Critical Care Explorations
November 2026 • Volume 8 • Number 11 • e1520

OBSERVATIONAL STUDY

OPEN

Preoperative Serum Biomarkers Improve Prediction of Poor Postoperative Outcome After Cardiopulmonary Bypass Compared With Clinical Variables in Neonates

Monique M. Gardner, MD, MSCE[1]
Daniel Balcarcel, MD, MBI[2,3]
Fang Chen, PhD[4]
Andrew Huff, BS[4]
Robert Olsen, BS[5]
Kyle Winser, BS[5]
Andrea Kennedy, BS[5]
Hillary R. Bogner, MD, MSCE[6]
Rui Feng, PhD[7]
Anh D. Mai, MS[8]
Vanessa Gonzalez, MBA[4]
Joseph Fraietta, PhD[4]
Asif Padiyath, MBBS[3,8,9]
Scott L. Weiss, MD, MSCE[10]
J. William Gaynor, MD[3,11]
Nadir Yehya, MD, MSCE[2,3]

IMPORTANCE: Neonates undergoing surgery for congenital heart disease (CHD) are at risk for poor postoperative outcomes after cardiopulmonary bypass (CPB), including death, prolonged ICU stay, and ICU readmission. Current prognostic models rely predominantly on nonmodifiable clinical factors.

OBJECTIVES: To utilize machine learning methods to compare the predictive utility of preoperative blood biomarkers vs. clinical features alone for ICU-30, a validated composite outcome after neonatal CPB that includes postoperative 30-day morbidity and mortality.

DESIGN, SETTING, AND PARTICIPANTS: Prospective observational single-center cohort study of consecutively enrolled neonates younger than 30 days undergoing CPB for CHD.

MAIN OUTCOMES AND MEASURES: Plasma and clinical data were collected from consecutively enrolled neonates younger than 30 days immediately before CPB. Thirty-seven biomarkers were measured by individual or multiplex enzyme-linked immunosorbent assay and tested for association with ICU-30, defined as 1) mortality within 30 days of CPB, 2) ICU stay longer than 30 days, or 3) ICU readmission within 30 days. Predictive performance of extreme gradient boosting (XGBoost), least absolute shrinkage and selection operator, and random forest (RF) models was compared using the area under the receiver operating characteristic curve (AUROC). For XGBoost, the top 20 features were identified based on model importance (gain).

RESULTS: Biomarkers were available for 144 neonates. Most were male (51%), White (64%), and non-Hispanic (75%). Most operations were Society of Thoracic Surgeons–European Association for Cardiothoracic Surgery (STAT) category 3 (32%) or STAT category 5 (29%). ICU-30 occurred in 29 neonates (20%). Using combined clinical and biomarker data, XGBoost demonstrated moderate discrimination for ICU-30 (AUROC, 0.74), outperformed RF (AUROC, 0.72), and LASSO (AUROC, 0.70). Biomarkers alone performed better (AUROC, 0.72) than clinical data alone (AUROC, 0.67) with XGBoost. The top two features in the combined XGBoost model were neutrophil gelatinase-associated lipocalin (NGAL) and trefoil factor 3 (TFF3). Using Shapley additive explanation to assess both predictive impact and directionality, the most influential features were NGAL, a disintegrin and metalloproteinase with a thrombospondin type 1 motif, member 13 (ADAMTS13), and angiopoietin-2 (Ang-2).

CONCLUSIONS AND RELEVANCE: Among neonates with CHD requiring CPB, multibiomarker models improved prediction of poor postoperative outcome compared with clinical variables alone. The most informative model features included markers of renal injury/perfusion and gastrointestinal epithelial injury (NGAL and TFF3), as well as vascular injury and inflammation (ADAMTS13 and

Copyright © 2026 The Authors. Published by Wolters Kluwer Health, LLC. on behalf of the Society of Critical Care Medicine. This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

DOI: 10.1097/CCE.0000000000001520

KEY POINTS

Question: Do preoperative blood biomarkers improve prediction of poor postoperative outcome after neonatal cardiopulmonary bypass compared with clinical variables alone?

Findings: In 144 neonates, biomarker-based machine learning models improved prediction of the composite postoperative outcome for morbidity and mortality (ICU-30) compared with clinical variables alone; neutrophil gelatinase-associated lipocalin, trefoil factor 3, a disintegrin and metalloproteinase with a thrombospondin type 1 motif, member 13 , and angiopoietin-2 were informative features.

Meaning: Biomarker-informed risk stratification may better capture subclinical organ injury and vascular pathways underlying adverse postoperative outcomes in neonatal congenital heart surgery.

Ang-2). These findings support further investigation of biomarker-informed risk stratification and may provide insight into pathophysiologic mechanisms underlying adverse postoperative outcomes.

KEYWORDS: biomarkers; cardiopulmonary bypass; congenital heart disease; machine learning; neonatal cardiac surgery

Neonates with critical congenital heart disease require cardiopulmonary bypass (CPB) surgery in the first days to weeks of life to establish stable circulation. Compared with older children undergoing congenital heart surgery, neonates are at higher risk for poor outcomes, including mortality and prolonged ICU and hospital lengths of stay (1). Commonly used risk prediction methods in this population rely predominantly on fixed demographic, genetic, and procedural factors. Although these variables are clinically important, they may incompletely capture potentially modifiable preoperative organ dysfunction or mechanism-specific injury pathways. Incorporation of circulating biomarkers has improved prognostic and predictive enrichment in oncology (2, 3), acute respiratory distress syndrome (4–6), and sepsis (7), but this approach has not been well studied in children with CHD (8–12), particularly among neonates, who carry the highest perioperative risk. In this study, we evaluated the utility of pre-CPB blood-based biomarkers, separately from and in addition to clinical factors, for improving prediction of poor postoperative outcome after neonatal CPB. We used our previously validated 30-day postoperative composite outcome for neonates, ICU-30, which is associated with 1-year survival (13). We hypothesized that novel biomarkers identified in prior neonatal or cardiac studies would improve prediction of ICU-30 compared with previously studied preoperative clinical factors.

MATERIALS AND METHODS

This prospective observational study included neonates younger than 30 days who required CPB for critical congenital heart disease and had not undergone prior cardiac intervention. Parents or guardians of eligible patients were approached consecutively between August 2020 and July 2024 for informed consent. The study was approved by the Children's Hospital of Philadelphia Institutional Review Board (IRB; "Development of a Plasma Biomarker Risk Profile for Neonates Requiring Cardiopulmonary Bypass for Congenital Heart Disease," IRB 20-017388, approved June 15, 2020). All procedures were performed in accordance with institutional ethical standards and the Declaration of Helsinki 1975 and its later amendments. Methods and results are reported in accordance with the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis guidelines. The objective of this analysis was model development with internal validation to estimate prognostic performance in this cohort rather than immediate clinical implementation.

