Classifying Infection Risk Following Pediatric Cardiac Surgery

By

Kaitlin Colleen Williamson, MD

Thesis

Submitted to the Faculty of the

Graduate School of Vanderbilt University

In partial fulfillment of the requirements

for the degree of

Master's of Science

In

Biomedical Informatics

June 30th, 2022

Nashville, Tennessee

Approved:

Daniel Fabbri, PhD

Ritu Banerjee, MD, PhD

Sharon Davis, PhD

ACKNOWLEDGEMENTS

Dr. Williamson received funding support from National Library of Medicine T15 LM007450 grant in completing this work.

We would like to acknowledge the contributions of Dr. Thomas Lasko, Dr. Kathryn Edwards, and Dr. Andrew Smith, who served on Dr. Williamson's pediatric infectious diseases scholarship oversight committee and contributed to this work.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0. INTRODUCTION

Congenital heart disease (CHD) is the most common class of birth defects, affecting one percent of live births; CHD is approximately sixty times more common than all forms of pediatric cancer combined.[1,2] Treatment for CHD often involves one or more surgeries, and unfortunately, postoperative infections are a major cause of morbidity and mortality.[3] Preventive measures are available, but most carry risks, and are best deployed on an individual basis, based upon the particular risk of the patient.

Unfortunately, determining which patients are at highest risk of postoperative infection is not always straightforward. As we will discuss in the following chapters, a few risk scores have been proposed, but each have weaknesses or limitations, including in the definition of infection, the performance difference in "proven" and "unproven" infections, the reliance on only regression without attempting any other machine learning methods, and in their ease of automation for deployment. Both models previously proposed used manually curated data, and as such could not be easily executed automatically in the electronic health record (EHR). This is a major disadvantage, as automatable models can be more accurate, easier to use, can save time, and can help guide targeted interventions in real time.[4] Because of the weaknesses of existing prediction rules for infection following pediatric cardiac surgery, our goal with this thesis was to generate a model or models to overcome the limitations of existing models, and support bedside decision making with a practical, accurate prediction tool.

To achieve this goal, I pursued the following 3 aims:

1. To create a cohort of pediatric cardiac surgery patients at Monroe Carell Jr. Children's Hospital with labeled & validated infection outcomes.
2. To compare performance of regression and machine learning models on validated and unvalidated data.
3. To create and evaluate a bedside prediction rule from the model with the highest performance, as measured by area under the receiver operating characteristic curve (AUC).

In the first manuscript (Chapter 2) of this thesis, we describe the methods and results for Aims 1 and 2, in which we create and validate the cohort, and compare the performance of regression and machine learning models on validated and unvalidated outcomes. This manuscript is submitted and under review for the American Medical Informatic Association (AMIA) Annual Symposium precedings and student paper contest. The second manuscript (Chapter 3) describes the methods and results for Aim 3, in which we propose a bedside decision rule. We then summarize the conclusions learned from completing all three aims.

2.0. Manuscript 1: Classifying Infection Risk Following Pediatric Cardiac Surgery

Corresponding Author:
Kaitlin C Williamson MD
Vanderbilt University Medical Center
Department of Pediatrics, Division of Infectious Diseases
D-7235 Medical Center North
1161 21st Avenue South
Nashville, TN 37232-2581
Kaitlin.williamson@vumc.org

Kaitlin C. Williamson MD[1,2]; Daniel Fabbri PhD[2]

[1]Department of Pediatrics, Division of Infectious Diseases, Vanderbilt University Medical Center, Nashville, TN, USA; [2]Department of Biomedical Informatics, Vanderbilt University, Nashville, TN, USA

2.1. ABSTRACT

Postoperative infections frequently complicate pediatric cardiac surgery, increasing morbidity and cost. If high risk patients could be identified early, preventive measures could mitigate infection risk. In this study, we used structured health data to generate a cohort of pediatric cardiac surgery cases from a single center and used billing codes to assign outcomes for postoperative sepsis, bacteremia, necrotizing enterocolitis, and a composite outcome incorporating all infections. We subsequently validated these outcomes manually using clinical notes and microbiology data. Using this cohort of 2080 surgeries, we trained models to classify the risk of postoperative infections using logistic regression and several machine learning methods. We compared the performance of the models trained on the validated outcomes to those trained on unvalidated outcomes. Manual validation revealed low accuracy of diagnosis codes as classifiers of postoperative infections. Despite significant differences in outcome assignments, similar model performance was achieved using unvalidated and validated outcomes.

2.2. BACKGROUND

Every year in the United States, approximately 40,000 children undergo cardiac surgery, with a postoperative fatality rate of approximately 3%.[5,6] Healthcare-associated infections, including bacteremia, sepsis, and wound infections, are a common postoperative complication which contribute significant morbidity and cost. Despite efforts at reduction, infection rates following cardiac surgery are as high as 30% in many series.[7] These infections are possibly associated with increased mortality[7] and cause significant morbidity, extended patient stays and increased costs,[8] and by driving antibiotic use, may contribute to antibiotic resistance. At Monroe Carell Jr. Children's Hospital at Vanderbilt, approximately 300 cardiac surgeries are performed annually, with surgical site infections complicating ~2-3%, and other confirmed healthcare-associated infections complicating ~3-5% (unpublished data).

If patients at higher risk of infection could be identified prospectively, they could be targeted for clinical interventions that may mitigate infection risk. For example, patients at higher risk of post-operative infectioncould receive enhanced antibiotic prophylaxis,[9] more aggressive screening for colonization and efforts at decolonization,[10] tighter perioperative glycemic control,[11] earlier central line removal or foley catheter removal,[12] prioritization of earlier sternal closure,[13] among other targeted interventions. For the patients at the lowest risk, providers could consider decreasing the use of laboratory tests that detect nonspecific inflammation and could consider reducing or foregoing prophylactic or empiric antibiotic treatment.

Multiple studies have evaluated risk factors for infection following cardiovascular surgery in adults.[14–16] Fewer studies have described risk of infection following cardiac surgery in children. Barker et all used 30,078 records from the Society of Thoracic Surgeons Congenital Heart Surgery Database to generate a multivariate logistic regression model to classify the risk of a composite outcome of septicemia, mediastinitis, or endocarditis following pediatric cardiac surgery.[17] Algra et al used single center data from the Netherlands, including 412 procedures, to train a multivariable logistic regression model to predict surgical site infection, bloodstream infections, urinary tract infections, gastroenteritis, skin infections, and respiratory tract infections.[18] Necrotizing enterocolitis (NEC) is an important cause of morbidity, mortality and infection following pediatric cardiac surgery, but we did not find any published models classifying or predicting the risk of post operative NEC in the pediatric cardiac surgery setting. No prediction rules for postoperative infection following pediatric cardiac use regression methodology. While a very useful and longstanding method for prediction rules, the utility of regression models may be limited when the relationship between features and outcomes is nonlinear. Though the data is somewhat mixed, many recent publications have highlighted the potential of machine learning methods for risk classification in areas ranging from cardiovascular risk,[19] outcomes following neurosurgery,[20] survival following traumatic brain injury,[21] development of AKI following liver transplant,[22] and delayed graft function following renal transplant,[23] to name a few. In our literature search, we found fewer examples of the application of machine learning in the pediatric

population and were unable to find any examples of applying machine learning methods to prediction of infection following cardiac surgery in children.

In developing prediction models for clinical use, the accuracy of outcome labels is critical. Much recent work has explored the difficulty of using structured data elements, including claims data, in generating clinical phenotypes.[24–27] While using structured billing codes is an efficient and automatable means of phenotyping, relying solely on codes can be insufficiently sensitive or specific. One method of improving the accuracy of outcome assignment is via manual validation of unstructured data including clinical notes, but such validations can be effort intensive. Given the effort required for manual validation, it is worth studying whether such validation translates to improved performance of resulting classification models. In this study, our aim was to determine if manual validation could improve the performance of postoperative infection risk classification models for pediatric cardiac surgery patients. To accomplish this aim, we generated a cohort of pediatric patients undergoing cardiac surgery at Monroe Carell Jr. Children's Hospital. We used structured diagnosis codes to make initial outcome label assignments for several postoperative infections. We then followed the initial assignments with manual validation and reassignment of outcome labels based upon review of clinical data. We used these datasets to train logistic regression and machine learning models and compared the performance of the models trained on the validated and unvalidated datasets.

2.3. METHODS

Cohort Creation:

This research was approved by the Institutional Review Board at Vanderbilt University Medical Center. Data for this work was obtained via query of the Vanderbilt Research Derivative (RD), a database of clinical and related data, derived from VUMC's clinical enterprise and repurposed for research.[28] The clinical cohort was generated using structured queries, using patient age, surgery date, and current procedural technology (CPT) codes pertaining to pediatric cardiothoracic surgeries of interest. Included CPT codes were derived from standardized Centers for Disease Control and Prevention (CDC) code sets.[29] Patients were included if their record included a procedure code for at least one procedure of interest, performed between January 1, 2015, and November 30, 2020, on patients up to 6573 days (18.0 years) of age at the time of their surgery.

Feature Selection:

Candidate features were obtained from literature review and discussion with experts in pediatric infectious diseases, pediatric cardiology, and pediatric cardiothoracic surgery. All features were obtained from values available prior to the end of the calendar day on which the surgery was performed. Candidate predictors included patient demographics, vital signs, surgical parameters, drug exposures, and laboratory values. Demographic variables were included to identify patients who might intrinsically be at higher risk of complications, including those at the extremes of age and body size. Demographics of interest included patient age at surgery, gender of record at the time of surgery, weight, and body mass index (BMI).

Vital signs served as an indicator of physiologic stress during surgery and included only values recorded during the operative encounter. Vital signs assessed included minimum and maximum values for heart rate, systolic blood pressure, diastolic blood pressure, and body temperature, as well as the range between the minimum and maximum intraoperative value of each vital sign. Surgical parameters were included to identify potentially more complicated and higher risk procedures, and included the total duration of the procedure, the number of procedure codes listed for the surgical encounter, the American Society of Anesthesiologists (ASA) class assignment,[30] and the case priority level, ranging from elective to emergent. Laboratory values included the maximum and minimum values, and the range between the two, collected on the day of surgery and the three prior days. Lab features were considered as markers of inflammation or poor hemostasis prior to surgery. Candidate lab features included total white blood cell count (WBC), absolute neutrophil count (ANC), hemoglobin level (HGB), platelet count (PLT), C-reactive protein (CRP), procalcitonin, albumin, aspartate aminotransferase (AST), alanine amino transferase (ALT), and blood glucose (GLUC). Prior drug exposures included the number of separate antibiotic prescriptions written in the 30 days prior to surgery, and a binary variable denoting presence or absence of systemic steroid exposure in the 30 days prior to surgery. An additional feature collected was the

number of distinct diagnoses present on the patient's problem list on the day of surgery, which was used as a marker of patient complexity.

