Received: 3 May 2022 | Revised: 28 June 2022 | Accepted: 11 July 2022

DOI: 10.1002/psp4.12848

ARTICLE

Machine learning-guided covariate selection for time-to-event models developed from a small sample of real-world patients receiving bevacizumab treatment

Eleni Karatza[1] | **Apostolos Papachristos**[2] | **Gregory B. Sivolapenko**[2] | **Daniel Gonzalez**[1]

[1]Division of Pharmacotherapy and Experimental Therapeutics, UNC Eshelman School of Pharmacy, The University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA

[2]Laboratory of Pharmacokinetics, Department of Pharmacy, School of Health Sciences, University of Patras, Rion, Patras, Greece

Correspondence
Eleni Karatza, UNC Eshelman School of Pharmacy, The University of North Carolina at Chapel Hill, 301 Pharmacy Lane, CB #7355, Chapel Hill, NC 27599-7569, USA.
Email: ekaratza@pharm.uoa.gr

Abstract

Therapeutic outcomes in patients with metastatic colorectal cancer (mCRC) receiving bevacizumab treatment are highly variable, and a reliable predictive factor is not available. Progression-free survival (PFS) and overall survival (OS) were recorded from an observational, prospective study after 5 years of follow-up, including 46 patients with mCRC receiving bevacizumab treatment. Three vascular endothelial growth factor (VEGF)-A and two intercellular adhesion molecule-1 genes polymorphisms, age, gender, weight, dosing scheme, and co-treatments were collected. Given the relatively small number of events (37 [80%] for the PFS and 26 [57%] for the OS), to study the effect of these covariates on PFS and OS, a covariate analysis was performed using statistical and supervised machine learning techniques, including Cox regression, penalized Cox regression techniques (least absolute shrinkage and selection operator [LASSO], ridge regression, and elastic net), survival trees, and survival forest. The predictive performance of each method was evaluated in bootstrapped samples, using prediction error curves and the area under the curve of the receiver operating characteristic. The LASSO penalized Cox-regression model showed the best overall performance. Nonlinear mixed effects (NLME) models were developed, and a conventional stepwise covariate search was performed. Then, covariates identified as important by the LASSO model were included in the base NLME models developed for PFS and OS, resulting in improved models as compared to those obtained with the stepwise covariate search. It was shown that having gene polymorphisms in *VEGFA* (rs699947 and rs1570360) and *ICAM1* (rs1799969) are associated with a favorable clinical outcome in patients with mCRC receiving bevacizumab treatment.

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?

Machine learning techniques can be helpful to identify predictors of therapeutic outcomes in oncology, provided there is an adequate sample size.

WHAT QUESTION DID THIS STUDY ADDRESS?

What is the best approach to identify significant predictors of overall survival and progression-free survival when the sample size is small, and how can it guide covariate selection for parametric time-to-event models?

WHAT DOES THIS STUDY ADD TO OUR KNOWLEDGE?

For a small sample of real-world patients receiving bevacizumab treatment, the least absolute shrinkage and selection operator method was an adequate way to screen for covariates when the sample is small.

HOW MIGHT THIS CHANGE DRUG DISCOVERY, DEVELOPMENT, AND/OR THERAPEUTICS?

Clinical trials in oncology include a small number of patients due to ethical restrictions. This study provides a tool to use to develop predictive models that allow for assessing covariates in small samples. By applying this methodology, the identification of better responders to bevacizumab treatment was enabled.

INTRODUCTION

Bevacizumab is a fully humanized monoclonal antibody that selectively inhibits the vascular endothelial growth factor (VEGF) binding to its receptors, reducing the microvascular growth of tumor blood vessels.[1] The process of angiogenesis is primarily mediated by the VEGF-A and secondarily by the intercellular adhesion molecule 1 (ICAM-1).[2] Bevacizumab is currently an essential therapeutic agent for treating metastatic colorectal cancer (mCRC). However, large variability in the progression-free survival (PFS) and overall survival (OS) of patients receiving bevacizumab has been reported, attributed to various factors, including genetics, demographic characteristics, disease status, and co-medications.[3–8]

The effect of these factors on PFS and OS has been studied using standard methods to analyze time-to-event (TTE) data, such as Kaplan–Meier estimator, log-rank tests, Cox regression, and nonlinear mixed effects (NLME) modeling.[9–11] Kaplan–Meier estimator and log-rank tests are nonparametric methods that do not require assumptions. They can qualitatively indicate the statistically significant factors affecting the survival probability. However, these methods can handle only categorical covariates, and they cannot quantify the magnitude of the effect or provide a predictive model. Cox regression is a semiparametric technique that does not require any assumptions for the hazard distribution. This technique uses a regression model that quantifies the effects of the covariates on the hazard ratio. As a result, it assumes that the hazards are proportional at any timepoint. It has been shown that a large number of patients are required to get accurate results with this technique as it tends to overfit the data when the ratio of events per covariates explored is low.[9–12] As a rule of thumb, 10 events per covariate have been proposed as an adequate sample size; however, some studies indicate that 10–20 events per covariate are necessary to obtain a Cox regression model with precise estimates of the effects of the key covariates.[13] Last, the development of parametric TTE models using NLME modeling can offer a fully specified hazard function. It can handle any type of covariate. Still, it is sensitive to misspecifications with small sample sizes as a specific distribution of the hazard for the population from which the data was retrieved is assumed.[9,14,15] When evaluating covariate effects using NLME models and the well-established stepwise forward addition and backward elimination, some disadvantages are that adequate data and a reliable model are needed to accurately estimate the random effects and low shrinkage of Empirical Bayes (EB) predictions toward the population values.[16]

Due to some of the disadvantages of the commonly used methods when applied to small sample sizes, supervised machine learning (ML) algorithms have been proposed, such as the survival forest algorithm or penalized Cox regression methods.[12,17,18] Thanks to its bagging feature (bootstrap aggregation), the survival forest algorithm can help identify trends in the data, even in relatively small datasets.[18,19] Then, semiparametric penalized Cox regression methods, such as the least absolute shrinkage and selection operator (LASSO), ridge regression, or elastic net, can be helpful to identify the covariates with the strongest effects, thanks to the penalty that these methods apply during the estimation of the coefficients.[12,13,17,20,21]

In this study, we sought to compare the standard stepwise covariate search and some ML models that can handle small sample sizes for covariate identification when developing TTE models with a limited dataset. We also aimed to identify significant predictors of bevacizumab's therapeutic outcome.

METHODS

Data

The data used were obtained from a previously published study.[22] Briefly, the study was an observational, prospective study with a 5-year follow-up that included 46 patients with mCRC. The study's primary end points were PFS and OS. The patients were treated as per standard clinical practice at the Department of Oncology, University Hospital of Patras, Greece. The study was conducted according to the Declaration of Helsinki and the International Conference on Harmonization (ICH) Good Clinical Practice Guideline. Approval was obtained from the Hospital's Ethics Committee. Prior to study enrollment, all patients provided signed informed consent.

