Performance assessment of the metastatic spinal tumor frailty index using machine learning algorithms: limitations and future directions

Elie Massaad, MD,1 Natalie Williams, BA,1 Muhamed Hadzipasic, MD, PhD,1 Shalin S. Patel, MD,2 Mitchell S. Fourman, MD, MPhil,2 Ali Kiapour, PhD,1 Andrew J. Schoenfeld, MD, MSc,3 Ganesh M. Shankar, MD, PhD,1 and John H. Shin, MD1

Departments of 1Neurosurgery and 2Orthopedic Surgery, Massachusetts General Hospital; and 3Department of Orthopedic Surgery, Brigham and Women’s Hospital, Harvard Medical School, Boston, Massachusetts

OBJECTIVE Frailty is recognized as an important consideration in patients with cancer who are undergoing therapies, including spine surgery. The definition of frailty in the context of spinal metastases is unclear, and few have studied such markers and their association with postoperative outcomes and survival. Using national databases, the metastatic spinal tumor frailty index (MSTFI) was developed as a tool to predict outcomes in this specific patient population and has not been tested with external data. The purpose of this study was to test the performance of the MSTFI with institutional data and determine whether machine learning methods could better identify measures of frailty as predictors of outcomes.

METHODS Electronic health record data from 479 adult patients admitted to the Massachusetts General Hospital for metastatic spinal tumor surgery from 2010 to 2019 formed a validation cohort for the MSTFI to predict major complications, in-hospital mortality, and length of stay (LOS). The 9 parameters of the MSTFI were modeled in 3 machine learning algorithms (lasso regularization logistic regression, random forest, and gradient-boosted decision tree) to assess clinical outcome prediction and determine variable importance. Prediction performance of the models was measured by computing areas under the receiver operating characteristic curve (AUROCs), calibration, and confusion matrix metrics (positive predictive value, sensitivity, and specificity) and was subjected to internal bootstrap validation.

RESULTS Of 479 patients (median age 64 years [IQR 55–71 years]; 58.7% male), 28.4% had complications after spine surgery. The in-hospital mortality rate was 1.9%, and the mean LOS was 7.8 days. The MSTFI demonstrated poor discrimination for predicting complications (AUROC 0.56, 95% CI 0.50–0.62) and in-hospital mortality (AUROC 0.69, 95% CI 0.54–0.85) in the validation cohort. For postoperative complications, machine learning approaches showed a greater advantage over the logistic regression model used to develop the MSTFI (AUROC 0.62, 95% CI 0.56–0.68 for random forest vs AUROC 0.56, 95% CI 0.50–0.62 for logistic regression). The random forest model had the highest positive predictive value (0.53, 95% CI 0.43–0.64) and the highest negative predictive value (0.77, 95% CI 0.72–0.81), with chronic lung disease, coagulopathy, anemia, and malnutrition identified as the most important predictors of postoperative complications.

CONCLUSIONS This study highlights the challenges of defining and quantifying frailty in the metastatic spine tumor population. Further study is required to improve the determination of surgical frailty in this specific cohort.

https://thejns.org/doi/abs/10.3171/2021.2.FOCUS201113

KEYWORDS metastasis; frailty; sarcopenia; spine surgery; complications; machine learning

Pain and disability related to spinal cord compression and mechanical instability can have profound consequences for patients with metastatic cancer.1,2 Although cancer is a major cause of death worldwide, innovations in radiotherapy, surgery, and systemic therapies have improved survival rates.3–5 Spine surgery has a palliative role, but complications and prolonged recovery can negatively impact quality of life, resumption or initiation of systemic therapies, and survival.3–8 Decision-making tools like the neurological, oncological, mechanical, and systemic (NOMS) system and Spine Instability Neoplastic Score (SINS) provide conceptual frameworks for surv-
Frailty, a concept generally associated with the elderly, describes the multifactorial deterioration of energy metabolism, strength, endurance, and function that is worse than expected for a given age. Chronological age has been studied as a marker of individual's vulnerability for increased dependency and death. Chronological age differs from chronological age, and frailty increases an individual’s vulnerability for increased dependency and death. Frailty is independently linked to adverse outcomes such as complications, increased LOS, and decreased survival. Irrespective of age, frailty has additional implications for immunologically and physiologically stressed cancer patients and their risks for surgery.