Outcome Assignments Using Diagnosis Codes:

Outcomes of interest included postoperative bacteremia, postoperative sepsis, postoperative NEC, and a composite outcome. Outcomes were assigned using a selection of International Classification of Diseases (ICD)-9 and ICD-10 codes derived from the CDC National Healthcare Safety Network (NHSN) code sets; each code was manually validated to ensure inclusion of pertinent codes for the outcomes of interest.[31,32] Cases were assigned an infection label if a corresponding code was added to the medical record at least once in the 30 days following surgery. Additional outcomes, including ventilator associated pneumonia, urinary tract infection, and surgical site infection, were considered, but were ultimately excluded given poor ability to validate reliable outcome assignments from available data, including clinical notes.

Manual Validation:

Following initial outcome assignments, a selection of cases was manually validated via review of clinical notes and blood and body fluid culture results obtained from the RD. Manual review was attempted by a senior pediatric infectious disease fellow on every case initially assigned an infection label based on ICD codes, and for 50 randomly selected cases initially assigned to the negative infection class. A label was confirmed when a clinically relevant culture result was obtained within 30 days postoperatively (for the bacteremia outcome), or when clinical notes confirmed occurrence of the outcome of interest within 30 days of surgery. For the sepsis outcome, documentation of temperature instability (fever or hypothermia) and vital signs changes including tachycardia or hypotension was required. Fever with or without laboratory changes, but without changes in other vital signs, was not considered to represent true sepsis. When medical records clearly indicated the absence of the outcome of interest in the 30 days following surgery, the label was reassigned, and the case was retained in the dataset. Likewise, when records revealed evidence of the condition emerging preoperatively, the label was reassigned to negative, and the case retained. When the record did not contain sufficient data to confirm or overturn a label, the initial outcome label was retained, and the case was retained in the data set.

Data Preprocessing:

Python scikit-learn was used for data preprocessing. Potential features with greater than 20% missingness were excluded; for included features, missing values were imputed to the mean. Scikit-learn's StandardScaler function[33] was used to standardize the dataset prior to model fitting. Potential features for inclusion were evaluated using backward selection, with the top features included in the logistic regression models.

Model Training:

Model fitting was performed using Python Scikit-learn. For each outcome of interest, an eighty-twenty test-train split was randomly performed on the full validated and unvalidated datasets. Models of interest included logistic regression with Ridge penalization, K-nearest neighbors, support vector machine, decision tree, and random forest models. Hyperparameter tuning was performed using Hyperopt with a search grid, as well as with Hyperopt-skl.[34] The best set of hyperparameters was retained for each model, improving the final performance of resultant models by over 10% compared to default hyperparameters. For each outcome of interest in the validated and unvalidated datasets, each model was fit to the training set using ten-fold cross validation within the 80% training set, and subsequently evaluated on the 20% test set. Performance was assessed using the area under the receiver operating characteristic curve (AUC), as well as accuracy, precision, and recall. Bootstrap resampling was performed 100 times with each model to obtain a mean and 95% confidence interval of each performance metric.

Assessment of Feature Importance:

Feature importance in the logistic regression models was analyzed by comparing the coefficients of the features in each model. For the decision tree models, feature importance was evaluated using Shapley (SHAP) scores from each model.

2.4. RESULTS

Cohort Features:

The database query generated a total of 2080 operative encounters. Following validation, 157 cases (7.55%) were positive for at least one infection outcome. The most common outcome was sepsis, which complicated 83 surgeries.

Results of Clinical Outcome Validation:

Manual validation revealed significant error rates in infection outcome assignments, as shown below (Table 1). Overall, a total of 203 outcomes were reassigned, of which 181 were reassigned from a positive label to negative. All but one of the cases reassigned from a negative label to positive had been preliminarily identified with at least one other positive label. The composite outcome was only reassigned from negative to positive in one case, in which the patient met criteria and was treated for culture negative sepsis but was not identified via the code query. In general, the use of billing codes for assignment led to overidentification of infection outcomes, as diagnostic codes were often applied when a diagnosis was merely possible or suspected, but never confirmed. Less frequently, codes were applied postoperatively for a condition that originated preoperatively.

Outcome	Labels Identified by Codes	Positive Labels Validated	Negative Labels Validated	True Positive	False Positive	False Negative	True Negative	Labels Included in Cohort
Sepsis	78 (3.75)	69 (88.46)	199 (9.94)	58 (98.31)	11 (15.94)	16 (8.04)	183 (91.96)	83 (3.99)
Bacteremia	116 (5.58)	116 (100.00)	152 (7/74)	73 (62.93)	43 (37.07)	2 (1.72)	150 (98.68)	75 (3.61)
NEC	90 (4.33)	88 (97.78)	180 (9.05)	34 (38.64)	54 (61.36)	3 (3.41)	177 (98.33)	39 (1.88)
Composite	229 (11.01)	218 (95.20)	50 (2.70)	145 (66.51)	73 (33.49)	1 (2.00)	49 (98.00)	157 (7.55)

Table 1: Results of validation of infection outcome assignments. Abbreviations: NEC (necrotizing enterocolitis). Values presented as n (%)

Feature Importance in Logistic Regression:

For all outcomes in both the validated and unvalidated datasets, the top two predictors identified via regression analysis were the number of diagnoses on the patient problem list at the time of surgery (diagnosis count) and the number of antibiotic courses in the 30 days prior to surgery (antibiotic courses). Other features of importance differed somewhat by outcome and between the validated and unvalidated datasets, but generally included measures of preoperative blood glucose, the patient's age at the time of surgery, intraoperative temperature and heart rate values, and the total duration of surgery. The absolute value of the coefficients for the top predictors for the composite outcome in the validated dataset and unvalidated dataset are shown below (Figure 1).

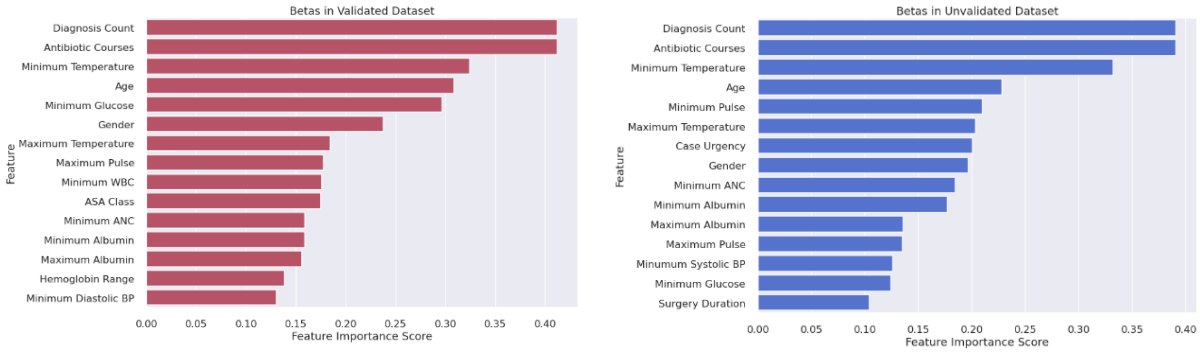

Figure 1. Top predictors of composite outcome for logistic regression models in validated and unvalidated datasets.

Feature Importance in Decision Tree:

In every outcome in the validated and unvalidated datasets, SHAP values identified the same top two predictors as the analyses performed on regression models: the number of diagnoses on the problem list at the time of surgery, and the number of antibiotic courses prescribed in the 30 days preceding surgery. Intraoperative temperature was of higher importance in the regression models, while albumin and WBC counts were of relatively higher importance in the decision trees. The SHAP values for the composite outcome decision trees trained on validated and unvalidated data are presented below (Figure 2).

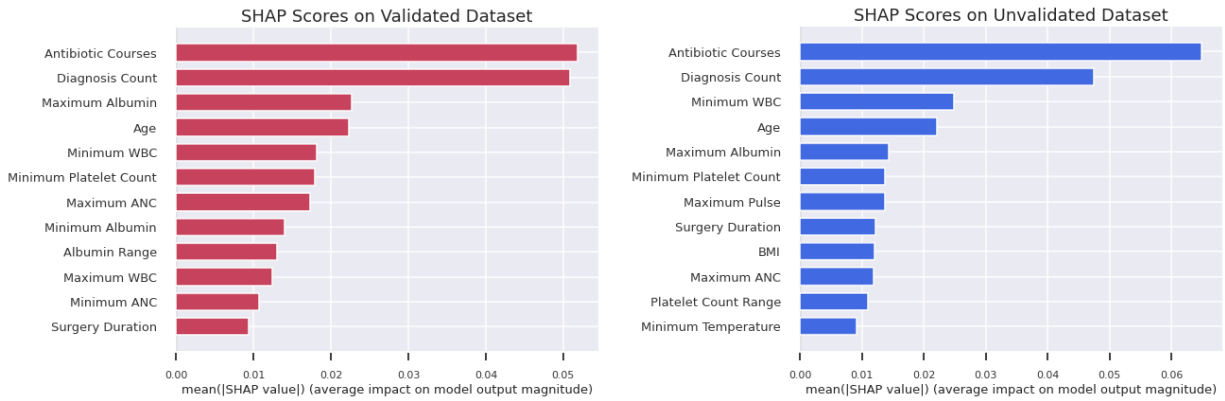

Figure 2. SHAP scores for top predictors of comprehensive outcome decision tree models for the validated and unvalidated datasets.

Model Performance:

For each outcome of interest, the logistic regression model achieved the highest AUC in both the validated and unvalidated datasets, compared to other model types, with one exception: the random forest model for the NEC outcome outperformed logistic regression in the unvalidated dataset only. Despite differences in top features, overall performance of the models in the validated and unvalidated datasets was roughly similar, with no consistent difference in performance between the validated and unvalidated datasets. A comparison of the performance of each model in the unvalidated and validated data sets is presented below (Figure 3).

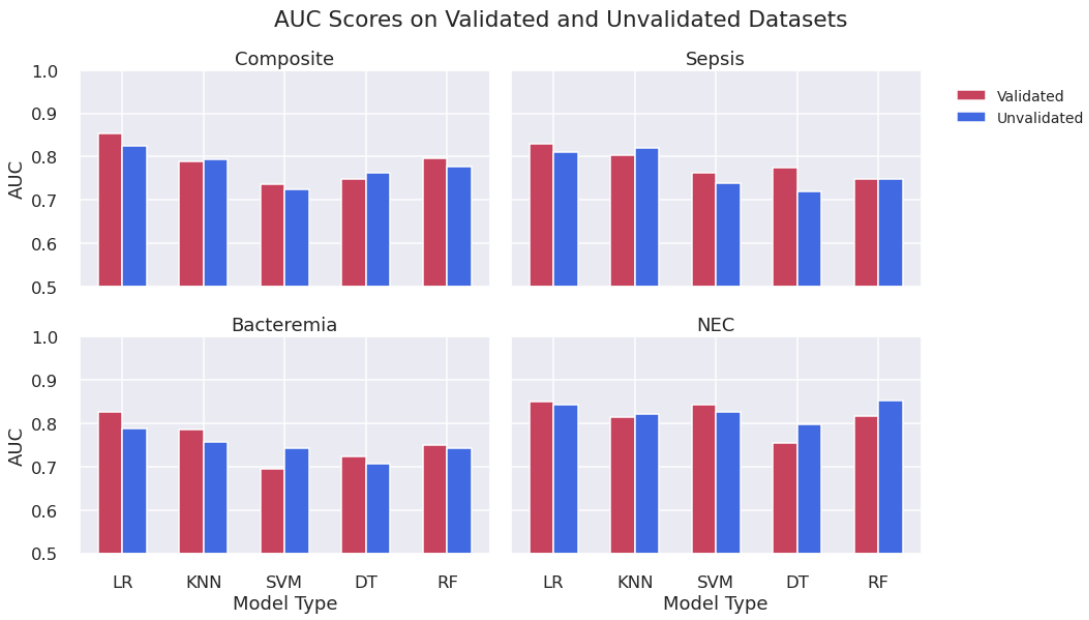

Figure 3. Comparison of AUC of models in validated and unvalidated datasets. LR: logistic regression; KNN: k-nearest neighbors, SVM: support vector machine, DT: decision tree, RF: random forest.

