



## CLINICAL RESEARCH ARTICLE

## Identification of prognostic factors for pediatric myocarditis with a random forests algorithm-assisted approach

Fu-Sheng Chou<sup>1</sup> and Laxmi V. Ghimire<sup>2</sup>

**BACKGROUND:** Pediatric myocarditis is a rare disease with substantial mortality. Little is known regarding its prognostic factors. We hypothesize that certain comorbidities and procedural needs may increase risks of poor outcomes. This study aims to identify prognostic factors for pediatric myocarditis.

**METHODS:** The national Kids' Inpatient Database was used in the study. A random forests algorithm was implemented for mortality prediction based on comorbidities and procedures. Linear regression analysis was then performed to quantify their associations with mortality and length of stay.

**RESULTS:** The prevalence of pediatric myocarditis among all pediatric hospitalizations doubled from 2003 to 2016. The mortality rate peaked in 2006 (6.7%) and declined steadily thereafter, with a rate of 3.2% in 2016. Brain injury (including encephalopathy, cerebral edema, and intracranial hemorrhage), acute kidney injury, dysrhythmias, coagulopathy, sepsis, and ECMO use were all independent prognostic factors associated with increased mortality and prolonged hospital stay.

**CONCLUSION:** Prognostic factor identification may not be straightforward in rare diseases such as pediatric myocarditis due to small cohort size in each treating facility. Findings from this report provide insights into the prognostic factors for pediatric myocarditis, and may allow clinicians to be better prepared when informing patients and their families regarding disease outcomes.

*Pediatric Research* (2021) 90:427–430; <https://doi.org/10.1038/s41390-020-01268-7>

**IMPACT:**

- The rate of hospitalization due to pediatric myocarditis was increasing but the mortality rate was declining over the past decade.
- End organ damage, including the brain and the kidney, was associated with mortality and prolonged hospital stay in pediatric myocarditis.
- Tachyarrhythmias and cardiac function compromise requiring ECMO were also associated with mortality and prolonged hospital stay.
- A data science approach combining machine learning algorithms and conventional regression modeling using a large dataset may facilitate risk factor identification and outcome correlation in rare diseases, as illustrated in this study.

**INTRODUCTION**

Pediatric myocarditis is a rare disease of children. The true incidence is unknown, but the number of cases follows a bimodal trend according to age, with the highest among infants and adolescents.<sup>1</sup> The etiologies of pediatric myocarditis are multiple, including viral, immune, rheumatological, and toxin mediated.<sup>2</sup> In recent studies using the Pediatric Health Information System database, the mortality rate of pediatric myocarditis was estimated to be around 7–8%, and was much higher in the infant population.<sup>1,3</sup>

It has been previously suggested that tachyarrhythmia is associated with poor outcomes in pediatric myocarditis, including higher mortality, longer length of hospital stay, and increased hospitalization costs.<sup>3,4</sup> Little is known whether additional prognostic factors exist in this devastating disease. One of the biggest limiting factors in the identification of outcome-associated risk factors in pediatric myocarditis was small cohort sizes due to its rarity.<sup>4–6</sup> With the advent of machine learning (ML) algorithms and

the availability of large hospitalization datasets spanning over multiple years, it may be feasible to take a data science approach to tackle this clinical question.

In this study, we borrowed an ML algorithm to conduct a search for prognostic factors for pediatric myocarditis. We hypothesize that certain comorbidities may increase risks of poor outcomes in pediatric myocarditis.

**METHODS**

## Dataset and data record extraction

The Kids' Inpatient Database (KID) is a survey-based de-identified database and is published every 3–4 years, with the latest release in 2016.<sup>7</sup> Datasets of 2003, 2006, 2009, 2012, and 2016 were purchased through the Healthcare Cost and Utilization Project (HCUP) online central distributor (ownership by L.V.G.). As the KID is a publicly available de-identified database, the study is considered

<sup>1</sup>Department of Pediatrics, Loma Linda University School of Medicine, Loma Linda, CA 92354, USA and <sup>2</sup>Lakes Region General Hospital, Laconia, NH 03246, USA  
Correspondence: Laxmi V. Ghimire (laxmivilasg@gmail.com)

Received: 12 September 2019 Revised: 26 August 2020 Accepted: 11 October 2020  
Published online: 18 November 2020

