Published in final edited form as:

Int J Radiat Oncol Biol Phys. 2020 November 01; 108(3): 554–563. doi:10.1016/j.ijrobp.2020.05.023.

Developing an Improved Statistical Approach for Survival Estimation in Bone Metastases Management: The Bone Metastases Ensemble Trees for Survival (BMETS) Model

Sara R. Alcorn, MD, MPH, PhD^{*}, Jacob Fiksel, BS[†], Jean L. Wright, MD^{*}, Christen R. Elledge, MD^{*}, Thomas J. Smith, MD[‡], Powell Perng, MD^{*}, Sarah Saleemi, MD, MPH^{*}, Todd R. McNutt, PhD^{*}, Theodore L. DeWeese, MD^{*}, Scott Zeger, PhD[†]

^{*}Department of Radiation Oncology and Molecular Radiation Sciences, Johns Hopkins School of Medicine, Baltimore, MD

[†]Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD

[‡]Department of Oncology, Johns Hopkins School of Medicine, Baltimore, MD

Abstract

Purpose: To determine whether a machine learning approach optimizes survival estimation for patients with symptomatic bone metastases (SBM), we developed the Bone Metastases Ensemble Trees for Survival (BMETS) to predict survival using 27 prognostic covariates. To establish its relative clinical utility, we compared BMETS with 2 simpler Cox regression models used in this setting.

Methods and Materials: For 492 bone sites in 397 patients evaluated for palliative radiation therapy (RT) for SBM from January 2007 to January 2013, data for 27 clinical variables were collected. These covariates and the primary outcome of time from consultation to death were used to build BMETS using random survival forests. We then performed Cox regressions as per 2 validated models: Chow's 3-item (C-3) and Westhoff's 2-item (W-2) tools. Model performance was assessed using cross-validation procedures and measured by time-dependent area under the curve (tAUC) for all 3 models. For temporal validation, a separate data set comprised of 104 bone sites treated in 85 patients in 2018 was used to estimate tAUC from BMETS.

Results: Median survival was 6.4 months. Variable importance was greatest for performance status, blood cell counts, recent systemic therapy type, and receipt of concurrent nonbone palliative RT. tAUC at 3, 6, and 12 months was 0.83, 0.81, and 0.81, respectively, suggesting excellent discrimination of BMETS across postconsultation time points. BMETS outperformed simpler models at each time, with respective tAUC at each time of 0.78, 0.76, and 0.74 for the C-3 model and 0.80, 0.78, and 0.77 for the W-2 model. For the temporal validation set, respective tAUC was similarly high at 0.86, 0.82, and 0.78.

Corresponding author: Sara Alcorn, MD, MPH, PhD; salcorn2@jhmi.edu.

Data sharing: Research data are stored in an institutional repository and will be shared upon request to the corresponding author.

Supplementary material for this article can be found at <https://doi.org/10.1016/j.ijrobp.2020.05.023>.

Conclusions: For patients with SBM, BMETS improved survival predictions versus simpler traditional models. Model performance was maintained when applied to a temporal validation set. To facilitate clinical use, we developed a web platform for data entry and display of BMETS-predicted survival probabilities.

Introduction

In the management of symptomatic bone metastases (SBM), selection of palliative treatments including radiation therapy (RT), surgery, and systemic therapy depends on accurate estimation of life expectancy. However, providers are notoriously inaccurate at estimating survival—particularly at the end of life¹—which can result in the delivery of high-cost, low-value care and reduced quality of life.²

To address this, a number of prognostic models have been developed to guide clinical decision-making for patients treated with palliative RT. Table E1 summarizes prediction models frequently cited for use in patients treated with palliative RT across treatment sites and primary cancer types. Numerous other models offer predictions for specific subpopulations, such as those with spinal metastases.^{3–5} Most of these models use Cox proportional hazards methodology, with final models using up to 7 prognostic covariates. Despite the breadth of options, 1 survey of radiation oncologists found that only 31% rated such models as moderately or very important to their estimation of life expectancy.⁶ Potential reasons for underuse include the complex and time-consuming nature of these prognostic tools.

Addressing such limitations, 2 Cox proportional hazards models summarized in Table E1 compared the predictive capacity of full versus reduced sets of predictor variables. Chow et al compared survival predictions from a 6- versus 3-covariate model.⁷ Their 3-variable number of risk factors (NRF) model (C-3) was comprised of nonbreast cancer, presence of nonbone metastases, and Karnofsky Performance Status (KPS) score ≥ 60 . This reduced model yielded a concordance statistic (C-statistic) of 0.65, versus a C-statistic of 0.67 for the full 6-variable model. Similarly, Westhoff et al⁸ compared the discriminative capacities of a 6- versus 2-variable model. Their 2-variable model (W-2) was comprised of primary tumor site and KPS and yielded a C-statistic of 0.71, which was comparable to 0.72 for the full 6-variable model. In both cases, the authors concluded that the reduced models resulted in similar predictive capacity and should be used instead of the full models due to ease of clinical application.