Clinical Data

Demographic, surgical, and postoperative outcome data were abstracted from the medical record. When available, validated variables collected for the Society of Thoracic Surgeons and Pediatric Cardiac Critical Care Consortium databases were used. The primary outcome, ICU-30, is a previously validated composite outcome defined as 1) mortality within 30 days of CPB, 2) ICU stay longer than 30 postoperative days (PODs), or 3) ICU readmission within 30 PODs (13).

Blood Samples

Blood was collected after induction of anesthesia but immediately before initiation of CPB in the operating

room. Samples were collected in citrated tubes, centrifuged within 30 minutes of collection (2000*g* for 20 min at 20°C), and stored at −80°C until analysis. Thirty-seven biomarkers (**Supplemental Table 1**, https://links.lww.com/CCX/B708) were measured using a combination of single- and multiplex enzyme-linked immunosorbent assays, customized at the University of Pennsylvania Translational & Correlative Studies Laboratory using products from R&D Systems, Inc. (Minneapolis, MN). Biomarkers were selected on the basis of the literature in adult cardiac disease, pediatric cardiac disease, and neonatal medicine and were chosen to represent diverse tissue sources (8–12, 14–40).

Initial Statistical Analyses

Statistical analyses were performed in Stata (Stata Statistical Software: Release 18; StataCorp, College Station, TX) and R (version 4.4.0; R Foundation for Statistical Computing, Vienna, Austria). Clinical and surgical characteristics and biomarker levels were summarized as *n* (%) or median (interquartile range), as appropriate, according to ICU-30 outcome. Fisher exact test or Wilcoxon rank-sum test was used to compare neonates with ICU-30 to those without the outcome. Differences in biomarker concentrations between groups were assessed using the Wilcoxon rank-sum test, with *p* value of less than 0.05 considered statistically significant.

Model Development and Evaluation

We developed and compared three supervised machine learning approaches: least absolute shrinkage and selection operator (LASSO), random forest (RF), and extreme gradient boosting (XGBoost). LASSO regression was used for feature selection and regularization, applying a penalty to shrink less informative coefficients toward zero. The regularization parameter was selected through 10-fold cross-validation to minimize mean squared error. For the RF model, an ensemble of decision trees was constructed using bootstrap aggregation, with the number of trees and maximum tree depth optimized via grid search with cross-validation. XGBoost models were built using gradient-boosted decision trees, with hyperparameters tuned via grid search with 10-fold cross-validation.

Baseline clinical variables and prebypass biomarker concentrations were selected a priori as candidate predictors. The complete list of candidate predictors includes sex, race, gestational age, birth weight, extracardiac abnormalities, chromosomal abnormalities, genetic syndrome, preoperative vasoactive/inotropic need, age at surgery, Society of Thoracic Surgeons–European Association for CardioThoracic Surgery (STAT) category, primary surgical procedure type. Preoperative prostaglandin exposure and mechanical ventilation were evaluated but were not included as adjustment variables in the final models. Prostaglandin use is closely linked to underlying cardiac diagnosis and surgical complexity, resulting in substantial overlap with variables already represented in the model. Preoperative mechanical ventilation was considered a marker of disease severity rather than an independent confounder. To account for baseline physiologic compromise, we considered established pediatric organ dysfunction measures such as Pediatric Logistic Organ Dysfunction, pediatric Sequential Organ Failure Assessment, Pediatric Organ Dysfunction Information Update Mandate, and oxygenation index; however, these instruments have not been validated for preoperative neonates with congenital heart disease and may poorly characterize illness severity in this population. Many components of these scores reflect expected physiologic consequences of congenital heart disease rather than acquired organ dysfunction, creating the potential for misclassification of disease severity. In contrast, preoperative vasoactive-inotropic need provides a more direct assessment of cardiovascular instability and hemodynamic support requirements and was therefore selected as a clinically relevant adjustment variable.

Categorical variables (sex, race, extracardiac abnormalities, chromosomal abnormalities, syndrome, STAT category, and primary surgical procedure type) were converted to binary indicator variables, yielding a high-dimensional predictor set relative to the number of outcome events. Continuous predictors were standardized using *z*-score transformation, whereas binary indicators were retained on their original scale. Missing data and scaling were handled according to model requirements: XGBoost was trained on the raw dataset because it can accommodate missingness; RF required median imputation; and LASSO required both median imputation and scaling to place variables

on a comparable scale. No external validation cohort was available; accordingly, all performance estimates reflect internal validation only.

We implemented stratified five-fold nested cross-validation to reduce bias in model evaluation. The dataset was first divided into five outer folds, with each fold serving once as the held-out test set. Within each training partition, model hyperparameters were optimized using an inner three-fold cross-validation loop with grid search. For XGBoost, tree complexity and learning parameters were varied; for RF, the number of candidate variables per split was adjusted; and for LASSO regression, the degree of penalization was selected from a grid of λ values.

For each model, predicted probabilities were generated for held-out test folds across the outer five-fold cross-validation loop. Performance was assessed using AUROC and area under the precision-recall curve (AUPRC), as well as sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) at a threshold of 0.5. Because ICU-30 occurred in 20% of the cohort, creating an approximately 4:1 ratio of nonevents to events, we considered the dataset moderately imbalanced. We therefore identified an optimal operating threshold for each model using Youden's *J* statistic (sensitivity + specificity − 1), which balances sensitivity and specificity. At this threshold, we report sensitivity, specificity, PPV, and NPV for each model.

To evaluate feature importance, we used the native function within the XGBoost package to quantify importance as the average gain in model performance attributable to each split, normalized as a percentage of total importance. We also computed Shapley additive explanation (SHAP) values to characterize the direction and magnitude of individual feature contributions to model predictions.

RESULTS

Preoperative biomarkers were available for 144 of 150 neonates (**Fig. 1**, **Table 1**). Most were male (n = 74; 51%), White (65%), and non-Hispanic/Latino (76%). Median gestational age was 39 weeks (interquartile range [IQR], 38–39 wk), median birth weight was 3.2 kg (IQR, 2.9–3.6 kg), and median age at surgery was 3.6 days (IQR, 2.6–5.2 d). Most operations were Society of Thoracic Surgeons–European Association for CardioThoracic Surgery (STAT) category 3 (32%), followed by STAT category 5 (29%). The composite ICU-30 outcome occurred in 29 neonates (20%): two died before POD 30, 22 remained in the ICU beyond POD 30, and five were readmitted to the ICU within 30 POD. Neonates with ICU-30 were more likely to have an extracardiac abnormality, chromosomal abnormality, genetic syndrome, and a higher STAT category operation (all $p < 0.05$).