Patients were receiving either 5 mg/kg bevacizumab (Avastin; Genentech/Roche) every 2 weeks co-administered with 5-fluorouracil/leucovorin and irinotecan (CPT-11) or oxaliplatin (OHP) or 7.5 mg/kg bevacizumab every 3 weeks co-administered with capecitabine and irinotecan (CPT-11) or OHP. Genotyping data for the following three VEGF single nucleotide polymorphisms (SNPs) were available: V2578 (*VEGFA*, rs699947), V1154 (*VEGFA*, rs1570360), and V634 (*VEGFA*, rs2010963) and for the following two ICAM-1 SNPs: ICAM469 (*ICAM1*, rs5498) and ICAM241 (*ICAM1*, rs1799969). The deviation from the Hardy–Weinberg equilibrium (HWE) was tested using the R package "HardyWeinberg"[23] (Table S5).

Other potential covariates collected were demographic data, including age, sex, and weight, and the patient's treatment (Table S1). Independence between two dichotomous covariates was tested using a chi-squared test and a Fisher's exact test. A lack of correlation between two continuous covariates was tested using a Pearson correlation test. A lack of correlation between a dichotomous and a continuous covariate was tested using a Mann Whitney *U* test in R.

The effect of 10 covariates (weight, age, sex, co-treatment, dosing scheme, 3 polymorphisms in *VEGFA*, and 2 polymorphisms in *ICAM1*) was explored in 46 patients. For the PFS, 37 events were recorded (events per covariate ratio: 3.7), and for the OS, a total of 26 events were recorded (events per covariate ratio: 2.6). The median survival time for PFS was 285 days, and for OS, it was 1016 days (Figure S1). In view of the small sample size, each SNP was evaluated separately as a dichotomous categorical variable, with one category being homozygous wild type (W) and the other carrying the variant allele (V) either as heterozygous or as homozygous. Thus, a dominant model was assumed in all cases (WV + VV vs. WW) for grouping, in line with previous studies.[24–27]

Machine learning models

ML algorithms (survival trees, survival forest, Cox-regression, LASSO, ridge-regression, and elastic net) were applied to explore the impact of the available covariates on PFS and OS. The ML models obtained were evaluated using bootstrapping. All the development and evaluation of the ML models were performed in R version 4.1.0. The development of the ML models is described in the Supplementary Material. The R code used for the final models and dummy data are also provided.

Machine learning models comparison

Given the small sample size, it was not possible to split the dataset into "training" and "test" datasets. Therefore, resampling and cross-validation techniques were applied to compare the predictive performance of the models.[28,29]

In this study, the models were initially evaluated for the ability to describe the observed data and based on statistical goodness of fit criteria, such as the Harrell's concordance index (C-index). The cross-validation prediction error using the R package "pec"[30] was estimated for 1000 bootstrapped samples, and prediction error curves were obtained. Then, using the package "risksetROC"[31] the area under the receiver operator characteristic curve (AUC of the ROC) to time was evaluated in a bootstrapped sample that included 1000 patients.

Parametric TTE modeling and stepwise covariate selection

A parametric TTE model for the PFS was developed using Monolix2019R2 (Lixoft). The exponential, Gompertz, log-logistic, uniform, Weibull, and gamma distribution were explored to describe the hazard function. The algorithm used to estimate the parameters was the stochastic approximation estimation method, whereas the objective

function value (OFV) was computed using the Importance Sampling Monte Carlo method with the final population parameter estimates. Parameters of the TTE model were assumed to follow a lognormal distribution, and an exponential model was used to describe interindividual variability, with the random effect (η) of each parameter following a normal distribution with a mean of zero and an estimated variance of ω.[1]

Initially, a one-way analysis of variance (ANOVA) for categorical covariates or a Pearson test for continuous covariates was performed with the random effects (η) to assess if a covariate should be included. Then, the impact of covariates on model parameters was evaluated using a stepwise forward addition and backward elimination. Continuous covariates (age and weight) were tested normalized to the sample median and using a power model. In contrast, the rest of the available categorical covariates were tested using a linear model. A likelihood ratio test (LRT) to test if the addition of the covariates results in a statistically significant change in the OFV and a Wald test to test if the coefficient is significantly different from zero were performed. A significance level of 5% was considered for both forward and backward steps.

Correlations between model parameters were also investigated. The models were evaluated based on the precision of the estimates and goodness-of-fit statistical criteria (−2 log-likelihood, Akaike information criterion, and Bayesian information criteria). The predictive performance and robustness of the models was assessed by visual predictive checks (VPCs). The VPCs were generated using Monte Carlo simulations of 1000 datasets and 90% prediction intervals.

Machine learning guided covariate selection

The LASSO was the ML model with the best predictive performance (higher C-index, lower prediction error, and higher AUC of the ROC). The covariates identified as significant for the outcome (coefficient ≠ 0) were used in the parametric models developed. Then, the covariate model obtained using the stepwise approach and the model obtained using the ML algorithm were compared in terms of stability (condition number), precision of the estimates (relative standard error), and interindividual variability terms.

ETHICS STATEMENT

The data have been previously published.[22] No new subject enrollment took place as part of this study. Approval was obtained by the University Hospital of Patras Ethics Committee (3221/7, 12.02.2013).

RESULTS

Machine learning models

A total of six different ML algorithms were explored in this study to evaluate the impact of the 10 available covariates on the PFS and OS. A summary of the models obtained with each technique is provided in Table 1. The tuning hyperparameters used for the final survival trees and survival forest models are presented in Tables S2 and S3.

Based on the survival tree analysis, none of the available covariates led to a statistically significant split for the PFS. Only ICAM241 affected the OS significantly (Figure S2). As no model was obtained for the PFS, this method was not included in the following steps of the analysis for comparative purposes.

For the PFS, all the tested algorithms showed that V2578 significantly affects the outcome. In terms of the C-index, the LASSO resulted in the best predictive performance, including the lowest number of parameters (Table 1). All the ML models developed described the data similarly well except for the Cox regression model, which was outperformed by the others (Figure 1a). Likewise, all the models gave a similar performance based on the prediction error curves, with the LASSO showing the lowest prediction error (Figure 2a). For the AUC of the ROC, the LASSO model had a slightly better predictive performance as compared to the other models (Figure 3a).

Similar results were also observed for the OS. All algorithms tested showed that V1154 and ICAM241 significantly affect the outcome. The LASSO resulted in the best performance based on the C-index, including the lowest number of parameters (i.e., the lowest predictive performance; Table 1). All the ML models developed described the data similarly except for the Cox regression model, which was outperformed by the others (Figure 1b). The models gave a similar performance based on the prediction error curves, with the LASSO and the elastic net showing the lowest prediction error (Figure 2b). After visual examination of the changes in the AUC of the ROC over time, the predictive performance of the LASSO and elastic net (overlapping curves) improved over time, outperforming the other models by the end of the observation period (Figure 3b). The similar predictive performance of the LASSO and the elastic net is probably because these two models had the same significant terms with very similar coefficients (Table 1).

