

1 **Title:** Application of Principal Component Analysis to Heterogenous Fontan Registry Data
2 Identifies Independent Contributing Factors to Decline

3 **Short Title:** PCA of Registry Data Predicts Fontan Decline

4 **Authors:** Margaret R. Ferrari, PhD¹, Michal Schäfer, MD PhD², Kendall S. Hunter PhD*^{3,4},
5 Michael V. Di Maria MD*⁵

6
7 **Affiliations:** ¹SafeBeat Rx Inc, Carson CA, United States

8 ²Division of Cardiothoracic Surgery, University of Utah Health, Salt Lake City, Utah, 84132,
9 United States

10 ³Department of Bioengineering, University of Colorado Anschutz Medical Campus, Aurora, CO,
11 80045, United States

12 ⁴Division of Cardiology, Heart Institute, Children's Hospital Colorado, University of Colorado
13 Anschutz Medical Campus, Aurora, CO, 80045, United States

14 ⁵Division of Pediatric Cardiology, Department of Pediatrics, School of Medicine, University of
15 Michigan, Ann Arbor, MI, 48109, United States

16 *Both senior authors contributed equally to this work

17 **Corresponding Author:** Kendall Hunter, Kendall.Hunter@cuanschutz.edu, Department of
18 Bioengineering, University of Colorado Anschutz Medical Campus, Aurora, CO, 80045, United
19 States

20
21 **Word Count:** 5231

22 **Abbreviations**

23 SVD, single ventricle disease; PLE, protein losing enteropathy; PB, plastic bronchitis; FALD,
24 Fontan-associated liver disease; PC MRI, phase contrast magnetic resonance imaging; VVCR,
25 ventricular vascular coupling ration; VO_2 , rate of oxygen consumption; AAo, ascending aorta;
26 SVC, superior vena cava; IVC, inferior vena cava; LPA, left pulmonary artery; cMRI, cardiac
27 MRI; TCPC, total cavopulmonary connection; PCA, principal component analysis; PCs,
28 principal components; EF, ejection fraction; EDVi, end diastolic volume index; ESVi, end
29 systolic volume index; CI, cardiac index; BNP, B-type natriuretic peptide; GGT, gamma-
30 glutamyl transferase; AST, aspartate aminotransferase; SaO_2 , arterial oxygen saturation; FEV1,
31 forced expiratory volume in one second; mSVCP, mean SVC pressure; mPAP, mean pulmonary
32 artery pressure; HLHS, hypoplastic left heart syndrome; TA, tricuspid atresia; double outlet right
33 ventricle, DORV; double inlet left ventricle, DILV; hypoplastic right heart syndrome, HRHS;
34 TCPC, total cavopulmonary circuit; FVC, forced vital capacity; RV, residual volume; TLC, total
35 lung capacity; Alk phos, alkaline phosphatase; BUN, blood urea nitrogen

36 **Abstract**

37 Single ventricle heart disease is a severe and life-threatening illness, and improvements in
38 clinical outcomes of those with Fontan circulation have not yet yielded acceptable survival over
39 the past two decades. Patients are at risk of developing a diverse variety of Fontan-associated
40 comorbidities that ultimately requires heart transplant. Our observational cohort study goal was
41 to determine if principal component analysis (PCA) applied to data collected from a substantial
42 Fontan cohort can predict functional decline (N=140). Heterogeneous data broadly consisting of
43 measures of cardiac and vascular function, exercise (VO_{2max}), lymphatic biomarkers, and blood
44 biomarkers were collected over 11 years at a single site; in that time, 16 events occurred that are
45 considered here in a composite outcome measure. After standardization and PCA, principal
46 components (PCs) representing >5% of total variance were thematically labeled based on their
47 constituents and tested for association with the composite outcome. Our main findings suggest
48 that the 6th PC (PC6), representing 7.1% percent of the total variance in the set, is greatly
49 influenced by blood serum biomarkers and superior vena cava flow, is a superior measure of
50 proportional hazard compared to EF, and displayed the greatest accuracy for classifying Fontan
51 patients as determined by AUC. In bivariate hazard analysis, we found that models combining
52 systolic function (EF or PC5) and lymphatic dysfunction (PC6) were most predictive, with the
53 former having the greatest AIC, and the latter having the highest c-statistic. Our findings support
54 our hypothesis that a multifactorial model must be considered to improve prognosis in the Fontan
55 population.

56 **Introduction**

57 Patients born with single ventricle heart disease (SVD), a severe and rare congenital heart
58 defect (CHD), are subjected to three palliative surgeries that culminate in the Fontan circulation
59 [1]. While staged palliation addresses the primary concerns of obstructed systemic blood flow
60 and cyanosis in a condition like hypoplastic left heart syndrome, a range of Fontan associated
61 comorbidities, including lymphatic, liver, and cardiac damage, are often apparent in adolescence
62 [2]. Because morbidity and mortality after Fontan surgery remain unacceptably high [2], new
63 approaches to predict patient decline are sorely needed. The goal of this work is to develop a new
64 prognostic model for patients with SVD and explore machine learning methods as a tool of risk
65 stratification in the Fontan population.

66 Principal component analysis (PCA) is a data reduction technique that is often applied to
67 large data sets in research [3-7], although it has not yet been applied to the Fontan population
68 outside of waveform analysis [8, 9]. Examples where PCA has been utilized to find novel
69 associations in large datasets that would have not have been amenable to more conventional
70 statistical approaches include: Scientists have used PCA to identify patterns of inflammatory and
71 adhesion molecules that contribute to muscle weakness acquired in the intensive care unit [4]. A
72 similar study in a population of adults used PCA to identify inflammatory markers that precede
73 major adverse cardiovascular events (MACE) following heart attack and found that the PC
74 influenced by interleukin-6 and interleukin-8 was a better predictor of MACE at one year than
75 univariate cytokine measures [6]. Additionally, PCA has recognized lymphocyte-monocyte-
76 neutrophil indices that contribute to disease severity in several cohorts of COVID-19 patients [5].
77 The aforementioned studies support our hypothesis that a PCA approach may be necessary to
78 understand and predict outcomes in heterogenous disease states like the Fontan population,

79 where a very large number of potential predictor variables exist that stem from anatomic,
80 surgical, imaging, laboratory, functional testing domains.

81 We have previously applied PCA to characterize non-pulsatile cavopulmonary flow
82 waveforms [8, 9]; here we expand that approach to include heterogenous biomarkers. Another
83 benefit of using a PCA approach in clinical data analysis is inclusion of correlative/colinear
84 parameters, of which many statistical outcomes models prohibit [5]. Our primary objective in
85 this study was to assess our previously defined novel waveform measures and other clinical
86 parameters in a heterogenous PCA approach, all in support of the overall hypothesis that
87 machine-learning extracted PCs will delineate patients with Fontan-associated comorbidities and
88 reveal parameters that indicate circulatory failure in patients with a Fontan circulation. More
89 specifically, a machine-learning extracted PC, which will consist of a pattern of abnormalities in
90 multiple of cardiac and non-cardiac test results, previously unrecognized as an important predictor
91 of outcomes, will be associated with a composite outcome of Fontan failure.

92 **Methods**

93 One-hundred and forty SVD patients that underwent cardiac MRI (cMRI) at Children's
94 Hospital Colorado between July 2011 and August 2022 were included in this retrospective
95 cohort study, permitted by the Colorado Multiple Institutional Review Board as a portion of
96 Fontan at Altitude Registry for Outcomes (FAROUT). All patients cared for in the Fontan
97 Multidisciplinary Clinic at the Children's Hospital Colorado have undergone surveillance testing
98 for end-organ damage and Fontan-associated comorbidities by way of a clinical practice
99 guideline since 2016 and were included. The FAROUT registry was queried and abstracted data
100 was used as a foundation for a study database, in addition to our single site venous flow patterns
101 [10]. Variables collected are shown in Table I. For the purposes of survival analysis, study

102 subject status was evaluated as of December 1, 2022, and a composite outcome was defined as
103 the development of plastic bronchitis (PB, n=1), protein-losing enteropathy (PLE, n=2), referral
104 to transplant (RTT, n=9), received a transplant (n=4), or death (n=0) from the time of cMRI to
105 time to follow up.

106 *cMRI Acquisition*

107 Phase images and corresponding magnitude images of the superior vena cava (SVC),
108 inferior vena cava (IVC), and left pulmonary artery (LPA) were obtained using a PC-MRI, ECG
109 gated sequence as previously described [11, 12] by applying a 1.5 or 3.0 Magnetom Avanto
110 (Siemens Medical Solutions, Erlangen, Germany) or Ingenia (Philips Medical System, Best,
111 Netherlands) Tesla magnet using a phased-array body surface coil. A free breathing PC-MRI
112 sequence was used under the following conditions: time to repetition, 14-28 milliseconds/25-40
113 cardiac phases; time to echo, 2.2-3.5 milliseconds; matrix, 160 x 256; flip angle, 25 degrees;
114 100% k space sampling; cross-sectional pixel resolution, 0.82 x 0.82 mm² and 1.56 x 1.56 mm²;
115 slice thickness, 5 millimeters. Heart rate dependent, PC-MRI acquisition varied 2-3 minutes for
116 each vessel. Aliasing was accommodated for using the following velocity-encoding values: SVC
117 and IVC, 75-100 cm/second; LPA, 50-100 cm/second. The AAO, SVC, and IVC images were
118 acquired in axial cine stack and the LPA in vertical long axis, all orthogonal to flow.