On univariate comparison, three biomarkers were lower preoperatively in neonates with ICU-30 than in those without the outcome (brain-derived neurotrophic factor, C-C motif chemokine ligand 5, and osteocalcin), whereas five biomarkers were higher in neonates with ICU-30 (suppression of tumorigenicity 2 [ST2], trefoil factor 3 [TFF3], growth differentiation factor 15 [GDF15], interleukin-6, and neutrophil gelatinase-associated lipocalin [NGAL]; **Supplemental Table 2**, https://links.lww.com/CCX/B708, **Fig. 2**).

For models that included both clinical variables and biomarkers, performance metrics are summarized in **Table 2**, **Fig. 3**, and **Supplemental Table 3** (https://links.lww.com/CCX/B708). Across all algorithms, AUROC values ranged from 0.66 to 0.73, with XGBoost and RF demonstrating higher overall discrimination. Because the outcome was moderately imbalanced, we also evaluated precision-recall performance. AUPRC values ranged from 0.58 to 0.62, with XGBoost achieving the highest value. Notably, all models achieved an AUPRC more than two times the outcome prevalence (0.16), indicating moderate discrimination for the positive class. To further account for imbalance in threshold-based classification, we evaluated performance at the Youden-optimal threshold, which balances sensitivity and specificity (Table 2). At this operating point, XGBoost achieved the highest Youden's *J* (0.42), with sensitivity of 0.66, specificity of 0.77, PPV of 0.41, and NPV of 0.90. In contrast, applying a fixed probability threshold of 0.5 (Supplemental Table 3, https://links.lww.com/CCX/B708) resulted in uniformly low sensitivities across all models (0.07–0.14). Overall, these internally cross-validated results suggest that XGBoost provided the most balanced performance among the evaluated models.

On review of variable importance (gain) in the combined clinical and biomarker XGBoost model, the top three features were NGAL, TFF3, and GDF15 (**Supplemental Fig. 1**, https://links.lww.com/CCX/B709). Of the top 20 features by gain, only one was

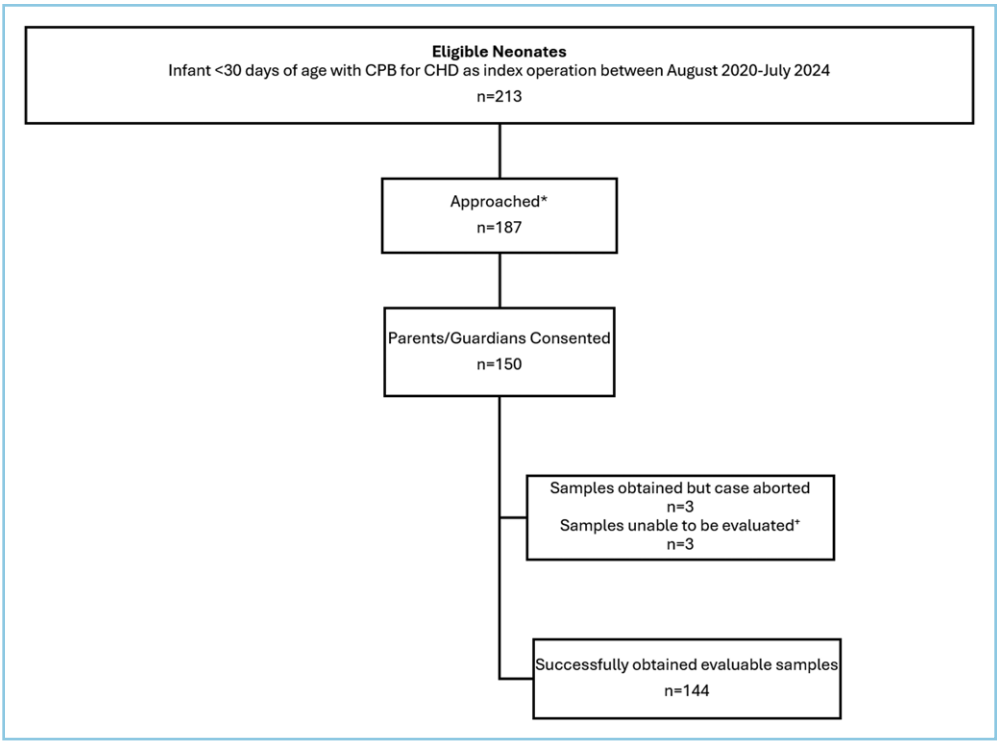

Figure 1. Consolidated Standards of Reporting Trials diagram for subject screening, approach, consent, and evaluable samples. *Parents/guardians were approached mostly during weekend daytime hours due to research coordinator availability. Patients who were screened but not approached were due to lack of staff or parent/guardian availability. +Samples were excluded from biomarker analysis because of insufficient sample quantity or concentrations outside the validated analytic range of the assay. CHD = congenital heart disease, CPB = cardiopulmonary bypass.

a clinical variable (age at surgery). Using SHAP to assess predictive impact and directionality (**Fig. 4**), the most influential features were NGAL, a disintegrin and metalloproteinase with a thrombospondin type 1 motif, member 13 (ADAMTS13), and angiopoietin-2 (Ang-2); higher values of NGAL and Ang-2 and lower ADAMTS13 were associated with ICU-30. Among clinical covariates, the presence of single-ventricle physiology and STAT 5 operations were associated with increased risk of ICU-30. Among biomarkers, higher cystatin C was associated with increased risk, whereas higher fibronectin and CCL5 were generally associated with lower predicted risk. Of the top 20 SHAP features, only four were clinical variables (single-ventricle physiology, STAT 5, age at surgery, and gestational age).

DISCUSSION

In this prospective study, we evaluated preoperative circulating biomarkers using machine learning techniques to improve prediction of poor postoperative outcome after CPB in neonates with CHD. To our knowledge, this is among the largest prospective studies examining biomarkers in this setting and one of the few focused exclusively on neonates, a population at particularly high risk for adverse outcomes. Among the evaluated machine learning approaches, XGBoost performed best overall, and biomarker-only models outperformed models based on clinical factors alone. Notably, NGAL and TFF3 were the top two features by gain in the XGBoost model, whereas SHAP analysis identified NGAL, ADAMTS13, and Ang-2 as the most influential features.