**Table 1.** ICD-9-CM and ICD-10-CM codes used to query diagnoses and procedures.

Group	Diagnoses/procedures	ICD-9 codes	ICD-10 codes
Primary diagnosis	Myocarditis	074.20, 074.23, 422.0, 422.99, 422.90, 422.91, 422.92, 429.0	B33.20, J10.82, I40.0, I40.1, I40.8, I40.9, I41, I51.4
Comorbidity risk factor	Brain injury	348.1, 348.30, 348.35, 348.39, 432.xx, 852.20, 852.22	G93.1, G93.40, G93.49, I67.83, G93.6, I62.1, I62.00, I62.9, S06.5
	Acute kidney injury	584.xx	N17.xx
	Dysrhythmias	427.xx	I47.1, I47.2, I47.9, I48.91, I48.92, I49.01, I46.9, I49.1, I49.3, I49.40, I49.49, I49.8, I49.9
	Coagulopathy	286.6, 286.7, 286.9, 776.2	D65, D68.4, D68.8, D68.9, P60
	Sepsis	995.91, 995.92, 771.81	A41.9, R65.2, P36.9
Procedural risk factor	Cardioversion	996.1, 996.2, 996.9	5A2204Z
	ECMO	396.5	5A15223

as non-human subject research. Therefore, a patient consenting process was not required.

Datasets from 2003 to 2012 contained International Classification of Diseases, Ninth Revision Clinical Modification (ICD-9-CM) diagnostic codes, whereas the 2016 dataset contained the ICD-10-CM diagnostic codes. Respective ICD-9-CM and ICD-10-CM codes for myocarditis, as well as the identified comorbidity and procedural risk factors, are listed in Table 1.

**ML approach to risk factor identification**

Datasets from 2003 to 2012 were combined for ML using a random forests algorithm. The 2016 dataset was not used because of technical difficulties translating ICD-10-CM codes in the 2016 dataset to their ICD-9-CM counterparts for combined use in ML. To prepare the datasets for ML, the ICD-9-CM diagnostic codes listed in the data records were reduced to category codes (first three digits of the ICD-9-CM codes). Subsequently, a feature (aka variable) was created for each ICD-9-CM category code (500 total), in which the absence or presence of the code in each data record was assigned. Supervised ML was then performed, with mortality as the outcome for prediction. All data were used for training, in which three repeats of 10-fold cross-validation were performed. After training, variable importance scores were obtained to identify category codes that were considered important in predicting mortality in the model. The same process was repeated for ICD-9-PRS procedural codes. ML was performed in R 3.6.3 in the RStudio 1.2 environment using the *caret* package.<sup>8</sup>

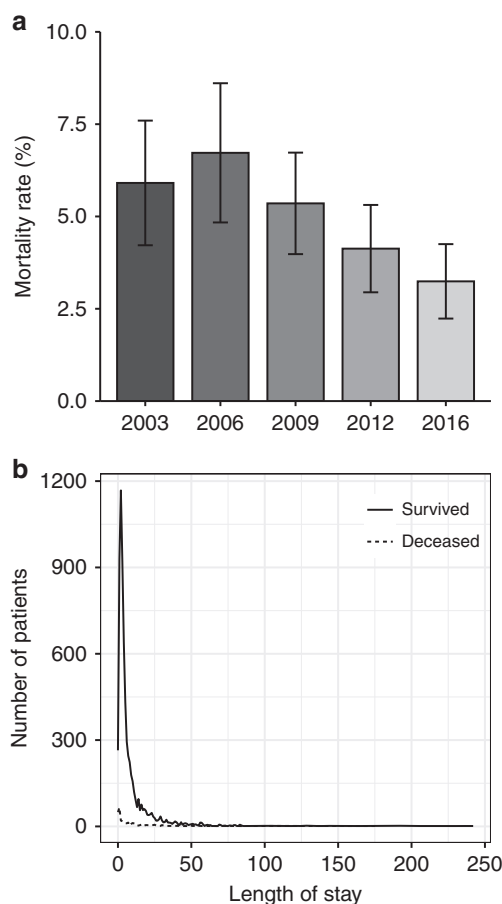
**Survey-weighted statistical analysis and linear regression modeling**

Survey-weighted statistical analysis was performed using the *survey* package for R.<sup>9</sup> All data presented were weight-adjusted. If not otherwise stated, data were presented as weighted numbers and their 95% confidence intervals (CI).

