
Prediction of ⁹⁰Y Radioembolization Outcome from Pretherapeutic Factors with Random Survival Forests

Michael Ingrischi¹, Franziska Schöppe¹, Karolin Paprottka¹, Matthias Fabritius¹, Frederik F. Strobl¹, Enrico N. De Toni², Harun Ilhan³, Andrei Todica³, Marlies Michl⁴, and Philipp Marius Paprottka¹

¹Department of Radiology, University Hospital Munich, Ludwig-Maximilians-University Munich, Munich, Germany; ²Department of Internal Medicine II, University Hospital Munich, Ludwig-Maximilians-University Munich, Munich, Germany; ³Department of Nuclear Medicine, University Hospital Munich, Ludwig-Maximilians-University Munich, Munich, Germany; and ⁴Department of Internal Medicine III, University Hospital Munich, Ludwig-Maximilians-University Munich, Munich, Germany

Our objective was to predict the outcome of ⁹⁰Y radioembolization in patients with intrahepatic tumors from pretherapeutic baseline parameters and to identify predictive variables using a machine-learning approach based on random survival forests. **Methods:** In this retrospective study, 366 patients with primary ($n = 92$) or secondary ($n = 274$) liver tumors who had received ⁹⁰Y 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 mo (95% confidence interval, 9.7–14.2 mo), with 228 deaths occurring during the observation period. The random-survival-forest analysis identified baseline cholinesterase and bilirubin as the most important variables (forest-averaged lowest minimal depth, 1.2 and 1.5, respectively), 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 0.657, with an SE of 0.02, comparable to the concordance index of 0.652 and SE of 0.02 for a previously published Cox proportional hazards model. **Conclusion:** Random survival forests are a simple and straightforward machine-learning approach for prediction of overall survival. The predictive performance of the trained model was similar to a previously published Cox regression model. The model has revealed a strong predictive value for baseline cholinesterase and bilirubin levels with a highly nonlinear influence for each parameter.

Key Words: radioembolization; bilirubin; cholinesterase; prediction; random survival forest

J Nucl Med 2018; 59:769–773

DOI: 10.2967/jnumed.117.200758

Radioembolization with ⁹⁰Y-loaded resin microspheres is an established and potentially life-prolonging treatment option for patients with hepatocellular carcinoma (1), cholangiocellular carcinoma (2), metastatic colorectal carcinoma (3), metastatic 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 the Karnofsky index and on carcinoembryonic antigen and cancer antigen 19-9 serum levels has been proposed (7). This score is easily applicable in clinical routine and has been shown to be strongly predictive. A similar approach was followed in a previous study (8), in which a nomogram based on the hazard ratios of risk factors was constructed. Because of their simplicity, these approaches may impose overly strong assumptions, such as on the linearity of the effect of predictive variables, and may not use the full information that is contained in baseline variables. An established statistical concept for prediction of risk is based on multivariate Cox proportional-hazards models (6,9). Here, a statistical model for the individual hazard ratio is derived; multivariate situations are addressed by, for example, using stepwise variable selection or 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. 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—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 evaluated whether random survival forests can predict response to radioembolization in a large cohort

Received Aug. 15, 2017; revision accepted Oct. 16, 2017.

For correspondence or reprints contact: Michael Ingrischi, Department of Radiology, University Hospital Munich, Ludwig-Maximilians-University Munich, Marchioninistrasse 15, 81377 München, Germany.

E-mail: michael.ingrischi@med.lmu.de

Published online Nov. 16, 2017.

