Surgical decision-making for long bone metastases

Janssen, S.J.

Publication date
2018
Document Version
Other version
License
Other

CHAPTER 10

Prognostication In Patients With Long Bone Metastases: Does A Boosting Algorithm Improve Survival Estimates?

S.J. Janssen
A.S. van der Heijden
M. van Dijke
J.E. Ready
K.A. Raskin
M.L. Ferrone
F.J. Hornicek
J.H. Schwab

Clinical Orthopaedics and Related Research 2015 Oct;473(10):3112-21.

Award:
Marshall Urist Young Investigator Award 2015

Presented at:
Association of Bone and Joint Surgeons Annual Meeting 2015, Eugene, Oregon, USA.
Musculoskeletal Tumor Society Annual Meeting 2015, Orlando, Florida, USA.

Poster at:
Massachusetts General Hospital Clinical Research Day 2014, Boston, Massachusetts, USA.

ABSTRACT

Objectives

To assess factors independently associated with decreased survival in patients with a fracture through a long bone metastasis and (2) compare the accuracy of a classic scoring system, nomogram, and boosting algorithms in predicting 30-, 90-, and 365-day survival.

Design

Retrospective cohort study.

Setting

Two tertiary care referral centers for orthopaedic oncology.

Participants

927 consecutive patients who underwent surgery for a pathological or impending fracture through a long bone metastasis between 1999 and 2013.

Outcome Measures

Overall survival.

Results

The following factors were associated with a decreased likelihood of survival after surgical treatment of a long bone metastasis, after controlling for relevant confounding variables: older age (hazard ratio [HR], 1.0; 95% CI, 1.0–1.0; $p < 0.001$), additional comorbidity (HR, 1.2; 95% CI, 1.0–1.4; $p = 0.034$), BMI less than 18.5 kg/m^2 (HR, 2.0; 95% CI, 1.2–3.5; $p = 0.011$), tumor type with poor prognosis (HR, 1.8; 95% CI, 1.6–2.2; $p < 0.001$), multiple bone metastases (HR, 1.3; 95% CI, 1.1–1.6; $p = 0.008$), visceral metastases (HR, 1.6; 95% CI, 1.4–1.9; $p < 0.001$), and lower hemoglobin level (HR, 0.91; 95% CI, 0.87–0.96; $p < 0.001$). The survival estimates by the nomogram were moderately accurate for predicting 30-day (area under the curve [AUC], 0.72), 90-day (AUC, 0.75), and 365-day (AUC, 0.73) survival and remained stable after correcting for optimism through fivefold cross validation. Boosting algorithms were better predictors of survival on the training datasets, but decreased to a performance level comparable to the nomogram when applied on testing datasets for 30-day (AUC, 0.69), 90-day (AUC, 0.75), and 365-day (AUC, 0.72) survival prediction. Performance of the classic scoring system was lowest for all prediction periods.

Conclusions

Comorbidity status and BMI are newly identified factors associated with decreased survival and should be taken into account when estimating survival. Performance of the

boosting algorithms and nomogram were comparable on the testing datasets. However, the nomogram is easier to apply and therefore more useful to aid surgical decision making in clinical practice.

INTRODUCTION

Estimated survival is an important factor in the decision to operate and surgical strategy in patients with bone metastasis.[1,2] Physicians often estimate survival based on their clinical assessments and previously described risk factors. Several tools in the form of scoring systems have been developed to assist clinicians with their estimation.[1-5] However, these tools lack accuracy and identification of additional and more-specific risk factors might improve survival estimation.[2,6]

Historically, tools like scoring systems that are used to provide survival probability are based on a summary score of weighted clinical or laboratory factors.[1,2,5] The Bauer score is such a classic scoring system commonly used for estimation of survival in patients with bone metastases.[1] It is a summary score of five prognostic factors: (1) no pathological fracture, (2) no visceral or brain metastases, (3) a solitary bone metastasis, (4) no lung cancer, and (5) multiple myeloma, lymphoma, breast or kidney carcinoma. Fulfilling four to five criteria corresponded to a 1-year survival probability of 0.5, two to three criteria to a 1-year survival probability of 0.25, and all patients who fulfilled none or only one criterion were deceased within 6 months after surgery.[1] Another frequently used tool to estimate survival in patients with cancer is the nomogram, which is a simple figure that generates an individualized numerical probability of survival based on a patient's unique set of characteristics; a number of points is assigned to each prognostic factor, which can be read from the nomogram and the sum of these points corresponds to a survival probability.[7-9] The nomogram can be seen as an extension of the classic scoring system. Advances in computer science has led to the development of more sophisticated boosting (machine learning) algorithms.[10,11] Machine learning is a method of automatically developing and constantly adjusting computer algorithms to recognize patterns in data and improve predictions (e.g. filtering spam email).[11] Boosted regression emerged from this field and is a method that iteratively applies classifiers (variables) in a sequential way –each step building on the previous step aiming to fit the residuals– and subsequently combines them to obtain predictions.[10,11] This can improve the accuracy of predicting an outcome based on weak learners (i.e. classifiers that are only slightly better than random guessing). A boosting algorithm provides outcome probabilities based on every possible combination of variables.[10]

We aimed to assess factors associated with survival in patients with long bone fracture through a bone metastasis. Based on those factors, we created a classic scoring algorithm,

a nomogram, and boosting algorithms to estimate survival. Specifically, we sought to (1) assess factors independently associated with decreased survival in patients with long bone metastasis, and (2) compare the accuracy of a classic scoring system, nomogram, and boosting algorithms in predicting 30-, 90-, and 365-day survival.

METHODS

Study Design

This retrospective study was approved by our institutional review board and a waiver of informed consent was granted. To identify patients with a long bone fracture through a bone metastasis, we retrieved all medical record data of patients who had an International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) code for a long bone fracture through a bone metastasis or a Current Procedural Terminology (CPT) code for prophylactic fixation of a long bone fracture at two tertiary care referral centers for orthopaedic oncology (Appendix 1).

After operative report and medical record screening, we included all 927 patients older than 18 years who had surgery for a pathological or impending long bone fracture between January 1999 and December 2013. We included only the first surgery per patient if patients underwent multiple operations on different occasions so as to not violate the statistical assumption of independence.[12] We defined long bones as the femur, humerus, tibia, fibula, radius, and ulna; multiple long bones operated on during the same procedure were categorized separately. Metastatic disease included, in addition to metastases from solid organs, multiple myeloma and lymphoma. We included patients regardless of followup duration. Exclusion criteria were (1) revision procedures, (2) metastatic involvement of the acetabulum or pelvis requiring reconstruction, and (3) surgical treatments other than endoprosthetic reconstruction, intramedullary nailing, open reduction and internal fixation with plate and screws or a dynamic hip screw.

