Development of Machine Learning Algorithms for Prediction of 30- Day Mortality After Surgery for Spinal Metastasis

Scholarly Report submitted in partial fulfillment of the MD Degree at Harvard Medical School

Date: 17 December 2018

Student Name: Aditya V. Karhade, B.E.

Scholarly Report Title: Development of Machine Learning Algorithms for Prediction of 30-Day Mortality After Surgery for Spinal Metastasis

Mentor Name(s) and Affiliations: Joseph H. Schwab, MD, MS, Chief of Spine Surgery, Departmentt of Orthopaedic Surgery, Massachusetts General Hospital

Abstract

Title: Development of Machine Learning Algorithms for Prediction of 30-Day Mortality After Surgery for Spinal Metastasis

Purpose: Preoperative prognostication of short-term postoperative mortality in patients with spinal metastatic disease can improve shared decision making around end- of-life care. The objective of this study was to (1) develop machine learning algorithms for prediction of short-term mortality and (2) deploy these models in an open access web application.

Methods: The American College of Surgeons, National Surgical Quality Improvement Program was used to identify patients that underwent operative intervention for metastatic disease. Four machine learning algorithms were developed, and the algorithm with the best performance across discrimination, calibration, and overall performance was integrated into an open access web application.

Results: The 30-d mortality for the 1790 patients undergoing surgery for spinal metastatic disease was 8.49%. Preoperative factors used for prognostication were albumin, functional status, white blood cell count, hematocrit, alkaline phosphatase, spinal location (cervical, thoracic, lumbosacral), and severity of comorbid systemic disease (American Society of Anesthesiologist Class). In this population, machine learning algorithms developed to predict 30-d mortality performed well on discrimination (c-statistic), calibration (assessed by calibration slope and intercept), Brier score, and decision analysis. An open access web application was developed for the best performing model and this web application can be found here: https://sorg-apps.shinyapps.io/spinemets/.

Conclusion: Machine learning algorithms are promising for prediction of postoperative outcomes in spinal oncology and these algorithms can be integrated into clinically useful decision tools. As the volume of data in oncology continues to grow, creation of learning systems and deployment of these systems as accessible tools may significantly enhance prognostication and management.

Contribution to work:

As first author on this study, I was responsible for leading the design, execution, analysis and writing. Together with Dr. Joseph Schwab and our co-authors, we designed this study to explore the utility of machine learning in spinal oncology. We chose spinal metastatic disease because decision making around the appropriateness of surgical intervention for this pathology is heavily dependent on preoperative scoring systems for estimating the predicted postoperative survival.

We searched for large registries of patients with spinal metastatic disease that would allow us to both obtain sufficient volumes of patients and ensure granular data. I applied for access to the American College of Surgeons National Surgical Quality Improvement Program and filled out the required administrative paperwork. After obtaining the data, I isolated the patients that met our inclusion criteria and then developed four machine learning algorithms to predict survival in this population by using the Python and R programming languages.

Finally, I lead the writing of the manuscript and took responsibility for submission to peer-reviewed journals and the revisions required for publication. I learned many skills through this process and I am grateful to the Scholars in Medicine office and my mentors for their guidance and support.

Citation:

Aditya V Karhade, Quirina C B S Thio, Paul T Ogink, Akash A Shah, Christopher M Bono, Kevin S Oh, Phil J Saylor, Andrew J Schoenfeld, John H Shin, Mitchel B Harris, Joseph H Schwab; Development of Machine Learning Algorithms for Prediction of 30-Day Mortality After Surgery for Spinal Metastasis, *Neurosurgery*, nyy469, https://doi.org/10.1093/neuros/nyy469

Appendix:

Full manuscript made available from next page onward.

Development of Machine Learning Algorithms for Prediction of 30-Day Mortality After Surgery for Spinal Metastasis

Aditya V. Karhade, BE*
Quirina C. B. S. Thio, MD*
Paul T. Ogink, MD*
Akash A. Shah, BS*
Christopher M. Bono, MD‡
Kevin S. Oh, MD§
Phil J. Saylor, MD¶
Andrew J. Schoenfeld, MD‡
John H. Shin, MD‖
Mitchel B. Harris, MD*
Joseph H. Schwab, MD, MS*

*Department of Orthopedic Surgery, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts; ‡Department of Orthopedic Surgery, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts; §Department of Radiation Oncology, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts; ¶Department of Hematology/Oncology, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts; ‖Department of Neurosurgery, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts

Correspondence:
Joseph H. Schwab, MD, MS,
Department of Orthopedic Surgery,
Massachusetts General Hospital,
Harvard Medical School,
55 Fruit Street,
Boston, MA 02114.
E-mail: jhschwab@mgh.harvard.edu

Received, April 19, 2018.
Accepted, August 31, 2018.

BACKGROUND: Preoperative prognostication of short-term postoperative mortality in patients with spinal metastatic disease can improve shared decision making around end-of-life care.

OBJECTIVE: To (1) develop machine learning algorithms for prediction of short-term mortality and (2) deploy these models in an open access web application.

METHODS: The American College of Surgeons, National Surgical Quality Improvement Program was used to identify patients that underwent operative intervention for metastatic disease. Four machine learning algorithms were developed, and the algorithm with the best performance across discrimination, calibration, and overall performance was integrated into an open access web application.

RESULTS: The 30-d mortality for the 1790 patients undergoing surgery for spinal metastatic disease was 8.49%. Preoperative factors used for prognostication were albumin, functional status, white blood cell count, hematocrit, alkaline phosphatase, spinal location (cervical, thoracic, lumbosacral), and severity of comorbid systemic disease (American Society of Anesthesiologist Class). In this population, machine learning algorithms developed to predict 30-d mortality performed well on discrimination (c-statistic), calibration (assessed by calibration slope and intercept), Brier score, and decision analysis. An open access web application was developed for the best performing model and this web application can be found here: https://sorg-apps.shinyapps.io/spinemets/.

CONCLUSION: Machine learning algorithms are promising for prediction of postoperative outcomes in spinal oncology and these algorithms can be integrated into clinically useful decision tools. As the volume of data in oncology continues to grow, creation of learning systems and deployment of these systems as accessible tools may significantly enhance prognostication and management.