Few studies have examined frailty as a predictor of outcomes in the context of metastatic spine tumor surgery. The problem with using frailty in this population is that it is a relative term and estimates are based on comparison to a referent. In this case, when a population is frail to some degree, determinations may be faulty. Recently, De la Garza Ramos et al. proposed a novel index called the metastatic spinal tumor frailty index (MSTFI) to predict postoperative complications, in-hospital mortality, and length of stay (LOS). The authors developed this index using the National (Nationwide) Inpatient Sample (NIS) and further validated it with the American College of Surgeons National Surgical Quality Improvement Program (ACS NSQIP). These authors found significant associations between frailty severity and risk of complications, mortality, and LOS, with a sevenfold increase in the odds for developing a major complication in the severe frailty class.

In this context, we sought to externally apply the MSTFI to our institutional database and study its ability to predict the outcomes of interest—postoperative complications, in-patient mortality, and LOS. We also sought to examine whether machine learning algorithms could improve on the methodological limitations of the index to identify the main predictors of frailty in spinal metastases.

### Methods

This retrospective study included medical record data from consecutive patients, older than 18 years of age, who were surgically treated for spinal metastases between 2010 and 2019 at the Massachusetts General Hospital. Included patients underwent surgery for decompression and stabilization of the spine for a diagnosis of epidural spinal cord compression, vertebral pathologic fracture, spinal instability, neurological deficit, and/or intractable pain. Patients were excluded if they were treated nonsurgically. Variables, including patient demographics, primary tumor type, details of surgery, laboratory values, oncological treatment history, performance status, complications, and mortality, were recorded in a longitudinal database. This study followed the Transparent Reporting of a Multi-variable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) reporting guideline for prognostic studies.

Institutional review board approval was obtained and informed consent for retrospective analysis of de-identified data was waived.

### MSTFI Parameters

The 9 parameters of the MSTFI include anemia, chronic lung disease, coagulopathy (thrombocytopenia, hemorrhagic conditions, genetic coagulopathies), electrolyte abnormalities (hyper-/hyponatremia, hyper-/hypokalemia, acidosis, alkalosis), pulmonary circulation disorders, renal impairment and failure, malnutrition, emergency/urgent admission, and an anterior or combined surgical approach.

The MSTFI was calculated for each individual, and frailty classifications were assigned according to published methodology. One point was assigned for each component, and 2 points were assigned for pulmonary circulation disorders, with a maximum score of 10. Patients were categorized as having no frailty (MSTFI = 0), mild frailty (MSTFI = 1), moderate frailty (MSTFI = 2), and severe frailty (MSTFI ≥ 3) (Table 1).

### Outcomes

The primary outcome measure was the occurrence of at least 1 major postoperative complication within 30 days of surgery. Complications included acute respiratory distress syndrome, cardiac arrest, deep vein thrombosis, myocardial infarction, neurological deficits, pneumonia, pulmonary embolism, renal failure, sepsis, shock, stroke, unplanned reintubation, and urinary tract infections. Secondary outcome measures included in-hospital mortality and postoperative LOS.
Statistical Analysis