The decision to operate and the selection of surgical strategy were made by the surgeon together with the patient. These decisions were based on factors including type of fracture, primary tumor type, extent of the metastatic lesion, level of disability and pain, and the estimated survival.

Outcome Measures

Our primary outcome was survival, defined as the time from surgical treatment until death resulting from any cause. Date of death was extracted from the medical records and Social Security Death Index database (database of death records created from the US Social Security Administration), last updated February 24, 2014.[13] This date also provided the moment of final followup for all patients alive in our study. Median followup was 8

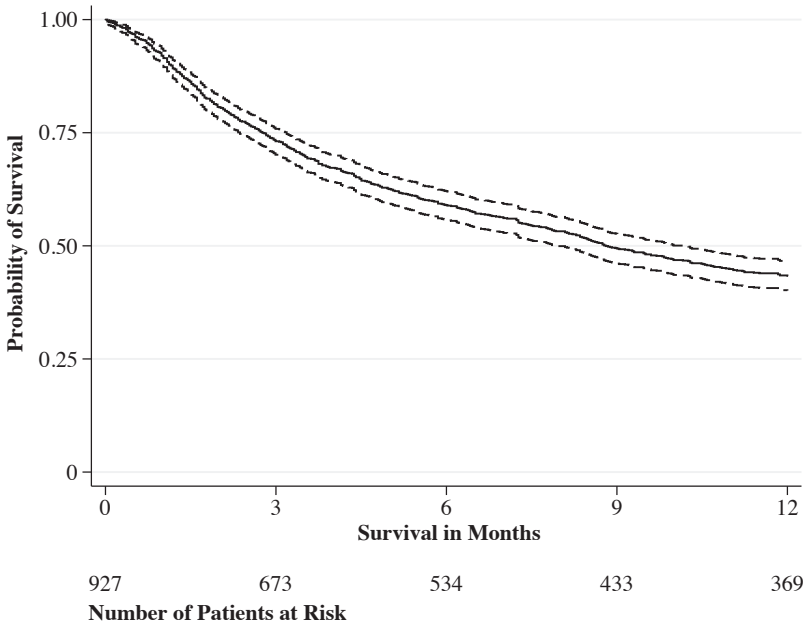

Figure 1: The Kaplan-Meier plot shows the probability of survival (solid line) with 95% CI (dashed line). The median survival is 268 days (95% CI, 241-309), with an interquartile range from 84 (95% CI, 72-97) to 1089 days (95% CI, 922-1262).

months (interquartile range, 3 to 25 months). All patients who were alive at 30 days were still in followup (n = 853), 673 of 678 patients (99%) who were alive at 3 months were still in followup, and 369 of 412 (90%) patients who were alive at 1 year were still in followup (Figure 1).

We selected the following explanatory variables based on the existing studies or theoretical association with survival: age, sex, BMI, comorbidity status, primary tumor type, type of fracture, anatomic location of fracture, time from diagnosis of the primary tumor until surgical treatment, other bone metastases, other previous pathological fractures, visceral metastases, previous systemic therapy, previous local radiation therapy of the affected long bone, and preoperative hemoglobin level, platelet level, white blood cell count, creatinine, and serum calcium level.[2,5,14]

We categorized BMI into: less than 18.5 kg/m² (underweight), between 18.5 and 30 kg/m² (normal weight), and 30 kg/m² or greater (obese) based on previously published cutoff points as we expect a nonlinear association of BMI with survival.[15,16]

We used the modified Charlson Comorbidity Index to indicate comorbidity status.[17,18] This index provides a score ranging from 0 to 24 with a higher score representing more severe comorbidity status based on 12 weighted comorbidities (congestive heart failure, dementia, chronic pulmonary disease, rheumatologic disease, mild liver disease, diabetes

with chronic complications, hemiplegia or paraplegia, renal disease, any malignancy, moderate or severe liver disease, metastatic solid tumor, and HIV/AIDS). We determined the modified Charlson Comorbidity Index through a previously described algorithm based on ICD-9-CM codes given before the day of surgery (Appendix 2).[19-21] The modified Charlson Comorbidity Index was dichotomized into any additional comorbidity (additional to the malignancy and metastatic disease) or none.

Based on a study by Katagiri et al.[5], we dichotomized primary tumor types into tumors with a relatively good prognosis (breast, kidney, prostate, thyroid, myeloma, and lymphoma) and tumors with a poor prognosis (lung and all other tumor types).

Fracture type was classified as pathological or impending. The latter was defined as bone with no visible fracture line, loss of height, rotation, or angulation, but the degree of destruction did mandate, in the surgeon's opinion, surgical treatment. Previous pathological fractures or prophylactically treated impending fractures were categorized into none, previous long bone fracture, and previous spine fracture (with or without previous long bone fracture).

We extracted the presence of bone metastases from bone scan, CT, and other imaging reports. Bone metastases were categorized into single bone metastasis, multiple bone metastases without spinal involvement, and multiple bone metastases with spinal involvement. The presence of visceral metastases was derived from CT and positron emission tomography scan reports. We regarded lung, liver, and brain metastases as visceral metastases and grouped lung and/or liver metastases together; brain metastases (with or without lung/liver metastases) were categorized separately.

We used laboratory values measured within 7 days before surgical treatment.

Statistical Analysis

Variables are presented with frequencies and percentages for categorical variables and as mean with SD for continuous variables. In bivariate analyses, the association between the response variable survival and the explanatory variables was assessed using Cox regression analysis (Appendix 3). The proportional hazards assumption was tested using Schoenfeld residuals and verified by assessing if Kaplan-Meier survival curves crossed. Our exploratory analysis identified the following variables: age ($p < 0.001$), BMI less than 18.5 kg/m^2 ($p < 0.001$), additional comorbidity ($p < 0.001$), multiple long bones surgically treated during same procedure ($p = 0.077$), poor prognosis tumor type (lung and all other tumor types) ($p < 0.001$), multiple bone metastases without spinal involvement ($p = 0.096$) and with spinal involvement ($p = 0.014$), lung and/or liver metastasis ($p < 0.001$) and brain metastasis ($p < 0.001$), previous systemic therapy ($p = 0.057$), hemoglobin level ($p = 0.001$), and platelet level ($p = 0.004$), which then were incorporated in our multivariable model (Appendix 3). We recategorized bone metastases as single and multiple metastases as exploratory analysis showed no difference in hazard ratios (HR) between patients with

multiple bone metastases without spinal involvement (HR, 1.2; 95% CI, 1.0–1.5) and those with spinal involvement (HR, 1.3; 95% CI, 1.0–1.5; $p = 0.67$). Visceral metastases also was recategorized as none and any visceral metastases as there was no difference in hazard ratios between lung/liver metastases (HR, 1.9; 95% CI, 1.6–2.2) and brain metastases (HR, 1.8; 95% CI, 1.5–2.2; $p = 0.82$) (Appendix 3).