Taken altogether, the models developed using the LASSO algorithm presented the best predictive performance. According to the LASSO model, individuals with a variant

TABLE 1 Significant covariates obtained with each machine learning method tested

	Progression-free survival						Overall survival					
	Tree	Random forest	Cox	LASSO	Ridge	Elastic net	Tree	Random forest	Cox	LASSO	Ridge	Elastic net
	Split	Imp.	Coeff.	Coeff	Coeff	Coeff	Split	Imp	Coeff	Coeff	Coeff	Coeff
Age	–	–	–	–	−0.009	–	–	–	–	–	0.000	–
Weight	–	–	–	–	−0.011	–	–	–	–	–	0.000	–
Sex	–	–	–	–	−0.004	–	–	–	–	–	0.000	–
V2578	–	0.0148	−0.991	−0.139	−0.047	−0.080	–	0.006	–	–	−0.001	–
V1154	–	–	–	–	−0.021	–	–	0.036	−0.845	−0.065	−0.002	−0.053
V634	–	–	−0.731	–	−0.010	–	–	–	–	–	0.000	–
ICAM469	–	–	–	–	−0.009	–	–	0.008	–	–	−0.002	–
ICAM241	–	0.0034	−0.711	–	−0.054	−0.035	Yes	0.028	−1.279	−0.159	−0.003	−0.120
Cotreatment	–	–	–	–	0.007	–	–	–	–	–	−0.001	–
Dosing scheme	–	0.0111	0.672	–	0.042	–	–	0.001	–	–	−0.001	–
C-index		0.566	0.629	0.632	0.579	0.611		0.473	0.606	0.699	0.665	0.665

Note: A endash (–) was used for covariates that did not produce a statistically significant split (tree), had negative relative importance (random forest), were excluded from the model as nonsignificant (Cox regression) or were excluded by the algorithm based on the lambda value used (LASSO, ridge, and elastic net). Split: whether in the survival tree the covariate resulted in a statistically significant split; Imp.: the relative importance of the covariate estimated by the random forest algorithm; Coeff.: The value of the coefficient obtained for each covariate; C-index: the concordance or Harrell's C-index, which is a measure of performance for survival models and ranges from 0 to 1, and where 1 indicates full accordance between observations and predictions. Dosing scheme: 5 mg/kg bevacizumab every 2 weeks + 5-fluorouracil/leucovorin or 7.5 mg/kg bevacizumab every 3 weeks + capecitabine, Cotreatment: irinotecan (CPT-11) or oxaliplatin (OHP). Genetic polymorphisms were evaluated as presence or absence of the variant allele for the single nucleotide polymorphisms (SNPs) V2578 (*VEGFA*, rs699947), V1154 (*VEGFA*, rs1570360), V634 (*VEGFA*, rs2010963), ICAM469 (*ICAM1*, rs5498), and ICAM241 (*ICAM1*, rs1799969).

Abbreviation: LASSO, least absolute shrinkage and selection operator.

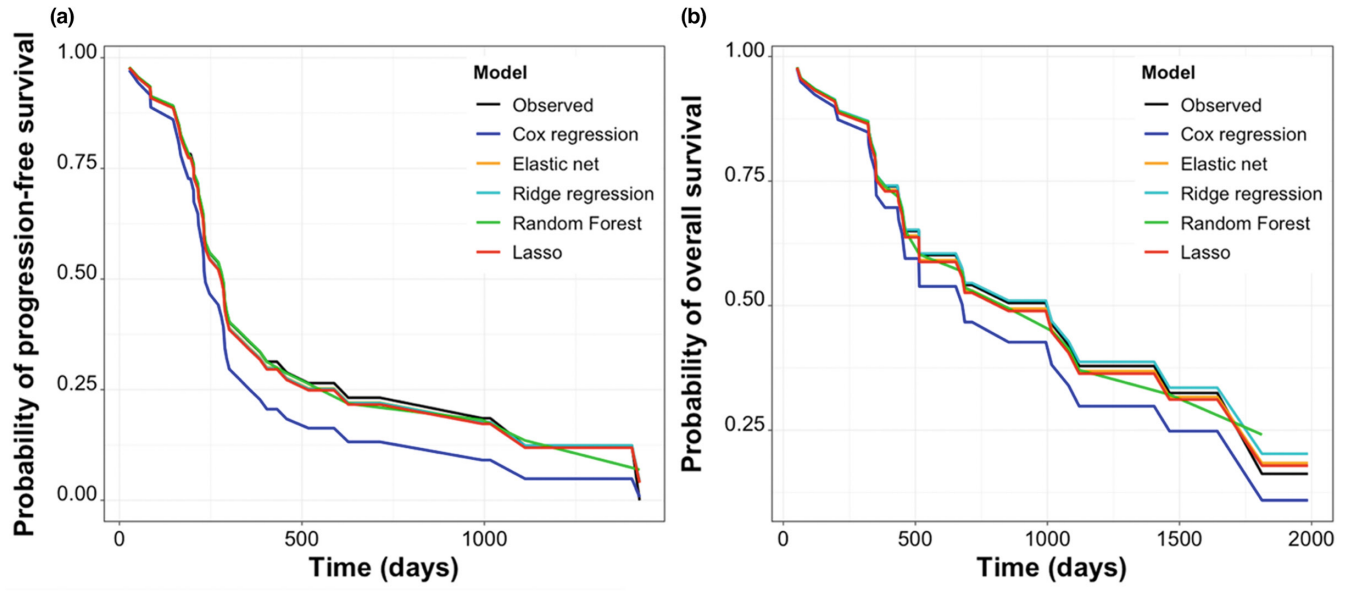

FIGURE 1 A comparison of the ability of the machine learning approaches to describe the observed data. After development of the models using the observed data (black line), the survival probability as a function of time obtained with the Cox regression model (blue line), elastic net model (yellow line), ridge regression model (turquoise line), least absolute shrinkage and selection operator (LASSO) model (red line), and survival forest (green line) were overlaid for (a) progression-free survival and (b) overall survival.