KEY WORDS: Artificial intelligence, Machine learning, Oncology, Prediction, Spinal metastases, Spine surgery

Neurosurgery 0:1–9, 2018 DOI:10.1093/neuros/nyy469 www.neurosurgery-online.com

Spinal metastatic disease develops in more than 40% of oncology patients and progresses to spinal cord compression in 20% of these cases.[1] The rate of surgical intervention for spinal metastatic disease has increased[2,3] since the randomized controlled trial by Patchell et al[4] demonstrating benefit of decompressive surgery and radiotherapy versus radiotherapy alone. However, short-term mortality after surgery for spinal metastatic disease is a marker for patients who did not benefit from this significant intervention for a primarily palliative result; preoperative prognostication of this adverse outcome can improve end-of-life care for these patients. In response to this need, numerous studies have created risk scores and nomograms for predicting outcomes in this population.[5-24] Nonetheless, relatively few studies have sought to apply machine learning algorithms to predict survival in spinal metastatic disease.[8,25] In addition, there are no studies focusing on operatively managed spinal metastatic disease and

ABBREVIATIONS: ACS, American College of Surgeons; **NSQIP,** National Surgical Quality Improvement Program

incorporating disease specific factors into open access decision tools for healthcare professionals.

Machine learning is an intersection of computer science and statistics used in oncology for pharmacogenomics, image classification, and decision support systems, among other areas.[26-28] Notably, the field has advanced to the extent that in 2017, Esteva et al[29] developed machine learning algorithms that rivaled 20 board-certified dermatologists in correctly identifying skin cancer from images alone.

The purpose of this study was (1) to explore the utility of machine learning algorithms for predicting short-term survival and (2) to develop accessible interfaces for healthcare professionals to use machine learning for prognosticating 30-d mortality in patients with spinal metastatic disease.

METHODS

Guidelines

Transparent Reporting of multivariable Prediction Models for Individual Prognosis or Diagnosis (TRIPOD) and JMIR Guidelines for Developing and Reporting Machine Learning Predictive Models in Biomedical Research were followed.[30,31] This was a retrospective machine learning classification study (outcome was binary categorical) for prognostication in spinal metastatic disease.

Data Source

The American College of Surgeons (ACS) National Surgical Quality Improvement Program (NSQIP) is a large, multi-institutional clinical registry of 30-d postoperative outcomes at US surgical centers and has been extensively used for outcomes research in spine surgery.[32] Institutional review board approval for this study was not sought as the de-identified NSQIP data have been previously exempt from individual review by our institutional review board.

Patient Selection

Patients were only included in the study if all of the following criteria were met (1) primary Current Procedural Terminology code for excision, osteotomy, decompression, fusion, or fixation (2) at the cervical, thoracic, or lumbosacral levels, (3) International Classification of Diseases diagnosis of secondary malignant neoplasm of bone, meninges, or spinal cord or diagnosis of pathological fracture, (4) confirmed comorbidity of disseminated cancer, (5) surgical subspecialty neurosurgery or orthopedics, (6) general anesthesia, (7) inpatient operation, and (8) year of operation between 2009 and 2016 (9) American Society of Anesthesiologist Classification indicating systemic disease (II-V).

Candidate Features

The following variables were extracted for each eligible patient based on prior work: (1) sex [male, female],[33] (2) age [continuous],[34] (3) body mass index [continuous], (4) functional status [independent, dependent],[7,8,15,34,35] (5) severity of comorbid systemic disease as assessed by the American Society of Anesthesiologists Classification [II, III, IV-V],[34] (6) spinal tumor location [cervical, thoracic, lumbosacral],[36] (7) corpectomy [yes, no], (8) laminectomy [yes, no], (9) fusion [yes, no], (10) instrumentation [yes, no], (11) number of levels, (12) preoperative albumin [continuous],[23,35] (13) preoperative serum alkaline phosphatase [continuous],[37-39] (14) preoperative hematocrit [continuous],[8] (15) preoperative white blood cell count [continuous],[8,40] (16) preoperative platelet count [continuous].[41,42] Multiple imputation with chained equations was used to impute missing preoperative laboratory characteristics with less than 25% missing data. Thirty-day mortality, as documented in NSQIP, was used as the dependent variable in this investigation.

Data Analysis

A stratified 80:20 split of the available data was carried out. The training set was used for algorithm training and assessment of performance by 10-fold cross validation. All study variables were entered into Random Forest algorithms, and recursive feature selection was used to identify the subset of features employed in final modeling.[43] Neural Network, Support Vector Machine, Bayes Point Machine, and Decision Tree models were subsequently trained to predict 30-d mortality.[43-46] The best performing model was used to predict 30-d mortality in the testing set.

Discrimination was assessed graphically with the receiver operating curve and numerically with c-statistic, also known as the area under the receiver operating curve for binary classification. Discrimination is the model's ability to distinguish patients who survived from those who died.[47-50] Models with perfect discrimination have c-statistic = 1, while models with performance no better than chance have c-statistic = 0.5.

Calibration was assessed graphically with calibration plots and numerically with calibration slope and calibration intercept.[49,50] Calibration measures how well the model's predicted probabilities concur with the observed probabilities in the study population. Calibration intercept measures whether on average the model tends to overestimate or underestimate the probability of the outcome; perfect models have a value of 0 for a calibration intercept. Calibration slope measures the difference between predictor effects for each model in the training and testing datasets. When predictor effects for the model are equivalent in the training and testing sets, the calibration slope is 1.

Overall model performance was assessed with the Brier score, the mean squared error between the observed values and the predicted probabilities.[50,51] The Brier score is a composite of discrimination and calibration that can also be used to benchmark model performance.[50] The Brier score for the null model, assigning a predicted probability for all patients equivalent to the prevalence of 30-d mortality in the population, was calculated and used to compare the Brier values attained by the machine learning algorithms. Brier scores closer to zero indicate better models (lower error between predictions and observed values).

Decision curve analysis was undertaken to determine the utility of the best model for clinical management.[52,53] Decision curve analysis allows for the assessment of net benefit over a range of probability thresholds. Net benefit is a function of true positives, false positives, and the relative weight assigned to false positives versus true positives based on the probability threshold. A single probability is used to identify both the threshold and the relative weight of true positives and false positives. Since probability thresholds may vary, decision curves are useful for examining the utility of prediction models over a range of probability thresholds in order to compare net benefit of changing management for no patients, changing management for all patients, changing management for patients based on an individual predictor, and for changing management based on the overall prediction model.

Application Development

The best algorithm across the model performance metrics for predicting 30-d mortality was incorporated into an interactive interface. The clinical decision tool was designed to collect the values entered by a healthcare professional, feed the values to the pre-trained algorithm, retrieve the result, and finally to output the result to the healthcare professional in real time. The clinical decision tool was deployed as an open-access web-based application and programmed to be accessible and adaptable for use on desktops, tablets, and smartphones. The Anaconda Distribution (Anaconda Inc, Austin, Texas), Microsoft Azure (Microsoft Corporation, Redmond, Washington), R version 3.4.3 (The R Foundation, Vienna, Austria), RStudio version 1.0.153 (RStudio, Boston, Massachusetts), and Python version 3.6 (Python Software Foundation, Wilmington, Delaware) were used for data analysis, model creation, and web application development.