We entered these explanatory variables with a p value less than 0.10 on bivariate analysis in a backward stepwise multivariable Cox regression analysis to assess the independent association with survival.[22,23] We did not test interactions of variables. Hazard ratios and beta regression coefficients are presented to quantify the association of explanatory variables with survival. The HR indicates the relative likelihood of death in one group compared with another group. Hazard ratios are adjusted for all explanatory variables included in the multivariable Cox regression analysis. We assume missing values –BMI (21%, 197 of 927 patients) and hemoglobin (6%, 59 of 927 patients)– to be at random and used multiple imputation to replace missing values 40 times based on the remaining explanatory variables.[24]

All statistical analyses were performed using Stata® 13.0 (StataCorp LP, College Station, TX, USA). A two-tailed p value less than 0.05 was considered statistically significant.

Development Of A Classic Scoring System, Nomogram, And Boosting Algorithms

We developed a classic scoring system to estimate survival probability by assigning a weighted score to every factor independently associated with survival by rounding its HR to the nearest integer.[1,5,14] To allow for scoring of continuous variables, we dichotomized age (65 years and older) and hemoglobin level (10 g/dL and less) and rounded the HRs of the mean difference between the dichotomized groups to the nearest integer (mean difference in age, 20 years [HR, 1.34] and mean difference in hemoglobin level 2.6 g/dL [HR, 1.27]).[2] The total score of the classic scoring system ranges from 0 to 10 (Table 1). We categorized scores as: good prognosis, (0–2 points), intermediate prognosis (3–5 points), and poor prognosis (6–10 points) based on the survival probability curves (Figure 2). Survival probability was demonstrated per prognostic group for each prediction period (30, 90, and 365 days; Table 2).[1,5]

We developed a nomogram by ranking the effect estimates (β regression coefficients) of all factors independently associated with survival to a scale ranging from 0 to 100 points.[8,9] The predicted probability of 30-, 90-, and, 365-day survival were calculated for each patient using the multivariable Cox regression model underlying the nomogram.[8,25]

Boosting algorithms to predict 30-, 90-, and 365-day survival were developed using the "multiple additive regression trees" gradient boosting technique implemented in Stata® 13.0.[10,11] Boosting is a machine learning technique that produces a prediction algorithm based on additive decision trees to classify outcome (30-, 90-, and 365-day survival) in a

Table 1: Classic scoring system

Variable	Points
Age 65 years or older	1
Additional comorbidity	1
Body mass index below 18.5 kg/m^2	2
Tumor type other than: breast, kidney, prostate, thyroid, myeloma, and lymphoma	2
Multiple Bone metastases	1
Visceral metastases*	2
Hemoglobin level 10 or below (g/dL)	1

*Visceral metastases include lung, liver, and/or brain metastases.

Table 2: Classic scoring system: survival probability per prognostic group and prediction period (n = 927)*

Prognostic groups	Probability of survival (95% confidence interval)
Good prognosis (score 0-2)	
30 days	0.98 (0.96-0.99)
90 days	0.93 (0.89-0.96)
365 days	0.66 (0.60-0.73)
Intermediate prognosis (score 3-5)	
30 days	0.92 (0.90-0.94)
90 days	0.73 (0.69-0.76)
365 days	0.39 (0.34-0.43)
Poor prognosis (score 6-10)	
30 days	0.84 (0.79-0.89)
90 days	0.49 (0.41-0.56)
365 days	0.17 (0.11-0.23)

*Missing values for body mass index and hemoglobin level were imputed using multiple imputation.

stepwise fashion.[10,11] We included the same set of factors independently associated with survival for development of the boosting algorithms. Each boosting algorithm allowed for two-way interactions. The algorithm provides an estimated survival probability for every possible combination (permutation) of the included variables for each prediction period, thereby taking into account the interaction of the included variables.[10]

All three prediction models were compared using fivefold cross validation on the 40 multiple imputed datasets, meaning that the models were created five times on randomly selected training subsets (80%) of the data and tested on the remaining 20%.[26,27] The average performance (the ability of a model to separate patients with different outcomes) was calculated over the five training and testing repetitions per multiple imputed dataset for all three models and prediction periods and subsequently pooled. Performance was

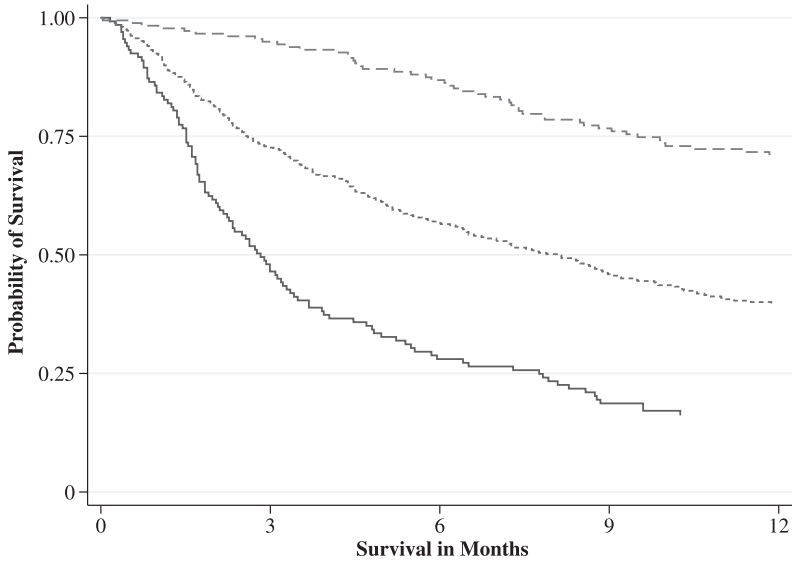

Figure 2: The Kaplan-Meier plot shows the probability of survival per prognostic group of the classic scoring algorithm: good prognosis (0–2; dashed line), intermediate prognosis (3–5; dotted line), and poor prognosis (6–10; solid line).

assessed using receiver operating characteristic (ROC) curves.[28-30] ROC curves are made by plotting the rate of false positives (1 – specificity) on the x-axis and the rate of true positives (sensitivity) on the y-axis for all threshold values. The area under the ROC curve (AUC) represents its discriminatory power; an AUC of 1.0 indicates perfect discrimination (100% sensitivity and 100% specificity), whereas an AUC of 0.50 represents no discriminatory power (a coin toss).