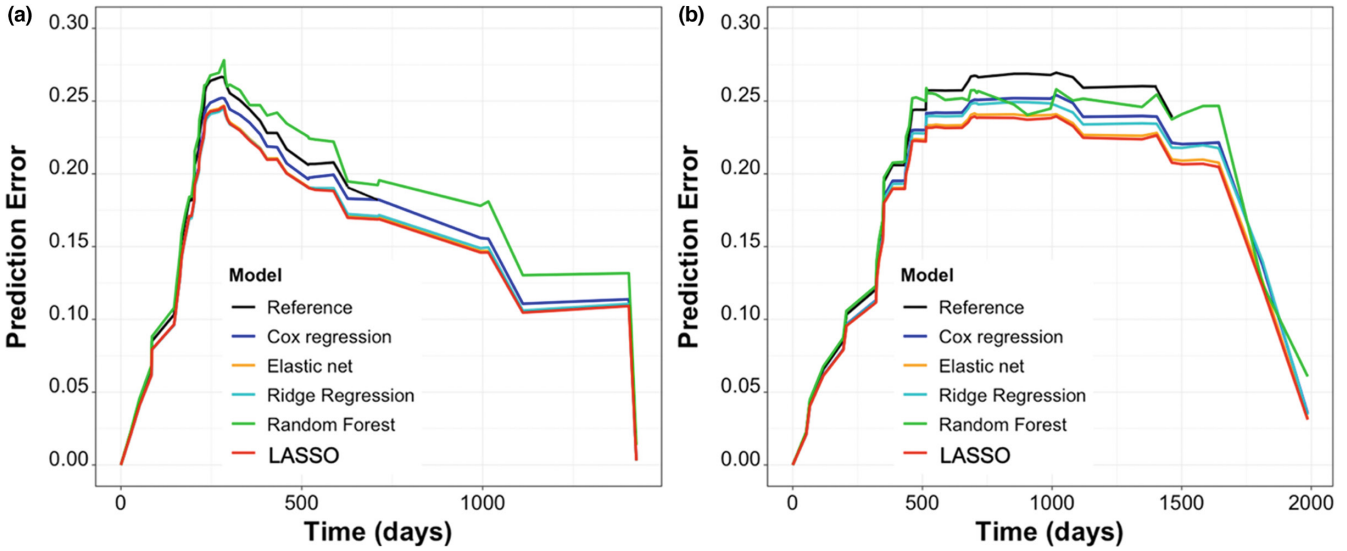

FIGURE 2 The prediction error curves for all the approaches applied and based on the Brier score followed over time. The reference model was a full Cox regression model with all the covariates (black line). In addition, the Cox regression model (blue line), elastic net model (yellow line), ridge regression model (turquoise line), least absolute shrinkage and selection operator (LASSO) model (red line) and survival forest (green line) were overlaid for (a) progression-free survival and (b) overall survival. A total of 1000 bootstrapped samples were used for cross validation.

allele V2578 have about a 13% reduced risk of disease progression than those with only wild type alleles. The most significant predictors for OS were V1154 and ICAM241. According to the LASSO model, individuals with a variant ICAM241 allele have about a 15% reduced risk of death. Those with only the wild type allele and individuals with a variant V1154 allele have a 6% reduced risk of death compared to those with only the wild type allele.

Parametric TTE models and stepwise covariate search

The hazard for PFS was best described using a Weibull distribution, whereas for OS, it was best described using a uniform distribution (Table S4 and Figure S3). After the base model development, a stepwise covariate search was performed on the scale parameters of the two

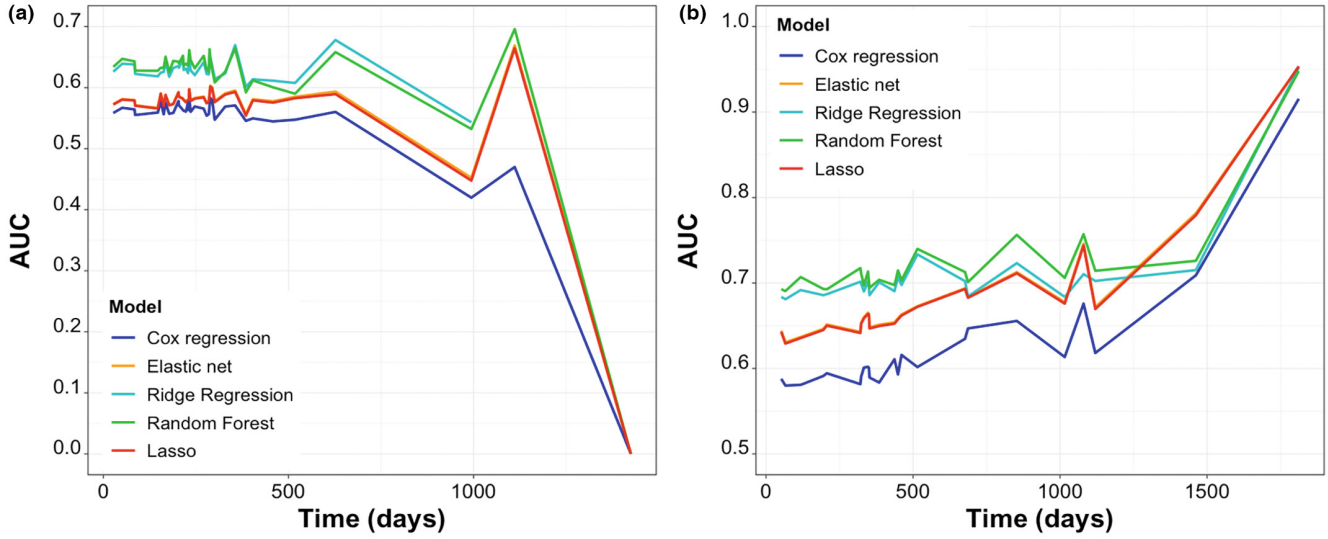

FIGURE 3 The area under the curve (AUC) to time of time-varying receiver operator curves (ROC) were generated using the observed data as a training dataset. As a test dataset, a bootstrapped sample containing 1000 volunteers was used to evaluate the performance of the machine learning approaches. The curves obtained with the Cox regression model (blue line), elastic net model (yellow line), ridge regression model (turquoise line), least absolute shrinkage and selection operator (LASSO) model (red line), and the survival forest model (green line) were overlaid for the (a) progression-free survival (PFS) and (b) overall survival (OS). The LASSO model and elastic net model (red and yellow lines) overlapped.

distributions. The results of the covariate search are presented in Table 2. It should be noted that the addition of any covariate did not promote any notable change in the VPC.

The covariates that seemed to affect the scale parameter of the hazard distribution significantly differed depending on the test performed. However, the final models were selected according to the results of the LRT, and given that the LRT is the most common way to select covariates in TTE parametric models. Thus, in the final models, weight and ICAM241 were included in the PFS model, and V1154, V2578, and ICAM241 were included in the OS model.

Covariate model based on LRT versus LASSO guided

The final covariate model obtained with the stepwise covariate search was compared to a model that included only the covariates shown to significantly affect the PFS or OS using the LASSO model. A comparison of the models obtained in each case is provided in Table 3.

In this analysis, where the number of events was relatively small compared to the covariates assessed, the covariate models suggested by the LRT were overparametrized. On the contrary, using the covariates suggested by the LASSO led to more stable models with a lower condition number and with improved precision of the estimates. Most importantly, the interindividual variability term of the scale parameter was significantly lower in the LASSO-guided covariate models, indicating that these covariates explained better the sources of interindividual variability. The final parametric TTE model suggests that patients with mCRC receiving bevacizumab treatment that have a variant V2578 allele have reduced the hazard of disease progression. In contrast, patients with the variant ICAM241 and/or V1154 alleles have reduced hazard of death.