Although these data offer compelling evidence that simpler models may be preferred when rendered from traditional statistical methods, newer machine learning approaches may offer a means to further optimize survival predictions using a larger number of covariates. Yet, no such machine learning model is currently available for clinical use in this setting. As such, we built the Bone Metastases Ensemble Trees for Survival (BMETS) model and web interface to provide survival estimates for patients with SBM using up to 27 prognostic variables. To establish its clinical utility relative to simpler, traditional statistical methods, we compared BMETS performance with that of C-2 and W-3. To address concerns about temporal changes in practice patterns such as rapid adoption of oral targeted agents in recent

years,⁹ we tested BMETS using a temporal validation set among patients treated for SBM in 2018.

Methods and Materials

Training set data, source, and study population

Patients seen in consultation for SBM in the Department of Radiation Oncology at the Johns Hopkins School of Medicine between March 1, 2007, and July 31, 2013, were identified through query of our departmental electronic medical records (EMR). The query was generated on the basis of age ≥ 18 years and ICD9/10 codes for bone site and/or treatments using ≥ 15 fractions. Owing to infrequent use of stereotactic body RT (SBRT) during the study period, patients seen in consultation for this approach were excluded from the query. The Johns Hopkins University Institutional Review Board approved this work (IRB00125143), with a waiver of informed consent.

The query yielded 404 patients. We limited analysis to patients with pathologically or radiologically confirmed metastatic cancer with SBM. To minimize the statistical implications of multiple treatments within the same patient, only data from the first palliative treatment consultation within the study period were included.

Training set patient, disease, and treatment characteristics

The EMR was retrospectively assessed for 27 factors considered to be prognostic for survival in this population. In addition to variables evaluated in Table E1, literature review and expert opinion identified other potential prognostic covariates: white blood cell (WBC) and lymphocyte counts within 1 month of consultation,^{10,11} steroid use,^{12,13} opiate pain medication use (as proxy for pain magnitude),¹⁴ type and timing of systemic therapy most recently delivered (including newer targeted oral agents^{15–17}), and presence of central spinal canal and/or neuroforaminal stenosis at the site of palliative RT.¹⁸ Detailed information on other sites of metastatic disease was also included. Table 1 lists operational definitions of BMETS covariates.

EMR-documented performance status (PS) was available for 72% of patients, including KPS in 60% and Eastern Cooperative Oncology Group PS in an additional 12%. To minimize missing values for PS, a single author (S.A.) reviewed all EMR notes ≥ 1 month from consultation to estimate a KPS based on documentation reflecting the patient's functional level at the time. For those with a recent EMR-recorded KPS, a clinically significant difference of ≥ 15 points between EMR-recorded and author-estimated KPS was identified in only 3% of patients. Given rare discordance, estimated KPS was used for all patients in our analysis. Seven patients did not have sufficient information for author estimation and were thus excluded.

The primary outcome was survival time between the date of palliative RT consultation and the date of death or last follow-up, assessed through December 31, 2018. Date of death was confirmed via the EMR and/or Social Security Death Index.

Temporal validation set

Patients included in the temporal validation set met the above inclusion criteria but were treated from January 1, 2018, to June 30, 2018, at the same institution. The temporal validation set yielded 87 patients, and 2 were omitted from evaluation on the basis of the above exclusion criteria.

The patient, treatment, and disease characteristics were collected for the validation set in the same manner specified for the training set. Notably, missing PS was significantly less frequent in the training set owing an institutional requirement for KPS or Eastern Cooperative Oncology Group PS documentation that was introduced in the time between data sets. Of 85 patients in the temporal validation set, 84 had a documented PS. The same author (S.A.) estimated KPS in the validation set, and a clinically significant difference of >15 points between EMR-recorded and author-estimated KPS was identified in 4% of patients.

The primary outcome of survival time was estimated as previously noted and assessed through February 1, 2020.

Statistical analysis

BMETS methodology—We used established random survival forests (RSF) methodology¹⁹ to model survival time after consultation for palliative RT using the 27 covariates noted. To do so, we used bootstrap aggregation to take 1000 bootstrap samples from the original data set. Binary survival trees were grown in each bootstrap sample via iterative splitting of the sample population into non-overlapping groups (nodes).¹⁹ Each split was created on the basis of a predictor covariate to maximize the log-rank statistic between the two nodes, creating clusters of patients with similar survival. To estimate the survival curve for a new individual based on the model, we first “dropped” the observation down each survival tree and obtained a Kaplan-Meier curve for each tree, based on the observations in the terminal node in which the dropped observation landed. The algorithm then averaged these Kaplan-Meier curves across trees for the final prediction. Specific model methodology and subsequent survival time predictions are described in Statistics E1. We named the final algorithm the BMETS model.