The final classic scoring system, nomogram, and boosting algorithms were developed on the 40 multiple imputed datasets and results were pooled.[26,27]

RESULTS

Baseline Characteristics

Among the 927 patients, 401 (43%) were men, and the mean age of the patients was 62 years (Table 3). There were 515 (56%) pathological fractures and 412 (44%) impending fractures. The femur (70%; 646 of 927 patients) and humerus (23%; 210 of 927 patients) were most commonly affected. Most tumors originated from the breast (23%; 216 of 927 patients), lung (23%; 215 of 927 patients), myeloma (16%; 148 of 927 patients), kidney (9%; 87 of 927 patients), and prostate (5%; 48 of 927 patients) (Table 4). Median overall survival was 9 months (Figure 1). Ninety-two percent of the patients survived 30 days

Table 3: Baseline characteristics

	n = 927
Demographics	**Mean (± SD)**
Age (years)	62 (± 13)
Body mass index (kg/m^2)*	27 (± 5.8)
Modified Charlson Comorbidity Index	6.7 (± 2.0)
	n (%)
Men	401 (43)
Pathological fracture	515 (56)
Anatomical location	
Femur	646 (70)
Humerus	210 (23)
Tibia	31 (3.3)
Radius	3 (0.32)
Ulna	2 (0.22)
Multiple locations§	35 (3.8)
Oncological status	**n (%)**
Bone metastases	
Single bone metastasis	202 (22)
Multiple bone metastases without spinal involvement	185 (20)
Multiple bone metastases with spinal involvement	540 (58)
Previous pathological fractures	
None	662 (71)
Previous long bone (impending) pathological fracture	79 (8.5)
Previous spine (impending) pathological fracture†	186 (20)
Visceral metastases	
None	504 (54)
Lung and/or liver metastases	273 (29)
Brain metastases (with or without lung/liver metastases)	150 (16)
Previous systemic therapy	577 (62)
Previous local radiotherapy of the affected long -bone	170 (18)
Laboratory values‡	**Mean (± SD)**
Hemoglobin (g/dL)	11 (± 1.6)
Platelets (1000/mm^3)	254 (± 116)
White blood cell count (1000/mm^3)	9.7 (± 4.9)
Creatinine (mg/dL)	0.94 (± 0.81)
Calcium (mg/dL)	8.8 (± 0.93)

*Body mass index was available in 730 cases; §Multiple metastatic fracture locations undergoing fixation during the same procedure were: bilateral femur (12 patients), femur and humerus (16 patients), bilateral humerus (2), tibia and femur (1), tibia and humerus (1), fibula and tibia (1), radius and humerus (1), ulna and radius (1); †with or without previous long bone pathological fracture; ‡hemoglobin level was available in 868 cases, platelet level in 866 cases, white blood cell count in 867 cases, creatinine in 812 cases, and calcium in 654 cases.

Table 4: Origin of primary tumors

Tumor distribution	n (%)
Breast	216 (23)
Lung	215 (23)
Myeloma	148 (16)
Kidney	87 (9)
Prostate	48 (5)
Lymphoma	43 (5)
Melanoma	25 (3)
Esophagus	18 (2)
Colorectal	16 (2)
Thyroid	15 (2)
Hepatocellular	12 (1)
Bladder	10 (1)
Other*	35 (4)
Unknown	39 (4)

*Neuroendocrine (n = 6), salivary gland (n = 5), nasopharyngeal squamous cell carcinoma (n = 5), pancreas (n = 4), ovaries (n = 4), endometrium (n = 3), skin squamous cell carcinoma (n = 3), stomach (n = 2), vulva (n = 2), and mesothelioma (n = 1).

(853 of 927 patients), 73% (676 of 922 patients) survived 90 days, and 42% (368 of 884 patients) survived 365 days. The median time from diagnosis of the primary tumor until surgical treatment of the fracture was 21 months. Five-hundred sixty (60%) patients underwent intramedullary nailing, 209 (23%) had endoprosthetic reconstruction, 140 (15%) had open reduction and internal fixation with plate and screws or a dynamic hip screw in 18 (2%).

Explanatory Variables Associated With Survival

The following factors were associated with a decreased likelihood of survival after surgical treatment of a long bone metastasis, after controlling for relevant confounding variables: older age (HR, 1.0; 95% CI, 1.0–1.0; $p < 0.001$), additional comorbidity (HR, 1.2; 95% CI, 1.0–1.4; $p = 0.034$), BMI less than 18.5 kg/m^2 (HR, 2.0; 95% CI, 1.2–3.5; $p = 0.011$), tumor type with poor prognosis (HR, 1.8; 95% CI, 1.6–2.2; $p < 0.001$), multiple bone metastases (HR, 1.3; 95% CI, 1.1–1.6; $p = 0.008$), visceral metastases (HR, 1.6; 95% CI, 1.4–1.9; $p < 0.001$), and lower hemoglobin level (HR, 0.91; 95% CI, 0.87–0.96; $p < 0.001$) (Table 5).

Table 5: Hazard ratios for survival from stepwise backward multivariable Cox regression analysis[§]

Explanatory variables	β regression coefficient	Standard error	Hazard ratio (95% confidence interval)	*p* value
Age (in years)	0.015	0.003	1.015 (1.008, 1.021)	***< 0.001***
Additional comorbidity	0.164	0.077	1.179 (1.013, 1.372)	***0.034***
Body mass index				
Below 18.5 kg/m^2[*]	0.707	0.276	2.027 (1.175, 3.480)	***0.011***
Between 18.5 and 30 kg/m^2[*]	Reference	Reference	Reference	*Reference*
Above 30 kg/m^2[*]	0.045	0.106	1.046 (0.549, 1.288)	*0.672*
Tumor type other than: breast, kidney, prostate, thyroid, myeloma, and lymphoma	0.609	0.080	1.839 (1.574, 2.150)	***< 0.001***
Multiple Bone metastases	0.254	0.095	1.290 (1.069, 1.555)	***0.008***
Visceral metastases	0.500	0.079	1.649 (1.412, 1.926)	***< 0.001***
Hemoglobin level (g/dL)[*]	-0.094	0.025	0.911 (0.867, 0.956)	***< 0.001***

bold font indicates a significant difference (two-tailed *p* value below 0.05); [*]body mass index was available in 730 cases; and hemoglobin level in 868 cases; missing values were imputed using multiple imputation. [§]Variables were selected using stepwise backward multivariable Cox regression analysis retaining variables with a *p* value below 0.10.