DISCUSSION

In the past few years, ML techniques have been increasingly applied along with pharmacometrics in oncology to identify underlying trends and relationships in the setting of big data (omics, imaging, or electronic health records)[32] and, most importantly, to identify prognostic and predictive factors associated with long-term OS and tumor growth dynamics in large clinical trials.[33–35] In the present analysis, the ability of ML techniques to identify predictive factors associated with improved OS and PFS was explored in a dataset including 46 patients. The effect of 10 covariates/predictors was investigated using six ML survival models developed, and their predictive performance was compared. The ML models were compared in terms of prediction error, AUC of the ROC, and C-index. The values of C-index obtained from the models were rather

TABLE 2 Results obtained with stepwise covariate selection

Progression-free survival							Overall survival						
	OFV	**ΔOFV**	**ΔωTe**	**LRT**	**Wald**	**Pearson or ANOVA**		**OFV**	**ΔOFV**	**ΔωTe**	**LRT**	**Wald**	**Pearson or ANOVA**
Base model	524.58	–	–	–	–	–	Base model	424.49	–	–	–	–	–
Age	524.11	0.47	−0.004	0.493	0.734	0.772	Age	424.28	0.21	0.010	0.647	0.447	0.525
Weight	520.03	4.55	−0.024	0.033	0.154	0.180	Weight	424.37	0.12	0.009	0.729	0.698	0.478
Sex	524.44	0.14	0.007	0.708	0.868	0.734	Sex	424.37	0.12	0.014	0.729	0.554	0.923
V2578	521.67	2.91	0.233	0.088	0.035	0.068	V2578	420.48	4.01	0.241	0.045	0.063	0.070
V1154	523.49	1.09	−0.005	0.296	0.386	0.354	V1154	418.68	5.81	0.101	0.016	0.060	0.017
V634	524.69	−0.11	0.014	1.000	0.698	0.669	V634	424.19	0.30	0.007	0.584	0.744	0.600
ICAM469	523.59	0.99	−0.057	0.320	0.808	0.734	ICAM469	420.80	3.69	0.135	0.055	0.123	0.217
ICAM241	520.27	4.31	0.070	0.038	0.069	0.145	ICAM241	417.96	6.53	0.174	0.011	0.042	0.029
Cotreatment	523.99	0.59	−0.033	0.442	0.671	0.120	Cotreatment	423.64	0.85	0.028	0.357	0.273	0.430
Dosing scheme	523.24	1.34	0.062	0.247	0.192	0.831	Dosing scheme	422.46	2.03	0.157	0.154	0.222	0.118
Weight + ICAM241	516.41	5.18	−0.004	0.027	–	–	V2578 + V1154	417.14	7.35	0.220	0.025	–	–
							V2578 + ICAM241	415.49	9.00	0.341	0.011	–	–
							V1154 + ICAM241	414.65	9.84	0.160	0.007	–	–
							V2578 + V1154 + ICAM241	414.50	9.99	0.341	0.019	–	–

Note: ΔOFV, OFV of the base model minus OFV of the model with the covariate; ΔωTe, interindividual variability term for the scale parameter Te in the base model minus the interindividual variability term for the scale parameter Te after including the covariate in the model; LRT, *p* value of a likelihood ratio test; Wald, *p* value of a Wald test; Pearson or ANOVA: *p* value of a Pearson test for continuous or a one-way ANOVA test for categorical covariates. Dosing scheme: 5 mg/kg bevacizumab every 2 weeks + 5-fluorouracil/leucovorin or 7.5 mg/kg bevacizumab every 3 weeks + capecitabine, Cotreatment: irinotecan (CPT-11) or oxaliplatin (OHP). Genetic polymorphisms were evaluated as presence or absence of the variant allele for the SNPs V2578 (*VEGFA*, rs699947), V1154 (*VEGFA*, rs1570360), V634 (*VEGFA*, rs2010963), ICAM469 (*ICAM1*, rs5498) and ICAM241 (*ICAM1*, rs1799969). The endash (–) means not available as either no covariate or multiple covariates were included in the respective model.

Abbreviations: ANOVA, analysis of variance; ICAM, intercellular adhesion molecule; OFV, objective function value.

TABLE 3 Comparison of the covariate models obtained with likelihood ratio test and machine learning-guided covariate selection

Progression-free survival									
Covariate model based on LRT					**Covariate model LASSO guided**				
OFV	516.41		C-index	0.972	OFV	521.67		C-index	0.979
Condition number	1200				Condition number	6.5			
		RSE (%)	IIV (CV%)	RSE (%)			RSE (%)	IIV (CV%)	RSE (%)
p	13.9	66.4	246.97	17.8	*p*	2.33	29.3	46.10	39.1
Te (days)	139	67.2	87.20	29.6	Te (days)	321	18.1	52.70	23.6
ICAM241 on Te	0.492	77.4			V2578 on Te	0.561	44.9		
Weight on Te	0.0115	66.4							
Overall survival									
Covariate model based on the LRT					**Covariate model based on the LASSO**				
OFV	414.5		C-index	0.727	OFV	414.65		C-index	0.754
Condition number	9.8				Condition number	3.6			
		RSE (%)	IIV (CV%)	RSE (%)			RSE (%)	IIV (CV%)	RSE (%)
Te (days)	1163	22.9	47.73	61.8	Te (days)	1320	20.2	41.43	34.6
ICAM241 on Te	0.778	95.5			ICAM241 on Te	1.11	42.3		
V1154 on Te	0.71	62.4			V1154 on Te	0.909	46.2		
V2578 on Te	0.476	63.1							

Note: CV%, coefficient of variation expressed as a percentage; IIV, interindividual variability term; LASSO, least absolute shrinkage and selection operator; LRT, likelihood ratio test; OFV, objective function value; *p*, shape parameter of a Weibull distribution; RSE, relative standard error expressed as a percentage; Te, scale parameter or characteristic time of a Weibull distribution for the progression-free survival models and scale parameter (time at which the survival probability is 0) or characteristic time of a Uniform distribution for the overall survival models.

low with the highest being 0.699. This magnitude of C-index is rather common in the highly variable space of oncology, as shown in a study including 136,719 patients.[36]

The results showed no marked differences among the algorithms tested except for the survival tree analysis. All the other models achieved adequate performance as determined by their prediction error (Brier scores). This finding is in line with a study that showed that penalized regression methods have a similar performance in small sample sizes.[13] In this prior study, a prospective cohort of patients with coronary artery disease was used to perform simulations and study the predictive accuracy, calibration, and discrimination of various Cox regression models by varying the events per variable ratio (i.e., the observations vs. covariates to evaluate). However, this study did not explore survival trees and survival forests, whereas the analyses were performed in simulated data. In our analysis, the survival tree identified only ICAM241 as a predictor of OS, whereas none of the available covariates had an effect strong enough to be identified as a predictor of PFS, indicating that probably a larger sample would be needed for this method.[37]