119 *Flow Profile Analysis*

120 Flow profile characteristics were assessed as described previously with slight
121 modifications [9]. Flow waveforms were acquired by precise parallel segmentation of 2D phase-
122 contrast image series in Circle CVi42 (Calgary, Canada). Flow data was captured for each
123 patient at the SVC, IVC, and LPA and imported into MATLAB (Natick, MA). Each waveform
124 was normalized by dividing each flow point by patient BSA to minimize size effect on the raw

125 data. Data was interpolated using cubic spline interpolation to 40 points, guaranteeing size-
126 matched array lengths for further analyses. Three data matrices were created containing single-
127 site flow data and the size of each data matrix varied and is as follows: SVC (124 x 40), IVC
128 (132 x 40), LPA (125 x 40).

129 *Clinical Biomarkers*

130 Global cardiovascular and ventricular indicators (EF, CO, CI, EDVi, ESVi, SVi) have
131 long been established as the benchmark for Fontan patient status [8, 13-15] and therefore were
132 included in the analysis for validation purposes. VVCR, mean catheterization pressures (mPAP
133 and SVC mean pressure), VO_{2max} , BNP_{max} and O_2 saturations were also included in the outcomes
134 analysis, all of which have been independently linked to Fontan circulation health and outcomes
135 [16-19] . We have previously determined that biomarkers indicative of lymphatic function and
136 PLE, specifically aspartate aminotransferase, alkaline phosphatase, cystatin-c and creatinine
137 were associated with caval flow patterns [9], strongly indicating that these parameters may
138 identify, or be a predictive of, which Fontan patients will experience circulatory failure.
139 Similarly, biomarkers such as albumin, total protein, blood urea nitrogen and platelet count were
140 included due to previous reports relating these to Fontan patient cyanosis and pulmonary blood
141 flow [20, 21].

142 *Principal Component Analysis*

143 PCA requires that the input data matrix has a value assigned to each position, and
144 therefore after the exported registry was read into MATLAB, patients with more than five
145 measures missing were removed from analysis. This was done in an effort to maintain a missing
146 data rate of less than 5% [22], and the resulting clinical parameters and demographic information
147 can be found in Table I. Patients that had missing values less than or equal to five were replaced

148 with the column median, as a value is required for each position in the input matrix. Columns
149 were normalized by subtracting the column mean from each sample and dividing by the
150 corresponding column standard deviation, and the resulting matrix was the input for PCA. Scree
151 plots were created to determine PCs representing greater than 5% of total variance.
152 Interpretation of which clinical parameters had the greatest influence on each PC were
153 graphically determined by visualizing PC eigenvectors.

154 *Statistics*

155 Statistical analyses were performed in GraphPad Prism and began by determining the
156 univariate Cox hazard ratio (HR) for PCs 1-10 and EF, a measure of systolic function that serves
157 as a benchmark diagnostic for patients with a Fontan circulation [23]. Akaike's information
158 criterion (AIC) and the c-statistic were also gathered, and measures with the greatest c-statistic
159 were used to create receiver operating characteristic (ROC) curves and determine the area under
160 the curve (AUC) and Youden's index, defined as $(\text{sensitivity}(x) - \text{specificity}(x)) - 1$. The
161 optimum sensitivity, specificity and clinical threshold for grouping was found and used to
162 defined groups for Kaplan-Meier survival analysis. The Mantel-Cox log-rank test was used to
163 determine if a significant difference existed between Kaplan-Meier curves. Univariate
164 parameters with the greatest c-statistics were used, up to two parameters at a time, for
165 multivariate (bivariate) regression analysis and AIC was used to compare univariate and
166 multivariate predictive models.

167 **Results**

168 *Principal Component Analysis*

169 The size of the original data matrix imported to MATLAB was 140 x 31 and was reduced
170 to 115 x 31 after removal of patients missing greater than five measures. The remaining matrix

171 had a 4.15% rate of missing data, and therefore fell within the 5% acceptable rate [22]. Columns
172 of the matrix included scores for single site SVC, IVC, and LPA flow patterns, each representing
173 a patient's contribution to that PC's waveform pattern, EF, EDVi, ESVi, SVi, CO, CI, BSA,
174 BNP max, AST, VVCR, lowest SaO₂, SVC mean pressure, mPAP, albumin, platelets, alkaline
175 phosphatase, total protein, creatinine, BUN, and cystatin-C (Table I).

176 Following PCA, a scree plot was used to identify the percent each PC contributed to the
177 overall variance in the original data set and can be seen in Figure 1. The first PC accounted for
178 about 17.5% of the original data matrix variance, followed by approximately 11.5% for PC2,
179 8.5% for PC3, and subsequently decreased as the PC number increased (Figure 1). Each of the
180 first 7 PCs accounted for more than 5% of the total variance, and together explained 70.3% of
181 that variance; PCs up to PC10 (3.6% of total variance, 77.7% cumulative variance) were
182 considered for survival analysis.

183 Interpretation of the first two PCs was aided by the biplot displayed in Figure 2, where
184 the blue lines represent the eigenvectors, or PC coefficients, and the length and direction
185 represents that parameter's influence on each PC. For example, further distance from the origin
186 on the x-axis means greater contribution to PC1, therefore EDVi, ESVi, EF, and VVCR
187 contributed greatest to PC1 (Figure 2). PC2 variance is explained by deviance from zero along
188 the y-axis, and major influencing parameters include CO, SVi, LPA PC1 and IVC PC1 (Figure
189 2). The red data points represent the scores, or how each patient sample contributes to the PCs.

190 The clinical implications of each PC were examined using bar graphs of the PC
191 eigenvectors, where the clinical parameters (x axis) and their relative contributions to each PC (y
192 axis) are displayed in Figure 3. The first PC was highly influenced by cardiac parameters,
193 including EF, EDVi, ESVi, CI, VVCR, AST, SVC PC1 Scores, and IVC PC1 scores, which

194 primarily describe cardiac function and the downstream effects in the Fontan circulation. The
195 fourth PC was highly influenced by IVC and LPA waveform patterns in addition to
196 cavopulmonary pressures and cystatin-C. PC5 was influenced by cardiac parameters
197 representative of systolic function, such as EF, SVi, CI, and VVCR, and waveform patterns IVC
198 PC2, SVC PC1 scores, and BUN. Albumin, alkaline phosphatase, total protein, BUN, BNP max,
199 and SVC waveforms scores influenced PC 6.

200 *Survival Analysis*

201 Univariate Cox proportional hazard ratio was determined for each PC and can be found in
202 Table II. The single best predictor of which patient is at a greater hazard is PC6 (AIC=109),
203 followed by the standard measure used for prediction in this population, EF (AIC=111) (Table
204 II). PC1 and PC5 also performed well, with AICs of 113 and 115. The hazard ratio for EF and
205 each PC is displayed in a forest plot in Figure 3, and the bars represent the 95% confidence
206 intervals. If a parameter's confidence interval crossed one, it was not statistically significant
207 (Figure 4).

208 Parameters with the greatest c-statistics were tested as classifiers of patients with Fontan
209 decline using ROC curves, and PC6 returned the greatest AUC at 0.767, followed by PC5
210 (AUC=0.740), and EF (AUC=0.696) (Figure 5). The accompanying optimum sensitivity and
211 specificity, determined using the greatest distance from the null hypothesis line or Youden's
212 index, was also found and shows that, while EF is highly specific (0.771) and therefore able to
213 designate patients with Fontan failure correctly (low EF is almost always accompanied by SVD
214 circulatory failure), its sensitivity is lacking at 0.643 (Figure 5). Sensitivity determines a
215 classifier's ability to label patients without Fontan decline correctly, and all PCs had the same, if
216 not superior, sensitivity compared to EF (Figure 5). PC6 had the greatest sensitivity (0.786) with

217 a reasonably balanced specificity at 0.686, suggesting it can rule healthy Fontan patients out as
218 having circulatory failure, and PC1 had both optimum sensitivity and specificity at 0.714 (Figure
219 4). PC6 also displayed the greatest maximum effective biomarker, represented by Youden's
220 index of 0.471 (Figure 5).

221 Grouping patients based on Youden's index allows for Kaplan-Meier curve generation,
222 displayed in Figure 6. Though EF was determined to have statistically significant differences in
223 survival using the Mantel Cox log rank test ($p=0.0006$), PC1 and PC5 performed better with p
224 values of 0.0003 and 0.0005 (Figure 6). PC6 also had highly significant differences in survival
225 ($p=0.0008$), though it was not as significant as EF, and PC4 also displayed a significant
226 difference in survival ($p=0.07$) (Figure 6).