Table 6: Area under the curve (AUC) for the classic scoring algorithm and boosting algorithm from receiver operating characteristic analysis

Prediction period	Classic Scoring System, AUC (95% confidence interval)	Nomogram, AUC (95% confidence interval)	Boosting Algorithm, AUC (95% confidence interval)	*p* value
Training subsets				
30 days	0.66 (0.60-0.72)	0.72 (0.66-0.78)	0.83 (0.78-0.88)	***< 0.001***
90 days	0.70 (0.66-0.73)	0.76 (0.72-0.80)	0.81 (0.78-0.85)	***< 0.001***
365 days	0.68 (0.65-0.72)	0.73 (0.70-0.77)	0.78 (0.75-0.81)	***< 0.001***
Testing subsets				
30 days	0.67 (0.55-0.78)	0.72 (0.59-0.84)	0.69 (0.55-0.83)	***0.328***
90 days	0.70 (0.62-0.77)	0.75 (0.68-0.83)	0.75 (0.67-0.83)	***0.075***
365 days	0.68 (0.61-0.75)	0.73 (0.65-0.80)	0.72 (0.65-0.80)	***0.125***

bold font indicates a significant difference (two-tailed *p* value below 0.05)

Comparing Performance Of The Classic Scoring, Nomogram, And Boosting Algorithms

The survival estimates by the nomogram were moderately accurate for predicting 30-day (AUC, 0.72), 90-day (AUC, 0.75), and 365-day (AUC, 0.73) survival and remained stable after correcting for optimism through fivefold cross validation (Table 6). Boosting algorithms were better predictors of survival at all prediction periods on the training datasets; however, after applying these to the testing datasets we found that accuracy of the boosting algorithms decreased substantially for the 30-day (AUC, 0.83 to 0.69), 90-day (AUC, 0.81 to 0.75), and 365-day (AUC, 0.78 to 0.72) prediction periods resulting in a performance comparable to the that of the nomogram (Table 6). Performance of the classic scoring system was lowest for all prediction periods.

DISCUSSION

Expected survival is an important factor in the decision to operate and in the selection of a surgical strategy for patients with long bone metastasis and a fracture.[1,2] Previous studies developed scoring algorithms to estimate survival,[1-5,14] but survival estimates remain imprecise.[2] In an attempt to improve survival estimation, we assessed which clinical factors and laboratory values were independent predictors of survival. Additionally, we developed a classic scoring system, nomogram, and boosting algorithms to estimate 30-, 90-, and 365-day survival and compared the accuracy of these methods. We found that older age, additional comorbidity, BMI less than 18.5 kg/m², primary tumor type with poor prognosis, multiple bone metastases, visceral metastases, and lower hemoglobin level, were independently associated with decreased likelihood of survival. The survival estimates by the boosting algorithm were most accurate on the training datasets, but comparable to those derived from the nomogram when applied to the testing datasets for 30-, 90-, and 365-day survival. We emphasize the use of the nomogram (Figure 3) for estimating survival as it is simpler to use in clinical practice.

This study has some limitations. First, there were no uniform criteria for surgical treatment because the study was retrospective. This might have resulted in selection bias and potentially influenced accuracy of the prediction models; for example, patients with a very poor overall health status might not have been considered for surgical treatment. Although this might limit the usefulness of the algorithms in these patients, we believe that this did not compromise the comparison of performance of the algorithms in our study. Second, we used diagnostic and billing codes to identify potentially eligible patients. We might have missed patients using this methodology; however, we expect this number to be low and therefore not influence our conclusions. Third, although we internally validated the algorithms through fivefold cross validation, predictive performance can worsen substantially

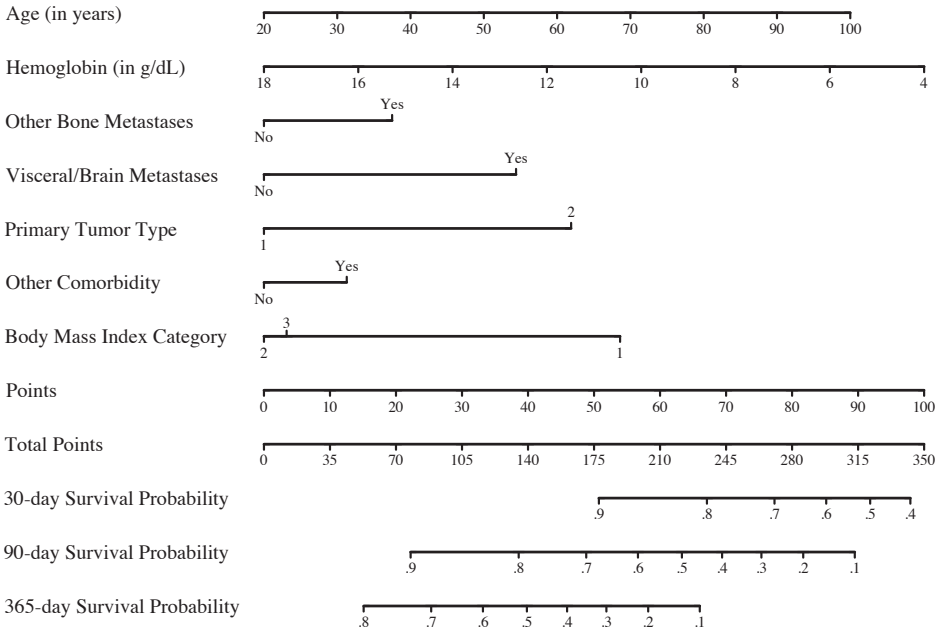

Figure 3: The nomogram for prediction of 30-, 90-, and 365-day survival is shown. Locate the patients age on the age axis and draw a straight line to the points axis. Repeat this process for all variables and sum the points obtained for each predictor. Locate the total points on the total points axis and draw a line straight down to find the 30-, 90-, and 365-day survival probabilities. BMI categories are: (1) less than 18.5 kg/m², (2) between 18.5 and 30 kg/m², and (3) 30 kg/m² or greater (obese). The primary tumor Group 1 includes breast, kidney, prostate, thyroid, myeloma, and lymphoma; Group 2 includes lung and all other primary tumor types. This nomogram is not applicable to a patient who otherwise is not a candidate for surgical treatment of a long bone metastasis. The outcome is a point estimate and the nomogram does not include the uncertainty of the estimate.