Despite the above, the LASSO method, which applies a penalty to shrink toward zero the coefficients of the covariates with minor contribution, showed the best predictive performance compared to the other ML models tested. This feature can be helpful when it is known that many of the covariates explored do not significantly impact the hazard ratio, helping to reduce the complexity of the model. This finding is in line with previous studies comparing the performance of ML algorithms in various settings.[12,17,19,20] Other penalized regression models were tested, namely the ridge regression and the elastic net. The ridge-regression penalty forces the covariates with a minor contribution to have their coefficients close to zero but never precisely zero, which can be helpful in cases where many of the covariates tested exert a significant effect on the outcome. The elastic net combines both types of penalty, a particularly useful feature in cases where some covariates explored are known to be correlated.[13,17,20,21] In this study, using the survival tree, it was not possible to obtain a model for PFS, and therefore for comparative purposes, it was not included in the subsequent comparisons. Despite being a very powerful and popular tool, the survival forest algorithm showed a modest performance, which might be due to the small sample size of the dataset or the shape of our survival curves.[18,38]

Population parametric TTE models were developed for the two end points explored, and a stepwise covariate search was performed. Three different tests were

performed to assess the influence of SNPs, demographics, and treatment on the scale factors of the hazard models developed (i.e., LRT, Wald test, and an ANOVA or a Pearson test between survival effects and covariates). A similar approach has been previously suggested by Bertrand et al.[39] for detecting the influence of genetic polymorphisms on pharmacokinetic parameters, with the only difference that, in that study, the ANOVA was performed on the EB estimates of the parameters. It was interesting to note that depending on the statistical test applied, different covariates were identified as significant, which suggests a need for more data to inform better the distribution of the parameters.[16] As the LRT is considered the most reliable test for including covariates in population TTE models,[16,40] the final covariate model was based on its results, which resulted in overparametrized models (Table 3). This finding was in line with the survival analysis literature, suggesting that in the setting of small sample sizes, parametric models can be sensitive to misspecifications as it is challenging to have an adequate characterization of the hazard function.[9,41]

Using the covariates suggested as significant by the LASSO, the model improved significantly (Table 3). This indicates that despite combining the inference from two differently conditioned models, it was more effective to identify the factors significantly affecting the interindividual variability on the outcome in the setting of small sample sizes. This is due primarily to the fact that the penalty shrinks nonimportant covariates to zero, helping to avoid overparameterization. In addition, it is a semiparametric technique, like Cox regression, which means that no distribution for the hazard is assumed, making it more robust to misspecifications due to the limited data.[9,12,17,20]

In terms of C-index (Tables 1 and 3), the NLME models were superior to the ML models, indicating that despite the limited data this technique can still offer a model with more accurate predictions. Comparing the results obtained with the LASSO with the tests performed during the stepwise covariate selection (Table 2), it was noted that the interindividual variability term after inclusion of the covariate ($\Delta\omega$Te) and the Wald test were in line with the LASSO for PFS. In contrast, for the OS, the ANOVA test and the Wald test were in line with the LASSO. This indicates that for small sample sizes it would be helpful to perform additional tests in addition to LRT.

This study is not without limitations. We could not use training and test datasets to further evaluate the ML algorithms because of the small dataset. However, it should be noted that it has been suggested, especially for smaller datasets, that splitting the original data into training and test datasets is not the most appropriate approach for validation as it reduces sample sizes of both training and testing datasets, and the results can be different by different splitting processes.[42] In addition, due to the limited data available, we could not perform an external evaluation of population TTE models developed with the stepwise approach and with the LASSO guidance. Moreover, many more ML algorithms could have been tested.[43,44] However, the present analysis included the methods known for their ability to work well in small sample sizes. In addition, many other covariates have been described in the literature to affect the outcome of patients with mCRC receiving bevacizumab treatment that was not evaluated in this study and could have a larger effect, such as other VEGF-A SNPs, primary tumor site, tumor histology, number of metastasis sites, albumin, disease control rate, and the baseline value of some angiogenic biomarkers.[3,5,8,45,46]

The LASSO penalty has been applied previously, to maximum likelihood and proposed as a solution for predictive covariate model building in NLME models in cases of small sample sizes. This covariate search method has been implemented in PsN to help covariate search using NONMEM.[47] If NONMEM had been used for this analysis, it would have been interesting to see if the covariates identified as significant using the penalized maximum likelihood are comparable to those obtained using the penalized Cox regression model or to those obtained with the stepwise covariate method.

The data used in this analysis have been previously analyzed by Papachirstos et al.[22] The effect of the SNPs on PFS and OS was studied using Kaplan–Meier estimate and log-rank test, whereas the effect of sex and age were studied using Cox regression. Each SNP was considered a categorical variable in this prior study with three categories, one for each possible genotype. Each SNP was treated as a dichotomous variable (variant or wild type) in the present study. V2578 was associated with increased PFS and ICAM241 with prolonged OS in both analyses. Bevacizumab and VEGF concentrations were also recorded in the same population and analyzed using NLME modeling. The study showed that patients with a variant ICAM241 allele have lower bevacizumab clearance, and patients with variant V2578 have lower free VEGF baseline levels,[2] possibly providing an explanation for the improved clinical outcome in these patients.

Even though deviations from the HWE were identified in this study (Table S5), indicating that one or more of the Hardy–Weinberg conditions are being violated or that there is a genotyping error, other studies have also shown the importance of these SNPs in patients under bevacizumab treatment. Interestingly, improved clinical outcomes in patients with a variant V2578 allele have also been reported in patients with glioblastoma,[4] nonsquamous non-small cell lung cancer,[48] and metastatic breast cancer.[49] Similarly, patients with metastatic breast cancer with a variant V1154 allele were shown to have

significantly longer OS,[49] whereas in another study including patients with mCRC, a variant V1154 allele was associated with reduced OS.[50]

In conclusion, the present study showed that the LASSO Cox regression model performed the best when analyzing data collected from a small sample of real-world patients receiving bevacizumab treatment. The LASSO method provided reliable results while avoiding model overparameterization. Interestingly, in the setting of small sample sizes, it might help guide covariate selection of parametric TTE models due to its ability to discern the significant effect without making assumptions on the underlying distribution. Finally, variant SNPs in V1154, V2578, and ICAM241 were associated with favorable clinical outcomes in patients with mCRC under bevacizumab treatment. As this methodology was applied only in the present study, further research is needed to evaluate the usefulness of ML models to guide covariate selection of population models in the setting of small sample sizes.

AUTHOR CONTRIBUTIONS

E.K. and D.G. wrote the manuscript. E.K. designed the research. E.K., A.P., and G.S. performed the research. E.K. analyzed the data.