227 The inputs for multivariate Cox hazard regression analysis were EF, ESVi, and PC1 and
228 PCs 4-6 and were chosen based on their univariate c -statistics. The models developed and each
229 covariate's HR estimate, 95% CI, p -value, c -statistic, and AIC are listed in Table III. The
230 greatest AIC, and therefore predictive model, was model B (0.807, AIC=97) and included EF
231 and PC6. However, model F, consisting of PC5 and PC6, has a greater c -statistic (0.845,
232 AIC=103) and therefore is more probable to randomly identify a patient that experienced an
233 event has a greater risk score than a patient that did not experience an event.

234 **Discussion**

235 In this study, we explored univariate and multivariate associations between composite outcomes
236 in a relatively large, single-center Fontan cohort, evaluating standard and ML-derived parameters
237 both singly as well as through heterogeneous feature reduction (PCA). Through use of PCA, we
238 determined PCs that appear to relate to specific features of Fontan decline, and that these features
239 were significant univariate and multivariate predictors of a composite event.

240 PC6 was identified as a highly significant measure of hazard in our cohort and was the
241 greatest univariate predictor of outcomes with an AIC of 109. The parameters that contributed
242 the greatest to this PC have also been associated with development of lymphatic dysfunction,
243 specifically PLE, in the Fontan population and include albumin, alkaline phosphatase, total
244 protein, BUN and BNP max [23-25]. Cavopulmonary flow patterns were also found to influence
245 PC6 and have previously been suspected as contributors to PLE [23]. PC6 was also the most
246 accurate classifier of patients with Fontan decline from those without, which supports our
247 understanding of PLE development and poorer prognosis in the Fontan population. It is also
248 worth noting that only two patients (of 16) experienced a composite outcome of PLE.
249 Additionally, with a sensitivity of 78.6% and specificity of 68.6%, this PC may be clinically
250 useful in categorizing (or risk stratifying) patients. For example, after undergoing surveillance
251 testing for end-organ damage and Fontan-related comorbidities, a patient could have their CMR-
252 derived flow waveforms and other biomarkers examined here projected into a known feature set
253 (heterogeneous PCs), after which these latter PCs would be used to classify the patient as either
254 at-risk or not, based on the present analysis. Additional work, in terms of data collection and
255 validation, as well as potentially longitudinal studies targeting causality, must be performed to
256 identify the role parameters represented by PC6 play in the development of PLE, though this
257 study provides a promising foundation for further research.

258 Survival was significantly predicted by all PCs identified as having a significant hazard
259 ratio (1, 4, 5, 6), though two performed superior to EF, the standard measure of health and
260 cardiac function in this population. PC1 performed better at prediction than EF alone and
261 represented components of general cardiac function, including EF, EDVi, ESVi, VVCR, CI, and
262 cavopulmonary flow measures. Our findings suggest that PCA is a supported method for

263 inclusion of colinear parameters, and inclusion of such measures does in fact improve prediction
264 of outcomes better than a single measure of ventricular function or multivariate testing. PC5 also
265 improved prediction of outcomes compared to EF and was influenced by EF, SVi, CI, BSA,
266 VVCR, BUN, SVC and LPA waveform patterns, most of which are affected by systolic
267 ventricular function. However, little improvement was noted in a multivariate model that
268 included EF and PC1 (AIC=112) and EF and PC5 (AIC=110) compared to the individual
269 parameter's AIC scores [EF (111), PC1 (115), and PC5 (109)] which may be due to redundant
270 information (i.e. EF, and thus, systolic function, is now accounted for twice in the model). These
271 results suggest that a PCA approach can improve outcomes prediction in the Fontan population
272 and continue to support the hypothesis that machine-learning extracted PCs will clearly delineate
273 SVD patients with Fontan-associated comorbidities to those without, in addition to our
274 previously published hypothesis that a multivariable approach, in this case PCA, improves
275 prediction of this heterogenous patient population with multiple organ systems in various stages
276 of failure [9].

277 The best predictive model explored in this study, determined by AIC, is the multivariate
278 Cox regression model B (Table III) including covariates EF and PC6. If PC6 is in fact linked to
279 PLE, our findings suggest a combination of systolic ventricular function and measures indicative
280 of lymphatic dysfunction may be an avenue for improved prognostication. This model, however,
281 did not have the greatest c-statistic, which suggests it may not be suitable for ranking patients
282 according to risk. Model F, including covariates PC5 and PC6, had the greatest c-statistic and is
283 most suitable in determining which patients are at a higher risk.

284 The limitations of this study, as previously described [8], include those inherent to PCA.
285 Linear data reduction does not consider non-linear reduction methods and, as the name suggests,

286 compresses the original data for usability and is accompanied by a loss of, ideally insignificant,
287 original data variance. Additionally, several patients were removed from analysis because PCA
288 requires that the input matrix has no missing data. The cohort may contain a selection bias,
289 because not all of our Fontan patients received a routine CMR examination in the past. Finally,
290 machine learning methods thrive on large datasets, and while the final set used for PCA (N=115)
291 is large for a pediatric population, clearly multicenter studies or learning networks that pool such
292 data will offer even greater insights into disease progression. Despite limitations, this work has
293 established that a heterogenous approach to PCA is beneficial to outcomes prediction in Fontan
294 patients, and that our novel single site venous waveform patterns contribute to PCs predictive of
295 decline.

296 *Conclusion*

297 The goal of this study was to determine if a heterogenous PCA approach applied to the Fontan
298 cohort can predict functional decline in this population. Our main findings suggest that PC6,
299 which represented roughly 7% of the overall variance and is greatly influenced by blood serum
300 biomarkers and SVC flow, is a superior measure of proportional hazard in this population
301 compared to EF. We also found that PC6 displayed the greatest accuracy for classifying Fontan
302 patients, as determined by AUC, and we identified two PCs that indeed predicted survival in this
303 population better than EF. Our findings support our suspicions that a multifactorial model must
304 be considered to improve prognosis in the Fontan population.

305 **Acknowledgements**

306 This research was supported by the Jayden DeLuca Foundation, NIH CTSA Grant UL1
307 TR002535, and the American Heart Association Children's Heart Foundation Predoctoral
308 Congenital Heart Defect Research Award 20PRE35260057 to MRF. The authors would like to

309 additionally thank Dr Dunbar Ivy for his support of the dissertation work that led to this
310 manuscript.

311 **Sources of Funding:** This research was supported by the Jayden DeLuca Foundation, NIH
312 CTSA Grant UL1 TR002535, and the American Heart Association Children's Heart Foundation
313 Predoctoral Congenital Heart Defect Research Award 20PRE35260057 to MRF.
314

315 **Disclosures:** All of the authors have nothing to disclose.
316

317 **References**

- 318 1. Kritzmire, S.M. and A.E. Cossu, *Hypoplastic Left Heart Syndrome*, in *StatPearls*. 2020:
319 Treasure Island (FL).
- 320 2. Rychik, J., *Forty years of the Fontan operation: a failed strategy*. *Semin Thorac*
321 *Cardiovasc Surg Pediatr Card Surg Annu*, 2010. **13**(1): p. 96-100.
- 322 3. Lee, Y.K., E.R. Lee, and B.U. Park, *PRINCIPAL COMPONENT ANALYSIS IN VERY*
323 *HIGH-DIMENSIONAL SPACES*. *Statistica Sinica*, 2012. **22**(3): p. 933-956.
- 324 4. Witteveen, E., et al., *Increased Early Systemic Inflammation in ICU-Acquired Weakness;*
325 *A Prospective Observational Cohort Study*. *Crit Care Med*, 2017. **45**(6): p. 972-979.
- 326 5. Qi, Y., et al., *Lymphocyte-monocyte-neutrophil index: a predictor of severity of*
327 *coronavirus disease 2019 patients produced by sparse principal component analysis*.
328 *Virol J*, 2021. **18**(1): p. 115.
- 329 6. Kristono, G.A., et al., *An IL-6-IL-8 score derived from principal component analysis is*
330 *predictive of adverse outcome in acute myocardial infarction*. *Cytokine X*, 2020. **2**(4): p.
331 100037.
- 332 7. Federolf, P.A., K.A. Boyer, and T.P. Andriacchi, *Application of principal component*
333 *analysis in clinical gait research: identification of systematic differences between healthy*
334 *and medial knee-osteoarthritic gait*. *J Biomech*, 2013. **46**(13): p. 2173-8.
- 335 8. Schafer, M., et al., *Flow profile characteristics in Fontan circulation are associated with*
336 *the single ventricle dilation and function: principal component analysis study*. *Am J*
337 *Physiol Heart Circ Physiol*, 2020. **318**(5): p. H1032-H1040.
- 338 9. Margaret R. Ferrari, M.S., Kendall S. Hunter, Michael V. Di Maria, *Coupled waveform*
339 *patterns in the arterial and venous fontan circulation are related to parameters of*
340 *pulmonary, lymphatic and cardiac function*. *International Journal of Cardiology*
341 *Congenital Heart Disease*, 2023. **11**(100429).
- 342 10. Ferrari, M.R., et al., *Central Venous Waveform Patterns in the Fontan Circulation*
343 *Independently Contribute to the Prediction of Composite Survival*. *Pediatric Cardiology*,
344 2023.
- 345 11. Schäfer, M., et al., *Characterization of CMR-derived haemodynamic data in children*
346 *with pulmonary arterial hypertension*. *European Heart Journal - Cardiovascular Imaging*,
347 2016. **18**(4): p. 424-431.
- 348 12. Schafer, M., et al., *Main pulmonary arterial wall shear stress correlates with invasive*
349 *hemodynamics and stiffness in pulmonary hypertension*. *Pulm Circ*, 2016. **6**(1): p. 37-45.

