Journal of Nuclear Medicine, published on November 16, 2017 as doi:10.2967/jnumed.117.200758

Prediction of ^{90}Y-Radioembolization Outcome from Pre-therapeutic Factors with Random Survival Forests

Michael Ingrisch[1,*], Franziska Schöppe[1], Karolin Paprottka[1], Matthias Fabritius[1], Frederik F. Strobl[1], Enrico De Toni[2], Harun Ilhan[3], Andrei Todica[3], Marlies Michl[4], Philipp Marius Paprottka[1]

[1]Department of Radiology
University Hospital Munich, Ludwig-Maximilians-University Munich

[2]Department of Internal Medicine II
University Hospital Munich, Ludwig-Maximilians-University Munich

[3]Department of Nuclear Medicine
University Hospital Munich, Ludwig-Maximilians-University Munich

[4]Department of Internal Medicine III
University Hospital Munich, Ludwig-Maximilians-University Munich

[*]Corresponding author:
Dr. Michael Ingrisch
Department of Radiology
University Hospital Munich, Ludwig-Maximilians-University Munich
Marchioninistr. 15
81377 München
Tel. +49 89 4400 74624
E-Mail michael.ingrisch@med.lmu.de
ORCID 0000-0003-0268-9078

Word count: 3297

Running title: Prediction of radioembolization outcome

ABSTRACT

To predict outcome of ^{90}Y radioembolization in patients with intrahepatic tumors from pre-therapeutic baseline parameters and to identify predictive variables using a machine-learning approach based on random survival forests (RSF).

Materials and methods: In this retrospective study, 366 patients with primary (n=92) or secondary (n=274) liver tumors who had received ^{90}Y radioembolization were analyzed. A random survival forest was trained to predict individual risk from baseline values of cholinesterase, bilirubin, type of primary tumor, age at radioembolization, hepatic tumor burden, presence of extrahepatic disease and sex. The predictive importance of each baseline parameter was determined using the minimal depth concept, and the partial dependency of predicted risk on the continuous variables bilirubin level and cholinesterase level was determined.

Results: Median overall survival was 11.4 months (95% C.I. 9.7-14.2 months) with 228 deaths observed during the observation period. The random survival forest analysis identified baseline cholinesterase and bilirubin as the most important variables with the forest-averaged lowest minimal depth of 1.2 and 1.5, followed by the type of primary tumor (1.7), age (2.4), tumor burden (2.8) and presence of extrahepatic disease (3.5). Sex had the highest forest-averaged minimal depth (5.5), indicating little predictive value. Baseline bilirubin levels above 1.5 mg/dl were associated with a steep increase in predicted mortality. Similarly, cholinesterase levels below 7.5 U/ predicted a strong increase in mortality. The trained random survival forest achieved a concordance index of c=0.657, with a standard error of 0.02, comparable to c=0.652 (0.02) of a previously published Cox proportional hazards model.

Conclusion: Random survival forests are a simple and straightforward machine learning approach for prediction of overall survival. Predictive performance of the trained model was similar to a previously published Cox regression model. The model has revealed a strong predictive value of baseline cholinesterase and bilirubin levels with a highly nonlinear influence of each parameter.

KEYWORDS

Radioembolization, bilirubin, cholinesterase, prediction, random survival forest

INTRODUCTION

Radioembolization with ^{90}Yttrium-loaded resin microspheres is an established and potentially life-prolonging treatment option for patients with hepatocellular carcinoma (HCC) (*1*), cholangiocellular carcinoma (*2*), metastases of colorectal carcinoma (*3*), metastases of neuroendocrine tumors (*4*,*5*) and metastatic breast cancer (*6*). When considering an aggressive therapy, one needs to balance cost and risk of complications against quality and potential prolonging of life. Prediction of treatment response is therefore highly relevant for patient selection and stratification.

For the stratification of patients eligible for radioembolization, a simple risk score model based on Karnofsky index and carcinoembryonic antigen and cancer antigen 19-9 serum levels has been proposed (*7*). This score is easily applicable in clinical routine and has been demonstrated to be strongly predictive. A similar approach was followed in (*8*), where a nomogram was constructed based on the hazard ratios of risk factors. Due to their simplicity, these approaches may impose overly strong assumptions, e.g. on linearity of the effect of predictive variables, and may not utilize the full information that is contained in baseline variables. An established statistical concept for prediction of risk is based on multivariate Cox proportional hazard models (*6*,*9*). Here, a statistical model for the individual hazard ratio is derived; multivariate situations are addressed e.g. by stepwise variable selection or by including interaction terms. In a recent study (*9*), multivariate Cox regression determined type of primary tumor, tumor burden, presence of extrahepatic disease and baseline level of cholinesterase as independent predictors of overall survival (OS). Cox regression relies on strong and potentially restrictive assumptions about linearity; and the selection of appropriate variables and interaction terms is an art in itself and often considered unintuitive and a 'black box' from the perspective of the clinician.