ACKNOWLEDGMENTS

The authors would like to thank Lixoft for providing an academic license of the Monolix software.

FUNDING INFORMATION

No funding was received for this work.

CONFLICT OF INTEREST

The authors declared no competing interests for this work.

DATA AVAILABILITY STATEMENT

Data used for the analyses described herein included sensitive information that could not be disclosed.

CONSENT TO PARTICIPATE

The data have been previously published.[22] No new subject enrollment took place as part of this study. Prior to study enrolment, all patients provided signed informed consent.

ORCID

Eleni Karatza https://orcid.org/0000-0001-8406-4121
Apostolos Papachristos https://orcid.org/0000-0003-0433-0931
Gregory B. Sivolapenko https://orcid.org/0000-0003-2388-362X
Daniel Gonzalez https://orcid.org/0000-0001-5522-5686

REFERENCES

1. U.S. Food and Drug Administration. *AVASTIN® prescribing information*. Drugs@FDA. https://www.accessdata.fda.gov/drugsatfda_docs/label/2011/125085s225lbl.pdf
2. Papachristos A, Karatza E, Kalofonos H, Sivolapenko G. Pharmacogenetics in model-based optimization of bevacizumab therapy for metastatic colorectal cancer. *Int J Mol Sci*. 2020;21(11):3753. doi:10.3390/ijms21113753
3. Loupakis F, Cremolini C, Yang D, et al. Prospective validation of candidate SNPs of VEGF/VEGFR pathway in metastatic colorectal cancer patients treated with first-line FOLFIRI plus bevacizumab. *PLoS ONE*. 2013;8(7):e66774. doi:10.1371/journal.pone.0066774
4. Galanis E, Anderson SK, Lafky JM, et al. Phase II study of bevacizumab in combination with sorafenib in recurrent glioblastoma (N0776): a North Central Cancer Treatment Group Trial. *Clin Cancer Res*. 2013;19(17):4816-4823. doi:10.1158/1078-0432.CCR-13-0708
5. Claret L, Gupta M, Han K, et al. Prediction of overall survival or progression free survival by disease control rate at week 8 is independent of ethnicity: Western versus Chinese patients with first-line non-small cell lung cancer treated with chemotherapy with or without bevacizumab. *J Clin Pharmacol*. 2014;54:253-257. doi:10.1002/jcph.191
6. Caulet M, Lecomte T, Bouché O, et al. Bevacizumab pharmacokinetics influence overall and progression-free survival in metastatic colorectal cancer patients. *Clin Pharmacokinet*. 2016;55:1381-1394. doi:10.1007/s40262-016-0406-3
7. Han K, Peyret T, Marchand M, et al. Population pharmacokinetics of bevacizumab in cancer patients with external validation. *Cancer Chemother Pharmacol*. 2016;78:341-351. doi:10.1007/s00280-016-3079-6
8. Papachristos A, Sivolapenko GB. Pharmacogenomics, pharmacokinetics and circulating proteins as biomarkers for bevacizumab treatment optimization in patients with cancer: a review. *J Pers Med*. 2020;10:1-20. doi:10.3390/jpm10030079
9. Hosmer DW, Lemeshow S, May S. *Applied Survival Analysis: Regression Modeling of Time-to-Event Data*. Wiley-Interscience; 2008:16-66 67-91, 244-285.
10. Vittinghoff E, Glidden DV, Shiboski SC, McCulloch CE. Regression methods in biostatistics: linear, logistic, survival, and repeated measures models. In: Gail M, Samet JM, eds. *Regression Methods in Biostatistics*. Springer Science & Business Media Inc; 2005:211-251. doi:10.1007/978-1-4614-1353-0
11. Mbogning C, Bleakley K, Lavielle M. Joint modeling of longitudinal and repeated time-to-event data using nonlinear mixed-effects models and the SAEM algorithm. *J Stat Comput Simul*. 2015;85:1512-1528. doi:10.1080/00949655.2013.878938
12. Pavlou M, Ambler G, Seaman SR, et al. How to develop a more accurate risk prediction model when there are few events. *BMJ (Online)*. 2015;351:h3868. doi:10.1136/bmj.h3868
13. Ojeda FM, Müller C, Börnigen D, et al. Comparison of Cox model methods in a low-dimensional setting with few events. *Genomics Proteomics Bioinformatics*. 2016;14:235-243. doi:10.1016/j.gpb.2016.03.006
14. Nardi A, Schemper M. Comparing Cox and parametric models in clinical studies. *Stat Med*. 2003;22:3597-3610. doi:10.1002/sim.1592