RESULTS

The 30-d mortality for the 1790 patients undergoing operative intervention for spinal metastatic disease was 8.49%. Patients who suffered 30-d mortality had lower albumin, higher white blood cell count, lower hematocrit, and higher alkaline phosphatase. Other baseline characteristics of the study population are displayed in Table 1 (continuous variables categorized for ease of interpretation and assessment of baseline data completeness). Random Forest algorithms identified albumin, functional status, white blood cell count, hematocrit, alkaline phosphatase, spinal location (cervical, thoracic, lumbosacral), and severity of comorbid systemic disease (American Society of Anesthesiologist class) as predictive factors for 30-d mortality. C-statistics of all models were similar in the training set, n = 1432, and ranged from 0.760 for the Support Vector Machine to 0.769 for the Neural Network (Table 2). The best model for predicting 30-d mortality as assessed by discrimination alone was the Neural Network with c-statistic 0.769. The calibration slope ranged from 0.728 for the Decision Tree to 1.013 for the Bayes Point Machine and the calibration intercept ranged from –0.009 for the Bayes Point Machine to 0.004 for the Decision Tree. Assessed graphically and numerically, the Bayes Point Machine was best calibrated over the full range of predicted probabilities.

Assessed by overall model performance, the Brier score ranged from 0.0701 for the Bayes Point Machine to 0.0711 for the Decision Tree. In comparison, the null Brier model performance (assigning a predicted probability to each patient equal to the prevalence of 30-d mortality in the population) was 0.079. Bayes Point Machine was chosen as the final model with superior performance on calibration and overall assessment. On evaluation in the testing set, n = 358, the model had c-statistic 0.782 (Figure 1), calibration slope 1.07, calibration intercept –0.062 (Figure 2), and Brier score 0.068 (null Brier score in the testing set = 0.078).

Decision curve analysis for the Bayes Point Machine model showed that changing management based on the Bayes Point Machine model would result in greater net benefit than changing management for no patients or for all patients undergoing operative intervention for spinal metastasis over all thresholds (Figure 3). In addition, the net benefit of changing management on the basis of the Bayes Point Machine model was greater than by changing management on the basis of ASA class alone.

The Bayes Point Machine was incorporated into a web application with a user interface and deployed as an open access tool for healthcare professionals (Figure 4). The web application can be accessed here: https://sorg-apps.shinyapps.io/spinemets/.

DISCUSSION

Preoperative evaluation of patients with spinal metastatic disease is imperative to minimize the risks of surgery (postoperative complications, decreased quality of life, shortened survival) for the subset of patients who are unlikely to attain the health benefits (longer survival, improved quality of life) for which spinal surgery in this population is intended.[8,13,34] In particular, prognosticating short-term mortality after surgery remains one of the most important benchmarks for patient counseling and shared decision making in the consideration of treatment pathways: surgery, radiotherapy, chemotherapy/immunotherapy, and palliative care.[54] This study evaluated the utility of multiple machine learning models for predicting 30-d mortality after surgical intervention for spinal metastatic disease and demonstrated the high performance of these models across discrimination, calibration, and decision analysis.

The ultimate intent of this effort was to develop a clinically useful predictive model for short-term mortality after operative intervention for spinal metastatic disease. The preoperative factors selected by Random Forest algorithms (and subsequently used for final modeling) reassuringly concurred with previous studies that have demonstrated that lower preoperative albumin, higher white blood cell count, and dependent functional status are important predictors for near-term mortality after intervention for spinal metastatic disease. For example, Schoenfeld et al[35] previously demonstrated that preoperative nutritional status is significantly associated with 30-d mortality.[23] In addition, Paulino Pereira et al[8] demonstrated that a higher preoperative white blood cell count and lower hemoglobin were associated with increased hazard of mortality.[8] Multiple studies have demonstrated that preoperative poor functional status is associated with worse outcomes following treatment for spinal metastatic disease.[5-8,23,35,40]

Paulino Pereiera et al[8] previously developed a boosting algorithm for predicting survival in patients with spinal metastatic disease treated at 2 tertiary care academic medical centers but did not deploy the boosting algorithm as an application for healthcare professionals. The multicenter origin of the data used in this analysis, as well as the size of our sample, advantages the results of this effort. Forsberg et al[25] also created a Bayesian Belief Network for predicting survival in patients with axial and appendicular skeletal metastasis using a single institutional registry.[55] However, the dataset used to create the algorithm had a majority of appendicular skeletal metastases patients and included only

TABLE 1. Baseline Characteristics of Patients Undergoing Operative Intervention for Spinal Metastatic Disease, n = 1790

Variable	Definition	Thirty-day mortality		
		Total, n = 1790	No, n = 1638	Yes, n = 152
Sex	Female	695 (38.9)	651 (39.8)	44 (29.1)
	Male	1093 (61.1)	986 (60.2)	107 (70.9)
Age (years)	<65	1017 (56.8)	935 (57.1)	82 (53.9)
	65-79	682 (38.1)	622 (38.0)	60 (39.5)
	>80	91 (5.1)	81 (4.9)	10 (6.6)
Body mass index (kg/m^2)	<18.5	143 (8.0)	126 (7.7)	17 (11.2)
	≥40	60 (3.4)	55 (3.4)	5 (3.3)
	18.5-29	1173 (65.5)	1070 (65.3)	103 (67.8)
	30-39	414 (23.1)	387 (23.6)	27 (17.8)
Functional status	Independent	1560 (87.2)	1448 (88.4)	112 (73.7)
	Dependent	230 (12.8)	190 (11.6)	40 (26.3)
American Society of Anesthesiologist Class	II	213 (11.9)	206 (12.6)	7 (4.6)
	III	1158 (64.7)	1077 (65.8)	81 (53.3)
	IV-V	419 (23.4)	355 (21.7)	64 (42.1)
Spine location	Cervical	327 (18.3)	286 (17.5)	41 (27.0)
	Lumbosacral	420 (23.5)	395 (24.1)	25 (16.4)
	Thoracic	1043 (58.3)	957 (58.4)	86 (56.6)
Corpectomy		416 (23.2)	379 (23.1)	37 (24.3)
Laminectomy		902 (50.4)	819 (50.0)	83 (54.6)
Fusion		1190 (66.5)	1097 (67.0)	93 (61.2)
Instrumentation		1107 (61.8)	1024 (62.5)	83 (54.6)
Number of levels	One or two	651 (36.4)	590 (36.0)	61 (40.1)
	Three or more	1139 (63.6)	1048 (64.0)	91 (59.9)
Albumin (g/dL)	<3.5	553 (30.9)	465 (28.4)	88 (57.9)
	≥3.5	834 (46.6)	796 (48.6)	38 (25.0)
	Not measured	403 (22.5)	377 (23.0)	26 (17.1)
Alkaline phosphatase (IU/L)	>115	528 (29.5)	467 (28.5)	61 (40.1)
	0-44	24 (1.3)	23 (1.4)	1 (0.7)
	45-115	832 (46.5)	769 (46.9)	63 (41.4)
	Not measured	406 (22.7)	379 (23.1)	27 (17.8)
Hematocrit (%)	<30	525 (29.3)	453 (27.7)	72 (47.4)
	≥30	1256 (70.2)	1176 (71.8)	80 (52.6)
	Not measured	9 (0.5)	9 (0.5)	0 (0.0)
White blood cell ($10^3/\mu$L)	<4	97 (5.4)	88 (5.4)	9 (5.9)
	≥11	509 (28.4)	441 (26.9)	68 (44.7)
	4-11	1173 (65.5)	1099 (67.1)	74 (48.7)
	Not measured	11 (0.6)	10 (0.6)	1 (0.7)
Platelets ($10^3/\mu$L)	<150	219 (12.2)	187 (11.4)	32 (21.1)
	>150	1563 (87.3)	1443 (88.1)	120 (78.9)
	Not measured	8 (0.4)	8 (0.5)	0 (0.0)