Recently, a particular statistical model termed *random survival forest* has emerged as an intuitive technique for predicting individual risk (*10*–*12*). By combining many individual decision trees, random survival forests form an ensemble method and as such have attractive properties: They require little input from the analyst and they can easily deal with nonlinear effects, correlated parameters and variable interactions. In addition, random survival forests allow for an intuitive assessment of variable importance (*13*) and allow insights into the partial dependency of predicted risk on individual variables.

In the present work, we evaluate whether random survival forests can predict response to radioembolization in a large cohort of patients with hepatic tumors and metastases who underwent ^{90}Y-radioembolization at our institute. In addition, we evaluate importance and predictive value of clinical variables for therapy outcome, and we compare results from the random survival forest analysis to a previously published Cox proportional hazards model with respect to prediction error and variable selection.

MATERIALS AND METHODS

Patients

This retrospective study analyzed patients from a previously described cohort (9). This cohort comprised consecutive patients with hepatocellular carcinoma, cholangiocellular carcinoma, metastases of colorectal carcinoma, neuroendocrine tumors, breast cancer and other hepatic liver metastases who underwent radioembolization at our institution between January 2009 and December 2012. The study has been approved by the institutional review board and the need for written informed consent was waived. One day before the first radioembolization procedure, the following pre-therapeutic parameters were recorded: bilirubin and cholinesterase levels in mg/dL and U/L, respectively, age at time of procedure, sex, type of primary tumor, extrahepatic disease, defined as presence of metastatic lymph nodes, or other non-life-limiting metastases, and hepatic tumor burden. The latter was assessed in three

categories (<25%, 25-50%, >50%) by means of pre-therapeutic contrast-enhanced magnetic resonance imaging using gadobenate dimeglumine after segmentation of tumor volume. Patients were followed up until December 2013 and were included into the retrospective analysis when all of the above pre-therapeutic baseline parameters were available. Patients where one or more of these parameters were unavailable were excluded from further analysis.

Before radioembolization, patients had undergone an angiographic procedure to detect and occlude relevant aberrant vessels, which otherwise would have led to extrahepatic deposition of microspheres. Approximately 100 MBq Tc-99m macro-aggregated albumin was applied at the arterial tree to assess relevant liver-lung and epigastric shunts by means of planar scintigraphy and single-photon emission computed tomography examination. At a second hepatic arterial catheterization conducted after therapy planning angiography, ^{90}Y resin microspheres (SIR-Spheres; Sirtex Medical Ltd, Sydney, Australia) suspended in water for injection were applied under intermittent fluoroscopic visualization. The prescribed activity was administered either in whole-liver, lobar or sequential lobar treatment. Within 24 hours after therapy, target deposition of microspheres was confirmed by SPECT/CT scans.

Statistical Analysis

A random survival forest is trained by growing a large number of individual trees (*10*,*11*). Each tree is trained on a random bootstrap sample from the original cohort. Starting with the entire sample at the tree trunk, a random set of variables is chosen as candidates for splitting the branch in two sub-branches, with the objective to maximize the difference in survival between sub-branches. The optimal splitting threshold is determined for each of the candidate variables, and the variable which maximizes the log-rank statistic between splitted data is chosen for splitting (*10*). This process is repeated until a pre-determined terminal nodesize is achieved. A trained random survival forest predicts an individual mortality, which is calibrated on the number of events. Specifically, if all patients shared the same characteristics, the predicted mortality equals the number of expected deaths.

All analyses were performed with R version 3.3.2 (www.R-project.org). A random survival forest with 2000 trees was trained on the entire dataset using the R package randomForestSRC (*14*), with a terminal nodesize of 5, mtry=3 variables were selected randomly in each iteration. Seven pre-therapeutic variables, described above, were used for analysis, with right-censored survival time as primary endpoint. To ensure unbiased evaluation, the individual risk was predicted from each tree only for the remaining data, which was not used during training (*out-of-bag* data, (*10*)).

To evaluate the predictive performance of the random survival forest, the *concordance index* (CI) of the final forest was calculated. The concordance index is a measure for the evaluation of statistical survival models and reports the fraction of allowed pairs of samples which is sorted in the right order. Hence, a concordance index of 0.5 indicates random sorting, and a concordance index of 1.0 perfect sorting. As a reference, a previously reported Cox proportional hazards model (*9*) was fitted to our dataset and the concordance index of this model was determined.