External Validation of the MSTFI

The main objective of this study was to assess performance of the MSTFI with observed and predicted complications. Summary statistics were performed and recorded for continuous and categorical variables. Calculated MSTFI scores were matched to each patient’s complications, in-hospital mortality, and LOS. To evaluate the association of frailty severity with complications and in-hospital mortality, the relative risk (RR) for each outcome was calculated. ANOVA was performed to test the association of frailty severity and prolonged LOS. All p values were from 2-sided tests, and results were statistically significant at p < 0.0027 after Bonferroni correction to account for multiple comparisons. We measured the prediction performance of each model by computing 1) the area under the receiver operating characteristic curve (AUROC) and 2) the calibration-in-the-large coefficient and calibration slope. The AUROC measures the ability of the MSTFI to differentiate between patients who reached the outcome of interest and those who did not; values close to 1 indicated very good discriminative ability, while values close to 0.5 indicated poor discriminative ability. A calibration slope of 1 in combination with a calibration-in-the-large coefficient of 0 indicated excellent overall calibration. Subgroup analysis by tumor type was also performed.

Model Building and Validation Using MSTFI Variables

Using this data set, we developed the reference MSTFI using logistic regression as well as 3 machine learning models to assess the following predictive probability of postoperative complications: 1) logistic regression with lasso regularization (lasso regression), 2) random forest, and 3) gradient-boosted decision tree. We performed 10-fold cross-validation to find the optimal regularization coefficient lambda that provided minimal misclassification error in lasso regression. Random forest and gradient boosting used an ensemble of decision trees based on random feature selection and gradient descent, respectively. We used the caret package in R (The R Foundation) to identify the best combination of hyperparameters for each model. Additional details are provided in Supplementary Methodology.

The discrimination ability of each model was calculated by the AUROC. A DeLong test was used to the compare AUROC in reference to the logistic regression model (p < 0.05). An internal validation of the 4 models was performed by bootstrapping with 500 replicates (with replacement) to estimate optimism (the level of model overfitting). Optimism was estimated by comparing the final model performance to the original data for each bootstrap sample. Classification performance was calculated including sensitivity, specificity, positive and negative predictive values, and positive and negative likelihood ratios. To address the class imbalance in the primary outcome (i.e., the low proportion of postoperative complications), the threshold probability for outcome classification was based on the AUROC at the point of optimal balance of false positives and false negatives to provide the value with the shortest distance to the perfect model.

To evaluate the contribution of each predictor to the machine learning models, variable importance was determined by the coefficient absolute value for the logistic regression model and selection frequency for the random forest and gradient-boosting models. All analyses were performed with R statistical software version 3.6.1 (The R Foundation).

Results

A total of 479 patients were included. The most common tumors were lung (18.8%), kidney (10.7%), breast (9.8%), prostate (8.8%), and thyroid (4.8%) (Table 2). The median age was 64 years (IQR 55–71 years), and 41.3% of patients were female. Of all patients, 88.5% (424/479) underwent posterior approaches, while anterior and combined (anterior and posterior) procedures accounted for

---

### Table 2. Characteristics of patients who had spine surgery for metastatic disease

| Variable                          | Value   |
|-----------------------------------|---------|
| Age, median, yrs (IQR)            | 64 (55–71) |
| Age, mean, yrs (SD)               | 62.4 (12.1) |
| Sex, n (%)                        |         |
| Female                            | 198 (41.3) |
| Male                              | 281 (58.7) |
| Primary tumor, n (%)              |         |
| Lung                              | 90 (18.8) |
| Kidney                            | 51 (10.7) |
| Breast                            | 47 (9.8) |
| Blood (HEME)                      | 44 (9.2) |
| Prostate                          | 42 (8.8) |
| Colon/rectal                      | 28 (5.9) |
| Bone sarcoma                      | 24 (5.0) |
| Thyroid                           | 23 (4.8) |
| Head & neck                       | 22 (4.6) |
| Skin                              | 21 (4.4) |
| Other                             | 86 (17.9) |
| MSTFI parameters, n (%)           |         |
| Emergency/urgent                  | 294 (61.5) |
| Anemia                            | 263 (59.1) |
| Electrolyte abnormalities          | 114 (23.8) |
| Coagulopathy                      | 86 (18.0) |
| Chronic lung disease              | 69 (14.4) |
| Anterior or combined (ant/post)    | 55 (11.5) |
| Renal failure                     | 54 (11.3) |
| Malnutrition                      | 51 (10.6) |
| Pulmonary circulation disorders    | 30 (6.3) |
| MSTFI category, n (%)             |         |
| No frailty                        | 31 (6.5) |
| Mild frailty                      | 111 (23.2) |
| Moderate frailty                  | 162 (33.8) |
| Severe frailty                    | 175 (36.5) |