TABLE 2. Machine Learning Model Performance for 30-d Survival Prediction in Patients Undergoing Operative Intervention for Spinal Metastatic Disease, Training Set, n = 1432

Method	Metric	Machine learning algorithm			
		Neural network	Support vector machine	Bayes point machine	Decision tree
Discrimination	C-statistic	0.769	0.758	0.768	0.760
Calibration	Calibration slope	0.941	0.938	1.013	0.728
	Calibration intercept	0.000	−0.002	−0.009	0.004
Overall	Brier score	0.0706	0.0709	0.0701	0.0711
	Null model Brier score		0.079		

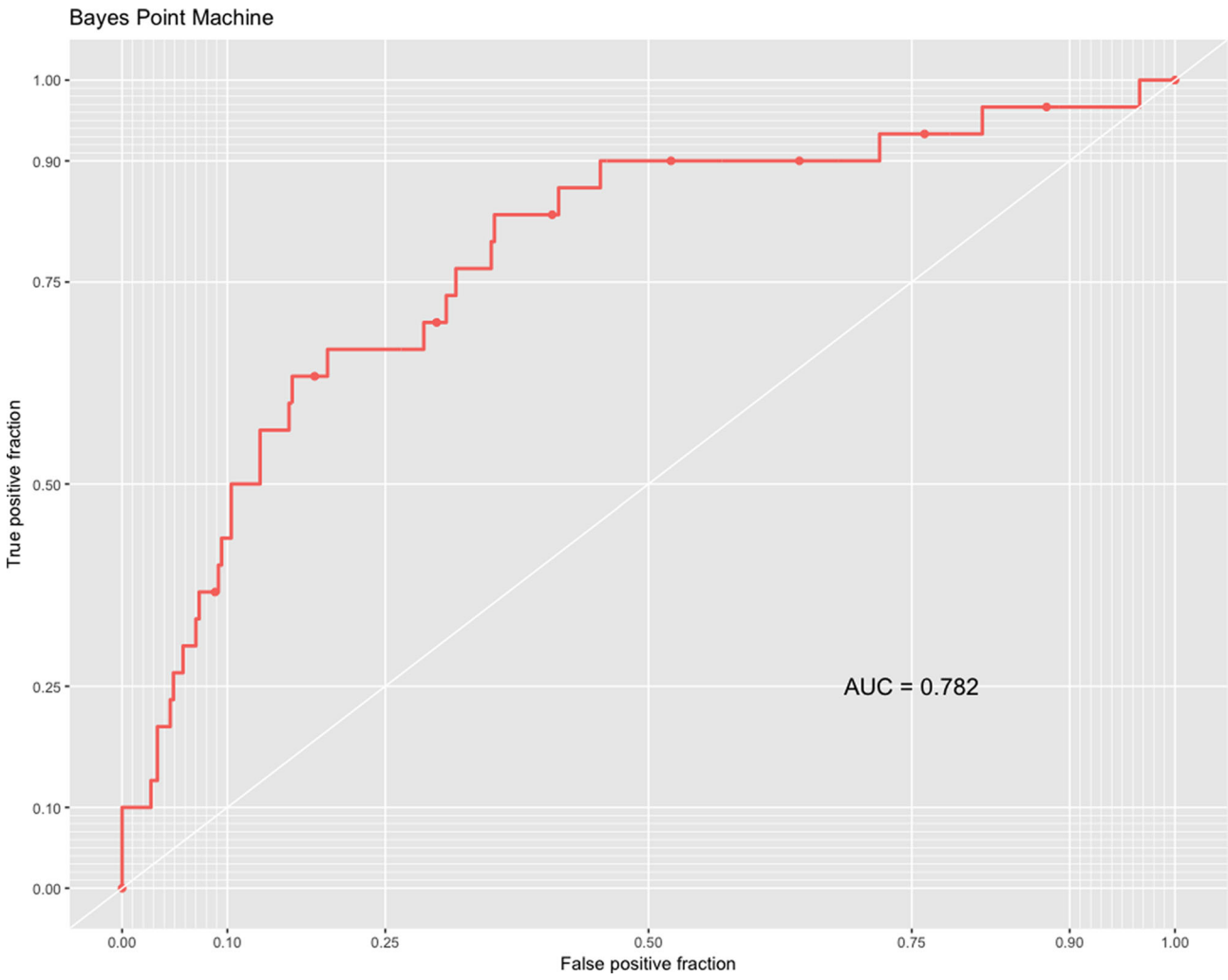

FIGURE 1. *Receiver operating curve by model for prediction of 30-d mortality; testing set, n = 358. AUC = area under the curve.*

33 (18%) spine patients.[55] Our study is the first algorithm for mortality specific to patients with spinal metastatic disease that relied on machine learning algorithms incorporating spine-specific variables derived from a large, multi-institutional dataset. This may enhance the generalizability of the algorithm presented here.

The discrimination of the predictive algorithms developed in this study approximated that of previous models proposed for predicting outcomes after operative intervention for spinal metastases.[7,8,23,25] However, as the purpose of this study was to create a clinically useful decision tool, model calibration was central to evaluating model performance in this study, numerically and graphically.[47] In comparison to previous studies, our study was one of the few that assessed model calibration with calibration plots and the only study that assessed calibration slope and intercept. Assessing model calibration graphically was crucial in that the model with the best discrimination, the Neural Network, was demonstrated to be inferior to the second best performing model, Bayes Point Machine, over the full range of predicted 30-d mortality. Future studies seeking to build predictive models should incorporate graphical and numerical assessment of model calibration as a key component of model performance.