- 350 13. Ghelani, S.J., et al., *Longitudinal changes in ventricular size and function are associated*
351 *with death and transplantation late after the Fontan operation.* J Cardiovasc Magn
352 Reson, 2022. **24**(1): p. 56.
- 353 14. Rathod, R.H., et al., *Cardiac magnetic resonance parameters predict transplantation-free*
354 *survival in patients with fontan circulation.* Circ Cardiovasc Imaging, 2014. **7**(3): p. 502-
355 9.
- 356 15. Ishizaki, U., et al., *Global strain and dyssynchrony of the single ventricle predict adverse*
357 *cardiac events after the Fontan procedure: Analysis using feature-tracking cine magnetic*
358 *resonance imaging.* J Cardiol, 2019. **73**(2): p. 163-170.
- 359 16. Saiki, H., et al., *Ventricular-Arterial Function and Coupling in the Adult Fontan*
360 *Circulation.* J Am Heart Assoc, 2016. **5**(9).
- 361 17. Diller, G.P., et al., *Exercise intolerance in adult congenital heart disease: comparative*
362 *severity, correlates, and prognostic implication.* Circulation, 2005. **112**(6): p. 828-35.
- 363 18. Mori, M., et al., *Catheter-measured Hemodynamics of Adult Fontan Circulation:*
364 *Associations with Adverse Event and End-organ Dysfunctions.* Congenit Heart Dis, 2016.
365 **11**(6): p. 589-597.
- 366 19. Loomba, R.S., et al., *Timing of Fontan Completion in Children with Functionally*
367 *Univentricular Hearts and Isomerism: The Impact of Age, Weight, and Pre-Fontan*
368 *Arterial Oxygen Saturation.* Pediatr Cardiol, 2019. **40**(4): p. 753-761.
- 369 20. Tomkiewicz-Pajak, L., et al., *Iron deficiency and hematological changes in adult patients*
370 *after Fontan operation.* J Cardiol, 2014. **64**(5): p. 384-9.
- 371 21. Cheng, A.L., et al., *Elevated Low-Shear Blood Viscosity is Associated with Decreased*
372 *Pulmonary Blood Flow in Children with Univentricular Heart Defects.* Pediatr Cardiol,
373 2016. **37**(4): p. 789-801.
- 374 22. Schafer, J.L., *Multiple imputation: a primer.* Stat Methods Med Res, 1999. **8**(1): p. 3-15.
- 375 23. Rychik, J., et al., *Evaluation and Management of the Child and Adult With Fontan*
376 *Circulation: A Scientific Statement From the American Heart Association.* Circulation,
377 2019: p. CIR00000000000000696.
- 378 24. Papadopoulou-Legbelou, K., M. Kavga, and M. Fotoulaki, *Protein-losing enteropathy*
379 *after Fontan operation: enteric capsule findings and management with atrial pacing.*
380 Hippokratia, 2017. **21**(4): p. 208.
- 381 25. Chin, A.J., et al., *Serum alkaline phosphatase reflects post-Fontan hemodynamics in*
382 *children.* Pediatr Cardiol, 2009. **30**(2): p. 138-45.
- 383

Table I. Measures used in heterogenous PCA, including novel single vessel waveform data, hemodynamic, global cardiovascular, blood, kidney, liver, and respiratory biomarkers, and their mean or median and corresponding standard deviation or interquartile range.

Measure	Mean or Median	St. dev or IQR
SVC Single PC1 Scores	-0.522	-2.87 - 3.12
SVC Single PC2 Scores	-0.478	-0.870 - 0.0638
IVC Single PC1 Scores	-0.172	-3.21 - 3.69
IVC Single PC2 Scores	0.21	-1.53 - 1.35
LPA Single PC1 Score	0.397	-3.01 - 2.07
LPA Single PC2 Score	0.181	-0.656 - 1.24
EF	47.9	8.11
EDVi	93.3	27
ESVi	49.8	20.5
SVi	43.5	10.4
CO	4.43	1.43
CI	3.41	1.04
BSA	1.32	0.415
BNP max	30	15 - 61.8
AST	45	38 - 57
VVCR	0.963	0.3
Lowest SaO2	86.4	5.56
SVC Mean	13	11 - 14
mPAP	12	10 - 13
Alb	4.54	0.685
Platelets	197	65.2
Alk Phos	158	80.3
Total Protein	7.53	1.1
Creatinine	0.61	0.49 - 0.763
BUN	14	12 - 17
Cystatin C	0.86	0.755 - 0.96

Table II. Univariate cox regression hazard ratio for each PC and EF, the gold standard for systolic function and patient decline in the Fontan population, and the accompanying 95% confidence intervals, their p-value, c-statistic and AIC.

Variable	HR	95% CI	P value	c stat	AIC
PC1	1.40	1.12 to 1.75	0.00280	0.705	115
PC2	0.836	0.607 to 1.13	0.254	0.555	121
PC3	0.941	0.649 to 1.38	0.753	0.453	123
PC4	1.43	1.00 to 1.99	0.0404	0.619	119
PC5	0.502	0.315 to 0.773	0.00250	0.718	113
PC6	0.449	0.280 to 0.691	0.000500	0.759	109
PC7	1.50	0.894 to 2.60	0.140	0.617	120.5
PC8	0.931	0.558 to 1.59	0.793	0.587	123
PC9	0.741	0.469 to 1.27	0.242	0.539	122
PC10	0.929	0.574 to 1.58	0.777	0.519	123
EF	0.873	0.803 to 0.944	0.000900	0.726	111

Table III. Multivariate models were developed, up to two measures at a time, for each PC and systolic measures EF and ESVi, the corresponding HR estimates, 95% CIs, p-values, c-statistic and AIC (used for model selection).

	Covariates	HR Estimate	95% CI	p value	c stat	AIC
A	EF	0.906	0.804 to 1.02	0.101	0.726	112
	ESVi	1.02	0.979 to 1.06	0.413		
B	PC6	0.455	0.287 to 0.682	0.0003	0.807	97
	EF	0.875	0.811 to 0.939	0.0003		
C	PC5	0.667	0.402 to 1.04	0.0931	0.738	110
	EF	0.901	0.822 to 0.988	0.0247		
D	PC4	1.37	0.960 to 1.93	0.0765	0.756	110
	EF	0.884	0.814 to 0.952	0.0019		
E	PC1	1.11	0.802 to 1.52	0.505	0.734	112
	EF	0.895	0.802 to 0.994	0.0413		
F	PC5	0.574	0.372 to 0.841	0.0071	0.845	103
	PC6	0.461	0.279 to 0.721	0.0012		
G	PC4	1.395	0.982 to 1.95	0.0558	0.803	108
	PC6	0.445	0.271 to 0.695	0.0007		
H	PC4	1.448	1.01 to 2.06	0.0426	0.764	111
	PC5	0.501	0.313 to 0.771	0.0024		
I	PC1	1.322	1.06 to 1.65	0.0123	0.799	105
	PC6	0.516	0.327 to 0.766	0.0022		
J	PC1	1.595	1.22 to 2.11	0.0006	0.784	103
	PC5	0.482	0.313 to 0.717	0.0005		
K	PC1	1.404	1.12 to 1.75	0.0028	0.781	113
	PC4	1.462	1.01 to 2.10	0.0425		