on external validation. External validation should be performed before widespread use of a prediction algorithm.[31,32] Fourth, we did not assess how discriminant the predicted probabilities by the different models were. We see this as a minor limitation and emphasized performance of the models as better performance (higher AUC) implies less uncertainty –more precision– of the points estimate of the predicted probability. Fifth, because the study was retrospective, we could not include performance status of the patient. Including this might have improved the predictive accuracy of our algorithms as previous studies showed a strong association with survival.[2,5] Sixth, we decided to select variables for inclusion in the boosting algorithms based on theory. Including all explanatory variables using a kitchen-sink approach (having the algorithm select variables) could improve its accuracy. However, this might have resulted in a large number of factors to consider when estimating survival, making it less useful in clinical practice, and potentially worsening its external validity. We aimed to compare the performance of prediction models based on the

same set of variables supported by theory. Seventh, we did not define minimum followup. We see this as a minor limitation as we used Cox regression analysis to account for right censoring (loss to followup) and followup was 90% for patients alive at 1 year.

Previous studies in patients with fractures through bone metastasis found that visceral metastases, primary tumor type, number of bone metastases, time from diagnosis of primary tumor to surgery, fracture type, performance status, previous chemotherapy, and hemoglobin level were independently associated with survival.[1,2,5,14,33] We identified additional factors associated with survival in patients with long bone metastasis, namely, comorbidity status and BMI. The association of comorbidity status with overall and cancer-specific survival has been shown in patients with primary malignancies[34-36] but not in patients with fracture through a long bone metastasis. BMI can be considered a surrogate marker of cancer severity because low body weight often is associated with more advanced cancer. Previous studies showed an association of BMI with survival in patients with cancer.[37,38] Furthermore, we divided multiple bone metastases into those with and those without spinal involvement and visceral metastases into lung and/or liver metastases and brain metastases to explore differences in their association with survival. However, we found no differences between these groups in terms of survival nor did we find an association of previous pathological fractures with survival. The difference in survival between impending and pathological fractures, found by Bauer and Wedin,[1] was not found in our study. Based on our findings, future studies should explore how specific comorbidities influence survival in patients with fractures and if optimizing perioperative nutritional status improves survival in these patients.

Although the boosting algorithm was most accurate in estimating survival on the training samples, its performance decreased when applied to the testing subsets of the data. This might be a result of overfitting of the boosting algorithm on the training data. Performance of the boosting algorithm was comparable to that of the nomogram when applied to the testing subsets. We therefore prefer using the nomogram in estimating survival as it is simpler to use in daily practice. However, the nomogram does not make treatment recommendations, it simply provides estimated survival probabilities and can enable a more informed decision-making process. Thirty-, 90-, and 365-day survival probabilities are based on the sum score of the points assigned to the prognostic factors of an individual patient (Figure 3). For example, a 77-year old patient with breast carcinoma, normal BMI, multiple bone metastases, but no visceral metastases, diabetes with chronic complications (additional comorbidity), and a preoperative hemoglobin of 9.4 g/dL gets assigned 157 points which corresponds to a 30-day survival probability of 0.93, 90-day survival probability of 0.74, and 365-day survival probability of 0.42. The mean total points in our cohort was 156 points (SD, 39). Forsberg et al.[3,4] described the development and testing of machine learning algorithms in determining survival of patients with long bone metastasis. The algorithms they developed, based on a prospective cohort of 189 patients, had an

AUC of 0.85 for 3-month survival and 0.83 for 1-year survival, indicating a higher accuracy compared with our algorithms.[3] The prospective collection of their data might have been more accurate and less prone to bias, resulting in higher accuracy. Future studies should externally validate survival prognostication models and assess which algorithm is most accurate in predicting survival for patients with extremity metastases.

Comorbidity status and BMI are two additional factors associated with survival and should be taken into account when estimating survival. These factors should be incorporated in survival prediction models. The nomogram remained most accurate in predicting survival after correcting for optimism and could be used on paper; however, its estimates will be more precise when implemented in an application. The nomogram could be made available on interfaces convenient in clinical practice (such as smartphone applications) to aid surgical decision making. We are working on external validation of the developed algorithms and aim to develop an online tool to estimate survival for use in clinical practice.

REFERENCES

1. Bauer HC, Wedin R. Survival after surgery for spinal and extremity metastases. Prognostication in 241 patients. Acta Orthop Scand. Apr 1995;66(2):143-146.
2. Nathan SS, Healey JH, Mellano D, et al. Survival in patients operated on for pathologic fracture: implications for end-of-life orthopedic care. J Clin Oncol. Sep 1 2005;23(25):6072-6082.
3. Forsberg JA, 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(5):e19956.
4. Forsberg JA, Wedin R, Bauer HC, et al. External validation of the Bayesian Estimated Tools for Survival (BETS) models in patients with surgically treated skeletal metastases. BMC Cancer. 2012;12:493.
5. Katagiri H, Takahashi M, Wakai K, Sugiura H, Kataoka T, Nakanishi K. Prognostic factors and a scoring system for patients with skeletal metastasis. J Bone Joint Surg Br. May 2005;87(5): 698-703.
6. Henderson R, Keiding N. Individual survival time prediction using statistical models. J Med Ethics. Dec 2005;31(12):703-706.
7. Oberije C, De Ruysscher D, Houben R, et al. A Validated Prediction Model for Overall Survival From Stage III Non-Small Cell Lung Cancer: Toward Survival Prediction for Individual Patients. Int J Radiat Oncol Biol Phys. Apr 30 2015.
8. Iasonos A, Schrag D, Raj GV, Panageas KS. How to build and interpret a nomogram for cancer prognosis. J Clin Oncol. Mar 10 2008;26(8):1364-1370.
9. Kattan MW, Eastham JA, Stapleton AM, Wheeler TM, Scardino PT. A preoperative nomogram for disease recurrence following radical prostatectomy for prostate cancer. J Natl Cancer Inst. May 20 1998;90(10):766-771.
10. Schonlau M. Boosted regression (boosting): An introductory tutorial and a Stata plugin. Stata Journal. 2005;5(3):330-354.

11. Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning. 2 ed. New York: Springer-Verlag New York; 2009.
12. Bryant D, Havey TC, Roberts R, Guyatt G. How many patients? How many limbs? Analysis of patients or limbs in the orthopaedic literature: a systematic review. J Bone Joint Surg Am. Jan 2006;88(1):41-45.
13. Huntington JT, Butterfield M, Fisher J, Torrent D, Bloomston M. The Social Security Death Index (SSDI) most accurately reflects true survival for older oncology patients. Am J Cancer Res. 2013;3(5):518-522.
14. Ratasvuori M, Wedin R, Keller J, et al. Insight opinion to surgically treated metastatic bone disease: Scandinavian Sarcoma Group Skeletal Metastasis Registry report of 1195 operated skeletal metastasis. Surg Oncol. Jun 2013;22(2):132-138.
15. US National Heart Lung and Blood Institute. The Practical Guide: Identification, Evaluation, and Treatment of Overweight and Obesity in Adults. Vol No. 00-4084: National Institutes of Health; 2000.
16. Goodwin PJ, Ennis M, Pritchard KI, Koo J, Trudeau ME, Hood N. Diet and breast cancer: evidence that extremes in diet are associated with poor survival. J Clin Oncol. Jul 1 2003; 21(13):2500-2507.
17. Quan H, Li B, Couris CM, et al. Updating and validating the Charlson comorbidity index and score for risk adjustment in hospital discharge abstracts using data from 6 countries. Am J Epidemiol. Mar 15 2011;173(6):676-682.
18. Poeran J, Rasul R, Suzuki S, et al. Tranexamic acid use and postoperative outcomes in patients undergoing total hip or knee arthroplasty in the United States: retrospective analysis of effectiveness and safety. BMJ. 2014;349:g4829.
19. Deyo RA, Cherkin DC, Ciol MA. Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. J Clin Epidemiol. Jun 1992;45(6):613-619.
20. Quan H, Parsons GA, Ghali WA. Validity of information on comorbidity derived rom ICD-9-CCM administrative data. Med Care. Aug 2002;40(8):675-685.
21. Voskuijl T, Hageman M, Ring D. Higher Charlson Comorbidity Index Scores are associated with readmission after orthopaedic surgery. Clin Orthop Relat Res. May 2014;472(5):1638-1644.
22. Hosmer D, Lemeshow S. Applied Survival Analysis: John Wiley & Sons, INC.; 1999.
23. Allison PD. Survival Analysis Using SAS: A Practical Guide. 2 ed2010.
24. Mackinnon A. The use and reporting of multiple imputation in medical research - a review. J Intern Med. Dec 2010;268(6):586-593.
25. Kim MS, Lee SY, Lee TR, et al. Prognostic nomogram for predicting the 5-year probability of developing metastasis after neo-adjuvant chemotherapy and definitive surgery for AJCC stage II extremity osteosarcoma. Ann Oncol. May 2009;20(5):955-960.
26. Harrell FE, Jr., Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med. Feb 28 1996;15(4):361-387.
27. Steyerberg EW, Harrell FE, Jr., Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. J Clin Epidemiol. Aug 2001;54(8):774-781.
28. Cleves MA. From the help desk: Comparing areas under receiver operating characteristic curves from two or more probit or logit models. Stata Journal. 2002;2(3):301-313.

29. Gupta S, Tran T, Luo W, et al. Machine-learning prediction of cancer survival: a retrospective study using electronic administrative records and a cancer registry. BMJ Open. 2014;4(3): e004007.
30. Pepe M, Longton G Fau - Janes H, Janes H. Estimation and Comparison of Receiver Operating Characteristic Curves. Stata Journal. 20100816 2009(1536-867X (Print)).
31. Siontis GC, Tzoulaki I, Castaldi PJ, Ioannidis JP. External validation of new risk prediction models is infrequent and reveals worse prognostic discrimination. J Clin Epidemiol. Jan 2015; 68(1):25-34.
32. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. BMJ. 2015;350:g7594.
33. Schneiderbauer MM, von Knoch M, Schleck CD, Harmsen WS, Sim FH, Scully SP. Patient survival after hip arthroplasty for metastatic disease of the hip. J Bone Joint Surg Am. Aug 2004;86-A(8):1684-1689.
34. Edwards BK, Noone AM, Mariotto AB, et al. Annual Report to the Nation on the status of cancer, 1975-2010, featuring prevalence of comorbidity and impact on survival among persons with lung, colorectal, breast, or prostate cancer. Cancer. May 1 2014;120(9):1290-1314.
35. Sogaard M, Thomsen RW, Bossen KS, Sorensen HT, Norgaard M. The impact of comorbidity on cancer survival: a review. Clin Epidemiol. 2013;5(Suppl 1):3-29.
36. Patnaik JL, Byers T, Diguiseppi C, Denberg TD, Dabelea D. The influence of comorbidities on overall survival among older women diagnosed with breast cancer. J Natl Cancer Inst. Jul 20 2011;103(14):1101-1111.
37. McTiernan A, Irwin M, Vongruenigen V. Weight, physical activity, diet, and prognosis in breast and gynecologic cancers. J Clin Oncol. Sep 10 2010;28(26):4074-4080.
38. Jung SH, Yang DH, Ahn JS, et al. Decreased body mass index is associated with poor prognosis in patients with multiple myeloma. Ann Hematol. May 2014;93(5):835-840.