The logistic regression models trained on both the validated and unvalidated datasets achieved an AUC >0.8 for the composite outcome. Cross validated receiver operator characteristic (ROC) curves for the logistic regression model in the validated and unvalidated datasets is presented below (Figure 4).

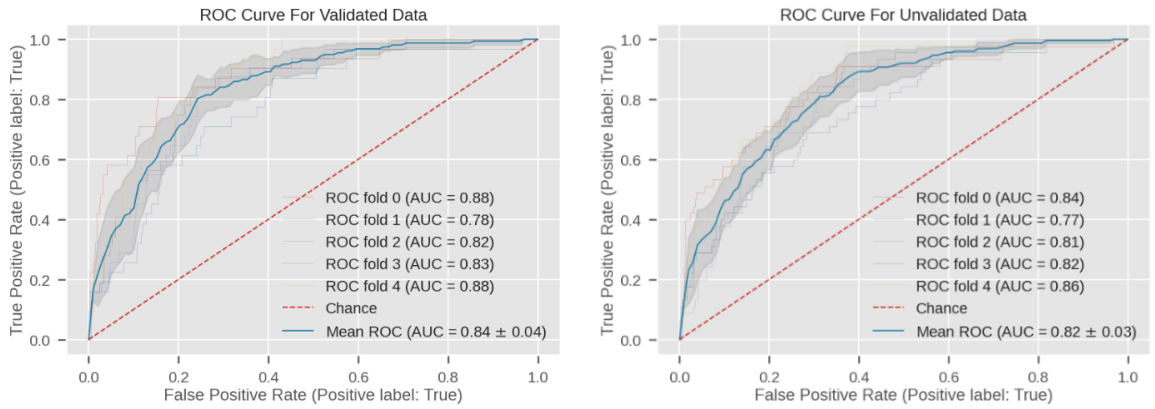

Figure 4. ROC curves for logistic regression models for the composite outcome in validated and validated datasets.

The performance of each model type trained in the validated and unvalidated datasets was evaluated using accuracy, precision, and recall. The full performance metrics of all models is presented below (Table 2).

Model	AUC		Accuracy		Precision		Recall	
	Validated	Un-validated	Validated	Un-validated	Validated	Un-Validated	Validated	Un-validated
Sepsis								
Regression	0.829	0.810	0.963	0.964	0.963	0.964	1.00	1.00
KNN	0.804	0.821	1.00	1.00	1.00	1.00	1.00	1.00
SVM	0.762	0.739	1.00	1.00	1.00	1.00	1.00	1.00
Decision Tree	0.774	0.719	0.966	0.971	0.967	0.971	0.998	0.999
Random Forest	0.748	0.749	1.00	1.00	1.00	1.00	1.00	1.00
Bacteremia								
Regression	0.827	0.789	0.957	0.936	0.960	0.939	0.997	0.997
KNN	0.786	0.756	1.00	1.00	1.00	1.00	1.00	1.00
SVM	0.694	0.742	1.00	0.965	1.00	0.965	1.00	1.00
Decision Tree	0.724	0.708	0.960	0.940	0.960	0.940	1.00	1.00
Random Forest	0.749	0.744	1.00	1.00	1.00	1.00	1.00	1.00
NEC								
Regression	0.850	0.843	0.979	0.954	0.979	0.956	1.00	0.998
KNN	0.814	0.821	1.00	1.00	1.00	1.00	1.00	1.00
SVM	0.843	0.826	1.00	1.00	0.981	1.00	1.00	1.00
Decision Tree	0.755	0.797	0.980	0.956	0.980	0.957	1.00	1.00
Random Forest	0.821	0.853	1.00	1.00	1.00	1.00	1.00	1.00
Composite								
Regression	0.853	0.824	0.928	0.897	0.928	0.928	1.00	0.991
KNN	0.788	0.795	1.00	1.00	1.00	1.00	1.00	1.00
SVM	0.737	0.725	1.00	1.00	1.00	1.00	1.00	1.00
Decision Tree	0.748	0.764	0.929	0.938	0.929	0.941	1.00	0.996
Random Forest	0.796	0.778	0.991	0.993	0.992	0.992	1.00	1.00

Table 2. Full performance metrics of all models in validated and unvalidated datasets.

2.5. DISCUSSION

Our models demonstrate fair to good performance in classifying risk of postoperative infections in children undergoing cardiac surgery, with several models achieving AUC > 0.8 and markers of accuracy, precision, and recall of >0.95. For each outcome of interest that we investigated, logistic regression had the highest AUC, in both the unvalidated and validated datasets, except for the NEC outcome in the unvalidated dataset, where the random forest model was superior. However, in clinical practice, AUC is not always the most useful indicator of model performance. For a potentially preventable complication with significant risks to morbidity and mortality, including the postoperative infections classified with these models, it is often desirable to avoid missing high risk cases while avoiding excess firing. For this reason, precision and recall may be more clinically useful performance metrics. For each outcome, the machine learning models achieved excellent performance in precision and recall, on par with logistic regression. Likewise, despite significant differences in the outcome labels in the validated and unvalidated datasets, the accuracy, precision, and recall were similar between models trained on the two datasets.

Though an increasing number of studies attempting to classify postoperative infection risk are being published, this current work is unique in several ways. Firstly, most studies in this space are focused on an adult population. Children, compared to adults, have different infectious outcomes and risk factors. Studies that do focus on infection risk classification in the pediatric population generally involve different surgical populations, including orthopedic surgeries and abdominal surgeries. Barker et al have published a model using regression to predict the risk of postoperative infection in this population, but their study excludes culture negative sepsis and necrotizing enterocolitis, which are important postoperative infectious complications in this population.[35] Additionally, their study uses only multivariable regression, and does not include other techniques. Lastly, Barker's study relied on a manually curated dataset. In our study, we have contrasted the performance of models in automatically generated versus manually validated outcome datasets. As noted above, though the outcomes in the two datasets varied significantly, the overall model performance was similar.

The models trained on validated and unvalidated data all identified the number of diagnosis codes on the problem list at the time of surgery as a strong predictor of postoperative infection. While it is perhaps unsurprising that this feature would be useful, given its role as a marker of patient complexity, it is novel in several ways. The problem list length feature is simple to extract from the medical record and ~~is~~ has not been previously described as a marker of infection risk in pediatric postoperative patients. This makes it an enticing candidate feature for models run automatically in electronic health record environments. Another useful feature identified by all models was the number of antibiotic courses prescribed in the 30 days preceding surgery. This feature shares similar advantages: it can be automatically extracted and has not been widely described in other studies. This feature may serve as a marker of patient complexity, but also prior antibiotic exposures can lead to microbiome dysbiosis and increase risk of later infection.

In this study, we used diagnosis codes as an initial method of outcome assignment, followed by manual validation. Our assignment criterion, namely a single entry of any relevant diagnosis code in the 30-day period following surgery, was an intentionally low threshold, intended to capture the maximum number of potentially positive cases by reducing false negatives. Our study demonstrates several patterns of error possible when using diagnosis codes for identification of postoperative infections. Some error resulted merely from postoperative coding of a condition that originated preoperatively. In other cases, codes were added when a diagnosis was merely suspected, but later excluded. Additionally, the negative impacts of hospital acquired condition (HAC) reporting may have led to under-use of billing codes associated with certain postoperative infections. Lastly, some postoperative infections can be difficult to diagnose definitively. Sepsis can be difficult to distinguish from other syndromes of postoperative instability. Mediastinitis (a type of surgical site infection) can be difficult to differentiate from a sterile collection of blood or fluid, and ventilator associated pneumonia can be confused with postoperative pulmonary edema or endotracheal tube colonization. Occasionally, different physicians or clinicians treating the same patient can disagree on whether a postoperative infection is present. Though accepted CDC consensus definitions exist, the difficulty in applying these definitions at the bedside contributes imprecision to models attempting to classify postoperative infection risk.

The fairly strong performance of our prediction models even on unvalidated data is an interesting result of this study. While the models trained on unvalidated data may not be the most accurate in classifying the risk of true, proven infection, they may provide useful risk classification for postoperative instability more generally. Additionally, it is encouraging that in validating a subset of cases not labeled as positive by diagnosis codes, only 2% of initially negative cases represented a false negative. This indicates that while single appearance of a billing code pertaining to infection may not be highly specific, the presence of such a code has excellent sensitivity in identifying patients with at least one clinically significant postoperative infection. This is an important finding, as such codes are easily mined from the health record and can be helpful in identifying higher risk patients who may deserve greater consideration by the clinical team.

Limitations of this study include a relatively small sample size, relatively rare infectious outcomes in the data set, the reliance on single center data, missing data, the inability to manually validate all cases, and the lack of an easily interpretable and hand-calculable pencil and paper tool. Missingness of structured data led to the exclusion of several potential features with strong literature support for predictive ability, including procalcitonin[36,37], preoperative mechanical ventilation status[35], and preoperative colonization status[38,39]. The exclusion of these predictors likely decreased the overall performance of the models we trained. Additionally, missing and incomplete documentation limited our ability to definitively confirm or reject every label that we attempted to validate.

Further work is needed to improve the clinical utility of these models. Because these models were trained and validated on single center data, external validation using data from other centers will be critical to assess the

generalizability of these models. To assess whether these models provide useful decision support that is not immediately apparent to human caregivers, it would be valuable to compare the classification performance of the models to that of human clinicians presented with the same preoperative and interoperative data. Finally, prior to deployment, it would be critical to prospectively validate the performance of these models by silently implementing in the electronic health record and evaluating performance over time.

2.6. CONCLUSIONS

Postoperative infections are an important source of morbidity and cost in pediatric cardiac surgery patients; identifying higher risk patients could allow targeted interventions to mitigate risk. In this work, we have generated a pediatric cardiac surgery patient cohort using structured data from a clinical data warehouse. Manual validation of outcome assignments revealed inaccuracies of using diagnosis codes for automatic detection of postoperative infections. However, overall model performance was similar for models trained on validated and unvalidated data sets. Most models achieved good performance, with many achieving AUC >0.8 and accuracy, precision, and recall >0.95, in both validated and unvalidated datasets. While logistic regression achieved the best AUC for each outcome apart from the NEC outcome in the unvalidated dataset, the accuracy, precision, and recall of the machine learning models was comparable to logistic regression. The performance of these models is promising. If successful in further validation, the models could be clinically useful in classifying postoperative infection risk in pediatric cardiac surgery patients, allowing targeted interventions to improve patient outcomes.