ant = anterior; post = posterior.
4.6% and 6.9%, respectively. Most surgeries (294/479, 61.5%) were emergency/urgent. In total, 31 patients (6.5%) were classified as having no frailty (MSTFI score 0), 111 (23.2%) as having mild frailty (MSTFI score 1), 162 (33.8%) as having moderate frailty (MSTFI score 2), and 175 (36.5%) as having severe frailty (MSTFI score ≥ 3).

### Complication Rates

The overall complication rate, excluding wound-related complications, was 28.4% (136/479). Deep vein thrombosis and PE were most common (36/136, 26.5%) (Table 3). Of note, wound-related complications were not included in the original MSTFI composite outcomes. Patients with severe frailty had the highest rate of complications after spine surgery (49/136, 36.1%) compared with patients with less severe frailty (Fig. 1). Variation in distribution of complications by frailty groups and tumor type was noted (Fig. 2). Frailty was associated with a nonsignificant increased risk of complications after spine surgery.

| Variable                                      | No. (%) |
|-----------------------------------------------|---------|
| Total no. of complications excluding wound complications | 136 (28.4) |
| Deep vein thrombosis/pulmonary embolism       | 36 (26.5) |
| Urinary tract infections                      | 31 (22.8) |
| Return to OR                                  | 18 (13.2) |
| Neurological deficit                          | 17 (12.5) |
| Respiratory tract infections                  | 16 (11.8) |
| Myocardial infarction                         | 2 (1.5)  |
| Total no. of complications including wound complications | 176 (36.7) |
| Wound complications                           | 40 (8.4)  |

OR = operating room.

**TABLE 3. Complications after spine surgery for metastatic disease**

![Complications, in-hospital mortality, and LOS according to frailty by MSTFI severity groups.](image)
Compared with patients with no frailty, patients with mild frailty (RR 1.05, 95% CI 0.51–2.20), moderate frailty (RR 1.28, 95% CI 0.64–2.57), and severe frailty (RR 1.48, 95% CI 0.74–2.93) had a nonsignificant increased risk of developing a major complication.

The MSTFI demonstrated poor discrimination for having complications after spine surgery, with an AUROC of 0.56 (95% CI 0.50–0.62) in the validation cohort. In addition, the MSTFI was poorly calibrated (calibration intercept 0.23, p < 0.001; calibration slope 0.25, p = 0.09) for the predicted and observed probabilities of having a postoperative complication. Subgroup analysis showed that MSTFI most effectively predicted postoperative complications in prostate cancer (AUROC 0.68) compared with breast (AUROC 0.52, p < 0.001), kidney (AUROC 0.54, p < 0.001), hematological (AUROC 0.54, p < 0.001), and lung cancer (AUROC 0.64, p < 0.001) metastases (Fig. 3).

**In-Hospital Mortality and LOS**

The in-hospital mortality rate was 1.9%. Frailty was associated with a nonsignificant increased risk of in-hospital mortality. There were no in-hospital deaths in the no-frailty group. Patients with severe frailty had a nonsignificant increased risk of in-hospital death compared with patients with moderate frailty (RR 6.46, 95% CI 0.81–52.09) and mild frailty (RR 4.44, 95% CI 0.55–35.6). The AUROC was 0.69 (95% CI 0.54–0.85) for predicting in-hospital mortality demonstrating that the MSTFI performed fairly across severity risk groups. The calibration slope (0.58, p = 0.06) was less than 1, indicating the MSTFI was overestimating the predicted risk of in-hospital mortality for high-risk patients (severe frailty), while underestimating for patients at low risk. In parallel, frailty was not associated with LOS (the mean LOS for no frailty was 6.2 days; mild frailty, 7.3 days; moderate frailty, 7.9 days; and severe frailty, 8.9 days).