387
388
389

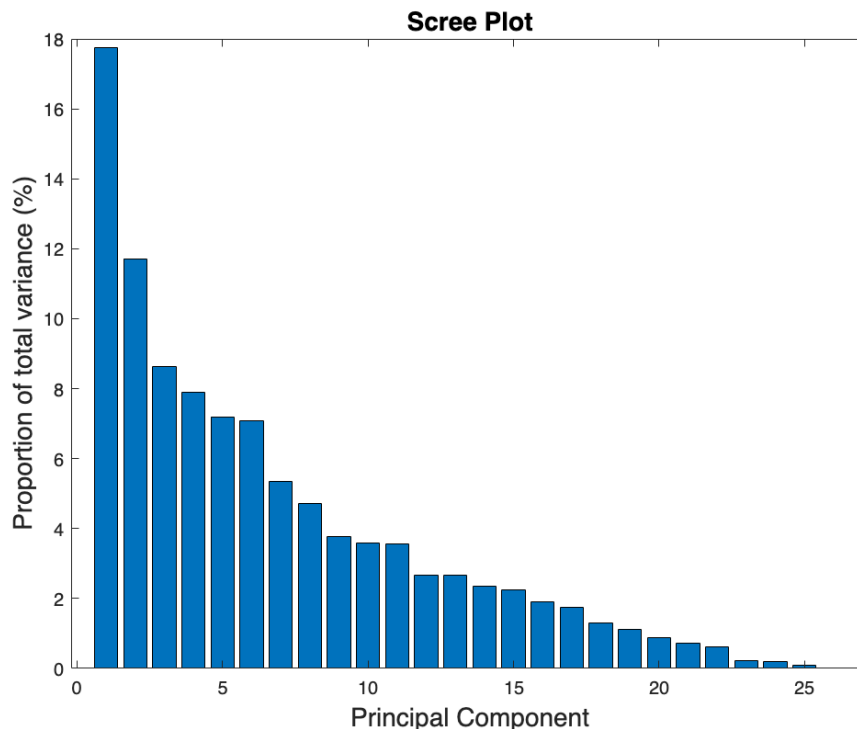


Figure 1. The scree plot displays each PC (x axis) and the percent variance it represents in the original data set (y axis).

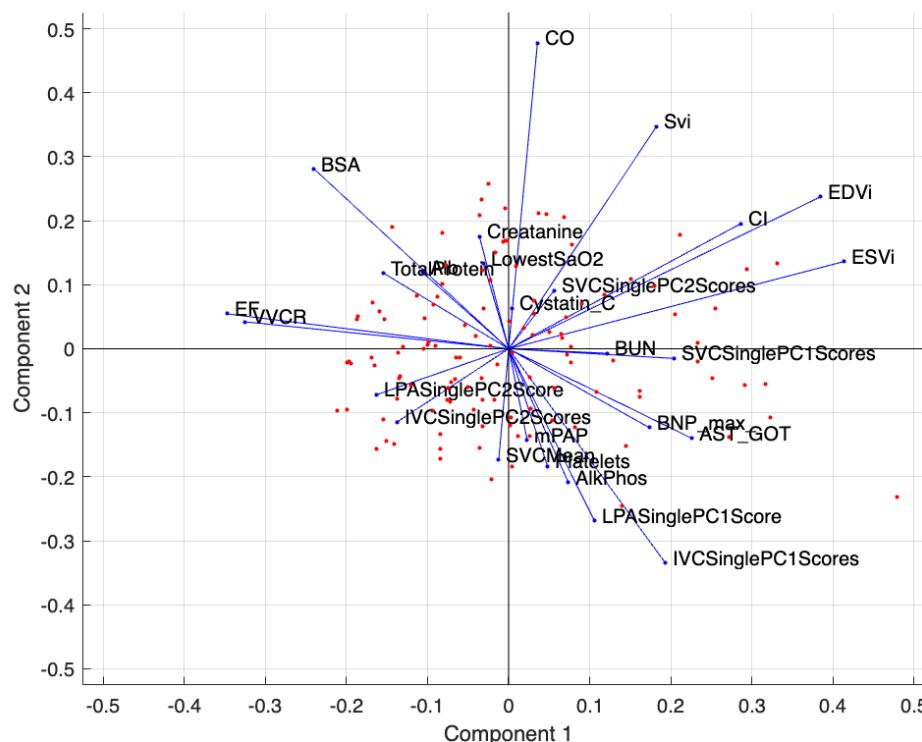


Figure 2. Principal component biplot that displays the scores returned from PCA, or each sample, in the original dimension and are represented by the red data points. The eigenvectors, or the coefficients, for each clinical parameter are displayed as the blue lines and the direction and length represent the influence each parameter has on PC1 (x axis) and PC2 (y axis).

39

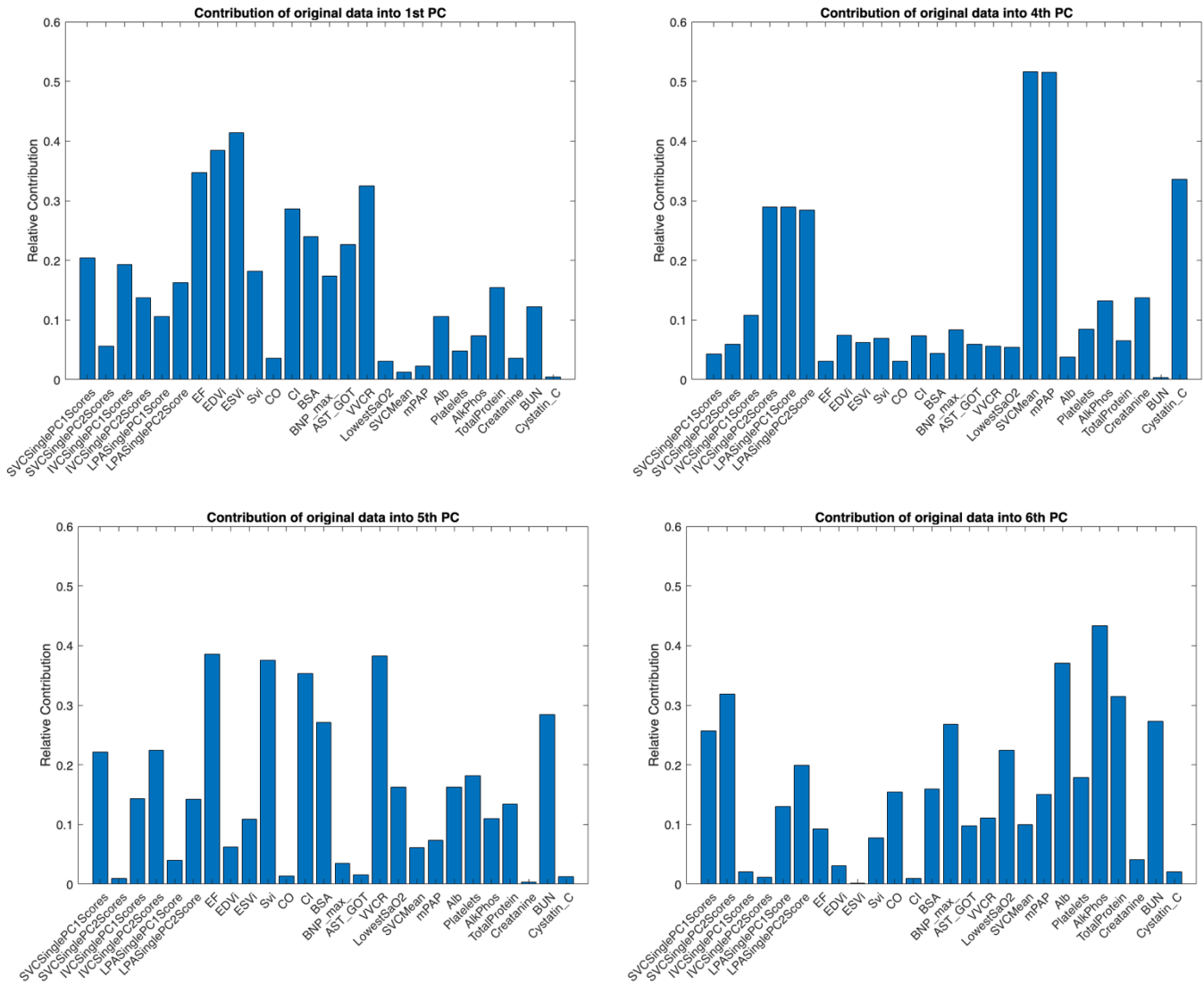


Figure 3. Bar graphs for PC1 and PCs 4-6 display each clinical parameter considered in PCA, or the column headers, and the amount (y axis, 0 up to 1) that parameter influences each PC.