COPYRIGHT © 2018 by the Society of Nuclear Medicine and Molecular Imaging.

of patients with hepatic tumors and metastases who underwent ^{90}Y radioembolization at our institute. In addition, we evaluated the importance and predictive value of clinical variables for therapy outcome, and we compared 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, or breast cancer; or other hepatic metastases who underwent radioembolization at our institution between January 2009 and December 2012. The study was approved by the institutional review board, and the need for written informed consent was waived. One day before the first radioembolization procedure, the following pretherapeutic 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 nonlife-limiting metastases; and hepatic tumor burden. The last of these was assessed in 3 categories (<25%, 25%–50%, >50%) by means of pretherapeutic contrast-enhanced MRI using gadobenate dimeglumine after segmentation of tumor volume. Patients were followed up until December 2013 and were included in the retrospective analysis when all the above pretherapeutic baseline parameters were available. Patients for whom 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 that otherwise would have led to extrahepatic deposition of microspheres. Approximately 100 MBq of $^{99\text{m}}\text{Tc}$ -macroaggregated albumin were applied at the arterial tree to assess relevant liver–lung and epigastric shunts by means of planar scintigraphy and SPECT examination. At a second hepatic arterial catheterization conducted after therapy-planning angiography, ^{90}Y -resin microspheres (SIR-Spheres; Sirtex Medical Ltd.) 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 h after therapy, deposition of microspheres at the target 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 into 2 subbranches, with the objective of maximizing the difference in survival between subbranches. The optimal splitting threshold is determined for each of the candidate variables, and the variable that maximizes the log-rank statistic between split data is chosen for splitting (10). This process is repeated until a predetermined terminal node size 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 would equal the number of expected deaths.

All analyses were performed with R, version 3.3.2 (www.R-project.org). A random survival forest with 2,000 trees was trained on the entire dataset using the R package randomForestSRC (14), with a terminal node size of 5; 3 variables were randomly sampled as candidates at each iteration. Seven pretherapeutic variables, described above, were used for analysis, with right-censored survival time as

the primary endpoint. To ensure unbiased evaluation, the individual risk was predicted from each tree only for the remaining data, which were 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 are 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 used. 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 that are most frequently

TABLE 1
Baseline Patient Characteristics

Characteristic	Data
Age (y)	64 (55.7–71.0)
Sex	
Male	217 (59)
Female	149 (41)
Primary tumor	
Colorectal cancer	128 (35.0)
Hepatocellular cancer	57 (15.6)
Neuroendocrine tumor	51 (13.9)
Metastatic breast cancer	40 (10.9)
Cholangiocarcinoma	35 (9.6)
Other*	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	
Bilirubin (mg/dL)	0.6 (0.5–0.9) (reference range < 1.2)
Cholinesterase (U/L)	6.35 (4.89–7.6) (reference range > 4.6)

*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 ($n = 2$), prostate cancer ($n = 2$), lung cancer ($n = 2$), thymus cancer ($n = 2$), base-of-tongue cancer ($n = 2$), squamous cell carcinoma ($n = 1$), endometrial cancer ($n = 1$), esophageal carcinoma ($n = 1$), thyroid carcinoma ($n = 1$), squamous cell carcinoma of maxillary sinus ($n = 1$), and testicular cancer ($n = 1$).

Qualitative data are expressed as numbers followed by percentages in parentheses; continuous data are expressed as median followed by interquartile range in parentheses.