Notably, if multiple SBM sites were considered for RT treatment in the same patient during the same consultation visit, all target SBM sites were included in the model. To account for this, we coded each target site location and the number of concurrent SBM sites treated for each patient and included these data as covariates in the model.

To offer insight into relative variable importance, we used the minimal depth statistic, with increasing minimal depth values signifying decreasing prognostic importance.¹⁹

Model cross-validation and performance—Estimation of the model’s expected performance on external data was achieved using 100 repeats of pooled 10-fold cross-validation. This method was preferred over separate training and validation sets due to concern for inefficient use of data and reduced accuracy for limited data sets with the latter method.²⁰ Model discrimination was measured using time-dependent area under the curve

(tAUC),²¹ in which tAUC of 0.5 would predict survival at a given time no better than chance and a tAUC of 1 would indicate perfect model discrimination. tAUC was measured for survival times from 0 to 12 months postconsultation (Statistics E2).

Comparative clinical utility versus existing models—W-2⁸ and C-3⁷ models were used to assess the relative utility of BMETS versus simpler, traditional statistical methods. Each Cox regression was refitted using our source data. Model discrimination using tAUC for each model was compared across time points using the cross-validation methodology previously described. To permit comparison at clinically relevant survival times, we selected time points of 3, 6, and 12 months, as per Krishnan et al.²² These times correspond to commonly used cutpoints for appropriateness of spine surgery,²³ hospice referral,²⁴ and consideration of ablative RT owing to prolonged survival, respectively.

Temporal model validation—For each time point after consultation, tAUC for the validation set was estimated using the riskRegression R package, based on estimated survival curves from the models built using the training set; 95% confidence intervals (CIs) were determined as per the procedure outlined by Blanche et al.²⁵

Analyses were performed using R statistical computing language, version 3.5.1.

Results

A total of 492 bone sites in 397 patients met the inclusion criteria and were evaluated in this analysis. Patient, disease, and treatment characteristics are summarized in Table 1. Median age was 62.3 years (standard deviation, 13.4), with median KPS of 80 (range, 40–100). The most common primary cancer site was lung (32% of cases), and the most frequent sites of palliative RT were spine and hip/pelvis (59% and 24%, respectively). A majority of patients (88%) had known metastatic disease outside of the target SBM site, most commonly within other bone (69% of cases). A total of 370 deaths were observed, and median survival time after consultation was 6.4 months (Fig. 1).

As previously noted, we built BMETS using 27 candidate prognostic covariates. KPS, WBC count, type of systemic therapy last used, concurrent delivery of palliative RT to nonbone sites, and primary cancer site showed the lowest minimal depth across survival trees, thus conferring the greatest prognostic information (Fig. 2). Among these variables, KPS and primary cancer site were included in both C-3 and W-2, whereas the other 3 were not previously assessed by any model in Table E1.

Given the complexity of survival trees produced by the BMETS algorithm, model output cannot be easily visualized in tree form for clinical use. For illustrative purposes, Figure E1 shows a single survival tree from 1 bootstrap sample limited to just the 5 variables with lowest minimal depth. To facilitate translation of BMETS into clinical use, we developed a web platform that collects patient information for the 27 prognostic covariates and displays predicted survival probabilities across time points based on these data. This platform was evaluated via user acceptance testing to ensure ease of clinical application. It can be

accessed at <https://oncospace.radonc.jhmi.edu/Tools/PalliationPrediction.aspx>. Figure 3 demonstrates BMETS web platform output for a sample patient.

Primary model cross-validation

Cross-validation techniques within the training set revealed excellent discrimination for BMETS across time points for at least 12 months after consultation (Fig. 4). Specifically, tAUC at 1, 3, 6, and 12 months postconsultation was 0.87, 0.83, 0.81, and 0.81, respectively.

Relative utility compared with simpler, traditional models

Table 2 shows Cox proportional hazards analyses for C-3 and W-2 models refit using the data from our source population. The hazard ratios and CIs for the reduced C-3 and W-2 models were not published. However, our hazard ratios were of similar magnitude compared with published values for the corresponding full 6-variables models.^{7,8}

tAUC remained 0.74 at several points up to 12 months postconsultation for all 3 models (Fig. 4). However, tAUC was highest for BMETS across all time points. For comparison, tAUC at 1, 3, 6, and 12 months postconsultation was 0.79, 0.78, 0.76, and 0.74 for C-3, respectively, and 0.82, 0.80, 0.78, and 0.77 for W-2, respectively. Whereas W-2 began to converge toward BMETS after 6 months, tAUC for C-3 tended to decline over time.