391

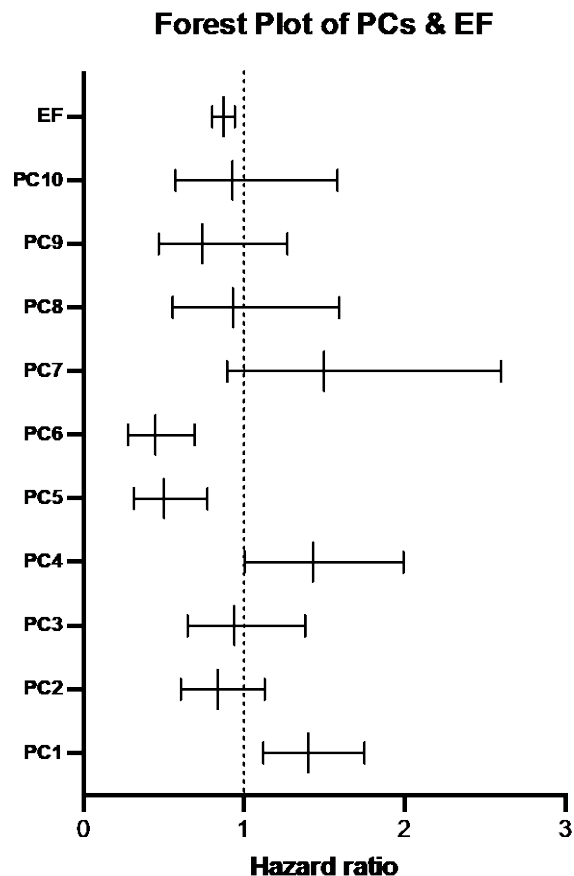


Figure 4. The forest plot displays the hazard ratio for each PC and EF and the corresponding 95% confidence intervals. Bars that cross 1, or the null hypothesis, represents no difference in hazards between patients that experienced an event versus those that did not.

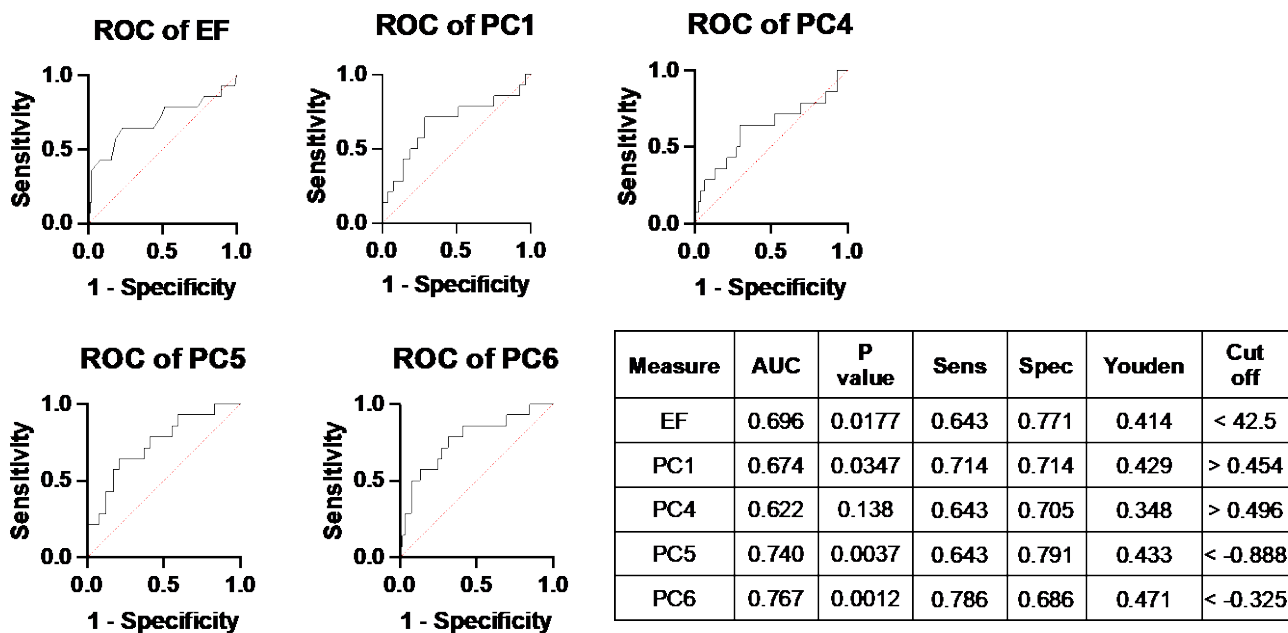


Figure 5. ROC curves for each PC that returned the greatest c-statistics and EF and the accompanying AUC, p-value, optimum sensitivity and specificity and the corresponding clinical cut off.