**Prediction of Postoperative Complications Using Machine Learning**

The discrimination ability of the 3 machine learning models as represented by AUROC is shown in Table 4. The random forest model had the highest discriminative ability (AUROC 0.62, 95% CI 0.56–0.68; p = 0.18), while the other models had a nonsignificant difference in AUROC (lasso regularization 0.53, 95% CI 0.50–0.58; and gradient-boosted decision tree 0.54, 95% CI 0.48–0.60). Additionally, compared with the reference model, all machine learning models had a higher sensitivity (0.21,
95% CI 0.14–0.28 in the reference model vs 0.37, 95% CI 0.29–0.45 in random forest) to predict postoperative complications (Table 4). The reference model had a slightly higher specificity (0.90, 95% CI 0.87–0.93) compared with the random forest machine learning model (0.87, 95% CI 0.83–0.90). Random forest had the highest positive predictive value (0.53, 95% CI 0.43–0.64), and highest negative predictive value (0.77, 95% CI 0.72–0.81). The top variables of importance for the random forest and gradient-boosting models (chronic lung disease, coagulopathy, anemia, and malnutrition) are shown in Fig. 4.

Discussion

In this study, we sought to externally validate the MSTFI with institutional data to determine if it accurately predicts postoperative complications, in-patient mortality, and LOS. The results of this analysis demonstrate that the predicted outcomes were not consistent with observed outcomes when stratified by frailty type. The results show that model performance of MSTFI for predicting at least 1 major complication after spine surgery (AUROC 0.56, 95% CI 0.50–0.62) did not match the model performance in the original development study (AUROC 0.67). Furthermore, prediction performance varied by tumor type, as greater discrimination was seen for complications related to prostate than breast cancer metastases (AUROC 0.68 vs 0.52, p < 0.001). This finding for prostate cancer (AUROC 0.68) is similar to the original study (AUROC 0.708). The reasons for this are unclear and may be attributable to biological sex and/or tumor-specific factors which are not directly incorporated as variables in the reference model. This could also be indicative of heterogenous frailty profiles across different tumor types.

The results of the external validation suggest that performance of the MSTFI is not consistent when applied to data outside of the cohort used in its development, and caution should be used when considering clinical application in spinal metastases. Improvements are needed to determine frailty thresholds that better align with observed performance. Using the same 9 input parameters as the MSTFI, we developed 3 machine learning models to assess factors that best predicted postoperative complications. The random forest algorithm had the highest positive predictive value and highest negative predictive value for predicting postoperative complications, identifying chronic lung disease, coagulopathy, anemia, and malnutrition as main predictors. However, these predictors were not consistent in stochastic modeling techniques, in this case gradient boosting and random forest, because gradient-boosting models have a tendency to concentrate variable importance in fewer variables, whereas random forest models tend to spread importance among variables more uniformly. Surgical approach was the least important variable to predict adverse clinical outcomes. The invasiveness of surgery is an important variable distinct from frailty that needs to be studied further. Future work is needed to further stratify the risk of complications by the invasiveness of surgery and physiological frailty of the patient.

The results of this study highlight the challenges of studying, defining, and quantifying frailty for patients with spinal metastases. Fundamentally, for this population, the questions are, what is frailty, and what are we trying to measure? Although frailty has been studied broadly across cancer and other surgical oncological fields, frailty as it relates to spinal metastases likely represents a surgeon’s perception of a patient’s physiological fitness for surgery with no clear consensus on criteria. However, a surgeon’s perception may not be aligned with objective frailty metrics or the patient’s own perception of their ability to recover from surgery. Many factors, including age, biological sex and/or tumor-specific factors which are not directly incorporated as variables in the reference model. This could also be indicative of heterogenous frailty profiles across different tumor types.

