

Clinical Research

# External Validation of PATHFx Version 3.0 in Patients Treated Surgically and Nonsurgically for Symptomatic Skeletal Metastases

Ashley B. Anderson MD, Rikard Wedin MD, PhD, Nicola Fabbri MD, Patrick Boland MD, John Healey MD, FACS, Jonathan A. Forsberg MD, PhD

Received: 14 August 2019 / Accepted: 19 November 2019 / Published online: 21 December 2019  
Copyright © 2019 by the Association of Bone and Joint Surgeons

## Abstract

**Background** PATHFx is a clinical decision-support tool based on machine learning capable of estimating the likelihood of survival after surgery for patients with skeletal metastases. The applicability of any machine-learning tool depends not only on successful external validation in unique patient populations but also on remaining relevant as more effective systemic treatments are introduced. With advancements in the treatment of metastatic disease, it is our responsibility to patients to ensure clinical support tools remain contemporary and accurate.

**Question/purposes** Therefore, we sought to (1) generate updated PATHFx models using recent data from patients treated at one large, urban tertiary referral center and (2) externally validate the models using two contemporary patient populations treated either surgically or nonsurgically with external-beam radiotherapy alone for symptomatic skeletal metastases for symptomatic lesions.

**Methods** After obtaining institutional review board approval, we collected data on 208 patients undergoing surgical treatment for pathologic fractures at Memorial Sloan

One author (JH) certifies that he is an assistant editor for *Clinical Orthopaedics and Related Research*®. Each author certifies that neither he or she, nor any member of his or her immediate family, has funding or commercial associations (consultancies, stock ownership, equity interest, patent/licensing arrangements, etc.) that might pose a conflict of interest in connection with the submitted article. Each author certifies that his or her institution approved the human protocol for this investigation and that all investigations were conducted in conformity with ethical principles of research.

The work was performed at Walter Reed National Military Medical Center, Bethesda, MD, USA; Johns Hopkins University, Baltimore, MD, USA; and Memorial Sloan Kettering Cancer Center, New York, NY, USA.

A. B. Anderson, J. A. Forsberg, Department of Surgery, Division of Orthopaedics, Walter Reed National Military Medical Center, Bethesda, MD, USA

A. B. Anderson, J. A. Forsberg, Uniformed Services University, Bethesda, MD, USA

R. Wedin, Department of Molecular Medicine and Surgery, K1, Orthopaedics, Karolinska Institutet, Stockholm, Sweden

P. Boland, N. Fabbri, J. Healey, Memorial Sloan Kettering Cancer Center, New York, NY, USA

J. A. Forsberg, Department of Orthopaedic Surgery, Johns Hopkins University, Baltimore, MD, USA

J. A. Forsberg (✉), Department of Orthopaedic Surgery, Johns Hopkins University, 601 North Caroline St., Baltimore MD 21205 USA, Email: jforsbe1@jhmi.edu

All ICMJE Conflict of Interest Forms for authors and *Clinical Orthopaedics and Related Research*® editors and board members are on file with the publication and can be viewed on request.

Kettering Cancer Center between 2015 and 2018. These data were combined with the original PATHFx training set ( $n = 189$ ) to create the final training set ( $n = 397$ ). We then created six Bayesian belief networks designed to estimate the likelihood of 1-month, 3-month, 6-month, 12-month, 18-month, and 24-month survival after treatment. Bayesian belief analysis is a statistical method that allows data-driven learning to arise from conditional probabilities by exploring relationships between variables to estimate the likelihood of an outcome using observed data. For external validation, we extracted the records of patients treated between 2016 and 2018 from the International Bone Metastasis Registry and records of patients treated nonoperatively with external-beam radiation therapy for symptomatic skeletal metastases from 2012 to 2016 using the Military Health System Data Repository (radiotherapy-only group). From each record, we collected the date of treatment, laboratory values at the time of treatment initiation, demographic data, details of diagnosis, and the date of death. All records reported sufficient follow-up to establish survival (yes/no) at 24-months after treatment. For external validation, we applied the data from each record to the new PATHFx models. We assessed calibration (calibration plots), accuracy (Brier score), discriminatory ability (area under the receiver operating characteristic curve [AUC]).

**Results** The updated PATHFx version 3.0 models successfully classified survival at each time interval in both external validation sets and demonstrated appropriate discriminatory ability and model calibration. The Bayesian models were reasonably calibrated to the Memorial Sloan Kettering Cancer Center training set. External validation with 197 records from the International Bone Metastasis Registry and 192 records from the Military Health System Data Repository for analysis found Brier scores that were all less than 0.20, with upper bounds of the 95% confidence intervals all less than 0.25, both for the radiotherapy-only and International Bone Metastasis Registry groups. Additionally, AUC estimates were all greater than 0.70, with lower bounds of the 95% CI all greater than 0.68, except for the 1-month radiotherapy-only group. To complete external validation, decision curve analysis demonstrated clinical utility. This means it was better to use the PATHFx models when compared to the default assumption that all or no patients would survive at all time periods except for the 1-month models. We believe the favorable Brier scores ( $< 0.20$ ) as well as DCA indicate these models are suitable for clinical use.

**Conclusions** We successfully updated PATHFx using contemporary data from patients undergoing either surgical or nonsurgical treatment for symptomatic skeletal metastases. These models have been incorporated for clinical use on PATHFx version 3.0 (<https://www.pathfx.org>). Clinically, external validation suggests it is better to use PATHFx version 3.0 for all time periods except when deciding whether to give radiotherapy to patients with the life

expectancy of less than 1 month. This is partly because most patients survived 1-month after treatment. With the advancement of medical technology in treatment and diagnosis for patients with metastatic bone disease, part of our fiduciary responsibility is to the main current clinical support tools.