Datasets from all 5 years were combined for linear regression analysis. Binomial regression models were developed to examine the association between mortality and various combinations of the risk factors identified in ML (127 mutations). Negative binomial regression models were developed to examine the association between length of hospital stay and the same 127 permutations of the identified risk factors. Only patients who survived to home discharge were used for length of stay modeling. Akaike information criteria (AIC) was used to select the best model. Odds ratios and the ratios of length of stay for each risk factor were then calculated.

**RESULTS**

Pediatric myocarditis is a rare disease with high mortality. There were a total 7241 hospitalizations with a diagnosis of myocarditis among a total of 35,279,684 pediatric hospitalizations



**Fig. 1 Mortality and length of stay in pediatric myocarditis.** **a** Weighted mortality rate estimates by year. Error bars represent upper and lower limits of the 95% confidence interval. **b** Weighted distribution of length of hospital stay for the survivors (solid line) and non-survivors (dashed line).

over 5 years. The prevalence of pediatric myocarditis doubled from 15.6 per 100,000 hospitalizations in 2003 to 31.6 per 100,000 hospitalizations in 2016, with an average of 21.4 myocarditis hospitalizations per 100,000 pediatric hospitalizations. The overall mortality rate was 4.9%. The mortality rate was the highest (6.7%) in 2006 and declined since then to 3.2% in 2016 (Fig. 1a). Median length of hospital stay was 4 days (interquartile range: 2–9 days) among the survivors, and was 5 days (interquartile range: 1–18 days) among the deceased (Fig. 1b).

**Table 2.** The estimated percentage of pediatric myocarditis patients with indicated risk factors stratified by whether they survived hospitalization or not.

Group	Risk factor	Survived	Deceased
Comorbidity	Brain injury	2.5% (2.0–3.0%)	16.9% (11.9–22.0%)
	Acute kidney injury	6.4% (5.6–7.3%)	33.9% (27.9–39.9%)
	Dysrhythmias	20.0% (18.6–21.3%)	51.0% (44.5–57.5%)
	Coagulopathy	3.5% (2.9–4.1%)	27.2% (21.6–32.7%)
	Sepsis	6.6% (5.9–7.4%)	27.3% (21.5–33.1%)
Procedure	Cardioversion	1.6% (1.2–1.9%)	12.0% (7.9–16.1%)
	ECMO	3.6% (3.0–4.2%)	28.8% (22.3–35.3%)

Data presented as percentage (95% confidence interval).

**Identification of risk factors for mortality by using a random forests algorithm**

Supervised ML was performed to identify mortality risk factors by using a random forests algorithm. ICD-9 diagnostic and procedural category codes were transformed into binary features (absent or present), and mortality was used as the binary outcome for training. For diagnostic codes, top five groups identified by variable importance scores included brain injury (including encephalopathy, cerebral edema, and intracranial hemorrhage), acute kidney injury, dysrhythmias/tachyarrhythmias, coagulopathy, and sepsis. For procedural codes, two procedures that received high variable importance scores included cardioversion and extracorporeal membrane oxygenation (ECMO). Based on their category codes, the full ICD-9 codes, along with their corresponding ICD-10 codes, were extracted, as listed in Table 1. The percentages of cases with each of the identified risk factors are listed in Table 2.