392

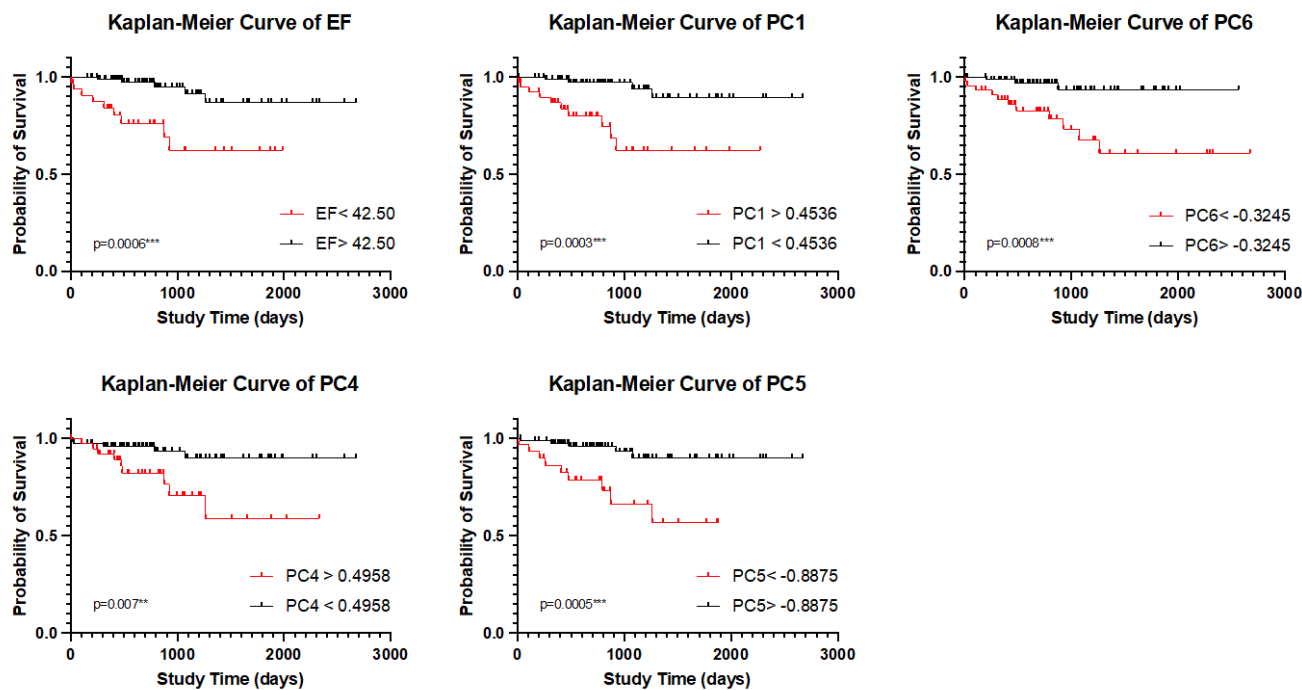
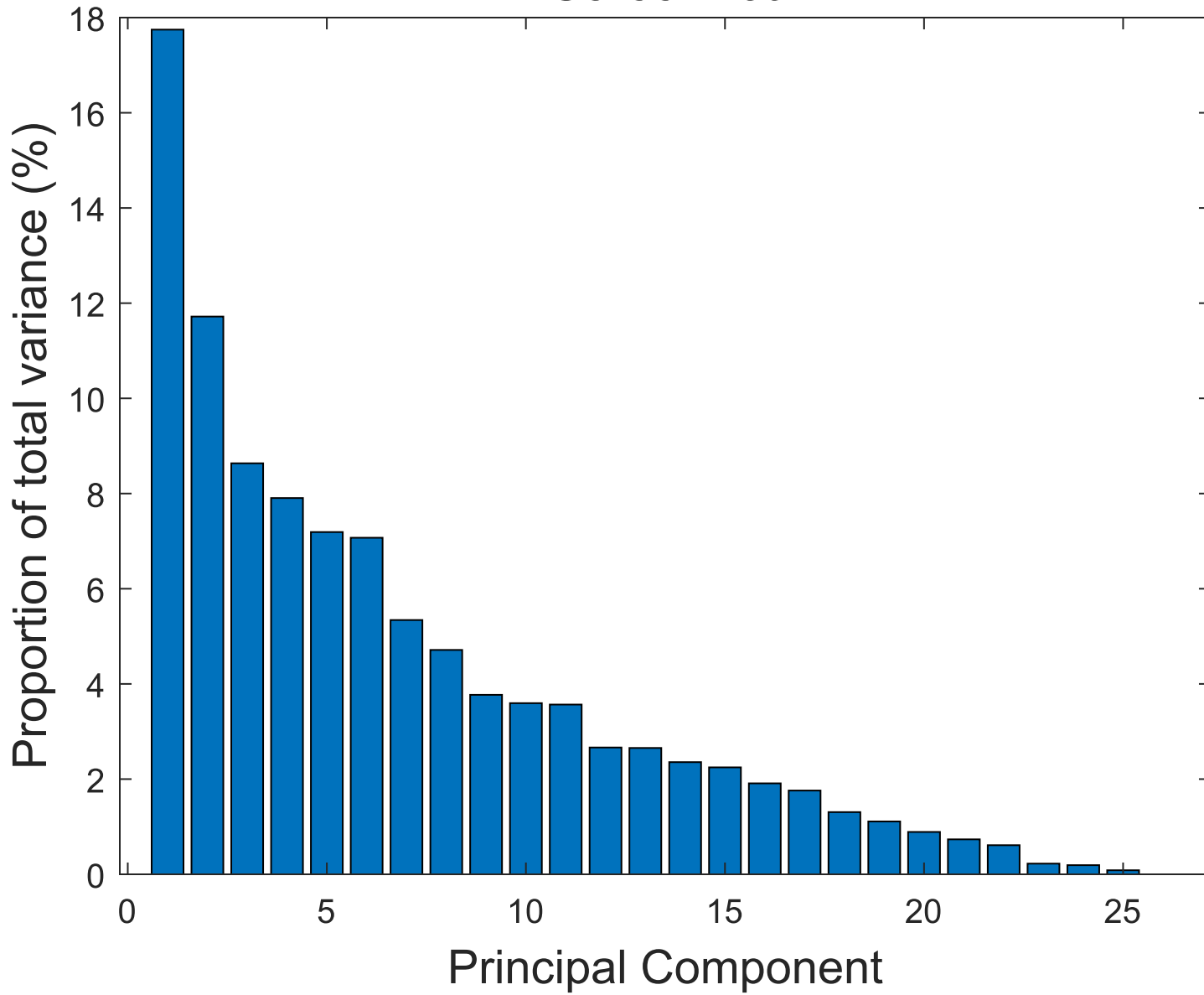
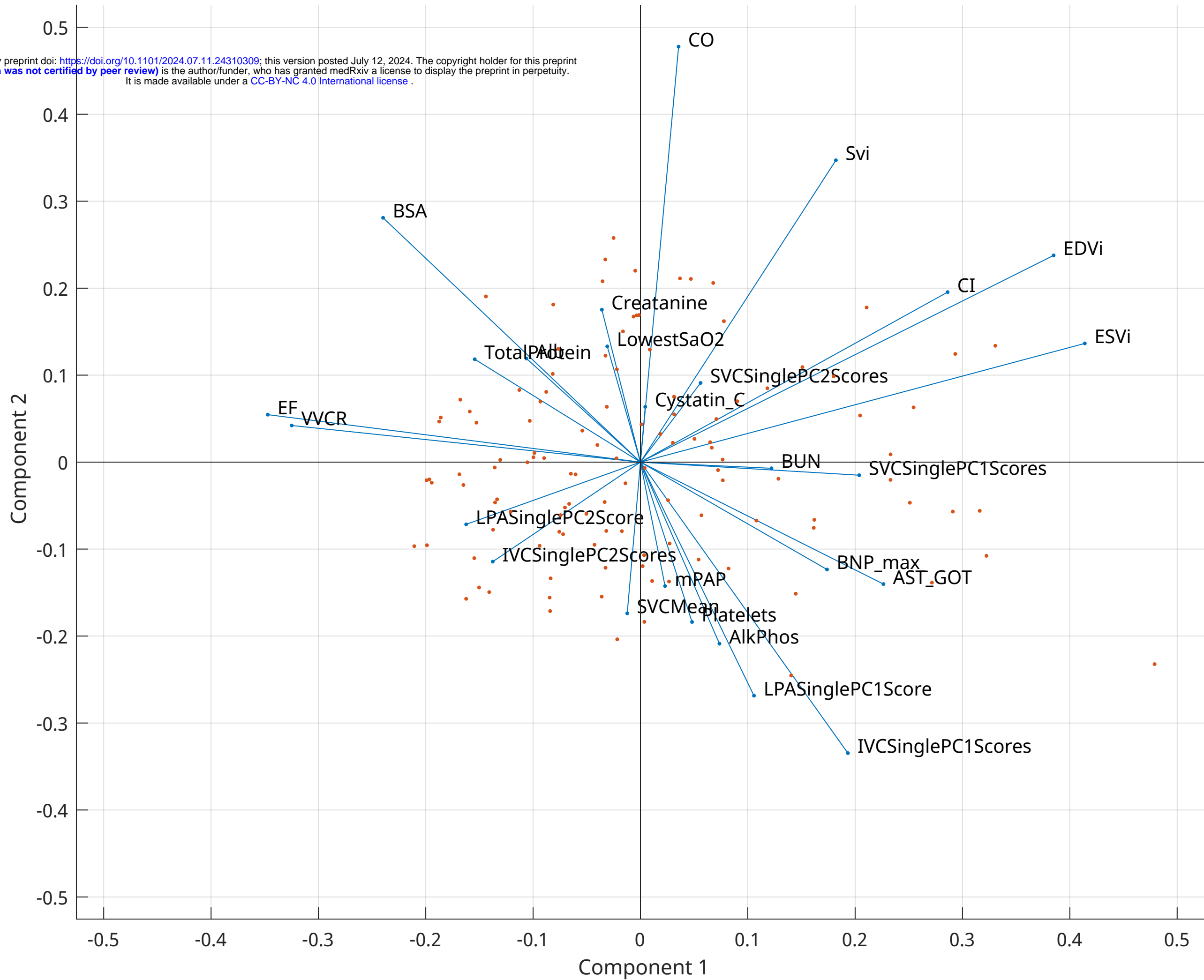
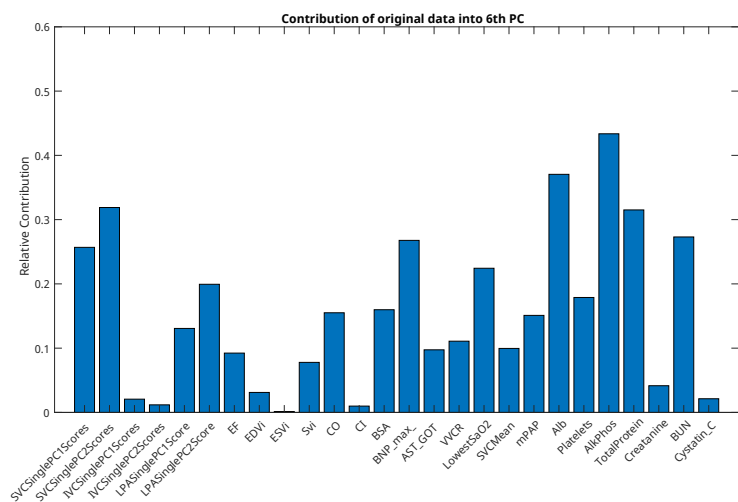
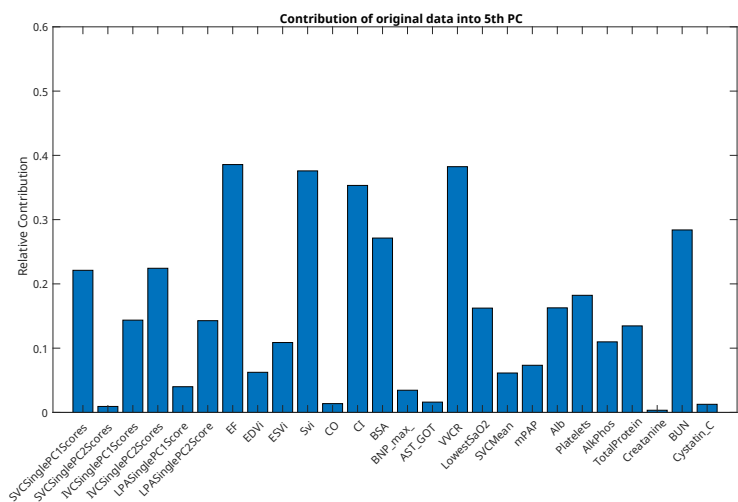
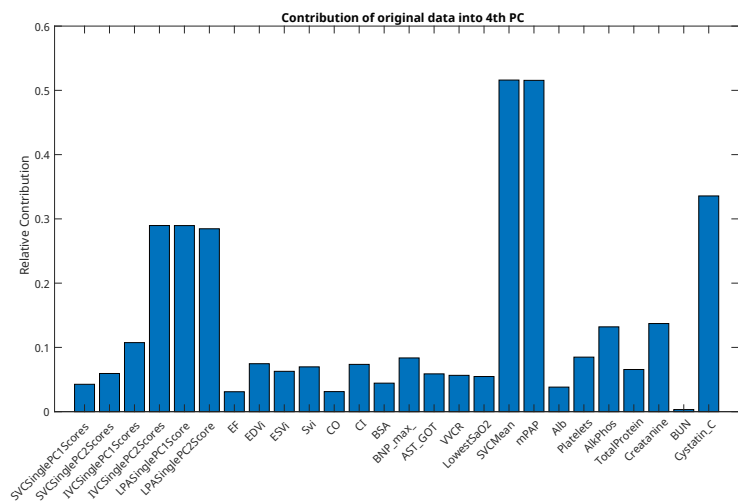
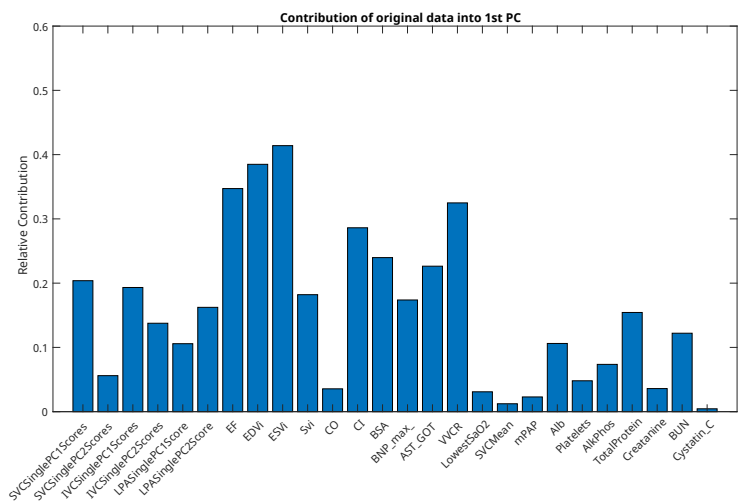