Limitations

There are several limitations to the work presented in this study. Although ACS-NSQIP has been widely employed in a number of contexts, the data veracity and data completeness are variable and may not be as stringent as data prospectively collected for a specific research protocol. This study was a retrospective analysis of the NSQIP database and the limitations of

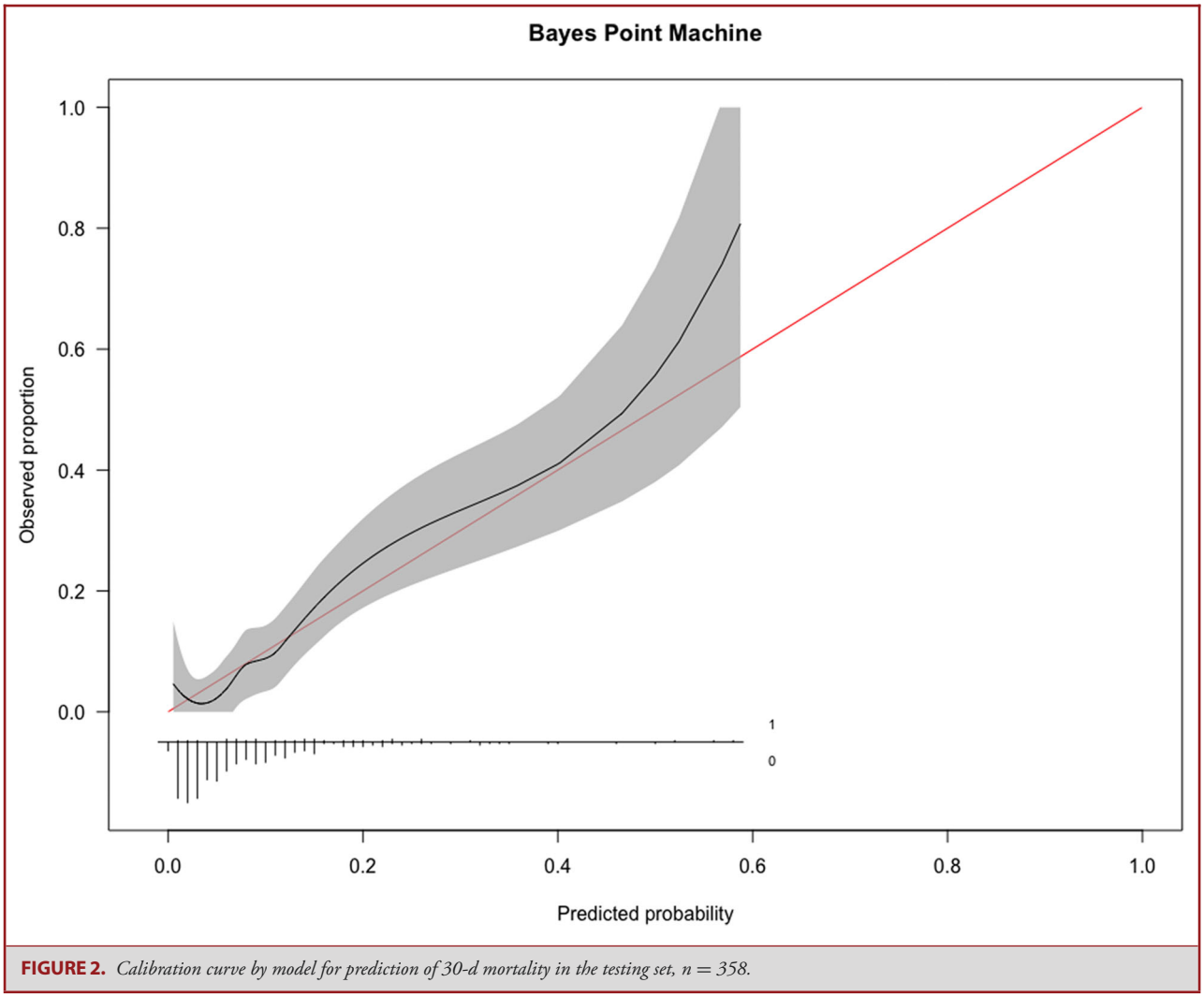

FIGURE 2. *Calibration curve by model for prediction of 30-d mortality in the testing set, n = 358.*

retrospective research must be considered when interpreting the findings presented here. In addition, predictors that may be pertinent to short-term survival prediction in this population, such as primary tumor histology and metastatic tumor burden (bone, lung, liver, brain), cannot reliably be extracted from ACS-NSQIP. Verlaan et al[34] studied 1266 patients in a prospective, longitudinal study at 23 international spine centers from 2001 to 2014 and did not find primary tumor histology or the presence of brain metastases significantly associated with survival less than 3 mo on multivariable analysis. Similarly, Schoenfeld et al[35] only found nutritional status and ambulatory status to be significantly associated with 30-d mortality. NSQIP also does not record the overall trajectory of metastatic disease prior to operative intervention; for example, history of local radiation, history of systemic therapy, and recurrence of tumor are not captured in the NSQIP database. These are significant factors in decision making and should be evaluated by future prospective studies. Although the ACS-NSQIP collects data from a variety of centers, with the 2016 data drawing from 600+ hospitals, the patients included in the ACS-NSQIP database may not reflect the demographic and clinical characteristics of patients for which these models are ultimately used; healthcare professionals should be aware of these differences while seeking to interpret the probabilities developed from this analysis.

Furthermore, the machine learning models built in this study are optimized for highly accurate prediction but not for explanation. Unlike logistic regression, the model parameters of the machine learning algorithms created in this study cannot be simply deployed for explanatory purposes of the independent effect of individual risk factors on 30-d mortality. In addition, this study did not examine multivariate logistic regression or proportional hazards models. There is a need for future studies to examine the predictive performance of these methods relative to the algorithms presented here. Logistic regression