Prediction of poor outcome in CHD is predominantly based on nonmodifiable risk factors, such as birth weight, gestational age, and the presence of chromosomal or genetic abnormalities (41), as well as the inherent risk of the required surgery as summarized by the Society of Thoracic Surgeons–European Association for Cardiothoracic Surgery (STAT) categories (42) or Risk Adjustment for Congenital Heart Surgery categories (43). Although these factors are strongly associated with worse outcomes, many are nonmodifiable and do not capture underlying mechanisms of organ-specific inflammation or injury. Previous biomarker studies have been limited, often from a single center, and have focused on more heterogeneous populations spanning wider age ranges and different cardiac procedures (8–12). These groups carry different baseline risks for mortality and poor postoperative outcomes, making it difficult to apply previously reported biomarkers (including ST2, Galectin-3, N-terminal pro-B-type natriuretic peptide, and glial fibrillary acidic protein) to neonatal populations. Data from neonatal-only studies are especially limited, and many have focused

TABLE 1.
Preoperative Clinical Characteristics of the Neonatal Cohort According to ICU-30 Status

Clinical Variable Name	Overall Cohort, *N* = 144	No ICU-30, *N* = 115	ICU-30, *N* = 29	*p*
Male sex, *n* (%)	74 (51.4)	61 (53.0)	13 (44.8)	0.429
Race, *n* (%) White Black or African American Indian/Asian Multiracial or other Refused/unknown	93 (64.6) 12 (8.3) 3 (2.1) 30 (20.8) 6 (4.2)	75 (65.2) 9 (7.8) 2 (1.7) 24 (20.9) 5 (4.3)	18 (62.1) 3 (10.3) 1 (3.4) 6 (20.7) 1 (3.4)	0.952
Ethnicity, *n* (%) Non-Hispanic or Non-Latino Hispanic or Latino Refused	109 (75.7) 31 (21.5) 4 (2.8)	87 (75.7) 24 (20.9) 4 (3.5)	22 (75.9) 7 (24.1) 0 (0.0)	0.570
Gestational age, wk, median (IQR)	39 (38–39)	39 (38–39)	38 (38–39)	0.061
Birth weight, kg, median (IQR)	3.21 (2.9–3.61)	3.24 (2.93–3.62)	3.16 (2.9–3.41)	0.45
Extracardiac abnormality, *n* (%)	31 (21.5)	22 (19.1)	9 (31.0)	0.045
Chromosomal abnormality, *n* (%)	40 (27.8)	27 (23.5)	13 (44.8)	0.049
Genetic syndrome, *n* (%)	31 (21.5)	22 (19.1)	9 (31.0)	0.051
Preoperative prostaglandin use, *n* (%)	120 (83.3)	96 (82.8)	24 (82.7)	0.785
Preoperative mechanical ventilation, *n* (%)	48 (33.3)	39 (33.9)	9 (31.0)	0.998
Preoperative vasoactive/inotropic support, *n* (%)	34 (23.6)	25 (21.7)	9 (31.0)	0.328
Age at surgery (d), median (IQR)	3.58 (2.58–5.23)	3.65 (2.59–5.33)	3.03 (2.50–4.86)	0.16
Society of Thoracic Surgeons–European Association for Cardiothoracic Surgery category, *n* (%) 1 2 3 4 5	7 (4.9) 32 (22.2) 46 (31.9) 18 (12.5) 41 (28.5)	7 (6.1) 30 (26.1) 37 (32.2) 16 (13.9) 25 (21.7)	0 (0.0) 2 (6.9) 9 (31.0) 2 (6.9) 16 (55.2)	0.004
Cardiopulmonary bypass times, min, median (IQR) Total bypass Cross clamp Deep hypothermic circulatory arrest	140 (96–176) 77 (43–117) 0 (0–27)	140 (91.8–182.5) 80 (43–118) 0 (0–25.3)	139 (101–103) 69 (46–97) 2 (0–38)	 0.884 0.576 0.080

IQR = interquartile range.

primarily on cardiac markers such as BNP and troponin, with only modest preoperative predictive performance (44, 45).

The prominence of NGAL and TFF3 in predicting ICU-30 suggests that preoperative end-organ vulnerability may be an important pathway linking neonatal CHD physiology to adverse postoperative outcomes. NGAL emerged as the most important feature in both SHAP and gain assessments. NGAL is a well-studied marker of early renal injury in both neonatal (46) and cardiac populations (47). Given the difficulty of assessing renal injury in newborns, because neonatal

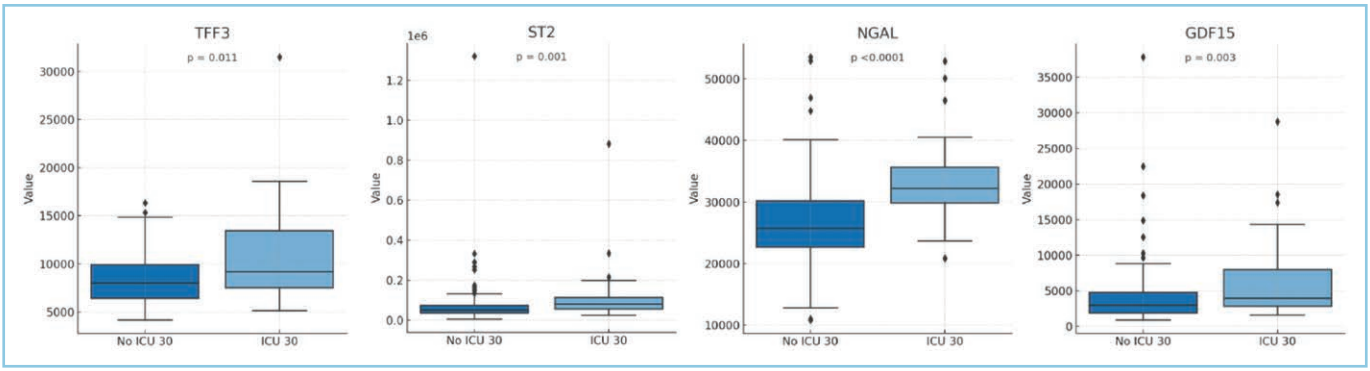

Figure 2. Box-and-whisker plots of unadjusted biomarker levels according to ICU-30 status for four of the eight biomarkers that differed significantly between groups ($p < 0.05$). The four shown here are growth differentiation factor 15 (GDF15), neutrophil gelatinase-associated lipocalin (NGAL), suppression of tumorigenicity 2 (ST2), and trefoil factor 3 (TFF3). Also significantly different but not shown: brain-derived neurotrophic factor, C-C motif chemokine ligand 5/regulated upon activation, normal T cell expressed and presumably secreted, interleukin-6, osteocalcin.