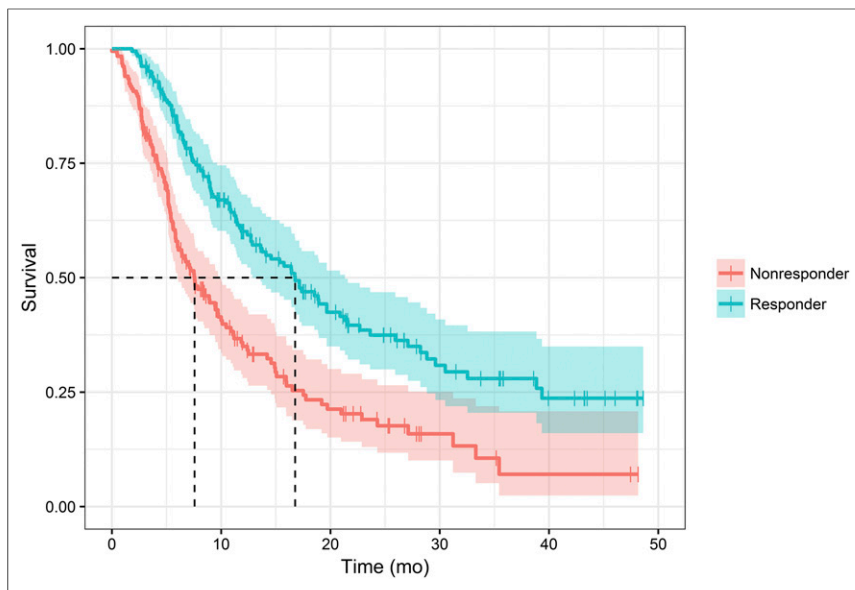


FIGURE 1. Splitting patient population on median of predicted mortality reveals strong and highly significant ($P = 6 \times 10^{-8}$) differences in overall survival. Median survival in nonresponder group was 6.6 mo, whereas median survival in responder group was 16.8 mo.

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

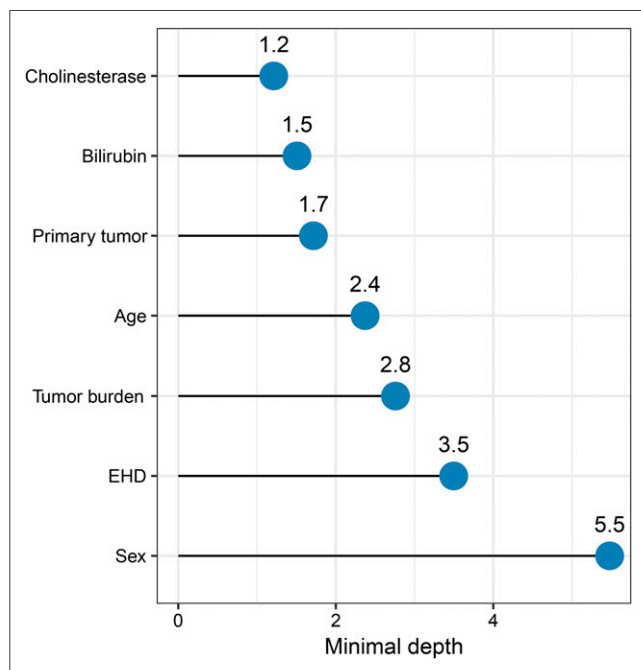


FIGURE 2. Minimal depth of baseline parameters, measuring the variable importance: low values indicate that variable is used early in tree growing and has stronger predictive value. Cholinesterase and bilirubin levels have lowest minimal depth, highlighting importance of liver function. EHD = extrahepatic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

out all other variables. This is a powerful tool for assessing nonlinear behavior of single variables and allows novel insights that are not accessible with the more traditional Cox proportional-hazards models. Partial dependencies were calculated for baseline levels of cholinesterase and bilirubin and for tumor type.

RESULTS

In total, 366 patients who had received radioembolization were included in the study (217 male, 149 female; mean age, 62 y; range, 31–91 y) and were analyzed retrospectively. Median overall survival was 11.4 mo (95% confidence interval, 9.7–14.2 mo). During the observation period, 228 deaths occurred. Details are provided in Table 1.

Bilirubin and cholinesterase values showed a moderately negative but significant correlation ($r = -0.38$, $P < 0.001$). ANOVA revealed a significant ($P < 0.001$) association of cholinesterase with tumor burden.

The trained random survival forest achieved a concordance index of 0.657, with an SE of 0.02. In comparison, the concordance index of the previously used parsimonious Cox model was 0.652 (SE, 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 patient population into 2 equal-sized groups with the median predicted mortality as the threshold, we found that survival was significantly longer in the group with a low predicted mortality than in the group with a high predicted mortality (16.8 vs. 6.6 mo, $P = 6.06 \times 10^{-8}$), as demonstrated in Figure 1.

The variable importance of pretherapeutic variables is illustrated in Figure 2. Cholinesterase and bilirubin levels had the lowest forest-averaged minimal depth (1.2 and 1.5, respectively), followed by type of primary tumor (1.7), age (2.4), tumor burden (2.8), and presence of extrahepatic disease (3.5). Sex had the highest averaged minimal depth (5.5), indicating little predictive value. Importantly, both cholinesterase and bilirubin were included in the model and had similar importance.