*Level of Evidence* Level III, therapeutic study.

## Introduction

Accurate predictions of life expectancy in patients with skeletal metastases are important to help guide appropriate treatment selection. We previously introduced PATHFx, a clinical decision-making support tool for physicians treating patients with metastatic disease [3]. We then externally validated it in Scandinavian [3] and Italian patients [12]. The models were expanded to include 1-month and 6-month survival estimates [6] and were externally validated in Japanese patients [11]. Finally, 18-month and 24-month models were added to provide a more complete survival trajectory on which to base treatment decisions, which were externally validated in Australian patients [9].

The applicability of any machine-learning tool depends not only on successful external validation in unique patient populations but also on remaining relevant as more effective systemic treatments are introduced. Importantly, the original PATHFx models were created before targeted therapies, such as checkpoint inhibition and adoptive immunotherapy [10], became widely used in patients with metastatic bone disease. Over time, estimates derived by PATHFx may underestimate survival, particularly for those with non-small cell lung cancer [8] and other diagnoses responding to targeted therapies. To date, validation studies have only focused on patients who have undergone surgical fixation to treat impending or actual pathologic fractures. However, it remains unclear whether PATHFx may be useful in patients undergoing palliative therapy such as external-beam radiotherapy for symptomatic lesions.

Therefore, we sought to (1) generate updated PATHFx models using recent data from patients treated at one large, urban tertiary referral center and (2) externally validate the models using two contemporary patient populations treated either surgically or nonsurgically with external-beam radiotherapy alone for symptomatic skeletal metastases for symptomatic lesions.

## Patients and Methods

After obtaining institutional review board approval, we collected demographic, disease-related, and outcome information from 208 patients undergoing surgical treatment for pathologic fractures at Memorial Sloan Kettering Cancer

Center between 2015 and 2018. These data were combined with the original PATHFx training set ( $n = 189$ ) [3] to create the final training set ( $n = 397$ ) used for this study. We created six Bayesian belief networks designed to estimate the likelihood of 1-month, 3-month, 6-month, 12-month, 18-month, and 24-month survival after treatment using the `bnlearn` package in R© Version 3.5.1 (R Foundation for Statistical Computing, Vienna, Austria).

We extracted the records of patients treated between 2016 and 2018 from the International Bone Metastasis Registry. Launched in 2016, the International Bone Metastasis Registry was designed to store data from patients treated for pathologic fractures internationally. In addition, we extracted the records of patients treated nonoperatively with external-beam radiation therapy for symptomatic skeletal metastases from 2012 to 2016, using the Military Health System Data Repository (MDR), a database containing patient-level details on all healthcare encounters for military beneficiaries. From each record, we collected the date of treatment, laboratory values at the time of treatment initiation, demographic data, and details of diagnosis and the date of death. All records reported sufficient follow-up to establish survival (yes/no) at 24-months after treatment.

The International Bone Metastasis Registry contained 197 complete records. In the MDR, we identified 315 records of patients undergoing radiotherapy for symptomatic skeletal metastasis during the time period chosen for this study; however, 123 lacked follow-up information to determine survival at 24-months or were missing more than four PATHFx variables and were therefore excluded. This left 192 radiotherapy (RT)-only records for analysis. For each group, follow-up was sufficient to establish survival at 1 month, 3 months, 6 months, 12 months, 18 months, and 24 months after surgery or radiotherapy, as appropriate. Other recorded variables included age at the time of surgery, race, sex, primary oncologic diagnosis, whether the pathologic fracture was impending or complete, number of bone metastases (solitary or multiple), presence or absence of visceral (organ) metastases, presence or absence of lymph node metastases, preradiation hemoglobin level (g/dL), preradiation absolute lymphocyte count (K/mL), and radiation oncologist-documented Eastern Cooperative Oncology Group Performance Status. We then generated calibration curves that plotted the predicted risk against the actual risk to determine the accuracy of each model.

We evaluated differences between continuous variables using the Bayes factor t-test and Welch's t-test to compare means using the Bayes factor library in R© Version 3.5.1 (R Foundation for Statistical Computing, Vienna, Austria). Differences between categorical variables were assessed using a Bayes factor contingency table comparison (contingency TableBF in R©). Similar to any other clinical decision support tool, the development of PATHFx involves internal validation [3] and multiple external validation sets [5, 6, 9,

11, 12]. These previous studies provide evidence of how and in what way the distributions vary, which can then be used to determine whether there are meaningful statistical differences in the distributions contained in the two validation sets used for this study. When comparing variables in the training and validation sets, we also included the Welch's t-test, chi-square, and Fisher's exact test as comparisons to Bayesian testing, given their familiarity to most readers. Bayes factor analysis of the model variables demonstrated the demographic and clinical features of patients in the validation sets differed from those of patients in the training set. The Bayes factor is not a probability but a ratio of the probability of the alternative hypothesis (meaningful differences) to the null hypothesis (no meaningful differences). We observed the impact of Bayes factors on various prior probabilities to understand what is strong versus weak evidence for the clinical question or hypothesis.