As a measure of the relative importance and hence the predictive value of variables, the *minimal depth* (*13*) was employed. This minimal depth of a variable in a single tree is the shortest distance from the tree trunk to the branch level of the first split of the variable. The most important variables are considered to be those variables, which are most frequently used for splits close to the tree trunks. Hence, the importance of each variable can be assessed as the forest-averaged minimal depth.

Random survival forests can be interpreted as a mapping of several independent variables to a combined outcome. An advantage of this concept is that the partial dependency of the predicted outcome on each independent variable can be assessed separately by integrating out all other variables. This is a powerful tool for assessing nonlinear behavior of single variables and allows novel insights which are not accessible with the more traditional Cox proportional hazard models. Partial dependencies were calculated for baseline levels of cholinesterase and bilirubin and for tumor type.

RESULTS

366 patients who had received radioembolization were included in the study (217 male, 149 female, mean age 62 years, range 31 to 91 years) and were analyzed retrospectively. Median overall survival was 11.4 months (95% C.I. 9.7-14.2 months). During the observation period, 228 deaths were observed. Details are provided in Table 1.

Bilirubin and cholinesterase values showed moderate negative, but significant correlation ($r=-0.38$, $p<0.001$). Analysis of variance revealed a significant ($p<0.001$) association of cholinesterase with tumor burden.

The trained random survival forest achieved a concordance index of $c=0.657$, with a standard error of 0.02. In comparison, the concordance index of the previously used parsimonious Cox model was 0.652 (0.02). The median of the individual predicted mortality was 93 – implying that, if all individuals had the same parameters as this patient, an average of 93 deaths would be expected. Predicted mortalities ranged from 14 to 254.

Splitting the patients in two equal-sized groups with the median predicted mortality as threshold, survival time in the group with low predicted mortality was significantly longer than in the group with high-predicted mortality (16.8 months vs 6.6 months, $p=6.06\cdot10^{-8}$), as demonstrated in Fig. 1.

The variable importance of pre-therapeutic variables is illustrated in Fig. 2. Cholinesterase and bilirubin levels have the lowest minimal depth with 1.2 and 1.5, followed by the type of primary tumor, with a forest-averaged minimal depth of 1.7. Age (2.4), tumor burden (2.8) and presence of extrahepatic disease (3.5) have lower minimal depth, and sex has the highest averaged minimal depth (5.5), indicating little predictive value. Importantly, both cholinesterase and bilirubin are included in the model and have similar importance.

Figs. 3 and 4 illustrate the dependency of predicted mortality on the pre-therapeutic levels of bilirubin and cholinesterase and illustrate highly non-linear behavior. Bilirubin levels below 1.5 mg/dL

have little influence on predicted mortality. Levels above 1.5 mg/dL are associated with a linear and steep increase in predicted mortality. Similarly, cholinesterase levels above 7.5 U/L are associated with a roughly constant predicted mortality, whereas a level falling below 7.5 U/L predicts strong increase in mortality.

Figure 5 demonstrates the influence of tumor type on the expected mortality. Predicted mortalities depend on the tumor type. The trained model predicts the highest mortality for metastatic breast cancer and the lowest mortality for neuroendocrine tumors.

DISCUSSION

In this study, we have used an advanced statistical method, random survival forests, to predict response to ^{90}Y-radioembolization in a cohort of patients with hepatic tumor burden. We have demonstrated that the predictive performance of the proposed random survival forest is similar to a previously published Cox model, without relying on restrictive assumptions. By means of the concept of 'minimal depth', we assessed the importance and hence the predictive value of pre-therapeutic variables. Confirming a previous finding (*9*), pre-therapeutic cholinesterase level emerged as a highly predictive factor, closely followed by pre-therapeutic bilirubin level and tumor type. This highlights the role of these parameters as marker of liver function. While elevated bilirubin levels indicate impaired hepatic bilirubin clearance, cholinesterase is an important biomarker of the synthetic liver function (15). As such, the cholinesterase level provides complementary information about liver function, especially in patients with primary tumors of the liver who have an underlying cirrhotic liver disease.

Importantly, the random survival forest model can accommodate both parameters with appropriate importance, whereas the conventional multivariate regression model excluded bilirubin level due to correlation with cholinesterase. In conventional analysis, such correlated parameters may

act as confounders, whereas random survival forests are able to circumvent this issue through the two-fold randomization in the training process (13).