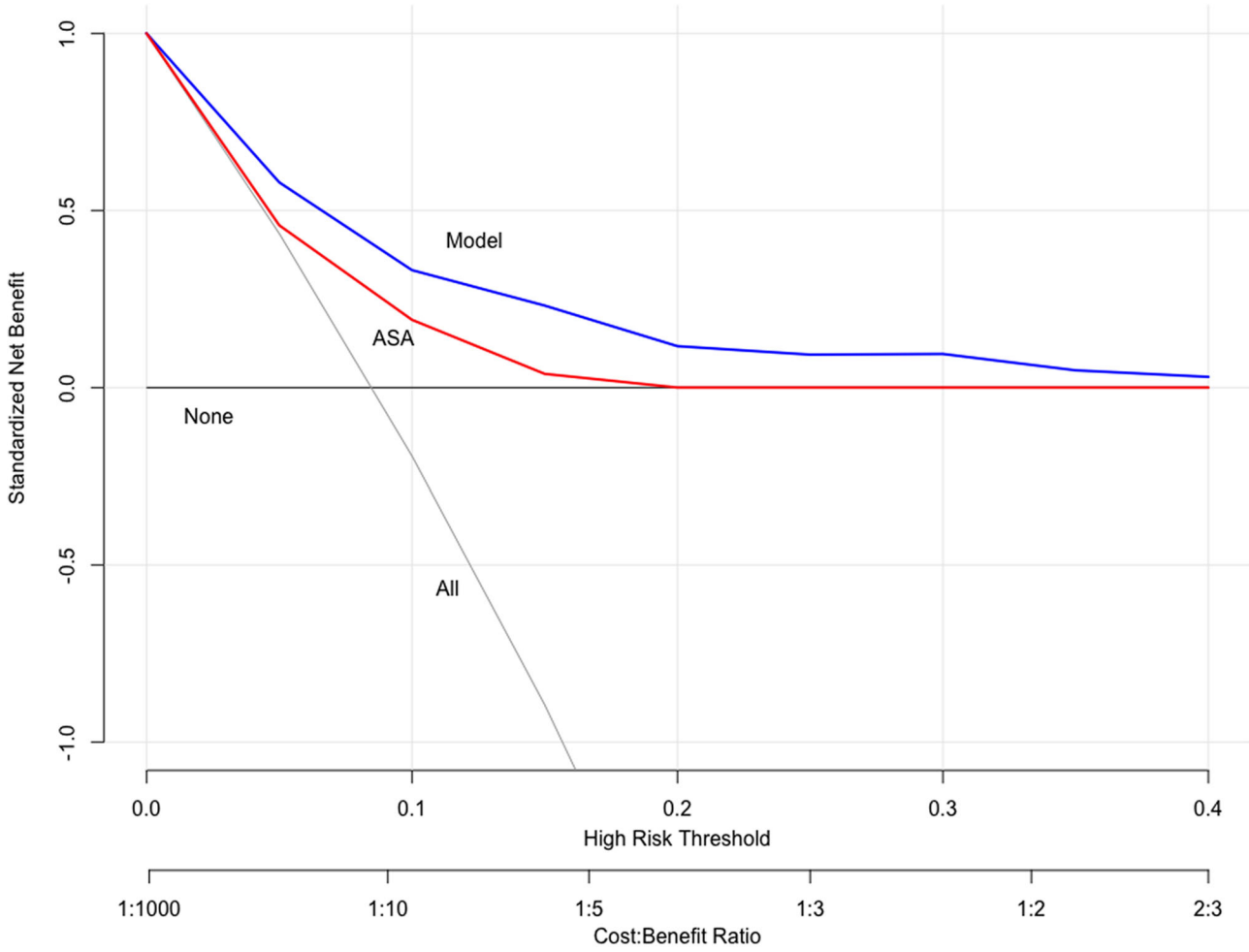

FIGURE 3. *Decision analysis curve for (1) changing management for all patients with spinal metastatic disease undergoing spine surgery, (2) no patients, and (3) based on American Society of Anesthesiologist (ASA) class alone, or (4) based on the Bayes model.*

and proportional hazards models have significant strengths in explanation and prediction, and future comparative studies can provide further recommendations on appropriateness of model selection in spinal oncology. Lastly, it should be acknowledged that there are other outcomes beyond short-term mortality that contribute to decision making for spine surgery including complication profile, postoperative functional status, and neurological function. The ability of the machine learning algorithms developed here to predict the impact of surgical intervention on these outcomes remains to be determined.

Nonetheless, this study fulfilled its primary objective of creating a discriminative and well-calibrated model for prediction of short-term mortality after surgery for spinal metastatic disease. This study achieved another milestone by creating an open access web application for healthcare professionals to access and use these computational models directly. For now, this web-based application exists as a separate tool similar to existing cardiovascular risk calculators and other surgical risk calculators based on regression analysis and nomograms.[56,57] However, one accomplishment of this study has been to preserve the complexity of the computational model while allowing the model to be accessed from a simple interface. Programming of computational interfaces in this manner may serve as a template for integration into modern electronic health systems and for this capability to be part and parcel of computationally and digitally enabled medicine. The creation of learning healthcare systems has been previously proposed, and this method of predictive algorithm creation and deployment can be one step in progress toward that goal.[58,59]