TABLE 2.
Classification Performance of the Combined Models at Youden-Optimal Thresholds

Model	Youden's *J*	Area Under the Receiver Operating Characteristic Curve	Area Under the Precision-Recall Curve	Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value
Least absolute shrinkage and selection operator	0.318	0.657	0.385	0.483	**0.835**	**0.424**	0.865
Random forest	0.367	**0.731**	**0.414**	**0.793**	0.574	0.319	**0.917**
Extreme gradient boosting	**0.420**	**0.731**	**0.358**	0.655	0.765	0.413	0.898

The highest-performing model is bolded for each metric.

creatinine frequently reflects maternal creatinine (48), biomarkers of renal injury or altered renal perfusion may be particularly important for risk stratification. TFF3, by contrast, is expressed in gastrointestinal epithelium and contributes to mucosal protection and repair; it has been studied in neonates with necrotizing enterocolitis (49, 50) and in children with sepsis (51). In neonates with ductal-dependent physiology, diastolic runoff, and altered systemic perfusion may plausibly contribute to this pattern of renal and gastrointestinal vulnerability. These findings support further investigation of preoperative splanchnic organ dysfunction and perfusion-sensitive biomarkers as potentially modifiable contributors to postoperative risk.

ADAMTS13 and Ang-2 are both markers of endothelial activation and dysfunction, and each ranked highly in SHAP analyses. Higher ADAMTS13 levels were associated with better outcomes, consistent with findings in pediatric critical illness, including sepsis (33, 52). Mechanistically, lower ADAMTS13 activity may reflect endothelial injury (53) or serve as a secondary marker of inflammation, both of which may be relevant to postoperative outcomes. Lower ADAMTS13 levels have also been associated with thrombus formation, providing a possible mechanistic link to organ injury from impaired blood flow (32, 54). Because ADAMTS13 levels vary with gestational age (55), further study in neonatal subgroups is warranted. Ang-2,

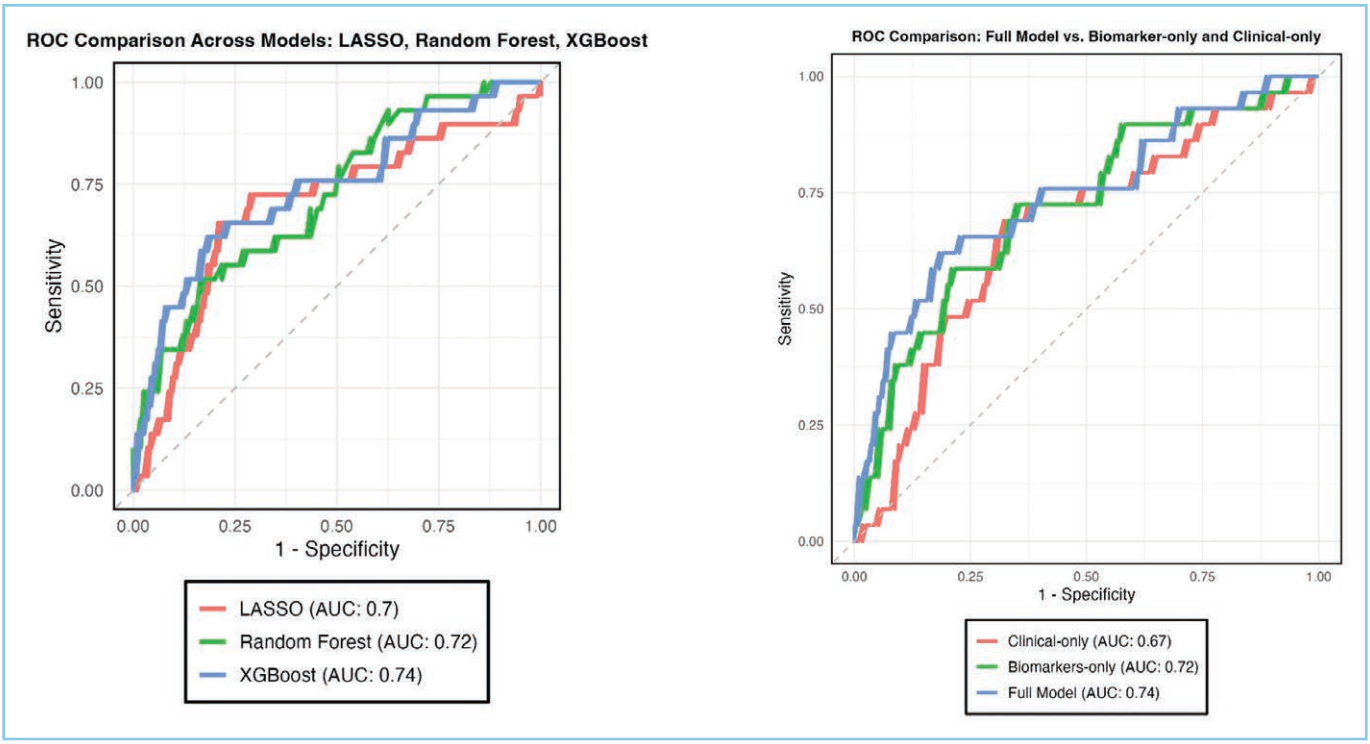

Figure 3. Left, Receiver operating characteristic (ROC) curves for the three machine learning models: least absolute shrinkage and selection operator (LASSO; *red*), random forest (*green*), and extreme gradient boosting (XGBoost; *blue*), with corresponding AUROC values. **Right**, ROC curves for the XGBoost models including all variables (full model, *blue*), biomarkers only (*green*), and clinical variables only (*red*).

which is involved in angiogenesis and maintenance of vascular stability, has been associated with outcomes in adult myocardial infarction (56), heart failure (57, 58), and renal disease (59). In congenital heart disease, Ang-2 has been associated with worse outcomes in adults with heart failure (26) and in Fontan physiology (60). A previous small neonatal study (n = 30) also linked Ang-2 elevation to longer bypass times and worse outcomes (61). Our findings are consistent with that observation and further support Ang-2 as a biomarker of interest in neonates undergoing CPB.

The application of machine learning enabled a novel integration of clinical and biomarker data. Because biomarkers are often interrelated and collinear with clinical factors, traditional approaches such as univariate or multivariable regression may not fully capture complex interactions. In contrast, machine learning methods can accommodate multidimensional relationships and may offer a more nuanced understanding of how biomarkers jointly contribute to risk prediction. As expected, different analytic approaches highlighted somewhat different features; however, the consistent prominence of NGAL across methods strengthens its relevance for future study. Additional work, including latent class analysis and evaluation of pre- to postoperative biomarker trajectories, may further clarify prognostic heterogeneity and identify potential targets for intervention.