3.0. Manuscript 2: Creating a Risk Prediction Rule for Infection Following Pediatric Cardiac Surgery

Corresponding Author:
Kaitlin C Williamson MD
Vanderbilt University Medical Center
Department of Pediatrics, Division of Infectious Diseases
D-7235 Medical Center North
1161 21st Avenue South
Nashville, TN 37232-2581
Kaitlin.williamson@vumc.org

Kaitlin C. Williamson MD[1,2]; Ritu Banerjee MD PhD[1]; Kathryn Edwards MD[1]; Daniel Fabbri PhD[2]

[1]Department of Pediatrics, Division of Infectious Diseases, Vanderbilt University Medical Center, Nashville, TN, USA; [2]Department of Biomedical Informatics, Vanderbilt University, Nashville, TN, USA

3.1. ABSTRACT

Objective:

To create an easily interpretable bedside prediction rule to classify the risk of infection following pediatric cardiac surgery, using data points automatically retrievable from the electronic health record.

Design:

Retrospective chart review.

Setting:

Patient records were abstracted from clinical data from Monroe Carell Junior Children's Hospital at Vanderbilt University Medical Center.

Patients:

Our cohort included 2080 patients from birth to 18.0 years of age, who underwent cardiac surgery between January 1, 2015 and November 30, 2020.

Interventions:

None.

Measurements and Main Results:

We used structured queries of a warehouse of clinical data repurposed for research to generate a cohort of pediatric cardiac surgery patients. We retrieved possible infection predictors. We used diagnosis codes and manual review of clinical notes to assign infection outcomes. We used logistic regression to create a prediction model from which we derived the risk score. Our bedside prediction rule achieved fair performance in classifying post operative infection risk, with an area under the Receiver Operating Characteristic curve of 0.753. The predictors in the final rule included the number of diagnoses on the problem list, minimum preoperative albumin level, the range between minimum and maximum preoperative white blood cell count, maximum platelet count, maximum hemoglobin, and minimum pulse.

Conclusions:

Our bedside prediction rule achieved fair performance in classifying the risk of infection following pediatric cardiac surgery. A pediatric cardiac surgery postoperative infection risk score could allow for targeted infection prevention interventions, improving resource utilization and potentially improving outcomes.

3.2. BACKGROUND

Congenital heart diseases are the most common birth defects, accounting for a third of major congenital anomalies.[1] Every year in the United States, approximately 40,000 children undergo cardiac surgery.[2] Postoperative infections, including surgical site infections, bacteremia, culture-negative sepsis, and abdominal infections, complicate up to 30% of surgeries.[2] Infections cause a substantial burden to patients in the form of excess morbidity,[3] increased length of stay,[4] and mortality.[5]

If patients at higher risk of infection could be identified early, targeted interventions could protect those at highest risk, while reducing unnecessary interventions in those at lower risk. Patients at high risk for subsequent infection could receive tighter glycemic control,[7] enhanced efforts at decolonization,[8] changes in isolation precautions or care clustering, lower risk central lines and earlier removal,[9,10] earlier sternal closure,[11] and changes in skin care and wound care.[12] For patients at lower risk, providers could consider decreasing the use of laboratory tests for nonspecific inflammation and could consider delaying or foregoing empiric antibiotics in situations unlikely to represent true post-operative infection.

Several previous studies have attempted to predict postoperative infection risk following pediatric cardiac surgery. Hatachi et al used 526 surgeries, with 81 cases of healthcare-associated infections, to identify risk factors for pediatric postoperative infection following cardiac surgery.[2] Their regression analysis identified mechanical ventilation greater than or equal to 3 days, dopamine use, genetic abnormality, and delayed sternal closure as risk factors, but did not convert these factors to a calculable score.[2]

Barker et al utilized the Society of Thoracic Surgeons (STS) Congenital Heart Surgery Database to establish a model using data from 30,078 children from 48 centers.[13] Their model was designed for pre-operative application to predict “major infections” which included strict criteria for septicemia, mediastinitis, and endocarditis.[13] Young age, high surgical complexity, previous cardiothoracic operation, preoperative length of stay greater than one day, preoperative requirement for ventilator support, and presence of a genetic abnormality were included in the final risk score. The model was applied using pre-operative variables; performance was good with area under the receiver operating curve (AUC) 0.79 for the final model.[13] Limitations of the model included the strict definition of “serious infection” which likely led to under-identification of clinically significant infections. Necrotizing enterocolitis (NEC) and abdominal infection were also excluded from the infection end point. Additionally, the model used a manually curated dataset, and was not designed or intended for automatic, background use within the electronic health record (EHR).

Algra et. al developed a model using 412 procedures with 102 subsequent infections, with procedures performed between April 2006 and May 2009 at a single center in the Netherlands.[14] They utilized three predictors (age less than six months, postoperative pediatric intensive care unit (PICU) stay greater than 48 hours, and postoperative

open sternum greater than 48 hours) to generate a model intended for application 48 hours following surgery. This model achieved an AUC of 0.78, or 0.72 when using only proven infections, defined as those confirmed by either culture or polymerase chain reaction.[14] This model did not investigate any laboratory values, which have shown utility in other analyses. Like Barker, this model did not use structured data amenable for automatic extraction from the EHR.

Any risk score will have limitations. An ideal score is accurate, interpretable, actionable, and easy to use. Many proposed scores lack sufficient accuracy to discriminate between cases with infection and cases without. Others use advanced machine learning methods, but subsequently are less interpretable; the "black box" problem of machine learning models obscures why a patient received a high or low score.[15] Others yield less actionable results; a score applied well after surgery might be less useful than a score available on the day of surgery, when immediate postoperative care can still be impacted. Additionally, a score trained on a very strict and rigid definition of infection may fail to predict many other clinically important cases that fail to meet the strict outcome definition. Lastly, many scores, such as Barker's and Algra's, require manually curated data, and are not designed to run autonomously within the EHR. A score using structured, retrievable data could run automatically in the EHR, reducing the burden on busy health workers. An automatic score can run in the background, generating an alert only when prompted or if a patient's score exceeds a certain threshold.

Our aims in this study were to address some of the limitations of past risk scores, by generating a risk score using predictors that could be automatically extracted from an EHR on the day of surgery, to classify the risk of clinically relevant postoperative infections following pediatric cardiac surgery.

3.3. METHODS

Cohort Creation

This research was approved by the Vanderbilt University Medical Center (VUMC) Institutional Review Board, with waiver of consent. In this retrospective study, we included all cardiac surgery procedures performed at VUMC between January 1, 2015, and November 30, 2020, on patients under 18 years of age. We obtained the data for this study via query of the Vanderbilt Research Derivative (RD),[28] a structured data warehouse, abstracted from the EHR, and repurposed for research. The RD contains data on patient demographics, encounters, diagnoses, laboratory values, and physician and ancillary notes. We included any record containing a current procedural technology (CPT) code for a cardiac surgery, derived from Centers for Disease Control and Prevention (CDC) code sets.[50] We excluded codes that would not require a surgical incision in the chest (such as transcutaneous catheterization procedures).

Clinical Predictor Selection and Retrieval

Candidate predictors of interest were selected via literature review. Only predictors that would be available on or before the day of surgery, and retrievable via automated queries of the structured data warehouse, were considered.

Possible predictors included patient demographics, intraoperative vital signs, surgical parameters, preoperative and intraoperative drug exposures, and preoperative and intraoperative laboratory values. Demographics of interest included patient age at surgery, biological sex of record at the time of surgery, weight, and body mass index (BMI).

Vital signs served as an indicator of physiologic stress and included only values recorded during the operative encounter. Vital signs included minimum and maximum intraoperative values for heart rate, systolic blood pressure, diastolic blood pressure, and body temperature. We hypothesized that vital sign lability, independent of extreme values, might be an important predictor of subsequent infection. For this reason, in addition to maximum and minimum values, we also calculated a range value, defined as the difference between the highest and lowest recorded value for each vital sign parameter within the operative encounter.

Surgical parameters included total duration of the surgery, the number of procedure codes listed for the surgical encounter, the American Society of Anesthesiologists (ASA) class assignment,[18] and the case priority level, ranging from elective to emergent. An additional binary feature was captured, reflecting whether the patient had undergone a previous surgical procedure in the 30 days preceding the surgery of interest.

Laboratory parameters included the maximum and minimum values, and the range between the two, of all values collected on the day of surgery and the three prior days. Candidate lab features included total white blood cell count, absolute neutrophil count (ANC), absolute lymphocyte count, hemoglobin level, platelet count, c-reactive protein (CRP), procalcitonin, albumin, aspartate aminotransferase, alanine amino transferase, and blood glucose.

Prior drug exposures included the total number of separate antibiotic prescriptions prescribed inpatient or outpatient in the Vanderbilt medical system in the 30 days prior to surgery, and two binary variables denoting presence or absence of any systemic steroid exposure, or insulin exposure, in the 30 days prior to surgery. An additional feature was the number of distinct diagnoses present on the patient's problem list on the day of surgery.

Preliminary Outcome Assignment

The primary outcome of this study consisted of a composite metric of any infection occurring from one day to 30 days postoperatively. Infections included in the composite outcome were bacteremia, fungemia, endocarditis, clinical (including culture negative or viral) sepsis, NEC, superficial and deep surgical site infections. A preliminary outcome assignment was made using a set of International Classification of Diseases (ICD)-9 and ICD-10 codes, derived from the CDC National Healthcare Safety Network (NHSN) code sets.[29] Each code in the target code set was individually reviewed for inclusion or exclusion based on relevance. Codes that were not adequately specific or accurate (i.e., hypotension) were excluded, and additional codes were added via keyword queries of the Athena Database.[31] For each procedure record included in the cohort, a query for any of the diagnosis codes in the code set was performed. If the clinical record included one or more entries of any code in the infection code set between 1 and 30 days postoperatively, the case was preliminarily labeled as "infection present." If no entry of any diagnosis code in the infection code set was created between one and 30 days postoperatively, the case was preliminarily labeled as "infection absent."

Manual Validation of Outcomes

Following preliminary outcome assignments, a selection of cases was manually validated via review of clinical notes and blood and body fluid culture results. All data used for manual review was obtained from the RD data warehouse, and the review was performed by Dr. Williamson. Manual review was attempted for every case preliminarily labeled as "infection present", and for 51 randomly selected cases initially labeled as "infection absent". Multiple notes and culture results were reviewed for each case, including progress notes, operative notes, consultation notes, and discharge summaries. A detailed description of infection definitions is included in the supplemental material. An assignment to the "infection present" label was confirmed when clinical notes indicated diagnosis of, and treatment for, one of the infection types included in the composite. The case was reassigned to the "infection absent" label when medical records clearly indicated the absence of the outcome of interest in the 30 days following surgery, such as a discharge summary 30 or more days following surgery noting "no infectious concerns" supported by absence of mention of infection on prior daily progress notes, or b) notes indicating, "sepsis rule out performed, but sepsis excluded" coupled with lack of mention of true infection in any other clinical notes. Likewise, when records revealed evidence of the condition emerging preoperatively, the label was reassigned to "infection absent" unless evidence was found of a separate infection meeting the above criteria,

arising postoperatively. When the record did not contain sufficient data to definitively confirm or overturn the preliminary outcome label per the above criteria, the preliminary outcome label was retained, and the case was retained in the data set.