Temporal validation

A total of 85 patients treated to 104 SBM sites were included in the temporal validation set. Patient, disease, and treatment characteristics are listed in Table E2. A total of 60 deaths occurred, and median survival was 5.0 months. Patients in the temporal validation set (vs the training set) were more likely to have reported black/African American race (53% vs 23%, $X^2 = 12.7$, $P < .001$); primary cancer site other than breast, prostate, lung, or hematologic sites (45% vs 32%, $X^2 = 8.6$, $P = .003$); received any form of systemic therapy (84% vs 69%, $X^2 = 55.6$, $P < .001$); received steroids (40% vs 25%, $X^2 = 11.9$, $P = .001$); and had prior surgery at the RT target site (20% vs 12%, $X^2 = 12.7$, $P < .001$). They were less likely to have central spinal canal and/or neuroforaminal stenosis (27% vs 38%, $X^2 = 16.2$, $P = .001$). Average WBC was also lower in the temporal validation set (6831 cells/ μ L [SD, 3885] vs 8878 cells/ μ L [SD, 5725], t test $P = .004$). There were no other significant differences between characteristics listed in Table E2. Missing data were notably less frequent in the temporal validation set. In part, this is due to adoption of a new EMR system in 2015; before this, part of the medical record from the earliest patients may have existed in paper form only, limiting complete data capture for our analyses.

Within the temporal validation set, BMETS model discrimination remained excellent across time points. tAUC at 1, 3, 6, and 12 months postconsultation was 0.92 (95% CI, 0.84–0.96), 0.86 (95% CI, 0.77–0.94), 0.82 (95% CI, 0.73–0.91), and 0.78 (95% CI, 0.68–0.88), respectively.

Discussion

In this study, we successfully developed the BMETS machine learning model for predicting survival after consultation for SBM. To our knowledge, BMETS is the first of its kind to use granular patient data and RSF methodology to create patient-specific predicted survival curves for clinical use in this patient population. BMETS maintains high predictive discrimination among patients with both short (<1 month) and long (>12 months) survival times, supporting its applicability to a range of clinical scenarios. Furthermore, BMETS predictions outperformed simpler, traditional statistical models in this setting, justifying its use. In a temporal validation set from the same institution, BMETS maintained excellent model performance despite differences in patient, disease, and treatment characteristics and survival times between the training and validation sets. To offset its added complexity, we have created a web-based, free-access platform to facilitate ease of data entry, display, and interpretation of BMETS predictions. Although the model was built using data from patients seen in consultation for palliative RT, BMETS predictions may have application in promoting evidence-based, prognosis-appropriate treatment decisions for systemic therapy, surgery, and advanced care planning for patients otherwise meeting study inclusion criteria.

Compared with traditional models from Table E1, BMETS methodology offers several potential advantages. First, unlike Cox regressions previously described, the RSF algorithm does not require a priori understanding of the relationships between variables. Thus, it may be able to handle complex interactions and nonlinear effects better than Cox regressions, in which these components must be prespecified when building models.²⁶ Also unlike traditional methods, RSF may be particularly robust to inclusion of nonprognostic and collinear covariates. Furthermore, the RSF algorithm handles missing data in a native fashion, by imputing missing values based on similar individuals within the same branch of the survival tree. All of these factors permitted us to include more covariates than used in past models, perhaps explaining why BMETS outperformed simpler approaches. It is noted that alternatives such as deep learning algorithms that use automated data mining (eg, free-text capture) may be less time-intensive and might elucidate completely novel predictors. However, our methodology may better avoid the criticism of “black box” statistics by ensuring that covariates are mindfully included due to mechanistic evidence of their prognostic capacity and that data collected are meaningful and accurate.

Moreover, previous studies have generally presented their prognostic estimates by sorting patients into binary or categorical survival groups. Such groupings are associated with loss of important clinical information about the survival trajectory that can be best interpreted by visualizing the shape and slope of the underlying survival curve. Furthermore, these categories do not provide information on the relative position of an individual patient within the ranges of survival provided, nor do they necessarily allow providers to estimate survival at specific time points that may be used as thresholds for selecting clinical interventions.²⁷ On the other hand, visualization of even a 5-covariate RSF model produces output that is too complex for standard clinical use (Fig. E1). Our web-based platform circumvents both of these issues by displaying a patient-specific predicted survival plot that provides useful prognostic details while maintaining relative simplicity of interpretation.

Despite these advantages, concerns may remain regarding the likelihood of clinical use for this model with 27 covariates—particularly in a context in which rates of missing data may be high. The platform underwent user acceptance testing during development to assess its usability, and choices were made to facilitate ease of clinical application. For example, certain granular data points such as sites of metastases default to 0 within the platform, as they do in the underlying model. As such, providers only need to click a value if it is present, saving time during data entry. When directly linked to an EMR, >50% of the covariate values can be autopopulated into the model platform, which also minimizes data entry time. Ideally, with technological advances such as native language processing, data abstraction and entry into such models will become increasingly seamless. Even without this shortcut, our user acceptance testing showed that once the platform was used 1 to 2 times, providers were able to enter all patient data in 1 minute or less.