In addition, our random survival forest analysis provided novel insights into the influence of individual predictive variables on the overall predicted risk: by assessing partial dependency, we were able to demonstrate non-linear behavior for baseline cholinesterase and bilirubin levels, confirming previous intuitions. Our analysis suggests that bilirubin levels below 1.5 mg/dl have little influence on the predicted risk, whereas bilirubin levels above 1.5 mg/dl are associated with a strong and approximately linear increase in risk. Likewise, a pre-therapeutic cholinesterase level above 7.5 U/L is associated with good prognosis, whereas lower levels are associated with high predicted mortality. Moreover, the trained random survival forest captures the influence of tumor type on overall survival: Metastatic breast cancer is associated with the highest predicted mortality, closely followed by colorectal carcinoma, whereas the model predicts the lowest mortality for neuroendocrine tumors. This model behavior is in excellent agreement with the median survival times in our patient cohort (9).

Recently, a simple scoring system for patient selection was proposed (*7*), in which tumor burden, Karnofsky index and serum levels of carcinoembryonic antigen and/or cancer antigen 19-9 were binarized and formed a combined score. This score discriminated overall survival in patients with metastatic colorectal carcinoma, suggesting potential for improved patient selection. In comparison, our random survival forest model was trained on a larger cohort, including more types of primary tumors, considers more variables and, importantly, does not rely on the definition of threshold values.

Our random survival forest model predicts individual mortality as a continuous variable. To select patients who benefit most from radioembolization, an optimal cutoff value needs to be found. To choose such a threshold level, one needs to balance overtreatment and risks of aggressive therapy against the benefits of life-prolonging therapy with acceptable quality of life. Briefly, one would dichotomize the predicted mortality and estimate the log-rank statistic and the hazard ratio in a Cox model (*16*). As a cutoff value, one could then choose either the value where the log-rank statistic has the highest

significance, or one could choose the maximal predicted mortality that still results in a significant difference in overall survival, with the objective to include a large number of patients.

In our analysis, the hepatic tumor burden expressed in three categories (below 25%, 25%-50% and above 50%) was moderately important. This measure was derived from pre-therapeutic magnetic resonance imaging. In light of the emerging role of radiomics (*17*), it can be expected that imaging biomarkers with much higher predictive performance will be identified as demonstrated recently for high-grade brain tumors (*12*,*18*). Hence, a stronger contribution of pre-therapeutic imaging to the prediction of mortality and stratification of patients is not unlikely.

The present study is not without limitations. First, it relies on data from a single institution only; our findings should be validated on a large database, ideally from multiple institutions. Secondly, we have not validated our random survival forest on an independent test dataset, nor have we used cross validation for prediction of mortality. However, unbiased predictions were ensured by 'out-of-bag' predictions – every single tree predicted outcome only for the data that were not used for tree growing. Moreover, our analysis was not focused on prediction alone; we aimed also to derive insights into the contribution of individual variables. For this purpose, validation in separate datasets is not required.

In conclusion, we have utilized a modern statistical approach for prediction of overall survival after ^{90}Y radioembolization. Predictive performance of our model was similar to a previously published Cox proportional hazard model and, in addition, the model has revealed a strong predictive value of baseline cholinesterase and bilirubin with a highly nonlinear influence of each parameter.

DISCLOSURE

This study received no funding. The authors declare that they have no conflicts of interest

REFERENCES

1. Malfertheiner P, Verslype C, Kolligs FT, et al. The effectiveness of selective internal radiation therapy in challenging cases of liver-predominant unresectable hepatocellular carcinoma. *Future Oncol Lond Engl*. 2014;10:17-27.

2. Kloeckner R, Ruckes C, Kronfeld K, et al. Selective internal radiotherapy (SIRT) versus transarterial chemoembolization (TACE) for the treatment of intrahepatic cholangiocellular carcinoma (CCC): study protocol for a randomized controlled trial. *Trials*. 2014;15:311.

3. Jakobs TF, Paprottka KJ, Raeßler F, et al. Robust evidence for long-term survival with (90)Y radioembolization in chemorefractory liver-predominant metastatic colorectal cancer. *Eur Radiol*. 2017;27:113-119.

4. Sommer WH, Ceelen F, García-Albéniz X, et al. Defining predictors for long progression-free survival after radioembolisation of hepatic metastases of neuroendocrine origin. *Eur Radiol*. 2013;23:3094-3103.

5. Paprottka PM, Hoffmann R-T, Haug A, et al. Radioembolization of symptomatic, unresectable neuroendocrine hepatic metastases using yttrium-90 microspheres. *Cardiovasc Intervent Radiol*. 2012;35:334-342.

6. Fendler WP, Lechner H, Todica A, et al. Safety, efficacy, and prognostic factors after radioembolization of hepatic metastases from breast cancer: A large single-center experience in 81 patients. *J Nucl Med*. 2016;57:517-523.