The results of the external validation suggest that performance of the MSTFI is not consistent when applied to data outside of the cohort used in its development, and caution should be used when considering clinical application in spinal metastases. Improvements are needed to determine frailty thresholds that better align with observed performance. Using the same 9 input parameters as the MSTFI, we developed 3 machine learning models to assess factors that best predicted postoperative complications. The random forest algorithm had the highest positive predictive value and highest negative predictive value for predicting postoperative complications, identifying chronic lung disease, coagulopathy, anemia, and malnutrition as main predictors. However, these predictors were not consistent in stochastic modeling techniques, in this case gradient boosting and random forest, because gradient-boosting models have a tendency to concentrate variable importance in fewer variables, whereas random forest models tend to spread importance among variables more uniformly. Surgical approach was the least important variable to predict adverse clinical outcomes. The invasiveness of surgery is an important variable distinct from frailty that needs to be studied further. Future work is needed to further stratify the risk of complications by the invasiveness of surgery and physiological frailty of the patient.

The results of this study highlight the challenges of studying, defining, and quantifying frailty for patients with spinal metastases. Fundamentally, for this population, the questions are, what is frailty, and what are we trying to measure? Although frailty has been studied broadly across cancer and other surgical oncological fields, frailty as it relates to spinal metastases likely represents a surgeon’s perception of a patient’s physiological fitness for surgery with no clear consensus on criteria. However, a surgeon’s perception may not be aligned with objective frailty metrics or the patient’s own perception of their ability to recover from surgery. Many factors, including age, biological sex and/or tumor-specific factors which are not directly incorporated as variables in the reference model. This could also be indicative of heterogenous frailty profiles across different tumor types.

The results of the external validation suggest that performance of the MSTFI is not consistent when applied to data outside of the cohort used in its development, and caution should be used when considering clinical application in spinal metastases. Improvements are needed to determine frailty thresholds that better align with observed performance. Using the same 9 input parameters as the MSTFI, we developed 3 machine learning models to assess factors that best predicted postoperative complications. The random forest algorithm had the highest positive predictive value and highest negative predictive value for predicting postoperative complications, identifying chronic lung disease, coagulopathy, anemia, and malnutrition as main predictors. However, these predictors were not consistent in stochastic modeling techniques, in this case gradient boosting and random forest, because gradient-boosting models have a tendency to concentrate variable importance in fewer variables, whereas random forest models tend to spread importance among variables more uniformly. Surgical approach was the least important variable to predict adverse clinical outcomes. The invasiveness of surgery is an important variable distinct from frailty that needs to be studied further. Future work is needed to further stratify the risk of complications by the invasiveness of surgery and physiological frailty of the patient.

