*Letter*

# When Numbers Mislead: Reassessing Statistical Foundations in PCA-Based Prognostication of Fontan Outcomes

**Zhihao Lei \***

Department of Statistics and Data Science, Cornell University, Ithaca, NY 14853, USA

\* Corresponding author. E-mail: zl2269@cornell.edu (Z.L.)

To the Editor: We read with interest the article by Ferrari et al. [1], which applied principal component analysis (PCA) to a Fontan registry cohort. While innovative in concept, several methodological concerns warrant discussion.

First, the study's events-per-variable ratio appears likely insufficient for reliable multivariable modeling. Only 16 composite outcome events (and zero deaths) occurred over 11 years [1]. Yet over 30 covariates were analyzed in Cox regression models, including multiple PCA-derived components. This violates the well-established rule-of-thumb requiring at least ~10 outcome events per predictor variable [2]. Peduzzi et al. demonstrated that low events-per-variable leads to biased coefficients and spurious results [2]. With an EPV far below 1 in some models, the risk of overfitting is extreme. The model likely capitalized on noise, and any apparent associations may be unreliable. In short, the prognostic findings in [1] should be interpreted with extreme caution, given the model's severe overfitting risk [2].

Second, we are concerned by the overinterpretation of PCA components. Ferrari et al. label certain principal components as representing specific biological phenomena (e.g., "lymphatic dysfunction" for PC6) based on the variables loaded [1]. However, principal components are abstract mathematical constructs—linear combinations of variables that maximize variance, and do not have inherent physiologic meaning [3]. Assigning pathophysiological labels without external validation is speculative. The authors did not report the full PCA loadings or perform an independent validation to confirm that, for example, "PC6" truly correlates with lymphatic failure in external metrics. PCA yields orthogonal dimensions of variance, not pre-defined clinical factors, and any thematic naming must be supported by evidence (e.g., correlation of a "lymphatic" PC with known lymphatic imaging or biomarkers). Absent such justification, attributing mechanistic significance to PCs is overreaching. Components should be treated as agnostic features unless validated.

Third, the study's composite endpoint is problematic. Ferrari et al. combined disparate events—from relatively milder complications like protein-losing enteropathy (PLE) to major endpoints like heart transplant listing or failure—into a single outcome [1]. Such heterogeneity in the composite outcome can undermine interpretability [4]. Composite endpoints are most valid when components are of similar clinical importance and have a common biological rationale [4]. Here, the included events likely differ markedly in pathophysiology and severity. For instance, PLE (only 2 cases in the cohort) is a chronic Fontan complication, whereas transplant represents end-stage failure [1]. Combining these into one outcome assumes they are equivalent signals of "decline", which is questionable. This violates the principle that large gradients in event importance can mislead composite results [4]. A positive composite finding might be driven by frequent but less-critical events, obscuring true effects on the most serious outcomes [4]. We urge caution: a composite of convenience may not translate into meaningful clinical insight.

Fourth, there are issues with missing data handling and potential temporal bias. The authors performed median imputation for missing values (implicitly, since "PCA requires no missing data" and patients were excluded or imputed) [1]. Single imputation methods (e.g., mean or median) can introduce bias and understate uncertainty, as each missing value is filled in deterministically [5]. Best practices recommend assessing the missingness mechanism and using multiple imputation or sensitivity analyses to ensure results are robust to missing data assumptions [5]. No such analyses were reported. Additionally, the data were collected over an 11-year span [1] with no adjustment for secular trends. In

Fontan populations, outcomes and management have improved over time, as evidenced by declining Fontan mortality rates in more recent eras [6]. Failing to adjust for cohort year or era could confound the results—e.g., earlier patients might have worse outcomes due to older management strategies, not because of the risk factors identified. The analysis in [1] did not account for these temporal confounders. We recommend that any prognostic model spanning a decade consider adjusting for calendar year or treatment era, or at least test whether the predictive effect of variables changed over time.

Finally, the paper's claims of predictive improvement and clinical relevance appear overstated, and aspects of the statistical inference are inconsistent. The authors compare models by area under the ROC curve (AUC/c-statistic) and other metrics (AIC, sensitivity/specificity), but did not report any statistical tests for differences in discrimination. For example, they note that one PCA-based model had a higher c-statistic than an ejection fraction model [1], but no formal comparison (e.g., DeLong's test) was performed to confirm that this difference is significant [7]. It is known that small increases in AUC may not translate to meaningful net reclassification of patients or improved decision-making. Claims of "most accurate classifier", therefore, may be premature without net reclassification or decision-curve analysis to gauge clinical impact. Moreover, the study reports multiple performance metrics on the same data (AUC, AIC, hazard ratios for several models) without correction for multiplicity. Cherry-picking the best results (e.g., highlighting whichever metric paints a given model in the best light) can inflate Type I error. In Ref. [1], one model is touted for the highest AIC performance and another for the highest AUC, with significance asterisked for various hazard ratios—an inconsistent approach to statistical inference that risks overstating findings. We encourage a more rigorous comparison using statistical tests for improvement and adjustment for multiple comparisons when assessing incremental predictive value.

In summary, Ferrari et al. [1] present an interesting application of machine learning to Fontan outcomes, but the above methodological limitations raise important questions about the robustness and clinical significance of their conclusions. Overfitting due to scant events, ambiguous interpretation of PCA factors, an inappropriate composite outcome, and analytical biases (missing data and uncorrected multiple comparisons) could potentially overstate the model's purported value. We urge the authors and readers to view the results with caution. Future studies should incorporate larger multicenter data (to increase event rates), more rigorous handling of missing values and temporal trends, and validated endpoints and analyses to ensure that any identified risk factors truly advance our ability to predict and improve Fontan patient outcomes.

## Ethics Statement

Not Applicable.

## Informed Consent Statement

Not Applicable.

## Data Availability Statement

Not Applicable.

## Funding

## Declaration of Competing Interest

The author declares no conflicts of interest.

## References

1. Ferrari MR, Schäfer M, Di Maria MV, Hunter KS. Application of principal component analysis to heterogeneous Fontan registry data identifies independent contributing factors to decline. *Cardiovasc Sci*. **2025**, *2*, 10005. doi:10.70322/cvs.2025.10005.
2. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. *J. Clin. Epidemiol*. **1996**, *49*, 1373–1379. doi:10.1016/S0895-4356(96)00236-3.
3. Jolliffe IT. *Principal Component Analysis*, 2nd ed.; Springer: New York, NY, USA, 2002.
4. Ferreira-González I, Permanyer-Miralda G, Domingo-Salvany A, Busse JW, Heels-Ansdell D, Montori VM, et al. Problems

with use of composite end points in cardiovascular trials: Systematic review of randomised controlled trials. *BMJ* **2007**, *334*, 786. doi:10.1136/bmj.39136.682083.AE.

5.  Jakobsen JC, Gluud C, Wetterslev J, Winkel P. When and how should multiple imputation be used for handling missing data in randomised clinical trials—A practical guide with flowcharts. *BMC Med. Res. Methodol*. **2017**, *17*, 162. doi:10.1186/s12874-017-0442-1.

6.  Iyengar AJ, Winlaw DS, Galati JC, Celermajer DS, Wheaton GR, Gentles TL, et al. Trends in Fontan surgery and risk factors for early adverse outcomes after Fontan surgery: The Australia and New Zealand Fontan Registry experience. *J. Thorac. Cardiovasc. Surg*. **2014**, *148*, 566–575. doi:10.1016/j.jtcvs.2013.09.074.

7.  DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* **1988**, *44*, 837–845. doi:10.2307/2531595.