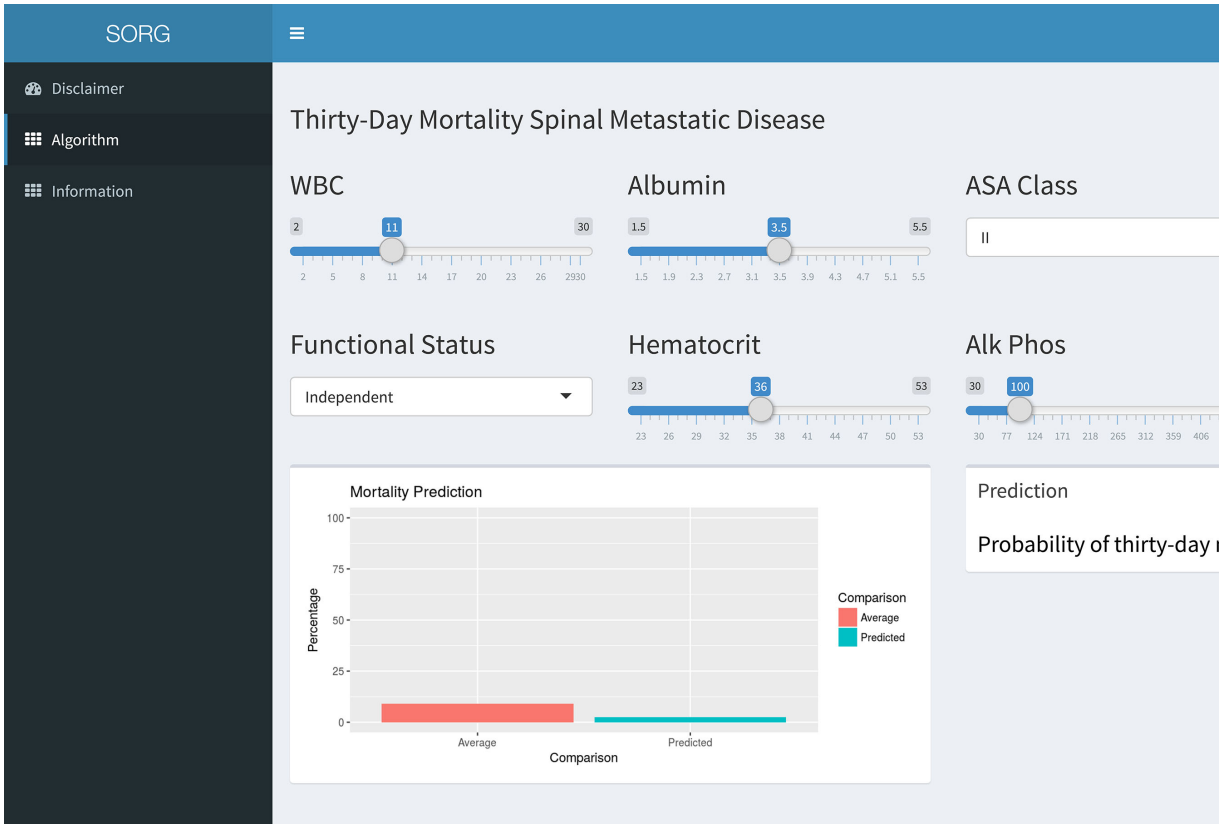

FIGURE 4. *Open access web application: https://sorg-apps.shinyapps.io/spinemets/.*

CONCLUSION

Machine learning algorithms are promising for prediction of postoperative outcomes in spinal oncology and these algorithms can be integrated into clinically useful decision tools. As the volume of data in oncology continues to grow, creation of learning systems and deployment of these systems as accessible tools may significantly enhance prognostication and management.

Disclosure

The authors have no personal, financial, or institutional interest in any of the drugs, materials, or devices described in this article.

REFERENCES

1. Barzilai O, Laufer I, Yamada Y, et al. Integrating Evidence-Based medicine for treatment of spinal metastases into a decision framework: neurologic, oncologic, mechanicals stability, and systemic disease. *J Clin Oncol*. 2017;35(21):2419-2427.
2. Kelly ML, Kshettry VR, Rosenbaum BP, Seicean A, Weil RJ. Effect of a randomized controlled trial on the surgical treatment of spinal metastasis, 2000 through 2010: a population-based cohort study. *Cancer*. 2014;120(6):901-908.
3. Yoshihara H, Yoneoka D. Trends in the surgical treatment for spinal metastasis and the in-hospital patient outcomes in the United States from 2000 to 2009. *Spine J*. 2014;14(9):1844-1849.
4. Patchell RA, Tibbs PA, Regine WF, et al. Direct decompressive surgical resection in the treatment of spinal cord compression caused by metastatic cancer: a randomised trial. *Lancet (London, England)*. 2005;366(9486):643-648.
5. Tokuhashi Y, Matsuzaki H, Oda H, Oshima M, Ryu J. A revised scoring system for preoperative evaluation of metastatic spine tumor prognosis. *Spine*. 2005;30(19):2186-2191.
6. Leithner A, Radl R, Gruber G, et al. Predictive value of seven preoperative prognostic scoring systems for spinal metastases. *Eur Spine J*. 2008;17(11):1488-1495.
7. Schoenfeld AJ, Le HV, Marjoua Y, et al. Assessing the utility of a clinical prediction score regarding 30-day morbidity and mortality following metastatic spinal surgery: the New England Spinal Metastasis Score (NESMS). *Spine J*. 2016;16(4):482-490.
8. Pereira NRP, Janssen SJ, van Dijk E, et al. Development of a prognostic survival algorithm for patients with metastatic spine disease. *J Bone Joint Surg-Am Vol*. 2016;98(21):1767-1776.
9. Lei MX, Liu YS, Tang CH, Yang SX, Liu SB, Zhou SG. Prediction of survival prognosis after surgery in patients with symptomatic metastatic spinal cord compression from non-small cell lung cancer. *BMC Cancer*. 2015;15:853. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4635615/pdf/12885_2015_Article_1852.pdf
10. Verlaan JJ, Choi D, Versteeg A, et al. Characteristics of patients who survived < 3 months or > 2 Years after surgery for spinal metastases: Can we avoid inappropriate patient selection? *J Clin Oncol*. 2016;34(25):3054-3061.
11. Luksanapruksa P, Buchowski JM, Hotchkiss W, Tongsai S, Wilartratsami S, Chotivichit A. Prognostic factors in patients with spinal metastasis: a systematic review and meta-analysis. *Spine J*. 2017;17(5):689-708.
12. Holman PJ, Suki D, McCutcheon I, Wolinsky J-P, Rhines LD, Gokaslan ZL. Surgical management of metastatic disease of the lumbar spine: experience with 139 patients. *J Neurosurg Spine*. 2005;2(5):550-563.
13. Bauer HC, Wedin R. Survival after surgery for spinal and extremity metastases: prognostication in 241 patients. *Acta Orthop Scand*. 1995;66(2):143-146.