Several study limitations merit consideration. First, this was a single-center study, and both the ICU-30 outcome and biomarker panel have thus far been evaluated only at our institution; external validation is therefore required before generalizability can be assumed. Second, the number of ICU-30 events was modest relative to the dimensionality of the predictor set, which increases the risk of model optimism despite the use of nested cross-validation and supports interpreting these findings as preliminary. Third, all reported model performance reflects internal validation only; calibration was not a primary focus of this analysis, and clinically actionable biomarker thresholds are not yet defined. Fourth, most biomarker measurements in this study were not available in real time and therefore could not have informed preoperative counseling or management during the study period. Although some markers, such as NGAL, may be adaptable to rapid testing platforms, clinical implementation would require timely assays, validated thresholds,

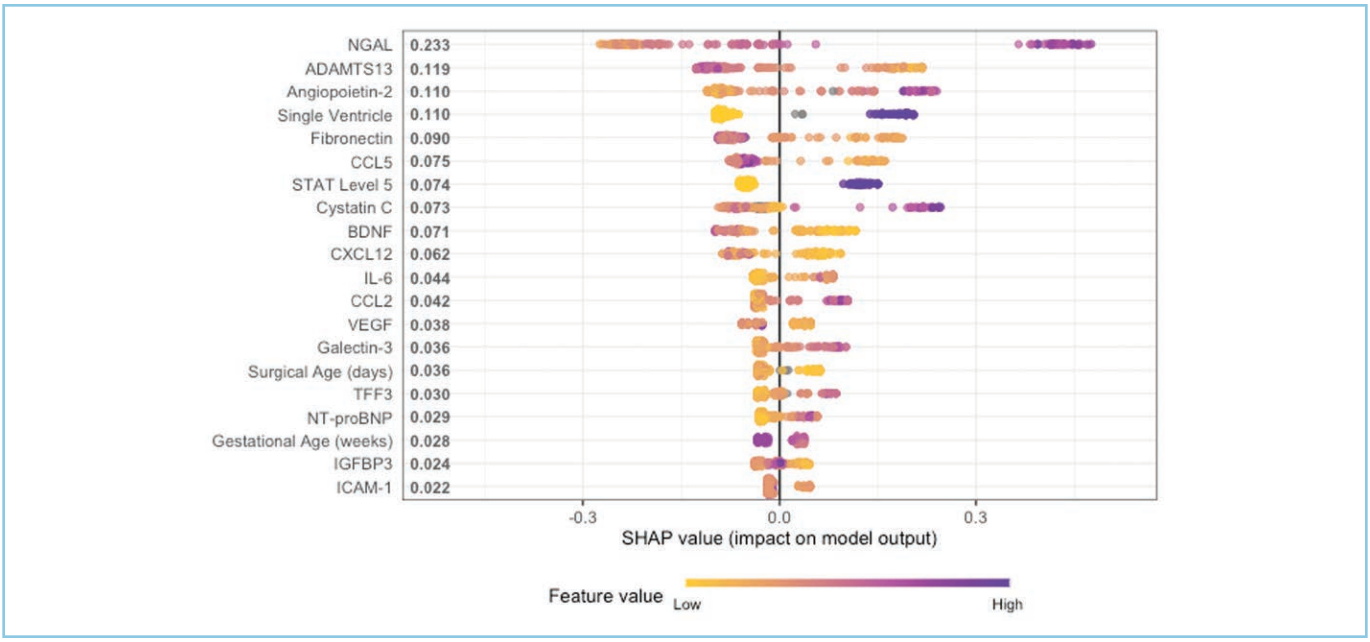

Figure 4. Shapley additive explanation (SHAP) summary plot of the top 20 features contributing to the extreme gradient boosting model. BDNF = brain-derived neurotrophic factor, CCL5 = C-C motif chemokine ligand 5, CXCL12 = C-X-C motif chemokine ligand 12, ICAM-1 = intercellular adhesion molecule-1, IGFBP3 = insulin-like growth factor-binding protein 3, IL-6 = interleukin-6, NGAL = neutrophil gelatinase-associated lipocalin, STAT = Society of Thoracic Surgeons–European Association for Cardiothoracic Surgery, TFF3 = trefoil factor 3, VEGF = vascular endothelial growth factor.

and demonstration that biomarker-informed risk stratification improves decision-making or outcomes. Fifth, although machine learning can identify important features and accommodate complex correlations, it does not provide effect estimates in the same manner as traditional regression models and should not be interpreted as establishing causality. Accordingly, these biomarkers should not yet be used to guide clinical decision-making. These limitations also define important directions for future work, including multicenter external validation, more comprehensive assessment of calibration and clinical utility, and mechanistic studies aimed at identifying modifiable pathways associated with adverse postoperative outcomes.

CONCLUSIONS

In this prospective neonatal cohort, biomarker-only machine learning prediction models improved prediction of poor postoperative outcome after CPB compared with models based on clinical risk factors alone. The identification of markers of renal and gastrointestinal epithelial injury (NGAL and TFF3), together with vascular injury and inflammation (ADAMTS13 and Ang-2), suggests that these pathways may be important contributors to risk in this vulnerable population. These findings support the potential value of biomarker-driven, machine learning-based approaches to improve prognostication, but they require external validation and further assessment of calibration and clinical utility before implementation.

1 *Division of Cardiac Critical Care Medicine, Department of Cardiovascular Services, Nemours Children's Hospital (Delaware), Wilmington, DE.*
2 *Division of Critical Care Medicine, Department of Anesthesiology and Critical Care, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA.*
3 *Children's Hospital of Philadelphia, Philadelphia, PA.*
4 *University of Pennsylvania Translational & Correlative Studies Laboratory, Philadelphia, PA.*
5 *Data and Analytics Group, The Cardiac Center, The Children's Hospital of Philadelphia, Philadelphia, PA.*
6 *Department of Family Medicine and Community Health, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA.*
7 *Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA.*
8 *Division of Cardiology, Children's Hospital of Philadelphia, Philadelphia, PA.*
9 *Division of Cardiac Anesthesiology, Department of Anesthesiology and Critical Care, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA.*

10 *Division of Critical Care Medicine, Department of Pediatrics, Nemours Children's Hospital (Delaware), Sidney Kimmel School of Medicine, Thomas Jefferson University, Philadelphia, PA.*

11 *Division of Cardiothoracic Surgery, Department of Surgery, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA.*

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's website (http://journals.lww.com/ccejournal).

This study is funded by The Brett Boyer Foundation and Big Hearts to Little Hearts.

Drs. Gardner, Weiss, and Yehya are supported through the National Institutes of Health (K23, R01, and R01s, respectively). The authors have disclosed that they do not have any potential conflicts of interest.