Notably, the prediction algorithm can be used even in the presence of a large amount of missing data. Given the method for imputing missing data as described in the Methods E1, the model is expected to converge with predictions similar to those of models with fewer variables, such as W-2 or C-3. Because model discrimination remains relatively high for W-2, the BMETS platform requires that at least KPS and primary cancer site be entered to produce a prediction. This permits users to select the desired complexity of the model based on their time, available data, and prediction goals—with the caveat that model discrimination may improve by small increments with the inclusion of additional covariates.

In addition to its clinical utility, the methodology for BMETS development holds important implications for future machine learning models built in this context. Specifically, our decision to compare BMETS performance to simpler approaches during model development is a unique means for justifying its clinical utility. It was our goal that including this feature as part of BMETS development may serve as a proof of concept for future machine learning efforts by demonstrating that unlike traditional models used in this clinical context, complexity may add to predictive accuracy with newer statistical approaches.

A significant limitation to our analysis is the retrospective nature of the data collection. Specific limitations relative to handling of missing KPS are delineated in the Methods and Materials section. It is noted that model performance was very similar in the temporal validation set, in which missing KPS was significantly less common. This supports the accuracy of our KPS estimation procedures and the validity of the model. Additionally, the retrospective design limited our ability to include patient-reported outcomes (PROs), which previous studies have identified as potentially useful for prognostication in advanced cancer.²⁸ In part, PROs were omitted in line with our goal of designing a model that could be applied using only information collected in standard clinical practice. Notably, in addition to their best reduced W-2 model, Westhoff et al also analyzed 2 other reduced models containing primary tumor site plus either patient-reported visual analog scale of general health or patient-reported verbal rating scale of overall valuation of life. Both models had worse predictive accuracy than W-2, and the authors concluded that these PROs were less prognostic than provider-reported KPS.⁸ At least 2 other studies for patients with metastatic cancer also found that inclusion of PROs did not substantially improve prediction over clinical and physician-reported factors.^{29,30} Nonetheless, the value of PROs in an RSF

algorithm has not been described, and future iterations of BMETS may include prospective collection of PROs in the model.

Given increasing use of SBRT, it should also be noted that the applicability of our model for such patients is unclear. However, as SBRT becomes increasingly common outside of select patients such as those with oligometastases, clinical features and thus expected survival times for patients treated with SBRT may begin to approximate those included in our study. If this becomes the case, our results should be applicable to this population as well.

Lastly, the use of cross-validation procedures and bootstrap sampling for tAUC estimations limits our ability to calculate 95% CI given the sample size of the training set. Although sophisticated approaches have been described,³¹ CIs obtained through these methods do not appear to have correctly specified coverage until the training set sample size reaches 5000 subjects. Consequently, we are also limited in our ability to perform hypothesis testing regarding the superiority of BMETS versus C-3 or W2 models. Our temporal validation set affords the means for establishing CIs, but the relatively small sample size prevents us from using these CIs to compare predictive discrimination between models. These limitations could be addressed in future efforts with larger external validation sets.

Conclusions

We have developed a machine learning model that substantially improves survival predictions versus simpler, traditional models, and we have optimized this model for its translation into clinical use for managing patients with SBM. Although well calibrated, BMETS performance in clinical practice must be tested. Because RSF may be especially susceptible to loss of validity when applied to nonsource populations,³² it has been recommended that researchers use external data to expand the training set and refit the model instead of using these outside data to perform traditional external validation procedures. This is indeed our next step, with preliminary data showing excellent model performance when BMETS is applied to and refit with data from external populations.³³ Moreover, proof of external validity for a survival model does not necessarily provide evidence of its clinical use. As such, we have also performed a pilot assessment of BMETS as part of a decision support tool within the source population.³⁴ This early evidence shows that the tool improves providers' survival predictions and selection of evidence-based and prognosis-appropriate palliative treatments for patients with SBM. Future directions include a multi-institutional, randomized assessment of the BMETS-based decision support tool in external environments.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments—

Dayflower.io assisted with creation of the BMETS web platform: Developing an Improved Statistical Approach for Survival Estimation in Bone Metastases Management: The Bone Metastases Ensemble Trees for Survival (BMETS) model.

This work was supported by a grant from the National Institutes of Health (5KL2TR001077) to S.A.

Disclosures: S.A., J.W., T.S., T.M., and T.D. are employed by the Johns Hopkins School of Medicine and S.Z. by the School of Public Health. S.A. reports grant from Elekta, nonfinancial support from Angiodynamics, and personal fees from Allegheny Health Network (AHN), all outside the submitted work. J.W. reports personal fees from AHN and is an editor of the *International Journal of Radiation Oncology, Biology, Physics*, outside the submitted work. T.S. reports nonfinancial support from EMD Serono and personal fees from Allergan, outside the submitted work. T.M. reports a grant from the Radiation Oncology Institute and is chairman of the board and a stockholder of a health-related start-up, Oncospace, Inc., outside the submitted work. T.D. is president of the American Society for Radiation Oncology and reports nonfinancial support from Elekta, Sanofi-Aventis, and Varian, Inc; outside the submitted work.