7. Damm R, Seidensticker R, Ulrich G, et al. Y90 Radioembolization in chemo-refractory metastastic, liver dominant colorectal cancer patients: outcome assessment applying a predictive scoring system. *BMC Cancer*. 2016;16:509.

8. Fendler WP, Ilhan H, Paprottka PM, et al. Nomogram including pretherapeutic parameters for prediction of survival after SIRT of hepatic metastases from colorectal cancer. *Eur Radiol*. 2015;25:2693-2700.

9. Paprottka KJ, Schoeppe F, Ingrisch M, et al. Pre-therapeutic factors for predicting survival after radioembolization: a single-center experience in 389 patients. *Eur J Nucl Med Mol Imaging*. 2017;44:1185-1193

10. Ishwaran H, Kogalur UB, Blackstone EH, Lauer MS. Random Survival Forests. *Ann Appl Stat*. 2008;2:841-860.

11. Hsich E, Gorodeski EZ, Blackstone EH, Ishwaran H, Lauer MS. Identifying important risk factors for survival in patient with systolic heart failure using random survival forests. *Circ Cardiovasc Qual Outcomes*. 2011;4:39-45.

12. Ingrisch M, Schneider MJ, Nörenberg D, et al. Radiomic analysis reveals prognostic information in T1-weighted baseline magnetic resonance imaging in patients with glioblastoma. *Invest Radiol*. 2017;52:360-366.

13. Ishwaran H, Kogalur UB, Chen X, Minn AJ. Random survival forests for high-dimensional data. *Stat Anal Data Min*. 2011;4:115-132.

14. Ishwaran H, Kogalur UB. Random forests for survival, regression and classification (RF-SRC) [Internet]. manual; 2016. Accessed 6/14/2016. Available from: https://cran.r-project.org/package=randomForestSRC

15. Mendel B, Rudney H. Studies on cholinesterase: 1. Cholinesterase and pseudo-cholinesterase. *Biochem J.* 1943;37:59-63.

16. Budczies J, Klauschen F, Sinn BV, et al. Cutoff Finder: a comprehensive and straightforward web application enabling rapid biomarker cutoff optimization. *PLOS ONE*. 2012;7:e51862.

17. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology*. 2015;278:563-577.

18. Kickingereder P, Götz M, Muschelli J, et al. Large-scale radiomic profiling of recurrent glioblastoma identifies an imaging predictor for stratifying anti-angiogenic treatment response. *Clin Cancer Res Off J Am Assoc Cancer Res*. 2016;22:5765-5771.