For information regarding this article, E-mail: monique.gardner@nemours.org

REFERENCES

1. O'Brien SM, Clarke DR, Jacobs JP, et al: An empirically based tool for analyzing mortality associated with congenital heart surgery. *J Thorac Cardiovasc Surg* 2009; 138:1139–1153
2. Haynes HR, Camelo-Piragua S, Kurian KM: Prognostic and predictive biomarkers in adult and pediatric gliomas: Toward personalized treatment. *Front Oncol* 2014; 4:47
3. Spyrou N, Akahoshi Y, Ayuk F, et al: The utility of biomarkers in acute GVHD prognostication. *Blood Adv* 2023; 7:5152–5155
4. Whitney JE, Feng R, Koterba N, et al: Endothelial biomarkers are associated with indirect lung injury in sepsis-associated pediatric acute respiratory distress syndrome. *Crit Care Explor* 2020; 2:e0295
5. Yehya N, Varisco BM, Thomas NJ, et al: Peripheral blood transcriptomic sub-phenotypes of pediatric acute respiratory distress syndrome. *Crit Care* 2020; 24:1–10
6. Yehya N, Wong HR: Adaptation of a biomarker-based sepsis mortality risk stratification tool for pediatric acute respiratory distress syndrome. *Crit Care Med* 2018; 46:e9–e16
7. Wong HR, Salisbury S, Xiao Q, et al: The pediatric sepsis biomarker risk model. *Crit Care* 2012; 16:R174
8. Brown JR, Stabler ME, Parker DM, et al: Biomarkers improve prediction of 30-day unplanned readmission or mortality after paediatric congenital heart surgery. *Cardiol Young* 2019; 29:1051–1056
9. Green MD, Parker DM, Everett AD, et al: Cardiac biomarkers associated with hospital length of stay after pediatric congenital heart surgery. *Ann Thorac Surg* 2021; 112:632–637
10. Parker DM, Everett AD, Stabler ME, et al: The association between cardiac biomarker NT-proBNP and 30-day readmission or mortality after pediatric congenital heart surgery. *World J Pediatr Congenit Heart Surg* 2019; 10:446–453
11. Parker DM, Everett AD, Stabler ME, et al: Novel biomarkers improve prediction of 365-day readmission after pediatric congenital heart surgery. *Ann Thorac Surg* 2020; 109:164–170
12. Parker DM, Everett AD, Stabler ME, et al: Biomarkers associated with 30-day readmission and mortality after pediatric congenital heart surgery. *J Card Surg* 2019; 34:329–336
13. Gardner MM, Keim G, Hsia J, et al: Characterization of "ICU-30": A binary composite outcome for neonates with critical congenital heart disease. *J Am Heart Assoc* 2022; 11:e025494
14. Baggen VJ, van den Bosch AE, Eindhoven JA, et al: Prognostic value of galectin-3 in adults with congenital heart disease. *Heart* 2018; 104:394–400
15. Jones AL, Faerber J, Huang J, et al: Biomarkers, socioeconomic factors, and right ventricular function after surgical repair for tetralogy of Fallot. *Pediatr Cardiol* 2023; 44:1232–1241
16. Bar-Yosef O, Greidinger D, Iskilova M, et al: Neurological deficit is predicted by S100B in children after cardiac surgery. *Clin Chim Acta Int J Clin Chem* 2018; 481:56–60
17. Bembea MM, Rizkalla N, Freedy J, et al: Plasma biomarkers of brain injury as diagnostic tools and outcome predictors after extracorporeal membrane oxygenation. *Crit Care Med* 2015; 43:2202–2211
18. Schmitt B, Bauersfeld U, Schmid ER, et al: Serum and CSF levels of neuron-specific enolase (NSE) in cardiac surgery with cardiopulmonary bypass: A marker of brain injury? *Brain Dev* 1998; 20:536–539
19. Ng EW, Poon TC, Lam HS, et al: Gut-associated biomarkers L-FABP, I-FABP, and TFF3 and LIT score for diagnosis of surgical necrotizing enterocolitis in preterm infants. *Ann Surg* 2013; 258:1111–1118
20. Montagnana M, Danese E, Lippi G: Biochemical markers of acute intestinal ischemia: Possibilities and limitations. *Ann Transl Med* 2018; 6:341
21. Herbert C, Patel M, Nugent A, et al: Serum cystatin C as an early marker of neutrophil gelatinase-associated lipocalin-positive acute kidney injury resulting from cardiopulmonary bypass in infants with congenital heart disease. *Congenit Heart Dis* 2015; 10:E180–E188
22. Wang J, Zou L, Song Z, et al: Meta-analysis of RAGE gene polymorphism and coronary heart disease risk. *PLoS One* 2012; 7:e50790
23. Negrin LL, Halat G, Kettner S, et al: Club cell protein 16 and cytokeratin fragment 21-1 as early predictors of pulmonary complications in polytraumatized patients with severe chest trauma. *PLoS One* 2017; 12:e0175303
24. Chang H-Y, Li F, Li F-S, et al: Genetic polymorphisms of SP-A, SP-B, and SP-D and risk of respiratory distress syndrome in preterm neonates. *Med Sci Monit* 2016; 22:5091–5100
25. Miyamoto K, Takeuchi D, Inai K, et al: Prognostic value of multiple biomarkers for cardiovascular mortality in adult congenital heart disease: Comparisons of single-/two-ventricle physiology, and systemic morphologically right/left ventricles. *Heart Vessels* 2016; 31:1834–1847
26. Lukasz A, Beutel G, Kümpers P, et al: Angiopoietin-2 in adults with congenital heart disease and heart failure. *PLoS One* 2013; 8:e66861
27. Yin D-l, Zhao X-h, Zhou Y, et al: Association between the ICAM-1 gene polymorphism and coronary heart disease risk: A meta-analysis. *Biosci Rep* 2019; 39:BSR20180923
28. Küster H, Degitz K: Circulating ICAM-1 in neonatal sepsis. *Lancet (London, England)* 1993; 341:506