Figures 3 and 4 illustrate the dependency of predicted mortality on the pretherapeutic levels of bilirubin and cholinesterase and illustrate highly nonlinear behavior. Bilirubin levels below 1.5 mg/dL had little influence on predicted mortality. Levels above 1.5 mg/dL were associated with a linear and steep increase in predicted mortality. Similarly, cholinesterase levels above 7.5 U/L were associated with a roughly constant predicted mortality, whereas a level falling below 7.5 U/L predicted a 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 predicted 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

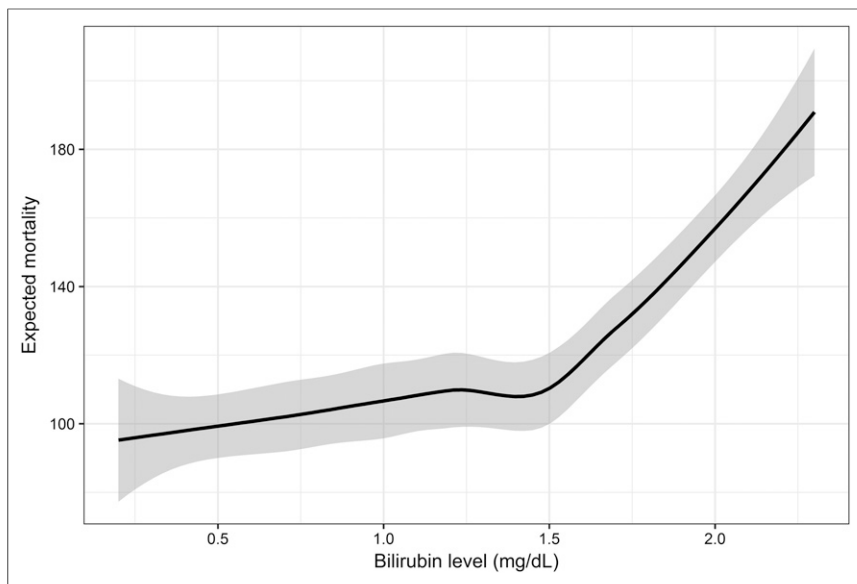


FIGURE 3. Partial dependency for bilirubin: expected mortality increases strongly once bilirubin levels exceed approximately 1.5 mg/dL.

in a cohort of patients with a 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 pretherapeutic variables. Confirming a previous finding (9), pretherapeutic cholinesterase level emerged as a highly predictive factor, closely followed by pretherapeutic bilirubin level and tumor type. This result highlights the role of these parameters as markers of liver function. Although elevated bilirubin levels indicate impaired hepatic bilirubin clearance, cholinesterase is an important biomarker of the synthetic liver function (15). As such, the cholinesterase

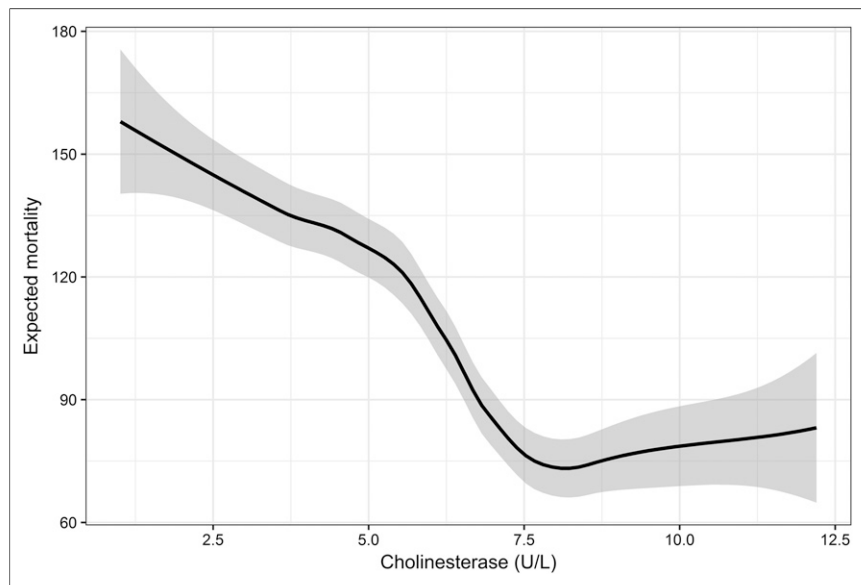


FIGURE 4. Partial dependency for cholinesterase: cholinesterase levels below approximately 7.5 U/L are associated with strong increase in expected mortality.