FIGURES

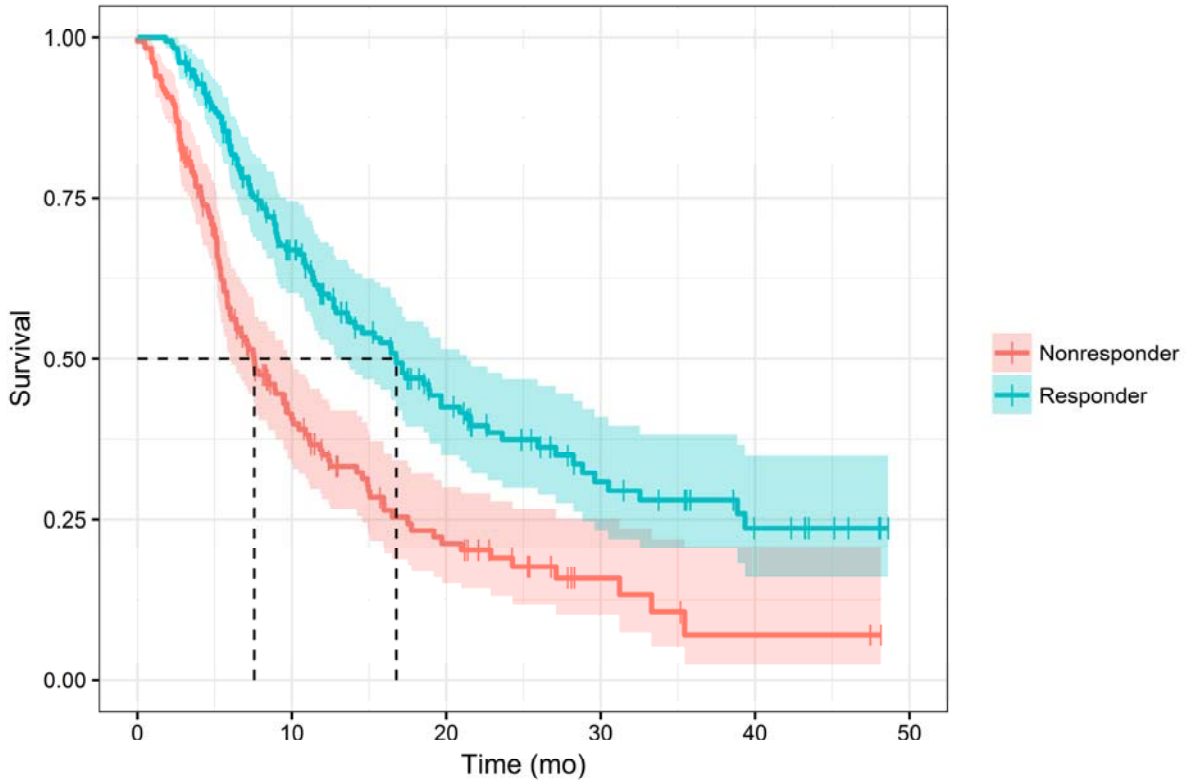

Figure 1: Splitting the patients on the median of the predicted mortality reveals strong and highly significant ($p = 6*10^{-8}$) differences in OS. Median survival time in the non-responder group is 6.6 months, whereas median survival in the responder group is 16.8 months.

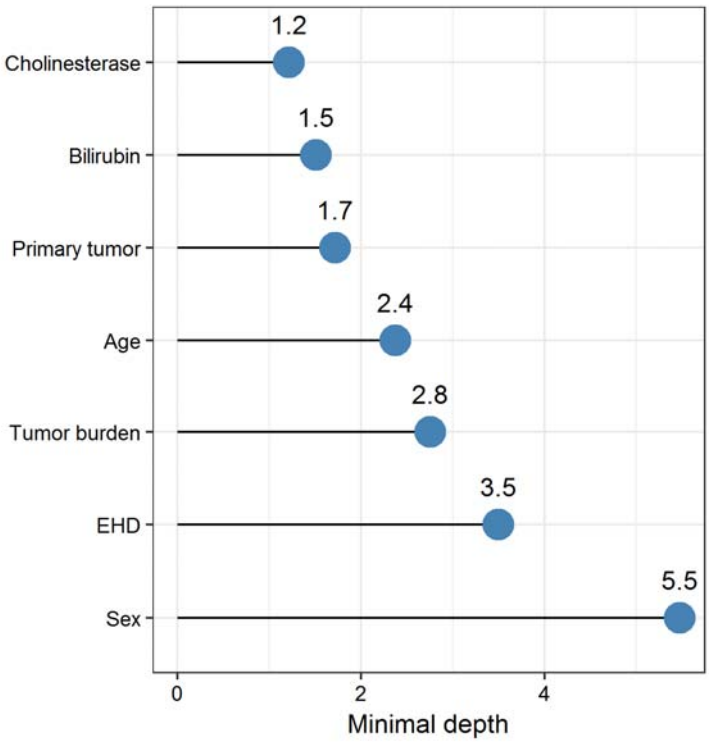

Figure 2: Minimal depth of baseline parameters, measuring the variable importance: Low values indicate that a variable is used early in tree growing and has stronger predictive value. Cholinesterase and bilirubin levels have the lowest minimal depth, highlighting the importance of liver function.