References

1. Jones JA, Lutz ST, Chow E, Johnstone PA. Palliative radiotherapy at the end of life: A critical review. *CA Cancer J Clin*. 64:296–310. [PubMed: 25043971]
2. Weeks JC, Cook EF, O'Day SJ, et al. Relationship between cancer patients' predictions of prognosis and their treatment preferences. *JAMA* 1998;279:1709–1714. [PubMed: 9624023]
3. Mizumoto M, Harada H, Asakura H, et al. Radiotherapy for patients with metastases to the spinal column: A review of 603 patients at Shizuoka Cancer Center Hospital. *Int J Radiat Oncol Biol Phys* 2011; 79:208–213. [PubMed: 20418024]
4. Chao ST, Koyfman SA, Woody N, et al. Recursive partitioning analysis index is predictive for overall survival in patients undergoing spine stereotactic body radiation therapy for spinal metastases. *Int J Radiat Oncol Biol Phys* 2012;82:1738–1743. [PubMed: 21489717]
5. Tokuhashi Y, Uei H, Oshima M, Ajiro Y. Scoring system for prediction of metastatic spine tumor prognosis. *World J Orthop* 2014;5:262–271. [PubMed: 25035829]
6. Tseng YD, Krishnan MS, Sullivan AJ, et al. How radiation oncologists evaluate and incorporate life expectancy estimates into the treatment of palliative cancer patients: A survey-based study. *Int J Radiat Oncol Biol Phys* 2013;87:471–478. [PubMed: 24074920]
7. Chow E, Abdoell M, Panzarella T, et al. Predictive model for survival in patients with advanced cancer. *J Clin Oncol* 2008;26:5863–5869. [PubMed: 19018082]
8. Westhoff PG, de Graeff A, Monninkhof EM, et al. An easy tool to predict survival in patients receiving radiation therapy for painful bone metastases. *Int J Radiat Oncol* 2014;90:739–747.
9. Carrington C Oral targeted therapy for cancer. *Aust Prescr* 2015;38: 171–176. [PubMed: 26648656]
10. Feng L, Gu S, Wang P, et al. White blood cell and granulocyte counts are independent predictive factors for prognosis of advanced pancreatic cancer. *Gastroenterol Res Pract* 2018;2018:8096234.
11. Fogar P, Sperti C, Basso D, et al. Decreased total lymphocyte counts in pancreatic cancer: An index of adverse outcome. *Pancreas* 2006;32:22–28. [PubMed: 16340740]
12. Hughes MA, Parisi M, Grossman S, et al. Primary brain tumors treated with steroids and radiotherapy: Low CD4 counts and risk of infection. *Int J Radiat Oncol Biol Phys* 2005;62:1423–1426. [PubMed: 16029802]
13. Dietrich J, Rao K, Pastorino S, et al. Corticosteroids in brain cancer patients: Benefits and pitfalls. *Expert Rev Clin Pharmacol* 2011;4:233–242. [PubMed: 2166852]
14. Halabi S, Vogelzang NJ, Kornblith AB, et al. Pain predicts overall survival in men with metastatic castration-refractory prostate cancer. *J Clin Oncol* 2008;26:2544–2549. [PubMed: 18487572]
15. Chionh F, Campbell A, Sukumaran S, et al. Oral versus intravenous fluoropyrimidines for colorectal cancer. *Cochrane Database Syst Rev* 2010;7:CD008398.
16. Chapman PB, Robert C, Larkin J, et al. Vemurafenib in patients with BRAFV600 mutation-positive metastatic melanoma: Final overall survival results of the randomized BRIM-3 study. *Ann Oncol Off J Eur Soc Med Oncol* 2017;28:2581–2587.
17. Shepherd FA, Rodrigues Pereira J, Ciuleanu T, et al. Erlotinibin previously treated non-small-cell lung cancer. *N Engl J Med* 2005;353:123–132. [PubMed: 16014882]
18. Loblaw DA, Laperriere NJ, Mackillop WJ. A population-based study of malignant spinal cord compression in Ontario. *Clin Oncol (R Coll Radiol)* 2003;15:211–217. [PubMed: 12846501]