As the size of the Bayes factor increases, so does the strength of the evidence, supporting our belief that there are meaningful differences between the distributions of features between the training set (Memorial Sloan Kettering Cancer Center) and the International Bone Metastasis Registry (Table 1) as well as the RT-only group (Table 2). The features with larger Bayes factors represent notable differences between the training set and the International Bone Metastasis Registry validation set, and they included the presence of visceral metastases, number of bone metastases, and survival longer than 3-months, 6-months, 12-months, and 18 months. On the other hand, differences in the RT-only validation set included sex (more male patients), presence of lymph node metastases, presence of visceral metastases, and number of bone metastases. In the International Bone Metastasis Registry validation sets, the variables with greater evidence against the null hypothesis, as indicated by larger Bayes factors, also had relatively low p values in corresponding Welch's t-test, chi-square, and/or Fisher's exact test (Table 1). This was also done in the RT-only validation set (Table 2).

Some features in the validation sets had a degree of missing data due to originating from large database sources. Notable features included lymph node metastases (missing in 100% of the International Bone Metastasis Registry validation set) and the surgeon's estimate of survival (missing in 100% of the RT-only validation set), all of which are important first-degree or second-degree predictors of survival in the PATHFx tool.

To perform external validation, we applied the data from each record, as is, to the new PATHFx models. We generated models containing the outcomes of 1-month, 3-month, 6-month, 12-month, 18-month, and 24-month survival (yes or no) and generated calibration curves that plotted the predicted risk against the actual risk to assess the accuracy of model predictions.

We determined the discriminatory ability of each model in estimating the likelihood of survival at each time interval

**Table 1.** Continuous variables contained in the training and validation sets

Feature	Training set		Validation set		Bayes factor		Welch's t-test (p value)	
	MSKCC (2003, 2018)		IBMR		RT			
	Value	% missing	Value	% missing	Value	% missing	IBMR vs train	RT vs train
Hemoglobin concentration (mg/dL)	11	< 1	11	0	12	0	0.16	0.30
Median	10 to 13		10 to 13		10 to 13		12.94	0.00
IQR								
Surgeon's estimate of survival (months)	9	4	6	38			1.33	0.01
Median	6 to 18		4 to 12					
IQR								

IBMR = International Bone Metastases Registry; MSKCC = Memorial Sloan Kettering Cancer Center; RT = radiotherapy-only group; IQR = interquartile range; train = training set.

by sensitivity, specificity, and receiver operating characteristic (ROC) curve analysis [2]. A minimum ROC area under the curve (AUC) of 0.7 was considered an acceptable predictive value [3, 5, 11]. We chose this threshold because we considered it to be the lowest acceptable limit of discriminatory ability [3]. We assessed the accuracy of each model by using the Brier score.

The concept of decision curve analysis is to determine clinical utility as expressed as the surgeon's estimation of survival in this study [13]. Thresholds vary by surgeon, patient, and clinical presentation; however, surgeons usually have a low threshold for treatment when treating healthy patients with straightforward problems and higher thresholds with sicker patients undergoing more complex care [3,13]. For example, in the context of this study, the decision curve analysis represents the surgeon's estimate of probability of a patient surviving to a certain time point would have to be met or exceeded for the surgeon to select particular treatment options, such as radiotherapy-alone versus doing nothing or surgical intervention versus doing nothing.

Finally, we included, where possible, information recommended in the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis [1].

**Results**

**Generating Updated PATHFx Models**

Our updated PATHFx version 3.0 models successfully classified survival at each time interval in both external validation sets and demonstrated appropriate discriminatory ability and model calibration.

**Externally Validating PATHFx v3.0**

The Bayesian models were reasonably calibrated to the Memorial Sloan Kettering Cancer Center training set (Fig. 1A-F). The models overestimate actual survival over the lower range probabilities and underestimate survival over the higher ranges (Fig. 2A-F). This is a subjective determination limited to the training data. Though methods exist to quantify the degree of calibration along a continuum, there is no consensus regarding what is, and is not, calibrated. However, we believe the favorable Brier scores and decision curve analysis indicate that these models are suitable for clinical use. The Brier scores were all less than 0.20 with upper bound of the 95% CIs all less than 0.25 for both the radiotherapy-only and International Bone Metastasis Registry groups. Additionally, AUC estimates were all greater than 0.70 with lower bounds of the 95% CI all greater than 0.68, except for the 1-month radiotherapy-only group