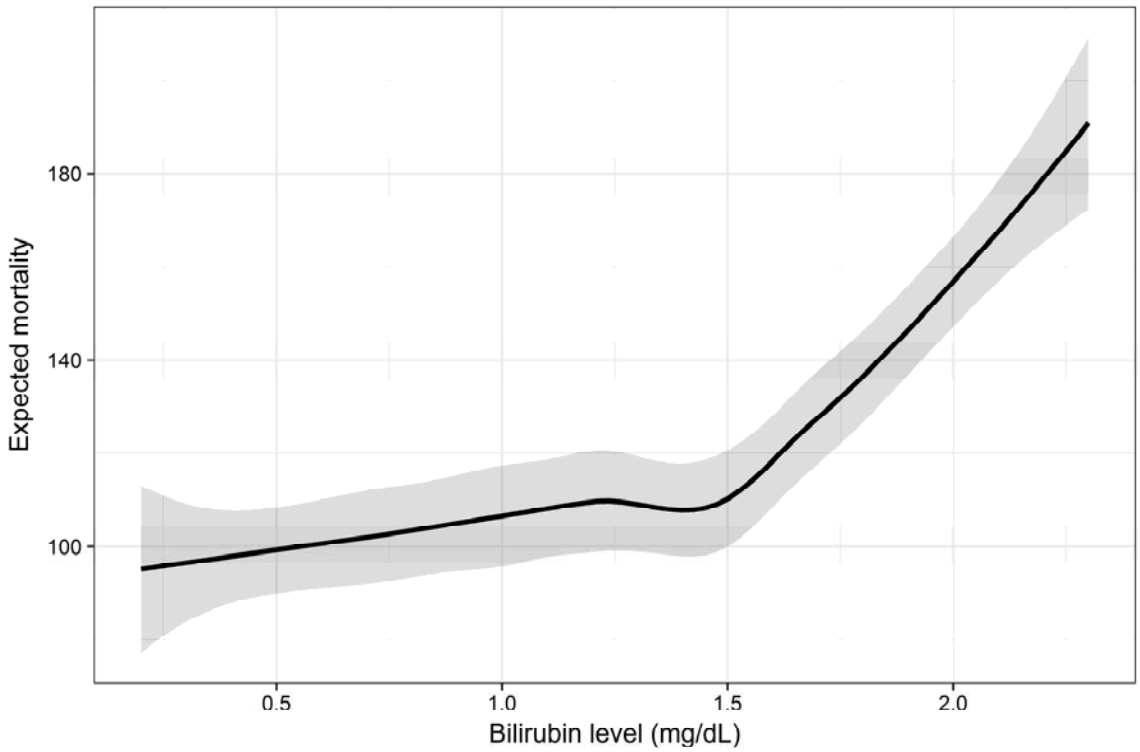

Figure 3: Partial dependency for bilirubin: Expected mortality increases strongly once bilirubin levels exceed a value of approximately 1.5 mg/dL.

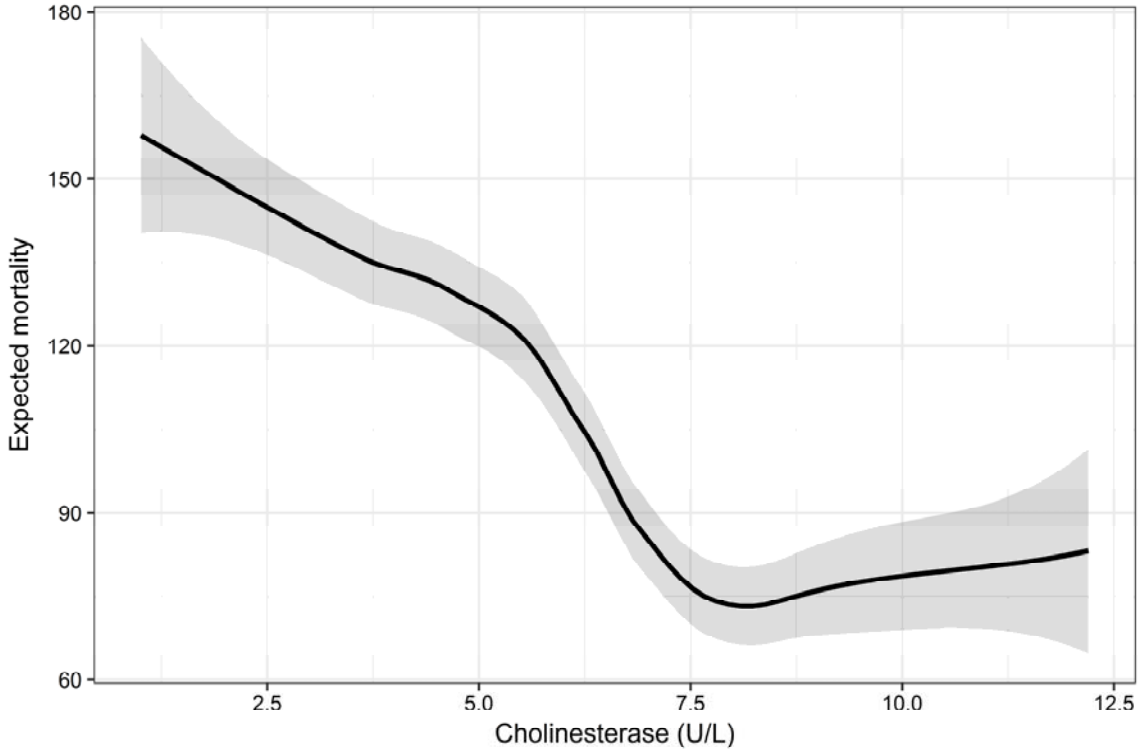

Figure 4: Partial dependency for cholinesterase: Cholinesterase levels below approx. 7.5 U/L are associated with a strong increase in expected mortality.