19. Ishwaran H, Kogalur UB, Blackstone EH, et al. Random survival forests. *Ann Appl Stat* 2008;2:841–860.
20. Kohavi R A study of cross-validation and bootstrap for accuracy estimation and model selection. *IJCAI (U S)* 1995;2:1137–1145.
21. Heagerty PJ, Zheng Y. Survival model predictive accuracy and ROC curves. *Biometrics* 2005;61:92–105. [PubMed: 15737082]
22. Krishnan MS, Epstein-Peterson Z, Chen Y-H, et al. Predicting life expectancy in patients with metastatic cancer receiving palliative radiotherapy: The TEACHH model. *Cancer* 2014;120:134–141. [PubMed: 24122413]
23. National Comprehensive Cancer Network. NCCN: Central nervous system cancers V1. 2019. Available at: https://www.nccn.org/professionals/physician_gls/pdf/cns.pdf. Accessed January 10, 2019.
24. Howell DD, Lutz S. Hospice referral: An important responsibility of the oncologist. *J Oncol Pract* 2008;4:303–304. [PubMed: 20856763]
25. Blanche P, Dartigues J-F, Jacqmin-Gadda H. Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. *Stat Med* 2013;32:5381–5397. [PubMed: 24027076]
26. Biau G Analysis of a random forests model. *J Mach Learn Res* 2010; 13:1063–1095.
27. Krishnan M, Temel JS, Wright AA, et al. Predicting life expectancy in patients with advanced incurable cancer: A review. *J Support Oncol* 2013;11:68–74. [PubMed: 23967494]
28. Gotay CC, Kawamoto CT, Bottomley A, et al. The prognostic significance of patient-reported outcomes in cancer clinical trials. *J Clin Oncol* 2008;26:1355–1363. [PubMed: 18227528]
29. Collette L, van Andel G, Bottomley A, et al. Is baseline quality of life useful for predicting survival with hormone-refractory prostate cancer? A pooled analysis of three studies of the European Organisation for Research and Treatment of Cancer Genitourinary Group. *J Clin Oncol* 2004;22:3877–3885. [PubMed: 15459209]
30. Efficace F, Biganzoli L, Piccart M, et al. Baseline health-related quality-of-life data as prognostic factors in a phase III multicentre study of women with metastatic breast cancer. *Eur J Cancer* 2004;40: 1021–1030. [PubMed: 15093577]
31. LeDell E, Petersen M, van der Laan M. Computationally efficient confidence intervals for cross-validated area under the ROC curve estimates. *Electron J Stat* 2015;9:1583–1607. [PubMed: 26279737]
32. Dankowski T, Ziegler A. Calibrating random forests for probability estimation. *Stat Med* 2016;35:3949–3960. [PubMed: 27074747]
33. LaVigne AW, Elledge CR, Fiksel J, et al. External validation of the Bone Metastases Ensemble Trees for Survival (BMETS) machine learning model to improve estimation of life expectancy. To be presented at the American Society of Radiation Oncology (ASTRO) Annual Meeting. October 2020; Miami, FL, USA.
34. Alcorn SR, Fiksel J, Hu C, et al. Pilot assessment of the BMET decision support platform: A tool to improve provider survival estimates and selection of prognosis-appropriate treatment for patients with symptomatic bone metastases. *Int J Radiat Oncol* 2019;105: S47.

Fig. 1. Kaplan-Meier survival estimate (solid line) and 95% confidence interval (dashed line) for the overall group (N = 397).

Fig. 2.

Minimal depth for each covariate within the BMETS model. This value represents the distance between the root node (at position 0) and the node first used to split each covariate, averaged across trees. A lower minimum depth indicates higher prognostic importance for a given variable. *Abbreviations:* BMETS = Bone Metastases Ensemble Trees for Survival; CC/NFS = central canal and/or neuroforaminal stenosis; KPS = Karnofsky Performance Status; RT = radiation therapy; WBC = white blood cells.

Fig. 3.

Example output for the web platform developed to collect covariate information and display the estimated survival probabilities from time of consultation to death as predicted by the Bone Metastases Ensemble Trees for Survival model. The patient was a 71-year-old black/African American woman with metastatic thyroid cancer who underwent outpatient consultation for a lumbar spine site, with initial cancer diagnosis made 5.25 years ago, most recent systemic therapy of oral (sorafenib) administered 1 month ago, no prior surgery to target site, weight loss in the past 6 months, Karnofsky Performance Status of 70, not taking either opiate pain medication or steroids, white blood cell count of 9160, and lymphocyte count of 2390. Imaging showed no definite spinal canal/neuroforaminal compromise, and she had metastasis at other bone sites but no plans for other concurrent palliative radiation therapy. An interactive plot displays the case patient's predicted survival probabilities after consultation for palliative radiation therapy (orange). For comparison purposes, the curves in the background demonstrate predicted survival probabilities for all other patients with symptomatic bone metastases in the database, displayed from lowest to highest percentile of survival time at 12 months (dark to light blue curves).

Fig. 4.

