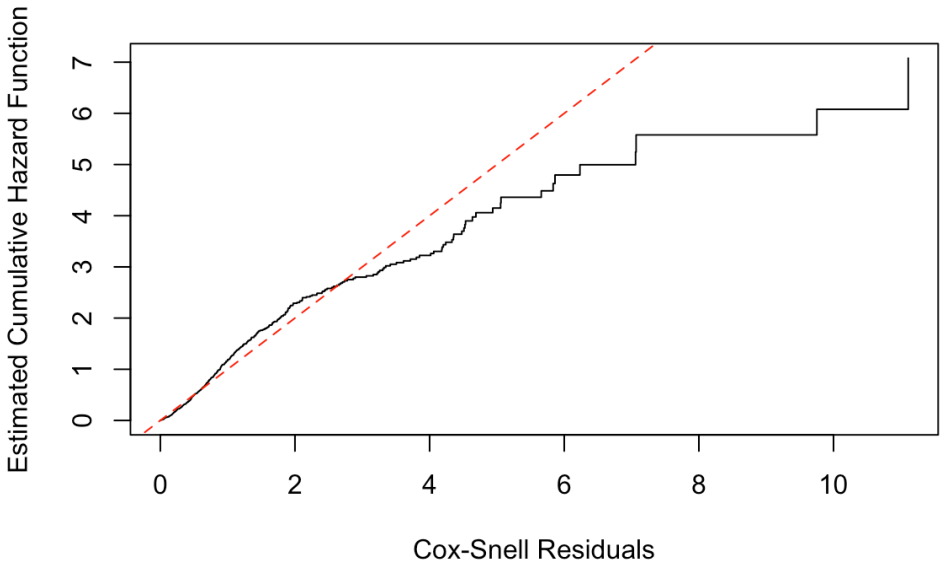

Figure 3: Receiver Operating Characteristics (ROC) Curve for Random Forest Model

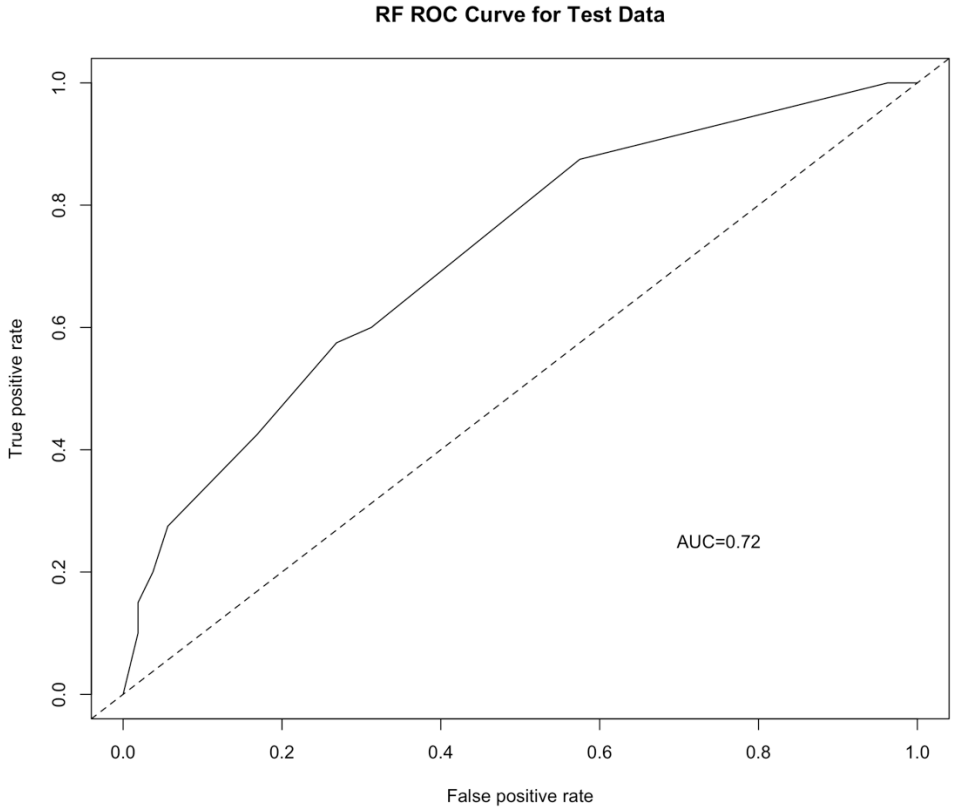

Figure 4: Online Clinical Tool Developed from Random Forest Predictive Model