14. Hirabayashi H, Ebara S, Kinoshita T, et al. Clinical outcome and survival after palliative surgery for spinal metastases. *Cancer*. 2003;97(2):476-484.
15. Tokuhashi Y, Matsuzaki H, Toriyama S, Kawano H, Ohsaka S. Scoring system for the preoperative evaluation of metastatic spine tumor prognosis. *Spine*. 1990;15(11):1110-1113.
16. Ghogawala Z, Mansfield FL, Borges LF. Spinal radiation before surgical decompression adversely affects outcomes of surgery for symptomatic metastatic spinal cord compression. *Spine*. 2001;26(7):818-824.
17. Wai EK, Finkelstein JA, Tangente RP, et al. Quality of life in surgical treatment of metastatic spine disease. *Spine*. 2003;28(5):508-512.
18. Sioutos PJ, Arbit E, Meshulam CF, Galicich JH. Spinal metastases from solid tumors. Analysis of factors affecting survival. *Cancer*. 1995;76(8):1453-1459.
19. Ulmar B, Naumann U, Catalkaya S, et al. Prognosis scores of Tokuhashi and Tomita for patients with spinal metastases of renal cancer. *Ann Surg Oncol*. 2007;14(2):998-1004.
20. Mizumoto M, Harada H, Asakura H, et al. Prognostic factors and a scoring system for survival after radiotherapy for metastases to the spinal column. *Cancer*. 2008;113(10):2816-2822.
21. Sciubba DM, Gokaslan ZL, Suk I, et al. Positive and negative prognostic variables for patients undergoing spine surgery for metastatic breast disease. *Eur Spine J*. 2007;16(10):1659-1667.
22. van der Linden YM, Dijkstra SP, Vonk EJ, Marijnen CA, Leer JW. Prediction of survival in patients with metastases in the spinal column: results based on a randomized trial of radiotherapy. *Cancer*. 2005;103(2):320-328.
23. Ghori AK, Leonard DA, Schoenfeld AJ, et al. Modeling 1-year survival after surgery on the metastatic spine. *Spine J*. 2015;15(11):2345-2350.
24. Tomita K, Kawahara N, Kobayashi T, Yoshida A, Murakami H, Akamaru T. Surgical strategy for spinal metastases. *Spine*. 2001;26(3):298-306.
25. 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.
26. Lambin P, Roelofs E, Reymen B, et al. 'Rapid Learning health care in oncology' - An approach towards decision support systems enabling customised radiotherapy'. *Radiother Oncol*. 2013;109(1):159-164.
27. Menden MP, Iorio F, Garnett M, et al. Machine learning prediction of cancer cell sensitivity to drugs based on genomic and chemical properties. *PLoS One*. 2013;8(4):e61318.
28. Polat K, Sahan S, Kodaz H, Gunes S. Breast cancer and liver disorders classification using Artificial Immune Recognition System (AIRS) with performance evaluation by fuzzy resource allocation mechanism. *Expert Syst Appl*. 2007;32(1):172-183.
29. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017;542(7639):115.
30. Luo W, Phung D, Tran T, et al. Guidelines for developing and reporting machine learning predictive models in biomedical research: A multidisciplinary view. *J Med Inter Res*. 2016;18(12):e323.
31. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMC Med*. 2015;13(1):1.
32. Karhade AV, Larsen AMG, Cote DJ, Dubois HM, Smith TR. National databases for neurosurgical outcomes research: Options, strengths, and limitations. *Neurosurgery*. 2018;83(3):333-344.
33. Finkelstein J, Zaveri G, Wai E, Vidmar M, Kreder H, Chow E. A population-based study of surgery for spinal metastases: survival rates and complications. *J Bone Joint Surg Br*. 2003;85(7):1045-1050.
34. Verlaan J-J, Choi D, Versteeg A, et al. Characteristics of patients who survived < 3 months or >2 years after surgery for spinal metastases: can we avoid inappropriate patient selection? *J Clin Oncol*. 2016;34(25):3054-3061.
35. Schoenfeld AJ, Leonard DA, Saadat E, Bono CM, Harris MB, Ferrone ML. Predictors of 30- and 90-Day survival following surgical intervention for spinal metastases: a prognostic study conducted at four academic centers. *Spine (Phila Pa 1976)*. 2016;41(8):E503-E509.
36. Hussain AK, Vig KS, Cheung ZB, et al. The impact of metastatic spinal tumor location on 30-day perioperative mortality and morbidity after surgical decompression. *Spine (Phila Pa 1976)*. 2018;43(11):E648-E655.
37. Seaman E, Goluboff ET, Ross S, Sawczuk IS. Association of radionuclide bone scan and serum alkaline phosphatase in patients with metastatic renal cell carcinoma. *Urology*. 1996;48(5):692-695.
38. Aruga A, Koizumi M, Hotta R, Takahashi S, Ogata E. Usefulness of bone metabolic markers in the diagnosis and follow-up of bone metastasis from lung cancer. *Br J Cancer*. 1997;76(6):760.
39. Roodman GD. Mechanisms of bone metastasis. *N Engl J Med*. 2004;350(16):1655-1664.
40. Paulino Pereira NR, McLaughlin L, Janssen SJ, et al. The SORG nomogram accurately predicts 3- and 12-months survival for operable spine metastatic disease: external validation. *J Surg Oncol*. 2017;115(8):1019-1027.
41. Suppiah R, Shaheen PE, Elson P, et al. Thrombocytosis as a prognostic factor for survival in patients with metastatic renal cell carcinoma. *Cancer*. 2006;107(8):1793-1800.
42. Gay LJ, Felding-Habermann B. Contribution of platelets to tumour metastasis. *Nat Rev Cancer*. 2011;11(2):123.
43. Kuhn M, Johnson K. *Applied Predictive Modeling*. Vol. 26. New York: Springer; 2013.
44. Wainer J. Comparison of 14 different families of classification algorithms on 115 binary datasets. arXiv preprint arXiv:1606.00930. 2016.
45. Maroco J, Silva D, Rodrigues A, Guerreiro M, Santana I, de Mendonça A. Data mining methods in the prediction of dementia: a real-data comparison of the accuracy, sensitivity and specificity of linear discriminant analysis, logistic regression, neural networks, support vector machines, classification trees and random forests. *BMC Res Notes*. 2011;4(1):299.
46. Fernández-Delgado M, Cernadas E, Barro S, Amorim D. Do we need hundreds of classifiers to solve real world classification problems. *J Mach Learn Res*. 2014;15(1):3133-3181.
47. Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation*. 2007;115(7):928-935.
48. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*. 1982;143(1):29-36.
49. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *Eur Heart J*. 2014;35(29):1925-1931.
50. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for some traditional and novel measures. *Epidemiology (Cambridge, Mass)*. 2010;21(1):128.
51. Brier GW. Verification of forecasts expressed in terms of probability. *Monthly Weather Review*. 1950;78(1):1-3.
52. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making*. 2006;26(6):565-574.
53. Kerr KF, Brown MD, Zhu K, Janes H. Assessing the clinical impact of risk prediction models with decision curves: guidance for correct interpretation and appropriate use. *J Clin Oncol*. 2016;34(21):2534-2540.
54. 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(4):1252-1261.
55. Piccioli A, Spinelli MS, Forsberg JA, et al. 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(1):424.
56. Bilimoria KY, Liu Y, Paruch JL, et al. Development and evaluation of the universal ACS NSQIP surgical risk calculator: a decision aid and informed consent tool for patients and surgeons. *J Am Coll Surg*. 2013;217(5):833-842. e833.
57. Lloyd-Jones DM, Huffman MD, Karmali KN, et al. Estimating longitudinal risks and benefits from cardiovascular preventive therapies among medicare patients: The million hearts longitudinal ASCVD risk assessment tool: A special report from the american heart association and american college of cardiology. *Circulation*. 2017;135(13):e793-e813.
58. McGinnis JM, Aisner D, Olsen L. *The Learning Healthcare System: Workshop Summary*. Washington, DC: The National Academies Press; 2007.
59. Friedman CP, Wong AK, Blumenthal D. Achieving a nationwide learning health system. *Sci Transl Med*. 2010;2(57):57cm29.

Acknowledgment

The American College of Surgeons National Surgical Quality Improvement Program and the hospitals participating in the ACS NSQIP are the source of the data used herein; they have not verified and are not responsible for the statistical validity of the data analysis or the conclusions derived by the authors.