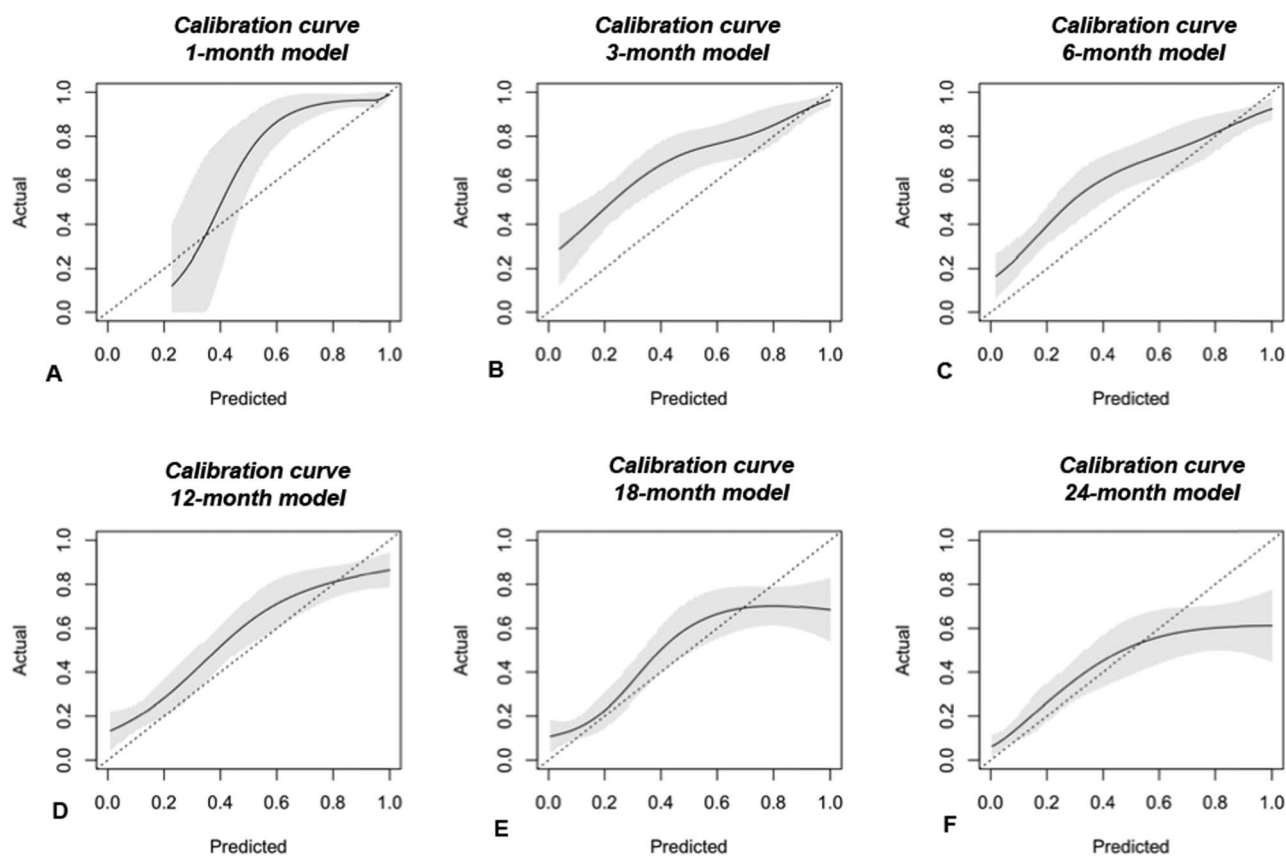
**Table 2.** Categorical variables in the training and validation sets

Feature	Training set			Validation set			Bayes factor		Pearson's chi-squared test (p value)		Fisher's test (p value)				
	MSKCC (2003, 2018)			IBMR			RT		IBMR vs train	RT vs train	IBMR vs train	RT vs train			
	(n = 397)			(n = 197)			(n = 192)								
	Number	%	% missing	Number	%	% missing	Number	%	% missing						
Sex															
Male	170	43	0	91	46	0	123	64	0	0.26	24235.27	0.49	< 0.00	0.48	< 0.00
Female	227	57		106	54		69	36							
Diagnosis group															
1	108	27	2	72	37	0	38	20	1	1.23	0.40	0.04	0.11	0.03	0.10
2	72	18		40	20		43	22							
3	211	53		85	43		110	57							
Visceral metastases															
Yes	247	62	< 1	61	31	3	111	58	4	11141562136.00	0.22	< 0.00	0.62	< 0.00	0.59
No	148	37		131	67		74	39							
Lymph node metastases															
Yes	152	38	0	0	0	100	111	58	16		340223034.00		<0.00		<0.00
No	245	62		0	0		51	27							
Number of bone metastases															
Solitary	112	28	0	42	21	8	26	14	0	0.39	581.04	0.23	< 0.00	0.22	< 0.00
Multiple	285	72		140	71		166	87							
ECOG performance status															
0, 1, or 2	222	56	3	112	57	0	117	61	0	0.20	0.26	0.95	0.49	0.93	0.47
3 or 4	164	41		85	45		75	39							
Survival > 1 month															
Yes	379	96	0	180	91	0	186	97	0	0.63	0.10	0.07	0.57	0.06	0.50
No	18	5		17	7		6	3							
Survival > 3 months															
Yes	309	78	0	126	64	0	157	82	0	93.56	0.28	< 0.00	0.32	< 0.00	0.28
No	88	22		71	36		35	18							
Survival > 6 months															
Yes	248	63	0	95	48	0	127	66	0	43.69	0.27	0.00	0.44	0.00	0.41

Table 2. continued

Feature	Training set			Validation set						Bayes factor		Pearson's chi-squared test (p value)		Fisher's test (p value)	
	MSKCC (2003, 2018)			IBMR			RT			IBMR vs train	RT vs train	IBMR vs train	RT vs train	IBMR vs train	RT vs train
	(n = 397)			(n = 197)			(n = 192)								
	Number	%	% missing	Number	%	% missing	Number	%	% missing						
No	149	38		102	52		65	34							
Survival > 12 months															
Yes	189	48	0	68	35	0	87	45	0	19.33	0.22	0.00	0.66	0.00	0.66
No	208	52		129	66		105	55							
Survival > 18 months															
Yes	134	34	0	46	23	0	63	33	0	5.36	0.19	0.01	0.89	0.01	0.85
No	263	66		151	77		129	67							
Survival > 24 months															
Yes	105	26	0	37	19	0	55	29	0	1.39	0.20	0.05	0.64	0.04	0.62
No	292	74		160	81		137	71							