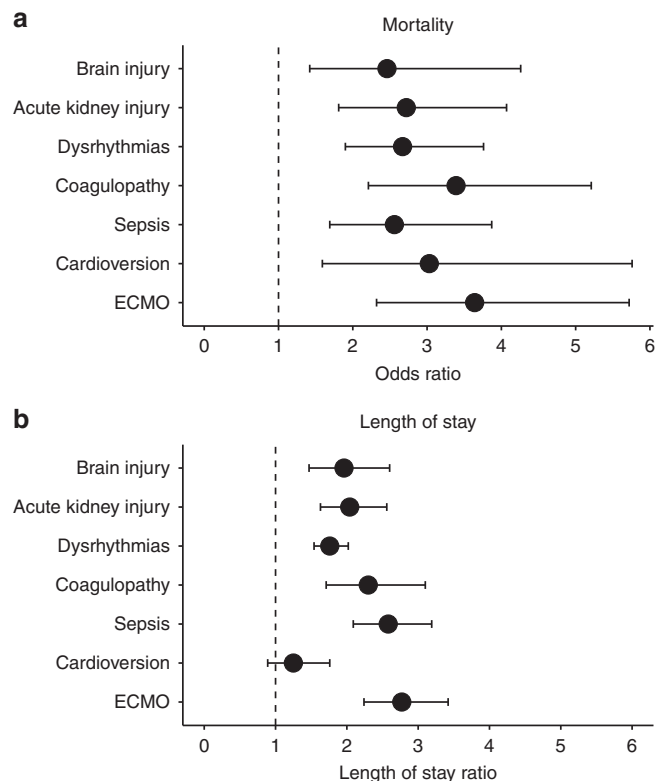
**Multiple linear regression modeling**

A binomial multiple regression analysis was performed to compare multiple models associating mortality with a combination of different risk factors identified above. The model with the best fit was determined by AIC. The selected model included all risk factors. The odds ratios of each factor after controlling for other factors were then calculated (Fig. 2a). We then asked if length of hospital stay is associated with any of the risk factors in question among the survivors. To this end, a negative binomial multiple regression analysis was performed. The best model selected by AIC also included all risk factors. In this model, the estimated length of stay without any risk factor was 5.8 (95% CI: 5.4–6.2) days. Cardioversion increased length of stay minimally by 25% (95% CI: –11 to 76%). All the other risk factors were independently associated with increased length of stay between two- to three-fold, with the need for ECMO showing the biggest effect on length of stay (2.8-fold, 95% CI: 2.2–3.4 fold) (Fig. 2b).

**DISCUSSION**

In this study, we aimed to identify prognostic factors for pediatric myocarditis. We found that comorbidities such as brain injury, acute kidney injury, tachyarrhythmias, coagulopathy, and sepsis, as well as the need for ECMO, were all independently associated with mortality and prolonged length of hospital stay. Additionally, cardioversion was also associated with increased mortality.

The study successfully utilized an ML algorithm to search for clinically important mortality risk factors from a total of 500 factors (ICD-9 category codes) present in the data records. Risk factors identified via ML were then validated in linear regression models to further quantify their risks. This approach facilitates risk stratification studies of rare diseases, and the findings may serve as the basis for future prospective studies.



**Fig. 2** Multivariable regression models for mortality and length of stay. **a** Odds ratios for mortality for each indicated risk factor after controlling for the others. Error bars represent upper and lower limits of the 95% confidence interval. **b** Ratios of length of hospital stay for each indicated risk factor, after controlling for the others, as compared to patients without any of the indicated risk factors. Error bars represent upper and lower limits of the 95% confidence interval.

The random forests algorithm is a decision tree-based algorithm that is particularly useful in studies dealing with categorical variables only. Additionally, the collinearity issue in linear regression algorithms is well tolerated in decision tree-based algorithms, eliminating the need to measure correlations between variables. Therefore, the random forests algorithm was chosen for ML training in this study.

The mortality rate of pediatric myocarditis was previously estimated to be around 7–8% in recent two studies using the Pediatric Health Information System database, which was higher than our findings of ~4.9% in the current study.<sup>1,3</sup> As shown in Fig. 1a, as mortality rates declined over recent years, it is possible

that the lower overall mortality rate in our study was due to inclusion of more recent datasets. Alternatively, the differences in mortality rate could also be due to differences in age criteria for inclusion in different databases. Specifically, in the KID, patients up to 20 years of age were included. As lower mortality rates were seen in old pediatric patients, age criteria could also contribute to a lower mortality rate in our study.

It was shown before that tachyarrhythmias are associated with poor outcomes in pediatric myocarditis.<sup>3,4</sup> Specifically, the study by Anderson et al.<sup>3</sup> showed that tachyarrhythmias, but not bradyarrhythmias, were associated with increased mortality, length of hospital stay, and daily hospitalization costs. Similarly, our study identified tachyarrhythmias as a risk factor for mortality and increased length of stay, further supporting the validity of the ML approach in risk factor identification. Consistent with previous studies, our analysis also did not find an association between conduction disorders/bradyarrhythmias and mortality (data not shown). To our knowledge, acute kidney injury (AKI) has not been reported in the literature as a risk factor for mortality in pediatric myocarditis, although a recent study in adult patients with acute myocarditis showed unfavorable outcomes in association with AKI.<sup>10</sup> AKI may cause disturbances in fluid, electrolytes, and acid–base balance. It is likely related to low cardiac output and poor renal perfusion. It may also be caused by afterload reduction therapy, which leads to a reduction in renal perfusion. Our finding warrants future prospective studies to further investigate the mechanisms and the measures to minimize its occurrence.