Statistical Analysis

All analyses were performed using Python, including the sckit-learn and statsmodel packages.

Before performing logistic regression, all possible predictors with greater than 20% missing data were excluded from further analysis. For the remaining predictors, missing data points were imputed to the mean. Because the risk score assigns points for a value above (or below) a given threshold, continuous variables were dichotomized to binary variables prior to further analysis. This was achieved by finding an optimal threshold value for each continuous variable, defined as the cut point that maximized the explained variance in the composite outcome. After all candidate predictors were converted to binary variables, univariate logistic regression was performed for each binary predictor variable, and those with p value <0.05 in univariate analysis were included as candidates for the multivariate logistic regression model. Stepwise backward logistic regression was performed to reduce the number of predictors to six, resulting in a reduced logistic regression model. This reduced model was trained on 80% of the full data set (the training set), repeating the analysis across 100 bootraps. Following training, the reduced logistic regression model was validated on the remaining 20% (the test set).

The reduced multivariate logistic regression model was then translated into a risk prediction rule. For each predictor in the model, risk score point values were determined by obtaining the coefficients for each predictor in the reduced multivariate logistic regression model, multiplying the coefficient by 10, and rounding the result to the nearest whole number. A score was calculated for every case in the cohort, by summing the points assigned (or not assigned) for each score criterion. The sensitivity, specificity, positive and negative predictive values, and overall rate of infection were calculated for different score thresholds and ranges. Balancing sensitivity and specificity at different total score thresholds, score ranges for high, medium, and low risk were determined.

3.4. RESULTS

An overview of full methods and abridged results is presented in Figure 5.

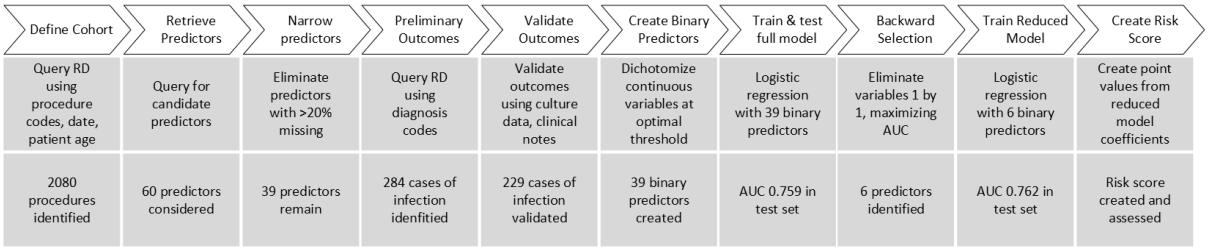

Figure 5. Overview of study methods. Description of methods and brief results from cohort creation to predictor retrieval, to preliminary and outcome assignment, conversion to binary predictors, logistic regression modeling, and rounding of coefficients to generate a calculable risk score.

A total of 2080 procedures were included, performed on a total of 1681 unique patients. Prior to outcome validation, 284 cases were preliminary labeled as “infection present”; after outcome validation, 229 (11%) were labeled as “infection present.” The most common category of validated postoperative infection was surgical site infection, which occurred following 104 cases. Additionally, there were 83 cases of sepsis, 77 cases of bacteremia, candidemia, or endocarditis, and 39 cases of NEC. Some procedures were followed by infections from multiple classes.

Originally, 59 predictors were considered; following the elimination of predictors with more than 20% missing data, 38 candidate predictors were considered for inclusion in the final model. Table A in the supplement presents these 38 candidate predictors, along with the identified optimal threshold cutoff, and the odds ratio for the binary variable between cases with and without infection. After analyzing p values in univariate regression, 37 of the 38 variables remained as candidate predictors for the risk score. The only binary predictor with p value >0.05 on univariate regression was maximum diastolic blood pressure <52.

The six binary predictors remaining after backward selection, and thus included as components in the final risk score, are presented in Table 3. For each component, the number and percentage of cases with and without infection that met the score criteria, as well as the odds ratio and p value for the association, are presented as well. The reduced model (the logistic regression model using these six predictors) achieved an AUC of 0.762, an accuracy of 0.898, precision of 0.903 and recall of 0.991 in the test set, averaged across 100 bootstraps.

Predictor	Procedures without Infection n =1851 (%)	Procedures with Infection n=229 (%)	Odds Ratio (95% CI)	P value
Diagnosis Count >54	83 (4.5)	68 (29.7)	9.00 (6.28 - 12.8)	<0.001
Minimum Albumin <2.7	467 (25.2)	135 (59.0)	4.26 (3.21 -5.65)	<0.001
White Blood Cell Range > 11.1	114 (6.2)	49 (21.4)	4.15 (2.87 -5.99)	<0.001
Maximum Hemoglobin > 17.27	177 (9.6)	59 (25.8)	3.28 (2.35 -4.58)	<0.001
Maximum Platelet Count < 216	532 (28.7)	110 (48.0)	2.29 (1.74 -3.03)	<0.001
Minimum Pulse >32	1140 (61.6)	167 (72.9)	1.68 (1.24 – 2.28)	<0.001

Table 3. Dichotomized predictors included in reduced model and risk score. Procedures exceeding the threshold in the "infection present" and "infection absent" groups are presented as n (%).

In Table 4, we present the coefficient of each predictor in the reduced, six-predictor logistic regression model, as well as the point values assigned in the risk score to cases meeting the condition. Having greater than 54 diagnoses on the problem list had the largest coefficient in the reduced model, and thus the most points assigned in the risk prediction score. Having a minimum pulse greater than 32 beats per minute had the smallest coefficient in the reduced logistic regression model, and thus had the fewest points awarded in the risk prediction score.

Predictor	Multivariate Model Coefficient	Risk Score Point Value
Diagnosis Count >54	1.12	11
Minimum Albumin <2.7	0.788	8
White Blood Cell Range > 11.1	0.611	6
Maximum Hemoglobin > 17.27	0.545	5
Maximum Platelet Count < 216	0.502	5
Minimum Pulse >32	0.498	5

Table 4. Multivariate model coefficients and resulting risk score point values.

The risk prediction score had an AUC of 0.753, accuracy of 0.897, precision of 0.901, and recall of 0.993 in the test set, averaged across 100 bootstraps. The receiver operating characteristic curve of the risk prediction score is presented in Figure 6.

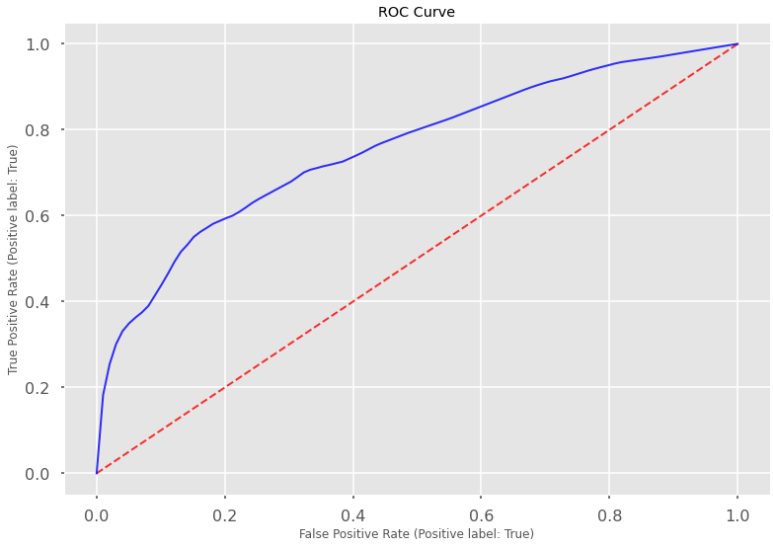

Figure 6. ROC curve for the risk prediction score.

The average score result in the dataset was 9.4, with a minimum value of 0 and a maximum value of 40. The average score with infection present was 16.34 (95% confidence interval 15.09-17.60), compared to 7.94 (95% confidence interval 7.65-8.22) with infection absent; p value <0.001. The distribution of scores in the infection present group and infection absent group are presented in Figure 7.

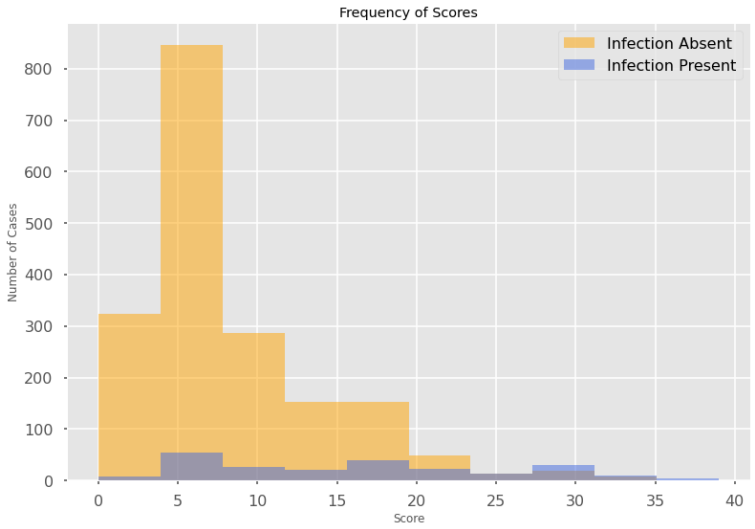

Figure 7. Distribution of score values for cases with infection present compared to infection absent.

Different total score thresholds yielded differences in sensitivity, specificity, positive predictive value, and negative predictive value for postoperative infection. The performance of the prediction rule for classifying infection risk, at various score thresholds is presented in Table 5.

Cutoff Value	Cases Above Threshold	Sensitivity	Positive Predictive Value	Specificity	Negative Predictive Value
>=5	1749	96.5%	12.6%	17.5%	97.5%
>=6	1584	93.4%	13.5%	26.0%	96.7%
>=7	933	74.7%	18.3%	58.8%	94.9%
>=9	847	72.5%	19.5%	63.2%	94.9%
>=11	834	71.6%	19.7%	63.8%	94.8%
>=12	572	63.3%	25.3%	76.9%	94.5%
>=14	487	58.5%	27.5%	80.9%	94.0%
>=15	383	53.7%	32.5%	86.0%	93.5%
>=17	353	51.1%	33.1%	87.3%	93.2%
>=20	168	34.9%	48.2%	90.1%	92.7%
>=25	88	23.1%	60.2%	98.1%	91.2%
>=30	60	15.7%	60.0%	98.7%	90.4%
>=31	22	5.68%	59.1%	99.5%	89.5%
>=35	12	2.62%	66.7%	99.8%	89.2%
>=36	5	1.75%	80.0%	99.9%	89.2%

Table 5. Performance of the infection risk score when total score exceeds different thresholds.