MSKCC = Memorial Sloan Kettering Cancer Center; IBMR = International Bone Metastases Registry; RT = radiotherapy-only group; Train = training set; Diagnosis Group 1 = Lung, Hepatocellular, Gastric Carcinoma, and Melanoma; Diagnosis Group 2 = Sarcoma and Other Carcinoma not in Groups 1 or 3; Diagnosis Group 3 = Breast, Prostate, Thyroid, Renal Cell Carcinoma, Myeloma, and Lymphoma



**Fig. 1** A-F The calibration curves of validation set data show the agreement between observed outcomes and those predicted by the (A) 1-month, (B) 3-month, (C) 6-month, (D) 12-month, (E) 18-month, and (F) 24-month PATHFx models. The shaded region depicts the 95% CI of the predictions. Perfect calibration to the training data overlies the 45° dotted line. Each model is reasonably well calibrated to the Memorial Sloan Kettering Cancer Center training data.

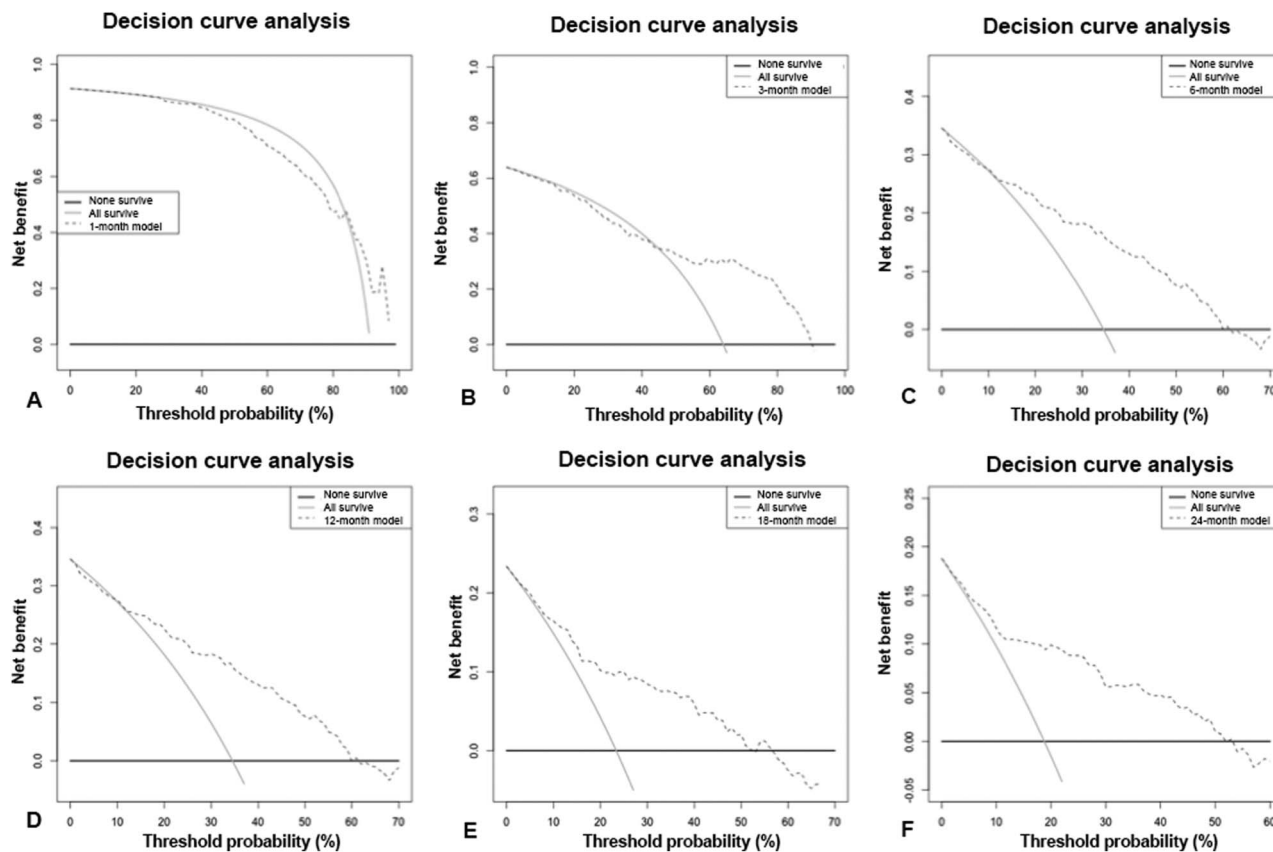
(Table 3). The decision curve analysis indicated that models possessed clinical utility in patients undergoing surgery and those not undergoing surgery for the International Bone Metastasis Registry group (Fig. 2A-F) and the RT-only group.

In this study, the models could be used rather than assume all or none of the patients with skeletal metastatic disease treated with surgical fixation will survive greater than 3-months, 6-months, 12-months, 18 months, and 24-months (Fig. 2A-F). However, for the RT-only group, the models could be used when making decisions about patients who survive longer than 12-months, 18 months, or 24-months (Fig. 3A-F). Not only are decision curves helpful in telling us when to use a model, but also they can tell physicians when support tools such as PATHFx should not be used. For example, for patients treated with RT alone, a clinician may achieve better outcomes by assuming all patients will survive longer than 1-month, 3-months, or 6-months unless their threshold for treatment exceeds 95%. This is the most evident for the 1-month

model when prescribing RT alone (Fig. 3A-F). This is also seen in the surgically treated patients (Fig. 2A-F). We found it is better to assume patients will survive 1-month than to rely solely on the PATHFx model, partly because more than 90% of patients in all of the sets tested in this study (Memorial Sloan Kettering Cancer Center, International Bone Metastasis Registry, and RT-only) survived for longer than 1-month.