#### Limitations

There are several limitations to the study. First, data in the KID were not collected for research purposes, and were for medical coding and billing. Therefore, incorrect or missing information may exist. Second, there is no information on the accuracy of the diagnoses that were used to identify prognostic factors, as well as their temporal association with the primary diagnosis. Nonetheless, the HCUP has stringent policies for quality assurance and the KID has been used for clinical observational studies which has resulted in more than 4000 publications to date, supporting its credibility in this type of research.

#### CONCLUSION

In summary, we implemented a random forests algorithm to identify risk factors for mortality and prolonged length of stay in pediatric myocarditis using a national database with data records spanning over more than a decade, followed by quantifying individual risks using linear regression analysis. We identified brain injury, acute kidney injury, dysrhythmias, coagulopathy,

sepsis, and the need for ECMO use to be independently associated with increased mortality and longer length of stay. Findings from this report provide insights into the prognostic factors for pediatric myocarditis, and may allow clinicians to be better prepared when informing patients and their families regarding disease outcomes.

#### ACKNOWLEDGEMENTS

The authors would like to acknowledge Dr. Hung-Wen Yeh, Ph.D., of Children's Mercy-Kansas City Health Services and Outcome Research for his consultation on machine learning and linear regression analysis. Purchase of KID datasets from HCUP was partially funded by Lakes Region General Hospital.

#### AUTHOR CONTRIBUTIONS

F.-S.C. conceptualized the study, performed machine learning and linear regression analyses, interpreted data, and prepared the original and the revised manuscripts. L. V.G. conceptualized the study, provided intellectual input to the project and to the manuscript, prepared the original and the revised manuscripts, and is the owner of the purchased KID datasets.

#### ADDITIONAL INFORMATION

**Competing interests:** The authors declare no competing interests.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

#### REFERENCES

1. Ghelani, S. J., Spaeder, M. C., Pastor, W., Spurney, C. F. & Klugman, D. Demographics, trends, and outcomes in pediatric acute myocarditis in the United States, 2006 to 2011. *Circ. Cardiovasc. Qual. Outcomes* **5**, 622–627 (2012).
2. Park, M. K. *Park's Pediatric Cardiology for Practitioners* (Elsevier Saunders, 2014).
3. Anderson, B. R., Silver, E. S., Richmond, M. E. & Liberman, L. Usefulness of arrhythmias as predictors of death and resource utilization in children with myocarditis. *Am. J. Cardiol.* **114**, 1400–1405 (2014).
4. Miyake, C. Y. et al. In-hospital arrhythmia development and outcomes in pediatric patients with acute myocarditis. *Am. J. Cardiol.* **113**, 535–540 (2014).
5. Sachdeva, S., Song, X., Dham, N., Heath, D. M. & DeBiasi, R. L. Analysis of clinical parameters and cardiac magnetic resonance imaging as predictors of outcome in pediatric myocarditis. *Am. J. Cardiol.* **115**, 499–504 (2015).
6. Rodriguez-Gonzalez, M., Sanchez-Codez, M. I., Lubian-Gutierrez, M. & Castellano-Martinez, A. Clinical presentation and early predictors for poor outcomes in pediatric myocarditis: a retrospective study. *World J. Clin. Cases* **7**, 548–561 (2019).
7. HCUP-US KID Overview. <https://www.hcup-us.ahrq.gov/kidoverview.jsp>.
8. Kuhn, M. The Caret Package. <http://topepo.github.io/caret/index.html> (2019).
9. CRAN—Package Survey. <https://cran.r-project.org/web/packages/survey/index.html>.
10. Yang, Y.-W. et al. Prevalence of acute kidney injury and prognostic significance in patients with acute myocarditis. *PLoS ONE* **7**, e48055 (2012).