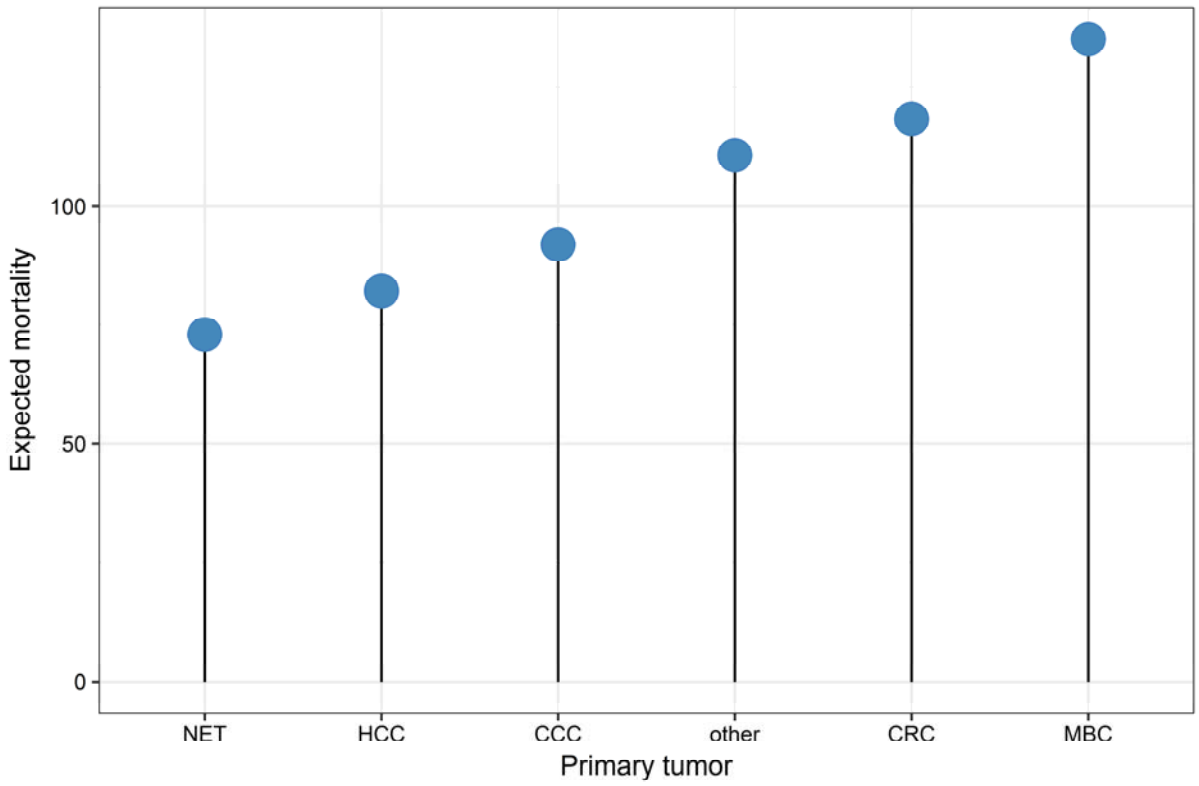

Figure 5: Expected mortality in dependence on tumor type. *NET* neuroendocrine tumor, *HCC* hepatocellular carcinoma, *CCC* cholangiocarcinoma, *CRC* colorectal carcinoma, *MBC* metastatic breast cancer.

TABLES

Table 1: Baseline patient characteristics

Characteristic	N(%) or median (IQR)
Total	
Age (years)	64 (55.7-71.0)
Sex	
Male	217 (59)
Female	149 (41)
Primary Tumor	
Colorectal Cancer (CRC)	128 (35.0)
Hepatocellular cancer (HCC)	57 (15.6)
Neuroendocrine tumor (NET)	51 (13.9)
Metastatic breast cancer (MBC)	40 (10.9)
Cholangiocarcinoma (CCC)	35 (9.6)
Other[a]	55 (15.0)
Hepatic tumor burden	
<25%	191 (52.2)
25%-50%	140 (38.3)
>50%	35 (9.6)
Extrahepatic disease	
Yes	253 (69.1)
No	113 (30.9)
Baseline liver function parameters	
Bilirubin, mg/dL	0.6 (0.5-0.9) (normal range <1.2)
Cholinesterase (U/L)	6.35 (4.89 – 7.6) (normal range >4.6)

[a]Pancreas (*n* = 13), uveal melanoma (*n* = 6), gastric cancer (*n* = 6), sarcoma (*n* =4), urothelial carcinoma (*n* = 4), ovarian cancer (*n* = 3), malignant melanoma (*n* = 3), cancer of unknown primary (CUP) (*n* = 2), prostate cancer (*n* = 2), lung cancer (*n* = 2), thymus cancer (n = 2), base of the tongue cancer (*n* = 2), squamous cell carcinoma (*n* = 1), endometrial cancer (*n* = 1), oesophageal carcinoma (*n* = 1), thyroid carcinoma (*n* = 1), squamous cell carcinoma of the maxillary sinus (*n* = 1), testicular cancer (*n* = 1)