## Discussion

With advancement in medical and surgical therapies, it is necessary to risk stratify patients with metastatic bone disease. Accurate estimates of patient survival are important because they help determine treatment choices. Poor decision-making impacts patient quality of life and resource utilization. PATHFx is a freely available clinical support tool (<https://www.pathfx.org>) that provides objective survival estimations for patients who undergo



**Fig. 2 A-F** The decision curve analysis (dashed line) of the predictive models based on International Bone Metastasis Registry data at (A) 1-month, (B) 3-months, (C) 6-months, (D) 12-months, (E) 18 months, and (F) 24 months indicates that all the models should be used rather than assuming all patients (continuous line) or none of the patients (thick continuous line) will survive longer than the period of each predictive model.

surgical fixation for skeletal metastases [4]. It helps guide surgical decision-making and avoids undertreatment or overtreatment of the disease [5]. Once a PATHFx model is developed, it undergoes rigorous evaluation for statistical accuracy and clinical utility. The introduction of a clinical support tool is only the beginning of the model lifecycle. For a tool to remain clinically relevant, it requires regular updating. To establish appropriate model lifecycle management, we successfully linked PATHFx to a large international registry, the International Bone Metastasis Registry. This linkage will allow for regular model improvements as new data become available. Finally, we successfully applied PATHFx to patients undergoing palliative radiotherapy for symptomatic metastatic bone disease.

When evaluating the results of this study, its limitations must be considered. Given the data used in this study, it is possible that other statistical techniques could be used to develop similar or superior prognostic models. Exploring other machine-learning techniques such as gradient

boosting algorithms and decision-tree and random forest models are slated for future projects. This may be indicated in future studies, given that the models appeared to be miscalibrated, slightly underestimating actual short-term survival over the lower range of probability estimates and overestimating long-term survival over the higher ranges. However, we have extensive experience using Bayesian inference for model development and application. Overfitting can occur with Bayesian belief networks, which would cause the results of the decision curve analysis to be overly optimistic. We sought to mitigate overfitting by including a larger training set than has been previously used for this application, in addition to two unique external validation sets. The techniques used to establish discriminatory ability and net benefit based on external validation have been used multiple times by our study group in the validation and external validation of PATHFx [3] and multiple external validation sets [5, 6, 11, 12]. However, these techniques are the minimum statistical requirements for a clinical decision-making support tool. Meares et al.



**Table 3.** Summary of the accuracy (AUC) and discriminatory ability (Brier Score) of the predictive model at each time period

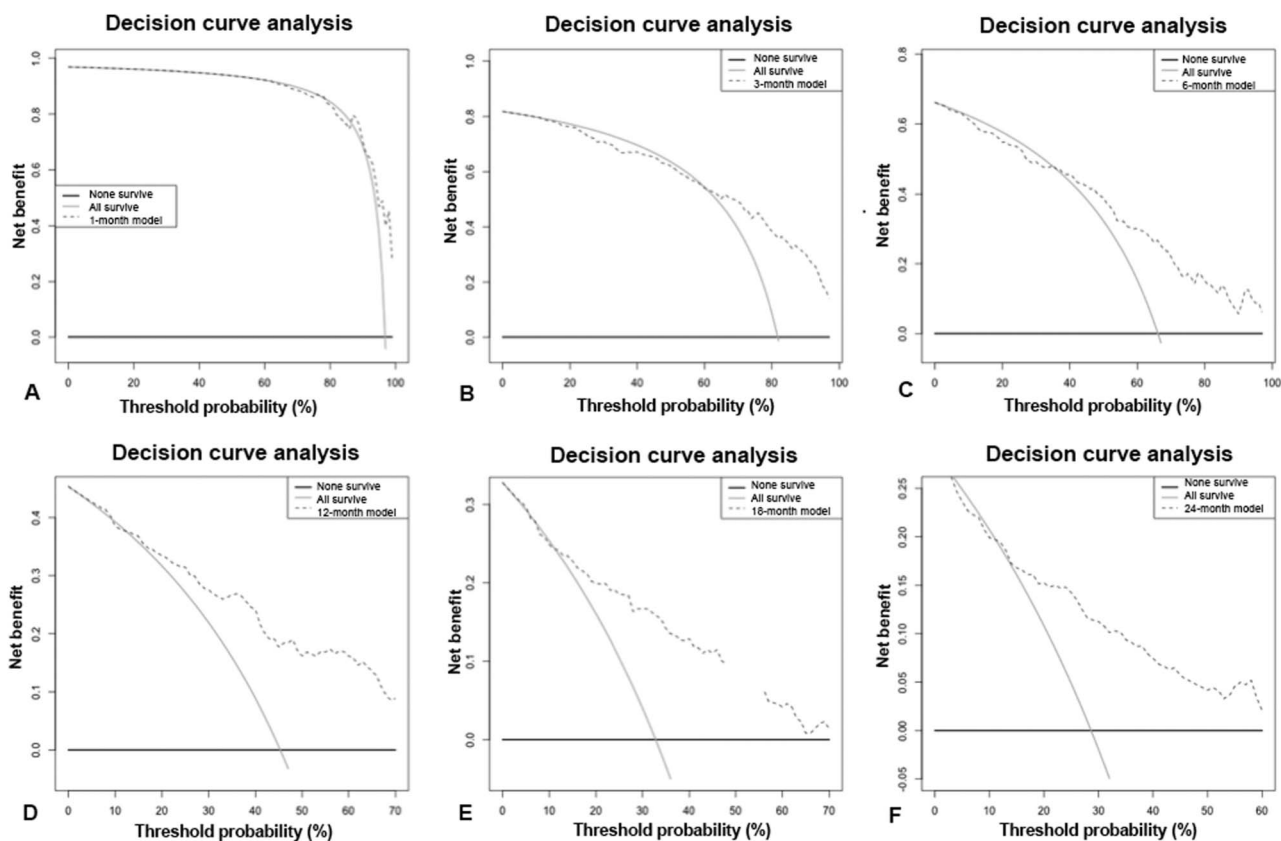
Predictive model	IBMR		Radiotherapy-only	
	AUC	Brier Score	AUC	Brier Score
1-month	0.70 (95% CI 0.58 to 0.82)	0.10 (95% CI 0.07 to 0.12)	0.82 (95% CI 0.68 to 0.95)	0.08 (95% CI 0.06 to 0.11)
3-month	0.77 (95% CI 0.70 to 0.84)	0.20 (95% CI 0.16 to 0.23)	0.83 (95% CI 0.77 to 0.90)	0.14 (95% CI 0.11 to 0.17)
6-month	0.77 (95% CI 0.70 to 0.83)	0.20 (95% CI 0.17 to 0.24)	0.79 (95% CI 0.73 to 0.86)	0.20 (95% CI 0.16 to 0.24)
12-month	0.78 (95% CI 0.71 to 0.85)	0.18 (95% CI 0.15 to 0.22)	0.79 (95% CI 0.73 to 0.86)	0.20 (95% CI 0.16 to 0.24)
18-month	0.79 (95% CI 0.71 to 0.86)	0.15 (95% CI 0.12 to 0.18)	0.79 (95% CI 0.72 to 0.86)	0.19 (95% CI 0.15 to 0.23)
24-month	0.82 (95% CI 0.75 to 0.90)	0.12 (95% CI 0.09 to 0.15)	0.76 (95% CI 0.69 to 0.84)	0.18 (95% CI 0.14 to 0.22)