15. Cox C, Chu H, Schneider MF, Muñoz A. Parametic survival analysis and taxonomy of hazard functions for the generalized gamma distribution. *Stat Med*. 2007;26:4352-4374. doi:10.1002/sim.2836
16. Hutmacher MM, Kowalski KG. Covariate selection in pharmacometric analyses: a review of methods. *Br J Clin Pharmacol*. 2015;79:132-147. doi:10.1111/bcp.12451
17. Pavlou M, Ambler G, Seaman S, de Iorio M, Omar RZ. Review and evaluation of penalised regression methods for risk prediction in low-dimensional data with few events. *Stat Med*. 2016;35:1159-1177. doi:10.1002/sim.6782
18. Weathers B, Cutler, R. *Comparision of survival curves between Cox proportional hazards, random forests, and conditional inference forests in survival analysis*. 2017. https://digitalcommons.usu.edu/gradreports
19. Wang H, Li G. A selective review on random survival forests for high dimensional data. *Quant Biosci*. 2017;36:85-96. doi:10.22283/qbs.2017.36.2.85
20. Tibshirani R. The lasso method for variable selection in the Cox model. *Stat Med*. 1997;16(4):385-389. doi:10.1002/(sici)1097-0258(19970228)16:4<385::aid-sim380>3.0.co;2-3
21. Huang J, Harrington D. Penalized partial likelihood regression for right-censored data with bootstrap selection of the. *Penalty Parameter*. 2002;58(4):781-791. doi:10.1111/j.0006-341x.2002.00781.x
22. Papachristos A, Kemos P, Katsila T, et al. VEGF-A and ICAM-1 gene polymorphisms as predictors of clinical outcome to first-line bevacizumab-based treatment in metastatic colorectal cancer. *Int J Mol Sci*. 2019;20(22):5791. doi:10.3390/ijms20225791
23. Graffelman J, Chang C, Puig X, Wigginton J, Ortoleva L, Engels WR. Statistical tests and graphics for Hardy-Weinberg equilibrium. *CRAN*. 2021. https://cran.r-project.org/web/packages/HardyWeinberg/HardyWeinberg.pdf
24. Han M, Murugesan A, Bahlis NJ, et al. Letters to blood maintenance therapy for multiple myeloma. *Blood*. 2016;128(5):732-735. doi:10.1182/blood-2016-06-716902
25. Kumar Yadav B, Yadav R, Chang H, et al. Genetic polymorphisms rs699947, rs1570360, and rs3025039 on the VEGF gene are correlated with extracranial internal carotid artery stenosis and ischemic stroke. *Ann Clin Lab Sci*. 2017;47(2):144-155.
26. Liu L, He D, Fang L, Yan X. Association between E469K polymorphism in the ICAM1 gene and the risk of diabetic nephropathy: a meta-analysis. *Lipids Health Dis*. 2018;17(1):293. doi:10.1186/s12944-018-0922-2
27. Chen XZ, Yu SJ, Wei MH, Li CY, Yan WR. Effects of maternal and fetal vascular endothelial growth factor a single nucleotide polymorphisms on pre-eclampsia: a hybrid design study. *Cytokine*. 2020;127:154995. doi:10.1016/j.cyto.2020.154995
28. Simon RM, Subramanian J, Li MC, Menezes S. Using cross-validation to evaluate predictive accuracy of survival risk classifiers based on high-dimensional data. *Brief Bioinform*. 2011;12:203-214. doi:10.1093/bib/bbr001
29. Sill M, Hielscher T, Becker N, Zucknick M. c060: extended inference with lasso and elastic-net regularized Cox and generalized linear models. *J Stat Softw*. 2014;62(5):1-22. 10.18637/jss.v062.i05
30. Gerds TA. *Package "pec" Title Prediction Error Curves for Risk Prediction Models in Survival*. 2022. https://cran.r-project.org/web/packages/pec/pec.pdf
31. Patrick Heagerty AJ, Saha-Chaudhuri P, Paramita Saha-Chaudhuri M. *Title Riskset ROC curve estimation from censored survival data*. 2015. https://cran.r-project.org/web/packages/risksetROC/risksetROC.pdf
32. Benzekry S. Artificial intelligence and mechanistic modeling for clinical decision making in oncology. *Clin Pharmacol Ther*. 2020;108:471-486. doi:10.1002/cpt.1951
33. Chan P, Zhou X, Wang N, Liu Q, Bruno R, Jin JY. Application of machine learning for tumor growth inhibition – overall survival modeling platform. *CPT Pharmacometrics Syst Pharmacol*. 2021;10:59-66. doi:10.1002/psp4.12576
34. Terranova N, French J, Dai H, et al. Pharmacometric modeling and machine learning analyses of prognostic and predictive factors in the JAVELIN gastric 100 phase III trial of avelumab. *CPT Pharmacometrics Syst Pharmacol*. 2022;11(3):333-347. doi:10.1002/psp4.12754
35. Meid AD, Gerharz A, Groll A. Machine learning for tumor growth inhibition: interpretable predictive models for transparency and reproducibility. *CPT Pharmacometrics Syst Pharmacol*. 2022;11(3):257-261. doi:10.1002/psp4.12761
36. Loureiro H, Becker T, Bauer-Mehren A, Ahmidi N, Weberpals J. Artificial intelligence for prognostic scores in oncology: a benchmarking study. *Front Artif Intell*. 2021;4:625573. doi:10.3389/frai.2021.625573
37. Bertsimas D, Dunn J, Gibson E, Orfanoudaki A. Optimal survival trees. *Mach Learn*. 2022. doi:10.1007/s10994-021-06117-0
38. Mogensen UB, Gerds TA. Evaluating random forests for survival analysis using prediction error curves. *J Stat Softw*. 2012;50(11):1-23. doi:10.18637/jss.v050.i11
39. Bertrand J, Comets E, Mentre F. Comparison of model-based tests and selection strategies to detect genetic polymorphisms influencing pharmacokinetic parameters. *J Biopharm Stat*. 2008;18:1084-1102. doi:10.1080/10543400802369012
40. Goulooze SC, Krekels EHJ, Hankemeier T, Knibbe CAJ. Covariates in pharmacometric repeated time-to-event models: old and new (pre)selection tools. *AAPS J*. 2019;21:11. doi:10.1208/s12248-018-0278-6
41. Holford N. A time to event tutorial for pharmacometricians. *CPT Pharmacometrics Syst Pharmacol*. 2013;2(5):e43. doi:10.1038/psp.2013.18
42. Zhang Z, Cortese G, Combescure C, et al. Overview of model validation for survival regression model with competing risks using melanoma study data. *Ann Transl Med*. 2018;6:325. doi:10.21037/atm.2018.07.38
43. Badillo S, Banfai B, Birzele F, et al. An introduction to machine learning. *Clin Pharmacol Ther*. 2020;107:871-885. doi:10.1002/cpt.1796
44. McComb M, Bies R, Ramanathan M. Machine learning in pharmacometrics: opportunities and challenges. *Br J Clin Pharmacol*. 2021;88(4):1482-1499. doi:10.1111/bcp.14801
45. Liu Y, Starr MD, Bulusu A, et al. Correlation of angiogenic biomarker signatures with clinical outcomes in metastatic colorectal cancer patients receiving capecitabine, oxaliplatin, and bevacizumab. *Cancer Med*. 2013;2:234-242. doi:10.1002/cam4.71
46. Demircan NC, Dane F, Ozturk MA, et al. Assessment of survival and prognostic factors in metastatic colorectal cancer patients treated with first-line bevacizumab-based therapy. *J BUON*. 2019;24:1494-1500.

47. Ribbing J, Nyberg J, Caster O, Jonsson EN. The Lasso – a novel method for predictive covariate model building in nonlinear mixed effects models. *J Pharmacokinet Pharmacodyn*. 2007;34:485-517. doi:10.1007/s10928-007-9057-1
48. Pallaud C, Reck M, Juhasz E, et al. Clinical genotyping and efficacy outcomes: exploratory biomarker data from the phase II ABIGAIL study of first-line bevacizumab plus chemotherapy in non-squamous non-small-cell lung cancer. *Lung Cancer*. 2014;86:67-72. doi:10.1016/j.lungcan.2014.07.019
49. Schneider BP, Wang M, Radovich M, et al. Association of vascular endothelial growth factor and vascular endothelial growth factor receptor-2 genetic polymorphisms with outcome in a trial of paclitaxel compared with paclitaxel plus bevacizumab in advanced breast cancer: ECOG 2100. *J Clin Oncol*. 2008;26:4672-4678. doi:10.1200/JCO.2008.16.1612
50. Koutras AK, Antonacopoulou AG, Eleftheraki AG, et al. Vascular endothelial growth factor polymorphisms and clinical outcome in colorectal cancer patients treated with irinotecan-based chemotherapy and bevacizumab. *Pharmacogenomics J*. 2012;12:468-475. doi:10.1038/tpj.2011.37

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Karatza E, Papachristos A, Sivolapenko GB, Gonzalez D. Machine learning-guided covariate selection for time-to-event models developed from a small sample of real-world patients receiving bevacizumab treatment. *CPT Pharmacometrics Syst Pharmacol*. 2022;11:1328-1340. doi:10.1002/psp4.12848