Appendix 1: International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) codes and Current Procedural Terminology (CPT) codes for identification of patients with pathologic or impending fractures

Category	Code	Code description
ICD-9-CM codes	733.1	Pathologic fracture, unspecified site
	733.11	Pathologic fracture of humerus
	733.12	Pathologic fracture of radius and ulna
	733.14	Pathologic fracture of neck of femur
	733.15	Pathologic fracture of other specified part of femur
	733.16	Pathologic fracture of tibia or fibula
	733.19	Pathologic fracture of other specified site
CPT codes	23491	Prophylactic treatment (nailing, pinning, plating or wiring) with or without methylmethacrylate; proximal humerus
	24498	Prophylactic treatment (nailing, pinning, plating or wiring), with or without methylmethacrylate, humeral shaft
	25490	Prophylactic treatment (nailing, pinning, plating or wiring) with or without methylmethacrylate; radius
	25491	Prophylactic treatment (nailing, pinning, plating or wiring) with or without methylmethacrylate; ulna
	25492	Prophylactic treatment (nailing, pinning, plating or wiring) with or without methylmethacrylate; radius and ulna
	27187	Prophylactic treatment (nailing, pinning, plating or wiring) with or without methylmethacrylate, femoral neck and proximal femur
	27495	Prophylactic treatment (nailing, pinning, plating or wiring) with or without methylmethacrylate, femur
	27745	Prophylactic treatment (nailing, pinning, plating or wiring) with or without methylmethacrylate, tibia

Appendix 2: Modified Charlson Comorbidity Index Algorithm Based on International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) Codes

Comorbidity	Weight*	Codes
AIDS/HIV	4	042
Any malignancy, including leukemia and lymphoma*	2	150.0-159.0, 162-173.59, 173.70-175.9, 180.0-183.9, 185-186.9, 188.0-188.6, 188.8-189.4, 189.9, 191.0-192.3, 192.9-194.4, 200.2-202.38, 202.70-202.81, 203.0-204.22, 204.90-208.22, 208.90-209.36, 209.70, 209.72-209.79, 230.2-230.6, 230.8, 231.2, 231.9, 232.5-232.7, 233.0, 233.1, 233.31, 233.32, 233.4, 233.7, 235.2-235.4, 235.7, 235.8, 236.2, 236.4, 236.5, 236.7-236.91, 237.1-237.4, 237.6, 238.0-238.3, 238.79, 239.0-239.4, 239.6, 239.7, 239.89, 239.9
Chronic pulmonary disease	1	416.8, 416.9, 490-491.0, 491.2-495.2, 495.4-505, 506.4, 508.1, 508.8
Congestive heart failure	2	398.91, 402.01, 402.11, 402.91, 404.01, 404.03, 404.11, 404.13, 404.91, 404.93, 425.4-425.9, 428-428.43
Dementia	2	290, 290.0, 290.3, 290.8-290.43, 294.1, 294.11-294.21, 331.2
Diabetes with chronic complications	1	249.40-249.91, 250.40-250.90
Hemiplegia or paraplegia	2	342.00-342.92, 344.00-344.5, 344.89-344.9
Metastatic solid tumor*	6	197.0-198.7, 198.81-190.9, 192.0-196.9, 199.0
Mild liver disease*	2	070.22, 070.23, 070.32, 070.33, 070.44, 070.54, 070.6, 070.9, 570, 570.1, 573.3, 573.4, 573.8, 753.9, V42.7
Moderate or severe liver disease*	4	456.0-456.2, 572.2-572.8
Renal disease	1	403.01, 403.11, 403.91, 404.02, 404.03, 404.12, 404.13, 404.92, 404.93, 582-583.7, 585-586, 588.0, V42.0, V45.1, V56-V56.8
Rheumatologic disease	1	466.5, 710.0-710.4, 714.0-714.2, 714.8, 725

*The following comorbidities were mutually exclusive: mild liver disease and moderate or severe liver disease, and any malignancy and metastatic solid tumor. For example, a patient with a metastatic solid tumor received 6 points total (not 6 points for metastatic solid tumor and 2 points for any malignancy).

Appendix 3: Hazard ratios for survival from bivariate Cox regression analysis*

Explanatory variables	β regression coefficient	Standard error	Hazard ratio (95% CI)	*p value*
Age (years)	0.010	0.003	1.010 (1.005-1.016)	*< 0.001*
BMI				
Less than 18.5 kg/m²	1.004	0.225	2.730 (1.756-4.244)	*< 0.001*
Between 18.5 and 30 kg/m²	Reference	Reference		*Reference*
Greater than 30 kg/m²	-0.057	0.100	0.945 (0.777-1.149)	*0.570*
Additional comorbidity	0.304	0.075	1.356 (1.171-1.570)	*< 0.001*
Men	0.013	0.075	1.013 (0.874-1.174)	*0.861*
Pathological fracture	0.032	0.075	1.033 (0.892-1.196)	*0.665*
Multiple locations operated during same procedure	0.316	0.179	1.371 (0.966-1.946)	*0.077*
Race				
White	Reference	Reference		*Reference*
Other	-0.206	0.144	0.814 (0.613-1.079)	*0.152*
Tumor type other than breast, kidney, prostate, thyroid, myeloma, and lymphoma	0.623	0.075	1.864 (1.608-2.161)	*< 0.001*
Time between primary tumor diagnosis and surgery for metastatic lesion	0.000	0.000	1.000 (1.000-1.000)	*0.415*
Bone metastases				
Single bone metastasis	Reference	Reference		*Reference*
Multiple bone metastases without spinal involvement	0.193	0.116	1.213 (0.966-1.522)	*0.096*
Multiple bone metastases with spinal involvement	0.233	0.095	1.263 (1.047-1.522)	*0.014*
Previous pathological fractures				
None	Reference	Reference		*Reference*
Previous long bone (impending) pathological fracture	0.147	0.128	1.158 (0.902-1.487)	*0.250*
Previous spine (impending) pathological fracture†	0.021	0.094	1.021 (0.849-1.228)	*0.826*
Visceral metastases				
None	Reference	Reference		*Reference*
Lung and/or liver metastases	0.618	0.084	1.855 (1.573-2.187)	*< 0.001*
Brain metastases (with or without lung/liver metastases)	0.593	0.104	1.809 (1.475-2.219)	*< 0.001*
Previous local radiotherapy of the affected long bone	0.129	0.093	1.138 (0.948-1.367)	*0.166*
Previous systemic therapy	0.147	0.077	1.158 (0.996-1.347)	*0.057*
Hemoglobin (g/dL)	-0.082	0.024	0.921 (0.879-0.965)	*0.001*
Platelets (1000/mm³)	0.001	0.000	1.001 (1.000-1.002)	*0.004*
White blood cell count (1000/mm³)	0.011	0.008	1.011 (0.996-1.027)	*0.157*
Creatinine (mg/dL)	0.033	0.053	1.034 (0.932-1.147)	*0.532*
Calcium (mg/dL)	-0.005	0.048	0.995 (0.906-1.094)	*0.923*

Total number of patients = 927; *patients with missing values were excluded per bivariate analysis: BMI was available in 730 patients; hemoglobin level in 868, platelet level in 866, white blood cell count in 867, creatinine in 812, and calcium in 654; †with or without previous long bone pathological fracture; variables with a *p* value less than 0.10 in bivariate analysis were included in multivariable Cox regression analysis.