AUC = area under the receiver operating characteristic curve; IBMR = International Bone Metastases Registry.

[9] compared PATHFx with six other published means of estimating survival in patients with skeletal metastases and found PATHFx to be most accurate for estimating 3-month and 6-month survival in patients with femoral metastatic bone disease. The same group found PATHFx to be the most consistent tool, providing accurate estimates of survival at all study time periods [9].

Another limitation is that the data in the training set were derived from a large tertiary referral center, which could limit universal application to community centers. To mitigate this and ensure the models remain applicable to the broadest international and community patient populations, we retained the original Memorial Sloan Kettering Cancer Center training data collected before the widespread use of immunotherapy and targeted therapy. The diverse nature of the training data may help explain the appearance of the calibration curves. Future models may need to acknowledge treatment details such as the use of immunotherapy or targeted therapy, including tyrosine kinase inhibitors, and whether the disease responds to initial treatment. Moreover, the International Bone Metastasis Registry is an international registry that was established in 2016. The validation and linkage of PATHFx to the International Bone Metastasis Registry may address the question of universal utility over time as the registry continues to grow. The RT-only group is from a highly selective, military tertiary center and may not represent other populations, specifically civilian populations. Further external validation studies in additional RT-only patient populations (such as those in the International Bone Metastasis Registry) is necessary before the PATHFx models may be used in additional settings. Both validation sets had model features with varying degrees of missing data. However, the models retained discriminatory ability with AUCs above the priori cutoff of 0.70 despite missing data. Previous validation sets [5, 6, 11, 12] all had missing input data. This current validation in the presence of missing data once again highlights an advantage of Bayesian belief networks as a machine learning method. This form of modeling uniquely performs well in the presence of missing data.

Additionally, Bayesian hypothesis testing allows investigators to incorporate prior knowledge when comparing groups of variables obtained from experimental observations. We chose to apply a Bayes factor analysis because conventional parametric and nonparametric tests tend to overstate the evidence against the null hypothesis by ignoring prior evidence and assuming that the null hypothesis is true. In contrast, the Bayes factor demonstrates whether and how beliefs derived from prior knowledge of this extensively studied patient population are altered by new data obtained in this study [7]. New information from validation sets allows for a better understanding of the distributions of variables, enabling one to extend the lifecycle of PATHFx through continued model improvement [7]. Similar to any



**Fig. 3** A-F The decision curve analysis (dashed line) of the predictive models based on RT data at (A) 1 month, (B) 3 months, (C) 6 months, (D) 12 months, (E) 18 months, and (F) 24 months indicates that all the models should be used rather than assuming all patients (continuous line) or none of the patients (thick continuous line) will survive longer than the period of each predictive model, except for the 1-month model where it is better to assume all survive than use the model; RT = radiotherapy-only group.

other clinical decision support tool, the development of PATHFx involves internal validation [3] and multiple external validation sets [5, 6, 11, 12]. However, the differences between each external validation set and the Memorial Sloan Kettering Cancer Center training set indicate that PATHFx continues to have room for improvement as a clinical support tool. As machine-learning methods continue to advance, the objective is to identify a modeling method that can be applied to diverse patient populations such as those chosen for this study.