Comparison of time-dependent area under the curve (tAUC) between prognostic models across survival time points after consultation for palliative radiation therapy. *Abbreviations:* BMETS = Bone Metastases Ensemble Trees for Survival model; C-3 = Chow's 3-variable number of risk factors model; W-2 = Westhoff's 2-variable model.⁸

Table 1

Patient, disease, and treatment characteristics for 397 patients included in the BMETS model

Covariate name	n (%) or mean (SD)	Missing
Patient-specific factors		
1. Age, y, mean (SD)	62 (12)	0
2. Sex, %		0
Female	48%	
Male	52%	
3. Race, %		0
White	72%	
Black/African American	23%	
Asian	2%	
Other	3%	
4. KPS in units of 10, median (range)	80 (40–100)	0*
5. WBC count within prior 1 mo in cells per μ L, mean (SD)	8878 (5725)	84
6. Lymphocyte count within prior 1 mo in cells per μ L, mean (SD)	1519 (2565)	106
7. Hospital inpatient status, % [†]		0
Yes	25%	
No	75%	
8. Any weight loss in prior 6 mo, %		67
Yes	67%	
No	33%	
Disease-specific factors		
9. Primary cancer site, %		0
Breast	19%	
Prostate	12%	
Lung	32%	
Leukemia, lymphoma, myeloma	5%	
Other	32%	
10. No. of concurrent palliative RT to other noncontiguous bone sites, % [‡]		0
0	81%	
1	15%	
2 ^b	4%	
11. Concurrent palliative RT to noncontiguous sites other than bone, % [‡]		0
None	92%	
Brain	5%	
Lung	2%	
Other or >1 type	1%	
12. Current steroid use, %		13
Yes	25%	
No	75%	

Covariate name	n (%) or mean (SD)	Missing
13. Current opiate analgesic use, %		7
Yes	71%	
No	29%	
14. Systemic therapy delivered within the previous 1 mo, %		95
Yes	55%	
No	45%	
15. Type of systemic therapy last delivered, % [§]		3
None	31%	
Intravenous	39%	
Nonhormonal oral	11%	
Hormonal	19%	
16. Prior surgery at RT target site, %		1
Yes	12%	
No	88%	
Treatment-specific factors		
17. RT target site, % ^{//}		0
Spine	53%	
Hip/pelvis	13%	
Extremity	18%	
Chest wall	13%	
Skull	4%	
18. CC/NFS, (%) [¶]		60
Yes	38%	
No	62%	
19. Time from initial diagnosis, mean (SD), mo	40 (55)	2
20. Other metastases (% yes) [#]		0
Brain	12%	
Lung	40%	
Liver	20%	
Adrenal gland	8%	
Lymph nodes ^{**}	42%	
Soft tissue	5%	
Other bone	69%	
Other sites	7%	

Abbreviations: BMETS = Bone Metastases Ensemble Trees for Survival; CC/NFS = central canal and/or neuroforaminal stenosis; KPS = Karnofsky Performance Status; SD = standard deviation; RT = radiation therapy; WBC = white blood cells.

^{*} See Methods and Materials section regarding handling of missing Karnofsky Performance Status data.

[†] Admission to offsite inpatient rehabilitation or nursing home facilities was not considered as *hospital inpatient*.

[‡] Does not include RT target sites requiring multiple contiguous fields due to large target size.

[§]If multiple types of systemic therapy were delivered concurrently, a single response was selected in the following order: intravenous > nonhormonal oral > hormonal.

// If RT target lesion encompassed multiple sites, site containing majority of target lesion was selected.

[¶]Defined as radiologic evidence of spinal cord, spinal canal, nerve root, or neuroforaminal impingement from direct involvement of the target lesion. Only considered for sites near the neuraxis (ie, spine, central pelvis, skull; n = 233).

[#]Includes all radiologically confirmed definite areas of metastatic disease outside the current palliative RT field. Indeterminate lesions or sites without radiologic evaluation were coded as “no.”

^{**}Includes locoregional nodal metastases for the primary site.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2

Multivariate Cox proportional hazards analyses for covariates from 2 validated Cox proportional hazards models, refitted using our source population data*

Model and covariates [†]	Hazard ratio	95% Confidence interval
Chow's 3-variable NRF (C-3) model ⁷		
Primary cancer site		
Breast	1.00	-
Nonbreast	1.75	1.34–2.29
KPS		
>60	1.00	-
<60	3.7	2.88–4.75
Site of metastases		
Bone only	1.00	-
Other	2.25	1.79–2.82
Westhoffs 2-variable (W-2) model ⁸		
Primary cancer site		
Breast	1.00	-
Prostate	1.17	0.78–1.75
Lung 2.57 1.89–3.49		
Other 2.05 1.53–2.77		
KPS		
90–100	1.00	-
70–80	1.81	1.37–2.40
20–60	6.17	4.41–8.58

Abbreviations: KPS = Karnofsky Performance Status; NRF = number of risk factors.

* A complete data set with no missing values for the covariates was used to refit the models.

[†] Covariate values are specified as per the source model's definitions.