Based upon the performance of the model, score cutoff ranges of 0-4, 5-24, and 25 and above were selected to define low, moderate, and high-risk groups respectively. With these groupings, the rate of infection ranged from 2.41% in the lowest group, to 60.2% in the highest group. The number of total cases and positive cases, and the probability of infection in each risk group, are presented in Table 6.

Score Range	Total Cases	Positive Cases	Rate of Infection (%)
0-4	331	8	2.41
5-24	1661	168	10.1
25+	88	53	60.2

Table 6. Probability of infection in different risk score ranges.

3.5. DISCUSSION

In this study, we propose a risk score to classify the risk of infection following pediatric cardiac surgery, capable of running autonomously in an EHR. Our model uses six predictors: the number of problems on the problem list, the minimum preoperative albumin value, the range between minimum and maximum white blood cell count, the maximum hemoglobin concentration, the maximum platelet count, and the minimum pulse. This risk score achieved fair performance, with an AUC of 0.753. In the past decade, Algra and colleagues and Barker and colleagues created predictive scores for postoperative infection following pediatric cardiac surgery.[13] Though these score rules performed well, the risk score we propose here is unique in several ways and offers distinct advantages.

The top predictors proposed in our model include novel additions, likely representing structured data fields that represent previously described associations. The number of diagnoses on the problem list may function similarly to the presence of genetic abnormality used by Barker et al, as a marker of patient complexity. Low albumin has been associated with postoperative infection following pediatric surgery[21] It may serve as an indicator of poor preoperative nutrition, or protein losing enteropathy, either of which may increase the risk of postoperative infection.[21–26] White blood cell count range could be a marker of physiologic stress, manifested by rapid demarginalization of leukocytes, or of an early infection not yet clinically apparent at the time of surgery. Thrombocytopenia after cardiopulmonary bypass has been associated with postoperative infection and mortality in adults,[27] and has been associated with early mortality in children receiving post-cardiotomy veno-arterial extracorporeal membrane oxygenation.[28] Interestingly, minimum operative pulse greater than 32 beats per minute may indicate the absence of cardiopulmonary bypass. It is surprising, based on prior literature, that the absence of bypass would be a predictor of postoperative infection, but this appears to be the case in our cohort.

Another unique feature of our risk score is the comprehensive outcome, with the inclusion of NEC and clinical sepsis in the predicted endpoint. NEC is an important complication of pediatric cardiac surgery, occurring in as many as 1.8% of children with congenital heart disease.[29–34] NEC involves bacterial translocation, is treated with antimicrobials, and can progress to sepsis or bloodstream infection if not detected early.[29] Despite the importance of this endpoint, it has been excluded from previous models in this space. Barker et al utilized an even more strict definition of infection, potentially missing many clinically important cases. Likewise, culture negative sepsis is an important endpoint; relying on a strict definition of culture positivity can miss relevant cases where cultures may be negative due to fastidious organisms, small culture volumes, or antibiotic pretreatment.

Additionally, our model is intended for application at a unique timepoint, which is at the completion of the operative encounter. Models intended for preoperative use, such as that proposed by Barker et al,[13] exclude intraoperative events which can serve as strong predictors. However, models such as that proposed by Algra et al for implementation 48 hours after surgery,[14] risk missing a critical window where perioperative intervention may

improve outcomes. In this study, we propose a time point that maximizes useful intraoperative predictors, but still allows a window of opportunity for intervention.

Finally, our model is unique in that the predictors, and thus the model itself, can be automatically abstracted from EHR data, without requiring manual implementation by human operators, reducing required clinician effort and potential errors. However, the model is simple enough to be implemented manually if preferred and is easily interpretable.

The risk score we propose here does have several limitations. First, our use of single center data may limit generalizability, due to institutional practice patterns and local patient demographics. Additionally, potentially important features were eliminated due to missing or inconsistent data. Finally, the final months of our study period occurred after the start of the COVID-19 pandemic, which significantly changed patterns of pediatric cardiac surgery as well as hospital staffing and infection control efforts. It is possible that this had unmeasured effects on the performance of our model.

Because our data was extracted from structured EHR fields and structured into an interpretable score, there are unique opportunities for clinical application of our risk score. The model is easily interpretable, which should increase confidence in the score compared to machine learning models subject to the “black box” phenomenon. However, unlike the risk scores previously published for this application, this model is designed to make use of structured data that could be automatically extracted from the EHR in real time. For this reason, this model could run automatically in the EHR, allowing for a best practice alert to be displayed to clinicians. Such an alert could be displayed to pediatric cardiac critical care physicians upon a patient’s transfer to the intensive care unit following surgery and could prompt a risk stratified intensive care unit order set. Orders for high-risk patients could include tight glycemic control,[7] careful management of enteral feeds,[35,36] enhanced decolonization efforts,[8] early sternal closure,[37] early line removal, or inflammatory laboratory monitoring. Limiting these interventions to the children at highest risk could reduce the risks of antibiotic resistance, hemodynamic instability,[11] false positive lab results and excess cost in lower risk patients.

Prior to deploying this risk score for clinical use, the generalizability and performance of this model will need to be confirmed via validation in different patient populations and different surgical centers. The model should also be validated using data from 2021 and later, to assess the impact of the COVID-19 pandemic and its aftermath, as well as evolving surgical practice, on model performance.

3.6. CONCLUSIONS

In this study, we have created an interpretable risk score intended to be applied at the conclusion of surgery~~,~~ and automated within the EHR, which can classify children as low, moderate or high risk of infection following pediatric cardiac surgery. The score achieved an AUC of 0.753 in internal validation. Our model contributes new potential predictors for postoperative infection following pediatric surgery, including the number of problems on the problem list, the minimum preoperative albumin value, the range between minimum and maximum white blood cell count, the maximum hemoglobin concentration, the maximum platelet count, and the minimum pulse. Our model uses structured clinical data, and as such could be automatically run within the EHR. Should the model perform well in external validation, this risk score could be a clinically useful tool to individualize infection prevention efforts to the risk of the individual child.

4.0. SUMARRY

We successfully completed all three aims of this thesis. In our first manuscript we addressed aims 1 and 2; we created a cohort of pediatric cardiac surgery patients at Monroe Carell Jr. Children's hospital, and manually validated a selection of the infection outcomes assigned via diagnosis code queries. Results of this first section demonstrated significant inaccuracies in the code-based infection outcome assignments, with both false positive and false negative labels being relatively common. Codes were most accurate for assignment of the sepsis outcome, and least accurate for the assignment of the NEC outcome. Often available clinical data was insufficient for manual review of the infection outcome, most notably in the case of surgical site infections.

In aim 2 of our study, we compared the performance of logistic regression and machine learning models trained on data with validated and unvalidated infection outcomes. This was a critical goal of this work, as validation via manual review is time intensive, but often is linked to improved model performance.[4,25,67] While our models trained on validated data trended toward improved classification performance, the difference was non-significant; for the logistic regression models predicting the composite outcome in aim 2, the mean (95% confidence interval) AUC in validated data was 0.853 (0-843-0.863), compared to 0.824 (0.802 -0.845) in unvalidated data. The lack of clinical significance of the performance difference is an interesting result of this study. Unvalidated data is much more straightforward to obtain and use; and our study shows that even data with unvalidated labels can be used to models with good performance in classifying the risk of infection follow pediatric cardiac surgery. Furather, it is possible that the "inaccurate" labels could even be clinically useful. Our false positive labels almost always identified a patient that was clinically unstable, with enough signs of infection to warrant a brief clinical "rule out" with blood cultures and empiric antibiotics. Though these patients did not ultimately meet the strict definition of infection we established, a model that predicts such cases could still be of clinical utility. This possibility warrants future study.

In the second manuscript, we addressed aim 3, by creating and internally validating an interpretable risk score to classify the risk of postoperative infection following pediatric cardiac surgery. This interpretable risk score differed from the models generated in the first manuscript in several ways. First, the top features differed between aim 2 and 3. In aim 2, the number of preoperative antibiotic courses, as well as the age of the patient, were very strong predictors of later infection. For the bedside decision rule, neither of these features were selected for inclusion in the final model. Likely this difference pertains to variable information loss from dichotomizing these continuous variables. In addition, there may also be differences introduced by the re-inclusion of the surgical site infection outcome in the composite outcome.

In addition to differing in top features, the interpretable model in aim 3 had lower overall performance than the best performing models in the first manuscript. A portion of the performance loss came from dichotomization of the continuous variables, and the resultant information loss; however, even a full model with all candidate predictors, prior to dichotomization, had lower performance when the surgical site infection was reentered into

the composite outcome (AUC = 0.766). This performance difference may partially be explained by differences in features predictive of surgical site infection compared to other infectious outcomes. Additionally, as surgical site infection was subject to the highest rate of failed validation, it is possible that this outcome remained particularly inaccurate even after attempted manual validation.

In aim 3, we created a hand calculable risk score, as this is a type of risk metric frequently used and accepted by most physicians and healthcare providers. However, because even our bedside model is optimized for automatic use by the electronic health record, it would likely be more valuable for later clinical deployment to utilize a logistic regression model that preserves continuous variables, eliminating information loss. A logistic model could still be largely interpretable to bedside providers, as features and coefficients yield insight into how a score is calculated. Future work will be needed prior to clinical implementation of these classification models. As our models were trained on single center data, the models will need to be validated on data from other centers with different practice patterns and patient demographics, prior to considering real-time clinical implementation. Additionally, it will be important to evaluate these models over time, to ensure that evolving medical practice or changes to infection epidemiology do not impact the accuracy of the model over time. The performance of the model could be assessed by running it as a silent best practice alert (BPA) within the EHR, using unvalidated data, and comparing performance in classification to the results presented here. Should the model perform well, it could be unmasked to provide a visible best practice alert, or order set recommendation, to clinicians caring for postoperative patients.

This body of works adds significant value to the field. First, we evaluated the performance of models trained on validated and unvalidated data, and found the difference to not be statistically significant. This result raises the possibility of running a model on unvalidated outcomes obtained via diagnosis code criteria. A model with good performance on unvalidated outcomes would save time and allow automatic implementation within the EHR. Likewise, we expanded the definition of postoperative infection to include NEC and culture negative sepsis, which are important post operative infections. Lastly, and perhaps most importantly, previously proposed models for this purpose do not use structured EHR data. By using structured data, our models may be run automatically within the EHR.

By addressing these limitations of past models, namely the reliance on manually validated outcomes, the limitation in the definition of infection that could miss clinically important cases, and the reliance on manual implementation, we open the door for a comprehensive model that could be automatically implemented by the EHR without manual calculation by clinicians. This benefit is significant, allowing the model to be integrated seamlessly and effortlessly into clinicians' workflow, and potentially fueling clinical decision support to clinicians at the bedside. Providing such timely support could help providers individualize care based upon patient risk, with the hope of reducing postoperative infections, and improving patient outcomes.