On external validation, the discriminatory ability of the PATHFx models for the International Bone Metastasis Registry dataset and the RT-only dataset was similar those of previous operative populations studied [3, 5, 6, 12]. Low Brier scores and favorable decision curve analysis results indicate that all models could be used for patients undergoing radiotherapy; however, the 1-month model should be used with caution and interpreted within the context of the other five models. For the RT-only population, with the decision curve analysis, we can derive clinical relevance and provide evidence that when deciding to treat a patient with radiotherapy, it is better to assume a patient will live 1-month than

to use PATHFx. With improved survival in patients with skeletal metastases and a heightened awareness of skeleton-related events guiding referrals, this study poses the question as to whether a 1-month model could be defined by other machine-learning methods or whether a 1-month model is clinically necessary. Nevertheless, the probability estimates for the 1-month PATHFx models are displayed in conjunction with the other timepoints, which allows the user to interpret them in the context of each patient’s survival trajectory. Despite this, we successfully showed the ability of the PATHFx Bayesian models to function in patients treated with radiotherapy for skeletal metastatic disease.

In summary, we found the data support upgrading the PATHFx web site, [www.pathfx.org](http://www.pathfx.org), with version 3.0 of the PATHFx models, to improve the lives of patients with metastatic bone disease. The study highlights our commitment to improving PATHFx over time as new therapies continue to extend the lives of patients with metastatic bone disease. Clinicians might use PATHFx version 3.0 to provide survival estimates for patients receiving palliative external-beam radiotherapy alone and those receiving surgical treatment for symptomatic

metastatic bone disease, except for 1-month survival in the radiotherapy-only group. Although PATHFx version 3.0 groups oncologic diagnoses together, new models are being developed to allow for patient-centered and disease-specific (for example, breast, lung, and prostate cancer) decision-making. Further studies continue to explore the application of the models to other patient populations and patients undergoing nonoperative or other palliative treatments, such as cryotherapy or radiofrequency ablation of metastatic lesions. With continued advancements in metastatic disease care, it is our fiduciary duty to maintain up-to-date clinical support tools to help patients and other providers navigate these complex treatment algorithms. As machine-learning methods advance, the objective is to identify a modeling method that can be applied to diverse patient populations such as those chosen for this study.

**Acknowledgments** We thank Clare F. Grazal MS, for her contributions to data analysis and Benjamin K. Potter MD, FACS, for arranging helpful institutional support.

## References

- Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): The TRIPOD Statement. *Ann Intern Med.* 2015;162:55.
- Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation.* 2007;115:928-935.
- Forsberg J, Eberhardt J, Boland PJ, Wedin R, Healey JH. Estimating survival in patients with operable skeletal metastases: An application of a Bayesian belief network. *PLoS ONE.* 2011;6:e19956-8.
- Forsberg J, Sjöberg D, Chen Q-R, Vickers A, Healey JH. Treating metastatic disease: Which survival model is best suited for the clinic? *Clin Orthop Relat Res.* 2013;471:843-850.
- Forsberg J, Wedin R, Bauer HC, Hansen BH, Laitinen M, Trovik CS, Keller JØ, Boland PJ, Healey JH. External validation of the Bayesian Estimated Tools for Survival (BETS) models in patients with surgically treated skeletal metastases. *BMC Cancer.* 2012;12:493.
- Forsberg JA, Wedin R, Boland PJ, Healey JH. Can we estimate short-and intermediate-term survival in patients undergoing surgery for metastatic bone disease. *Clin Orthop Relat Res.* 2017; 475:1252-1261.
- Goodman SN. Toward evidence-based medical statistics. 2: The Bayes factor. *Ann Intern Med.* 1999;130:1005.
- Horn L, Mansfield AS, Szczęśna A, Havel L, Krzakowski M, Hochmair MJ, Huemer F, Lośonczy G, Johnson ML, Nishio M, Reck M, Mok T, Lam S, Shames DS, Liu J, Ding B, Lopez-Chavez A, Kabbinar F, Lin W, Sandler A, Liu SV; IMpower133 Study Group. First-line atezolizumab plus chemotherapy in extensive-stage small-cell lung cancer. *New Engl J Med.* 2018;6;379:220-2229.
- Meares D, Badran D, Dewar D. Prediction of survival after surgical management of femoral metastatic bone disease – a comparison of prognostic models. *J Bone Oncol.* 2019;15:100225.
- Newick K, Moon E, Albelda SM. Chimeric antigen receptor T-cell therapy for solid tumors. *Mol Ther Oncolytics.* 2016;3: 16006.
- Ogura K, Gokita T, Shinoda Y, Kawano H, Takagi T, Ae K, Kawai A, Wedin R, Forsberg JA. Can a multivariate model for survival estimation in skeletal metastases (PATHFx) be externally validated using Japanese patients? *Clin Orthop Relat Res.* 2017;475:2263-2270.
- Piccioli A, Spinelli SM, Forsberg JA, Wedin R, Healey JH, Ippolito V, Daolio PA, Ruggieri P, Maccauro G, Gasbarrini A, Biagini R, Piana R, Fazioli F, Luzzati A, Di Martino A, Nicolosi F, Camnasio F, Rosa MA, Campanacci DA, Denaro V, Capanna R. How do we estimate survival? External validation of a tool for survival estimation in patients with metastatic bone disease—decision analysis and comparison of three international patient populations. *BMC Cancer.* 2015;15:424.
- Vickers AJ, Elkin EB. Decision curve analysis: A novel method for evaluating prediction models. *Med Decis Making.* 2006;26: 565-574.