29. Wang W, Xu A, Xu H: The roles of vascular endothelial growth factor gene polymorphisms in congenital heart diseases: A meta-analysis. *Growth Factors* 2018; 36:232–238
30. Hao L, Wang N: Changes in plasma thrombomodulin and D-dimer levels and their clinical significance in neonates with sepsis. *Zhongguo Dang dai er ke za zhi= Chin J Contemp Pediatr.* 2013; 15:841–844
31. Lupia E, Goffi A, Bosco O, et al: Thrombopoietin as biomarker and mediator of cardiovascular damage in critical diseases. *Mediators Inflamm* 2012; 2012:390892
32. Katneni UK, Holcomb DD, Hernandez NE, et al: In silico features of ADAMTS13 contributing to plasmatic ADAMTS13 levels in neonates with congenital heart disease. *Thromb Res* 2020; 193:66–76
33. Papadogeorgou P, Boutsikou T, Boutsikou M, et al: A global assessment of coagulation profile and a novel insight into ADAMTS-13 implication in neonatal sepsis. *Biology* 2023; 12:1281
34. De Paepe ME, Wong T, Chu S, et al: Stromal cell-derived factor-1 (SDF-1) expression in very preterm human lungs: Potential relevance for stem cell therapy for bronchopulmonary dysplasia. *Exp Lung Res* 2020; 46:146–156
35. Tian C, Cheng L, Gu X: Cord blood TNF-α and IL-6 levels as diagnostic indicators of brain damage in neonates with non-asphyxia fetal distress. *Arch Gynecol Obstet* 2017; 295:337–342
36. Stojewska M, Wąsek-Buko M, Jakub B, et al: Evaluation of serum chemokine RANTES concentration as a biomarker in the diagnosis of early-onset severe infections in neonates. *Adv Hyg Exp Med/Postepy Hig Med Dosw* 2016; 70:272–279
37. Zhang R, Gong J, Wang S, et al: Relationship between serum B7-H3 levels and prognosis of congenital heart disease in children. *Pediatr Cardiol* 2019; 40:177–181
38. Richter B, Uray T, Krychtiuk KA, et al: Growth differentiation factor-15 predicts poor survival after cardiac arrest. *Resuscitation* 2019; 143:22–28
39. Gerdes JS, Polin RA: Sepsis screen in neonates with evaluation of plasma fibronectin. *Pediatr Infect Dis J* 1987; 6:443–446
40. Gerdes JS, Yoder MC, Douglas SD, et al: Decreased plasma fibronectin in neonatal sepsis. *Pediatrics* 1983; 72:877–881
41. Jacobs JP, O'Brien SM, Hill KD, et al: Refining the Society of Thoracic Surgeons congenital heart surgery database mortality risk model with enhanced risk adjustment for chromosomal abnormalities, syndromes, and Noncardiac congenital anatomic abnormalities. *Ann Thorac Surg* 2019; 108:558–566
42. O'Brien SM, Clarke DR, Jacobs JP, et al: An empirically based tool for analyzing mortality associated with congenital heart surgery. *J Thorac Cardiovasc Surg* 2009; 138:1139–1153
43. Jenkins KJ, Gauvreau K: Center-specific differences in mortality: Preliminary analyses using the Risk Adjustment in Congenital Heart Surgery (RACHS-1) method. *J Thorac Cardiovasc Surg* 2002; 124:97–104
44. Carmona F, Manso PH, Vicente WVA, et al: Risk stratification in neonates and infants submitted to cardiac surgery with cardiopulmonary bypass: A multimarker approach combining inflammatory mediators, N-terminal pro-B-type natriuretic peptide and troponin I. *Cytokine* 2008; 42:317–324
45. Mori Y, Nakashima Y, Kaneko S, et al: Risk factors for cardiac adverse events in infants and children with complex heart disease scheduled for Bi-ventricular repair: Prognostic value of pre-operative B-type natriuretic peptide and high-sensitivity troponin T. *Pediatr Cardiol* 2020; 41:1756–1765
46. Stoops C, Gavigan H, Krallman K, et al: The utility of urinary NGAL as an alternative for serum creatinine to detect acute kidney injury in infants exposed to nephrotoxic medications in the neonatal intensive care unit. *Neonatol* 2024; 121:203–212
47. Mishra J, Dent C, Tarabishi R, et al: Neutrophil gelatinase-associated lipocalin (NGAL) as a biomarker for acute renal injury after cardiac surgery. *Lancet (London, England)* 2005; 365:1231–1238
48. Kastl JT: Renal function in the fetus and neonate – the creatinine enigma. *Semin Fetal Neonatal Med* 2017; 22:83–89
49. Guthmann F, Börchers T, Wolfrum C, et al: Plasma concentration of intestinal- and liver-FABP in neonates suffering from necrotizing enterocolitis and in healthy preterm neonates. *Mol Cell Biochem* 2002; 239:227–234
50. Derikx JP, Evennett NJ, Degraeuwe PL, et al: Urine based detection of intestinal mucosal cell damage in neonates with suspected necrotising enterocolitis. *Gut* 2007; 56:1473–1475
51. Žurek J, Kýr M, Vavřina M, et al: Trefoil factor 3 as a marker of intestinal cell damage during sepsis. *Open Med* 2015; 10:261–266
52. El-Mekkawy MS, El-Deeb SM, El-Hanafy MF: ADAMTS13 in pediatric sepsis: A prognostic biomarker with potential therapeutic implications. *Egypt Paediatr Assoc Gaz* 2023; 71:71
53. Russell RT, McDaniel JK, Cao W, et al: Low plasma ADAMTS13 activity is associated with coagulopathy, endothelial cell damage and mortality after severe paediatric trauma. *Thromb Haemost* 2018; 118:676–687
54. Lambers M, Goldenberg NA, Kenet G, et al: Role of reduced ADAMTS13 in arterial ischemic stroke: A pediatric cohort study. *Ann Neurol* 2013; 73:58–64
55. Strauss T, Elisha N, Ravid B, et al: Activity of Von Willebrand factor and levels of VWF-cleaving protease (ADAMTS13) in preterm and full term neonates. *Blood Cells Mol Dis* 2017; 67:14–17
56. Lee S-J, Lee C-K, Kang S, et al: Angiopoietin-2 exacerbates cardiac hypoxia and inflammation after myocardial infarction. *J Clin Invest* 2018; 128:5018–5033
57. Eleuteri E, Di Stefano A, Giordano A, et al: Prognostic value of angiopoietin-2 in patients with chronic heart failure. *Int J Cardiol* 2016; 212:364–368
58. Pöss J, Ukena C, Kindermann I, et al: Angiopoietin-2 and outcome in patients with acute decompensated heart failure. *Clin Res Cardiol* 2015; 104:380–387
59. Tsai Y-C, Lee C-S, Chiu Y-W, et al: Angiopoietin-2 as a prognostic biomarker of major adverse cardiovascular events and all-cause mortality in chronic kidney disease. *PLoS One* 2015; 10:e0135181
60. Shirali AS, Lluri G, Guihard PJ, et al: Angiopoietin-2 predicts morbidity in adults with Fontan physiology. *Sci Rep* 2019; 9:18328
61. Watkins KJ, Geisler S, Doman T, et al: Rise in angiopoietin-2 following neonatal cardiac surgery is associated with adverse clinical outcomes. *Pediatr Crit Care Med* 2020; 21:e827–e833