5.0. REFERENCES

1. Get the Facts About Congenital Heart Disease | Children's Hospital of Philadelphia. Accessed May 30, 2022. https://www.chop.edu/pages/get-facts-about-congenital-heart-disease
2. What are Congenital Heart Defects? | CDC. Accessed May 30, 2022. https://www.cdc.gov/ncbddd/heartdefects/facts.html
3. Hatachi T, Tachibana K, Inata Y, et al. Risk Factors for Healthcare-Associated Infections After Pediatric Cardiac Surgery*. *Pediatric Critical Care Medicine*. 2018;19(3):237. doi:10.1097/PCC.0000000000001445
4. Goldstein BA, Navar AM, Pencina MJ. Risk Prediction With Electronic Health Records: The Importance of Model Validation and Clinical Context. *JAMA Cardiol*. 2016;1(9):976. doi:10.1001/JAMACARDIO.2016.3826
5. Hatachi T, Tachibana K, Inata Y, et al. Risk Factors for Healthcare-Associated Infections After Pediatric Cardiac Surgery. *Pediatr Crit Care Med*. 2018;19(3):237-244. doi:10.1097/PCC.0000000000001445
6. Jacobs JP, He X, Mayer JE, et al. Mortality Trends in Pediatric and Congenital Heart Surgery: An Analysis of The Society of Thoracic Surgeons Congenital Heart Surgery Database. *Ann Thorac Surg*. 2016;102(4):1345-1352. doi:10.1016/J.ATHORACSUR.2016.01.071
7. Grisaru-Soen G, Paret G, Yahav D, Boyko V, Lerner-Geva L. Nosocomial infections in pediatric cardiovascular surgery patients: a 4-year survey. *Pediatr Crit Care Med*. 2009;10(2):202-206. doi:10.1097/PCC.0B013E31819A37C5
8. Tweddell S, Loomba RS, Cooper DS, Benscoter AL. Health care-associated infections are associated with increased length of stay and cost but not mortality in children undergoing cardiac surgery. *Congenital Heart Disease*. 2019;14(5):785-790. doi:10.1111/CHD.12779
9. Antibiotic Prophylaxis for Surgical Procedures: A Systematic Review [Internet]. *Antibiotic Prophylaxis for Surgical Procedures: A Systematic Review*. Published online 2010. Accessed February 10, 2022. https://pubmed-ncbi-nlm-nih-gov.proxy.library.vanderbilt.edu/28876768/
10. Savary L, de Luca A, el Arid JM, et al. Systematic skin and nasal decolonization lowers Staphylococcus infection in pediatric cardiac surgery. *Archives de Pédiatrie*. 2022;29(3):177-182. doi:10.1016/J.ARCPED.2022.01.009
11. Agus MSD, Asaro LA, Steil GM, et al. Tight Glycemic Control after Pediatric Cardiac Surgery in High-Risk Patient Populations: A Secondary Analysis of the Safe Pediatric Euglycemia after Cardiac Surgery Trial. *Circulation*. 2014;129(22):2297. doi:10.1161/CIRCULATIONAHA.113.008124
12. Andrioli ER, Furtado GHC, Medeiros EA. Catheter-associated urinary tract infection after cardiovascular surgery: Impact of a multifaceted intervention. *Am J Infect Control*. 2016;44(3):289-293. doi:10.1016/J.AJIC.2015.09.030
13. Nelson-McMillan K, Hornik CP, He X, et al. Delayed Sternal Closure in Infant Heart Surgery-The Importance of Where and When: An Analysis of the STS Congenital Heart Surgery Database. *Ann Thorac Surg*. 2016;102(5):1565-1572. doi:10.1016/J.ATHORACSUR.2016.08.081
14. Gatti G, Rochon M, Raja SG, Luzzati R, Dreas L, Pappalardo A. Predictive models of surgical site infections after coronary surgery: insights from a validation study on 7090 consecutive patients. *Journal of Hospital Infection*. 2019;102(3):277-286. doi:10.1016/j.jhin.2019.01.009
15. Raja SG, Rochon M, Jarman JWE. Brompton Harefield Infection Score (BHIS): Development and validation of a stratification tool for predicting risk of surgical site infection after coronary artery bypass grafting. *International Journal of Surgery*. 2015;16(Part A):69-73. doi:10.1016/j.ijsu.2015.02.008
16. Nieto-Cabrera M, Fernández-Pérez C, García-González I, et al. Med-Score 24: A multivariable prediction model for poststernotomy mediastinitis 24 hours after admission to the intensive care unit. *J Thorac Cardiovasc Surg*. 2018;155(3):1041-1051.e5. doi:10.1016/J.JTCVS.2017.09.160
17. Barker GM, O'Brien SM, Welke KF, et al. Major Infection After Pediatric Cardiac Surgery: A Risk Estimation Model. *The Annals of Thoracic Surgery*. 2010;89(3):843. doi:10.1016/J.ATHORACSUR.2009.11.048
18. Algra SO, Driessen MMP, Schadenberg AWL, et al. Bedside prediction rule for infections after pediatric cardiac surgery. *Intensive Care Medicine*. 2012;38(3):474. doi:10.1007/S00134-011-2454-3
19. Weng SF, Reps J, Kai J, Garibaldi JM, Qureshi N. Can Machine-learning improve cardiovascular risk prediction using routine clinical data? *PLoS ONE*. 2017;12(4). doi:10.1371/journal.pone.0174944

20. Senders JT, Staples PC, Karhade A v, et al. Machine Learning and Neurosurgical Outcome Prediction: A Systematic Review. *World Neurosurgery*. 2018;109:476-486.e1. doi:10.1016/j.wneu.2017.09.149
21. Feng J zhou, Wang Y, Peng J, Sun M wei, Zeng J, Jiang H. Comparison between logistic regression and machine learning algorithms on survival prediction of traumatic brain injuries. *Journal of Critical Care*. 2019;54:110-116. doi:10.1016/j.jcrc.2019.08.010
22. Lee HC, Yoon S, Yang SM, et al. Prediction of Acute Kidney Injury after Liver Transplantation: Machine Learning Approaches vs. Logistic Regression Model. *Journal of Clinical Medicine*. 2018;7(11):428. doi:10.3390/jcm7110428
23. Decruyenaere A, Decruyenaere P, Peeters P, Vermassen F, Dhaene T, Couckuyt I. Prediction of delayed graft function after kidney transplantation: Comparison between logistic regression and machine learning methods Standards, technology, and modeling. *BMC Medical Informatics and Decision Making*. 2015;15(1). doi:10.1186/s12911-015-0206-y
24. Kashyap M, Seneviratne M, Banda JM, et al. Development and validation of phenotype classifiers across multiple sites in the observational health data sciences and informatics network. *Journal of the American Medical Informatics Association*. 2020;27(6):877-883. doi:10.1093/jamia/ocaa032
25. Ostropolets A, Reich C, Ryan P, Shang N, Hripcsak G, Weng C. Adapting electronic health records-derived phenotypes to claims data: Lessons learned in using limited clinical data for phenotyping. *J Biomed Inform*. 2020;102. doi:10.1016/J.JBI.2019.103363
26. Gibson TB, Nguyen MD, Burrell T, et al. Electronic phenotyping of health outcomes of interest using a linked claims-electronic health record database: Findings from a machine learning pilot project. *J Am Med Inform Assoc*. 2021;28(7):1507-1517. doi:10.1093/JAMIA/OCAB036
27. Rhee C, Dantes R, Epstein L, et al. Incidence and Trends of Sepsis in US Hospitals Using Clinical vs Claims Data, 2009-2014. *JAMA*. 2017;318(13):1241-1249. doi:10.1001/JAMA.2017.13836
28. Research Derivative | Human Research Protections Program. Accessed December 10, 2020. https://www.vumc.org/irb/research-derivative
29. Code System: CDCNHSN | NHSN | CDC. Accessed February 9, 2022. https://www.cdc.gov/nhsn/cdaportal/terminology/codesystem/cdcnhsn.html
30. Doyle DJ, Goyal A, Garmon EH. American Society of Anesthesiologists Classification. *StatPearls*. Published online October 9, 2021. Accessed March 8, 2022. https://www.ncbi.nlm.nih.gov/books/NBK441940/
31. Athena. Accessed December 10, 2020. https://athena.ohdsi.org/search-terms/terms
32. SSI | PSC | NHSN | CDC. Accessed February 10, 2022. https://www.cdc.gov/nhsn/psc/ssi/index.html?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Fnhsn%2Facute-care-hospital%2Fssi%2Findex.html
33. 6.3. Preprocessing data — scikit-learn 1.0.2 documentation. Accessed February 9, 2022. https://scikit-learn.org/stable/modules/preprocessing.html
34. Komer B, Bergstra J, Eliasmith C. Hyperopt-Sklearn. Published online 2019:97-111. doi:10.1007/978-3-030-05318-5_5
35. Barker GM, O'Brien SM, Welke KF, et al. Major Infection After Pediatric Cardiac Surgery: A Risk Estimation Model. *Annals of Thoracic Surgery*. 2010;89(3):843-850. doi:10.1016/j.athoracsur.2009.11.048
36. Aryafar A, Marzio · A Di, Guillard · O, Pontailler · M, Vicca · S, Bojan · M. Procalcitonin Concentration Measured Within the First Days of Cardiac Surgery Is Predictive of Postoperative Infections in Neonates: A Case-Control Study. *Pediatric Cardiology*. 2019;40:1289-1295. doi:10.1007/s00246-019-02150-y
37. Sponholz C, Sakr Y, Reinhart K, Brunkhorst F. Diagnostic value and prognostic implications of serum procalcitonin after cardiac surgery: A systematic review of the literature. *Critical Care*. 2006;10(5). doi:10.1186/cc5067
38. Macher J, Gras Le Guen C, Chenouard A, et al. Preoperative Staphylococcus aureus Carriage and Risk of Surgical Site Infection After Cardiac Surgery in Children Younger than 1 year: A Pilot Cohort Study. *Pediatric Cardiology*. 2017;38(1):176-183. doi:10.1007/s00246-016-1499-z
39. Paling FP, Olsen K, Ohneberg K, et al. Risk prediction for Staphylococcus aureus surgical site infection following cardiothoracic surgery; A secondary analysis of the V710-P003 trial. *PLoS ONE*. 2018;13(3). doi:10.1371/journal.pone.0193445