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 because of 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 2-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 nonlinear 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 pretherapeutic 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 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 thresholds.

Our random-survival-forest model predicts individual mortality as a continuous variable. To select patients who will benefit most from radioembolization, an optimal cutoff needs to be found. To choose such a threshold, one needs to balance overtreatment and risks of aggressive therapy against the benefits of life-prolonging therapy with an 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, one could then choose either the value where the log-rank statistic has

DISCLOSURE

No potential conflict of interest relevant to this article was reported.

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.* 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, Raefler F, et al. Robust evidence for long-term survival with ⁹⁰Y 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 chemorefractory metastatic, 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, Ingris 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. Hsieh 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. Ingris 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). The Comprehensive R Archive Network website. <https://cran.r-project.org/package=randomForestSRC>. Published October 17, 2017. Accessed February 22, 2018.
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.* 2016;278:563–577.
18. Kickingeder 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.* 2016;22:5765–5771.

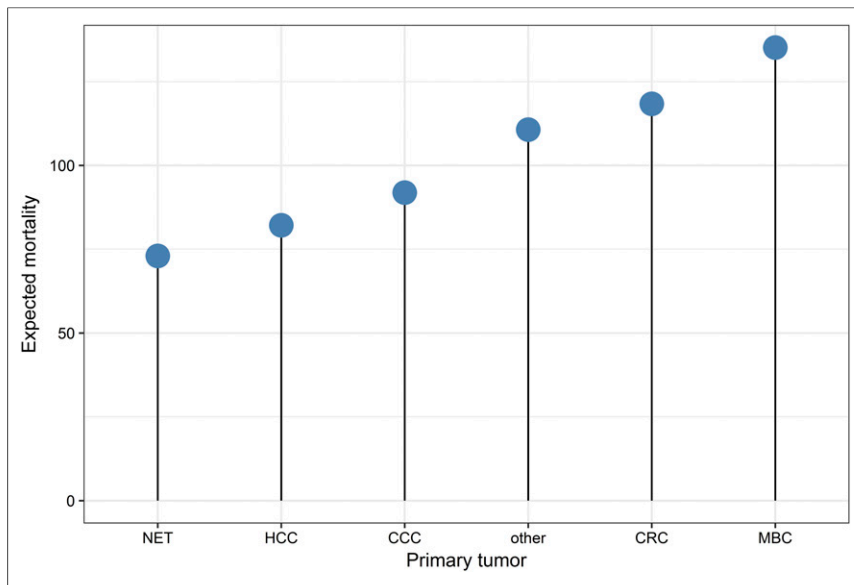


FIGURE 5. Expected mortality in dependence on tumor type. NET neuroendocrine tumor; HCC = hepatocellular carcinoma; CCC = cholangiocarcinoma; CRC = colorectal carcinoma; MBC = metastatic breast cancer.

the highest significance or the maximal predicted mortality that still results in a significant difference in overall survival, with the objective of including a large number of patients.

In our analysis, the hepatic tumor burden expressed in 3 categories (below 25%, 25%–50%, and above 50%) was moderately important. This measure was derived from pretherapeutic MRI. 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 pretherapeutic imaging to the prediction of mortality and stratification of patients is likely.

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. Second, 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.

CONCLUSION

We have used a modern statistical approach for prediction of overall survival after ⁹⁰Y radioembolization. The predictive performance of our model was similar to a previously published Cox proportional hazards model, and in addition, the model revealed a strong predictive value for baseline cholinesterase and bilirubin, with a highly nonlinear influence for each parameter.