Figure 6. Kaplan-Meier curves for each PC that returned the greatest c-statistics and EF, and the p value returned from the log-rank Mantel Cox p-value. Groups were created using the Youden's index defined cut-off values.

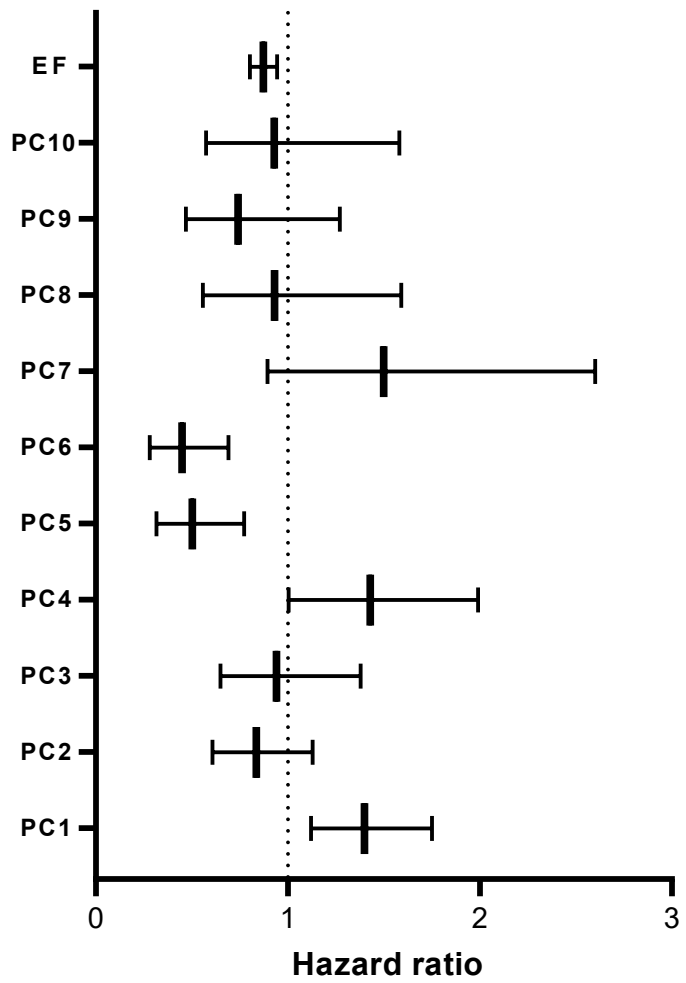
Scree Plot

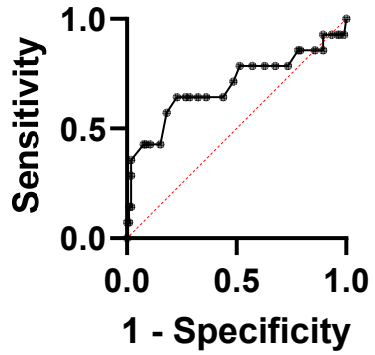
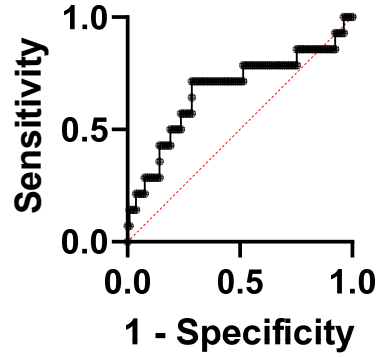
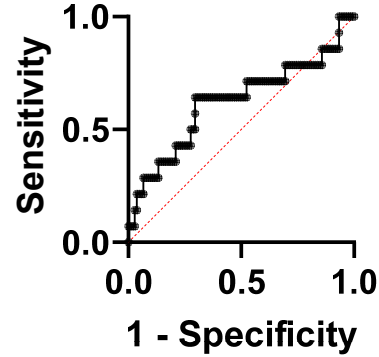
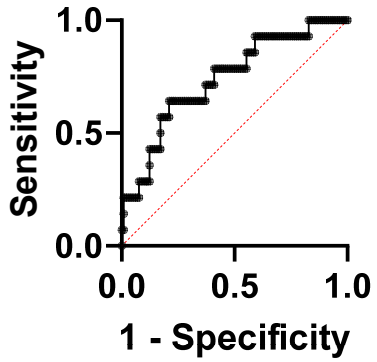
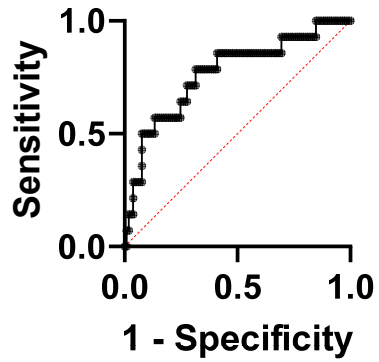






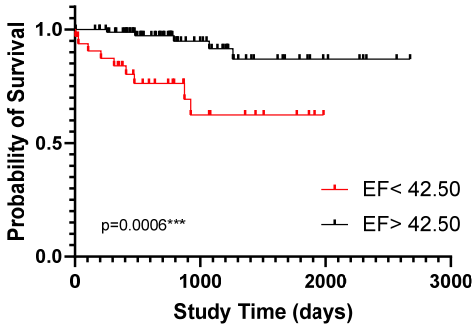
Forest Plot of PCs & EF



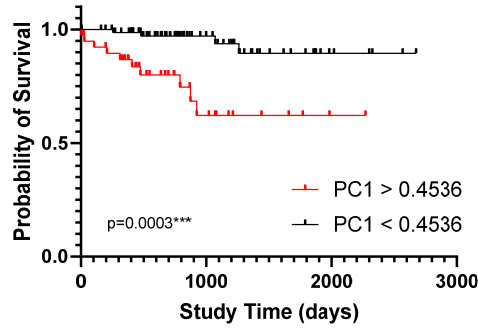
ROC of EF**ROC of PC1****ROC of PC4****ROC of PC5****ROC of PC6**

Measure	AUC	P Value	Sens	Spec	n	Cut Off
EF	0.696	0.0177	0.643	0.771	0.414	< 42.5
PC1	0.674	0.0347	0.714	0.714	0.429	> 0.454
PC4	0.622	0.138	0.643	0.705	0.348	> 0.496
PC5	0.740	0.0037	0.643	0.791	0.433	< -0.888
PC6	0.767	0.0012	0.786	0.686	0.471	< -0.325

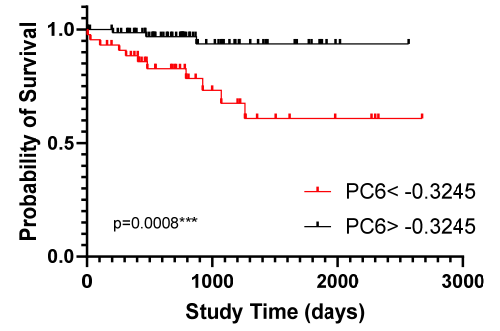
Kaplan-Meier Curve of EF



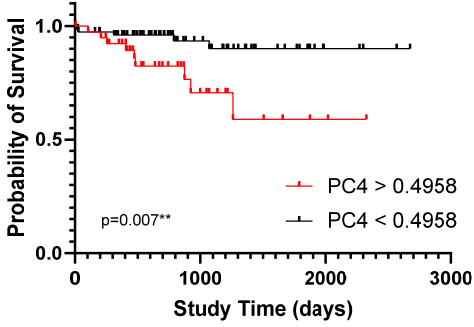
Kaplan-Meier Curve of PC1



Kaplan-Meier Curve of PC6



Kaplan-Meier Curve of PC4



Kaplan-Meier Curve of PC5