40. van der Linde D, Konings EEM, Slager MA, et al. Birth Prevalence of Congenital Heart Disease Worldwide: A Systematic Review and Meta-Analysis. *J Am Coll Cardiol*. 2011;58(21):2241-2247. doi:10.1016/J.JACC.2011.08.025
41. Agarwal HS, Wolfram KB, Saville BR, Donahue BS, Bichell DP. Postoperative complications and association with outcomes in pediatric cardiac surgery. *J Thorac Cardiovasc Surg*. 2014;148(2). doi:10.1016/J.JTCVS.2013.10.031
42. Mangukia C V., Agarwal S, Satyarthy S, Datt V, Satsangi D. Mediastinitis following pediatric cardiac surgery. *J Card Surg*. 2014;29(1):74-82. doi:10.1111/JOCS.12243
43. Jaworski R, Kansy A, Dzierzanowska-Fangrat K, Maruszewski B. Antibiotic Prophylaxis in Pediatric Cardiac Surgery: Where Are We and Where Do We Go? A Systematic Review. *Surg Infect (Larchmt)*. 2019;20(4):253-260. doi:10.1089/SUR.2018.272
44. Savary L, de Luca A, el Arid JM, et al. Systematic skin and nasal decolonization lowers Staphylococcus infection in pediatric cardiac surgery. *Archives de pediatrie : organe officiel de la Societe francaise de pediatrie*. Published online 2022. doi:10.1016/J.ARCPED.2022.01.009
45. Garcia X, Pye S, Tang X, Gossett J, Prodhan P, Bhutta A. Catheter-Associated Blood Stream Infections in Intracardiac Lines. *Journal of Pediatric Intensive Care*. 2017;06(03):159-164. doi:10.1055/S-0036-1596064
46. Elassal AA, Eldib OS, Dohain AM, Abdelmohsen GA, Abdella AH, Al-Radi OO. Delayed Sternal Closure in Congenital Heart Surgery: A Risk-Benefit Analysis. *Heart Surg Forum*. 2019;22(5):E325-E330. doi:10.1532/HSF.2471
47. Vos RJ, van Putte BP, Kloppenburg GTL. Prevention of deep sternal wound infection in cardiac surgery: a literature review. *Journal of Hospital Infection*. 2018;100(4):411-420. doi:10.1016/j.jhin.2018.05.026
48. Algra SO, Driessen MMP, Schadenberg AWL, et al. Bedside prediction rule for infections after pediatric cardiac surgery. *Intensive Care Medicine*. 2012;38(3):474-481. doi:10.1007/s00134-011-2454-3
49. Azodi CB, Tang J, Shiu SH. Opening the Black Box: Interpretable Machine Learning for Geneticists. *Trends Genet*. 2020;36(6):442-455. doi:10.1016/J.TIG.2020.03.005
50. FAQs: Surgical Site Procedure Codes | NHSN | CDC. Accessed February 10, 2022. https://www.cdc.gov/nhsn/faqs/faq-ssi-proc-codes.html
51. Roberson ML, Egberg MD, Strassle PD, Phillips MR. Measuring malnutrition and its impact on pediatric surgery outcomes: A NSQIP-P analysis. *J Pediatr Surg*. 2021;56(3):439-445. doi:10.1016/J.JPEDSURG.2020.10.001
52. Cheung YF, Tsang HYH, Kwok JSY. Immunologic profile of patients with protein-losing enteropathy complicating congenital heart disease. *Pediatr Cardiol*. 2002;23(6):587-593. doi:10.1007/S00246-001-0078-Z
53. Magdo HS, Stillwell TL, Greenhawt MJ, et al. Immune Abnormalities in Fontan Protein-Losing Enteropathy: A Case-Control Study. *Journal of Pediatrics*. 2015;167(2):331-337. doi:10.1016/j.jpeds.2015.04.061
54. Zaupper LB, Nielsen BW, Herlin T. Protein-losing enteropathy after the total cavopulmonary connection: impact of intravenous immunoglobulin. *Congenit Heart Dis*. 2011;6(6):624-629. doi:10.1111/J.1747-0803.2011.00568.X
55. Yilmaz Ferhatoglu S, Yurdakok O, Yurtseven N. Malnutrition on admission to the paediatric cardiac intensive care unit increases the risk of mortality and adverse outcomes following paediatric congenital heart surgery: A prospective cohort study. *Aust Crit Care*. Published online 2021. doi:10.1016/J.AUCC.2021.07.004
56. Ross F, Latham G, Joffe D, et al. Preoperative malnutrition is associated with increased mortality and adverse outcomes after paediatric cardiac surgery. *Cardiol Young*. 2017;27(9):1716-1725. doi:10.1017/S1047951117001068
57. Griffin BR, Bronsert M, Reece TB, et al. Thrombocytopenia After Cardiopulmonary Bypass Is Associated With Increased Morbidity and Mortality. *Ann Thorac Surg*. 2020;110(1):50-57. doi:10.1016/J.ATHORACSUR.2019.10.039
58. Jin Y, Feng Z, Zhao J, et al. Outcomes and factors associated with early mortality in pediatric postcardiotomy veno-arterial extracorporeal membrane oxygenation. *Artif Organs*. 2021;45(1):6-14. doi:10.1111/AOR.13773
59. Pickard SS, Feinstein JA, Popat RA, Huang L, Dutta S. Short- and long-term outcomes of necrotizing enterocolitis in infants with congenital heart disease. *Pediatrics*. 2009;123(5). doi:10.1542/peds.2008-3216

60. Kashif H, Abuelgasim E, Hussain N, Luyt J, Harky A. Necrotizing enterocolitis and congenital heart disease. *Ann Pediatr Cardiol*. 2021;14(4):507-515. doi:10.4103/apc.apc_30_21
61. Giannone PJ, Luce WA, Nankervis CA, Hoffman TM, Wold LE. Necrotizing enterocolitis in neonates with congenital heart disease. *Life Sciences*. 2008;82(7-8):341-347. doi:10.1016/j.lfs.2007.09.036
62. Kargl S, Maier R, Gitter R, Pumberger W. Necrotizing enterocolitis after open cardiac surgery for congenital heart defects - A serious threat. *Klinische Padiatrie*. 2013;225(1):24-28. doi:10.1055/s-0032-1331724
63. Watson JD, Urban TT, Tong SS, et al. Immediate Post-operative Enterocyte Injury, as Determined by Increased Circulating Intestinal Fatty Acid Binding Protein, Is Associated With Subsequent Development of Necrotizing Enterocolitis After Infant Cardiothoracic Surgery. *Front Pediatr*. 2020;8. doi:10.3389/FPED.2020.00267
64. Mukherjee D, Zhang YY, Chang DC, Vricella LA, Brenner JI, Abdullah F. Outcomes analysis of necrotizing enterocolitis within 11 958 neonates undergoing cardiac surgical procedures. *Arch Surg*. 2010;145(4):389-392. doi:10.1001/ARCHSURG.2010.39
65. Martini S, Beghetti I, Annunziata M, et al. Enteral nutrition in term infants with congenital heart disease: Knowledge gaps and future directions to improve clinical practice. *Nutrients*. 2021;13(3):1-13. doi:10.3390/nu13030932
66. Cognata A, Kataria-Hale J, Griffiths P, et al. Human Milk Use in the Preoperative Period Is Associated with a Lower Risk for Necrotizing Enterocolitis in Neonates with Complex Congenital Heart Disease. *Journal of Pediatrics*. 2019;215:11-16.e2. doi:10.1016/j.jpeds.2019.08.009
67. Stemerman R, Bunning T, Grover J, Kitzmiller R, Patel MD. Identifying Patient Phenotype Cohorts Using Prehospital Electronic Health Record Data. *Prehospital Emergency Care*. 2022;26(1):78-88. doi:10.1080/10903127.2020.1859658

APPENDIX 1: SUPPLEMENTAL MATERIAL

Infection definitions: The infection definitions included: bacteremia (requiring documentation of a positive blood culture following surgery and treatment of at least 5 days with antimicrobials for same; a single peripheral culture positive for coagulase negative Staphylococcus or diphtheroids was excluded), fungemia (requiring documentation of a positive blood culture for fungal or yeast species), endocarditis (requiring documentation of fulfillment of modified Duke criteria and of treatment initiation for endocarditis), sepsis (requiring a) temperature instability plus b) vital sign instability such as tachycardia or hypotension beyond baseline, plus c) treatment with antimicrobials at least 5 days with antimicrobials, unless a documented viral source was identified), surgical site infection (requiring a) description of erythema, purulent drainage, or positive wound culture, plus b) mention of treatment for surgical site infection, pocket site infection or mediastinitis, plus c) documentation of improvement following treatment initiation), and NEC (requiring a) a clinical assessment of NEC in the clinical notes plus b) documentation of antibiotic treatment of at least 5 days duration, plus c) documentation of at least one of: bloody stool, pneumatosis on imaging, or drained purulent abdominal collection).

Predictor	Infection Present	Infection Absent	OR	LL	UL	P Value
Diagnosis Count Binary >54	83	68	9.00	6.28	12.88	<0.001
Minimum Platelet Count <43	75	42	5.32	3.54	7.99	<0.001
Platelet Count Range >449	7	4	4.68	1.36	16.12	0.014
CRP Range >26	69	33	4.35	2.80	6.75	<0.001
Minimum Albumin <2.7	467	135	4.26	3.21	5.65	<0.001
Age <67 days	277	97	4.18	3.12	5.59	<0.001
WBC Range >11.1	114	49	4.15	2.87	5.99	<0.001
Minimum WBC >3.15	169	65	3.94	2.84	5.48	<0.001
Minimum CRP <17.47	195	72	3.89	2.84	5.34	<0.001
Maximum WBC >29.46	218	73	3.51	2.57	4.79	<0.001
Glucose Range >181	244	79	3.47	2.56	4.70	<0.001
Maximum Hemoglobin >17.3	177	59	3.28	2.35	4.58	<0.001
ANC Range >3.41	354	100	3.28	2.46	4.36	<0.001
Albumin Range >0.6	298	88	3.25	2.42	4.36	<0.001
Maximum CRP >45	201	63	3.12	2.25	4.31	<0.001
Maximum Glucose >236	332	92	3.07	2.30	4.10	<0.001
Maximum Pulse >191	106	36	3.07	2.05	4.61	<0.001
Minimum Diastolic BP <27	1183	193	3.03	2.09	4.37	<0.001
Hemoglobin Range >6.3	447	112	3.01	2.27	3.98	<0.001
Maximum ANC >7.97	425	108	2.99	2.26	3.97	<0.001
Minimum Systolic BP <50	1159	190	2.91	2.04	4.16	<0.001
Minimum Hemoglobin <6.5	47	16	2.88	1.61	5.17	<0.001
Procedure Count >8	27	9	2.76	1.28	5.95	0.009
Minimum Temperature <31.0	653	135	2.63	1.99	3.49	<0.001
Surgery Duration >502	123	35	2.53	1.69	3.80	<0.001
Maximum Albumin <3.2	418	95	2.43	1.83	3.23	<0.001
Maximum Systolic BP <97	140	37	2.36	1.59	3.48	<0.001
Steroid Exposed	134	35	2.31	1.55	3.45	<0.001
Antibiotic Courses >4	537	111	2.30	1.74	3.04	<0.001
Maximum Platelet Count <215	532	110	2.29	1.74	3.03	<0.001
Minimum ANC >4.67	389	86	2.26	1.69	3.02	<0.001
Minimum Glucose <57	202	47	2.11	1.48	3.00	<0.001
Maximum Diastolic BP <52	1819	221	2.06	0.94	4.52	0.072
Previous Surgery	113	26	1.97	1.26	3.09	0.003
ASA Class >3	949	148	1.74	1.30	2.31	<0.001
Minimum Pulse >32	1140	167	1.68	1.24	2.28	0.001
Maximum Temperature <37.2	581	98	1.64	1.24	2.16	0.001
Male Gender	961	138	1.40	1.06	1.86	0.017

Table A: Full list of candidate binary predictors. Infection absent and present note the number of cases in each category meeting the condition. OR: odds ratio, LL: lower limit of odds ratio confidence interval, UL upper limit of

odds ratio confidence interval, CRP: C reactive protein, WBC: white blood cell count in blood, ANC: absolute neutrophil count, ASA: American Society of Anesthesiologists, BP: blood pressure.