The results of this study highlight the challenges of studying, defining, and quantifying frailty for patients with spinal metastases. Fundamentally, for this population, the questions are, what is frailty, and what are we trying to measure? Although frailty has been studied broadly across cancer and other surgical oncological fields, frailty as it relates to spinal metastases likely represents a surgeon’s perception of a patient’s physiological fitness for surgery with no clear consensus on criteria. However, a surgeon’s perception may not be aligned with objective frailty metrics or the patient’s own perception of their ability to recover from surgery. Many factors, including age, biological sex and/or tumor-specific factors which are not directly incorporated as variables in the reference model. This could also be indicative of heterogenous frailty profiles across different tumor types.
cal sex, performance status, neurological and ambulatory function, medical comorbidities, and laboratory markers, have been studied as preoperative indicators of frailty, but these may represent only a sample of variables to consider; cancer markers, psychological stressors, and sarcopenia may also contribute.36–39 The concept of frailty and its association with complications after spine surgery has been explored in spine surgery, but developing practical tools for assessment has proved difficult.40–45 Most frailty measurements are expressed as a ratio of health deficits present to the total number of deficits considered.46,47 More than 70 different tools exist to measure frailty, ranging from a single item being measured to more than 90.34 With the exception of the 15-item modified cervical deformity frailty index developed by the International Spine Study Group, most surgical frailty indices, like the MSTFI and the modified frailty index (mFI), were developed from and/or tested on ACS NSQIP and other national databases.48–51 Although many prognostic calculators for metastatic spinal disease have been developed, their clinical application is limited without prospective validation.52 This study suggests that the MSTFI may not capture the physiological effects of cancer-related disease burden with the variables studied.14 Schoenfeld et al. reported that a cancer-specific model that considers serum albumin, the New England Spinal Metastasis Score (NESMS), significantly predicted 30-day mortality and major complications.53 These authors further validated the NESMS as a utility for survival at 1 year. In a cohort of 180 patients, an NESMS score of 0 was associated with perfect prediction, as 100% of individuals with this score were deceased at 1 year.54 In a retrospective analysis, Mohd Rothi et al. applied the mFI to 41 patients who underwent surgery for spinal metastases at a single center and found that the index correlated poorly with complications and survival.31 Bourassa-Moreau et al. also found that neither the mFI nor the MSTFI was associated with early mortality or adverse events in 108 patients who underwent emergent spine surgery.30 Moreover, sarcopenia, measured by the ratio of the L3 psoas muscle area divided by the area of the L3 vertebral body, was a better predictor of adverse events and 1-month mortality. Sarcopenia as a marker of frailty has been studied in numerous settings, including spinal metastases.55,56 It is conceptually distinct from frailty and relates to loss of skeletal muscle mass, strength, and endurance.50 In a retrospective, multicenter cohort study investigating muscle mass, the cross-sectional area of the psoas muscle at the L3–4 disc space was measured within 200 days before surgery and was divided into gender-specific tertiles. Psoas size predicted mortality better than Tokuhashi score, Tomita score, and Karnofsky Performance Scale score at 90 days, but this was not significant at 30 days. At 90 days, patients in the smallest tertile were 6.25 times more likely to die compared with patients in the largest tertile. Psoas size, as a surrogate for frailty, predicted 90-day and overall mortality independent of demographic, functional, oncological, and surgical characteristics. The European Working Group on Sarcopenia in Older People (EWGSOP2) recently published a new operational definition of sarcopenia that puts much more emphasis on muscle strength than

| Outcome | Graphical Model | Bootstrap Validation AUROC (95% CI) | p Value* | Sensitivity (95% CI) | Specificity (95% CI) | PPV (95% CI) | NPV (95% CI) | PLR (95% CI) | NLR (95%) |
|---------|----------------|-----------------------------------|----------|----------------------|----------------------|-------------|-------------|-------------|-----------|
| Postop complications | Logistic regression with lasso regularization | 0.53 (0.50–0.58) | 0.20 | 0.69 (0.66–0.73) | 0.37 (0.29–0.45) | 0.67 (0.53–0.80) | 0.50 (0.44–0.56) | 0.34 (0.26–0.42) | 0.75 (0.70–0.80) | 0.99 (0.97–1.01) |
| | Random forest | 0.62 (0.56–0.68) | 0.18 | 0.69 (0.66–0.73) | 0.37 (0.29–0.45) | 0.67 (0.53–0.80) | 0.50 (0.44–0.56) | 0.34 (0.26–0.42) | 0.75 (0.70–0.80) | 0.99 (0.97–1.01) |
| | Gradient-boosted decision tree | 0.54 (0.48–0.60) | 0.27 | 0.72 (0.65–0.77) | 0.34 (0.26–0.42) | 0.75 (0.70–0.80) | 0.50 (0.44–0.56) | 0.34 (0.26–0.42) | 0.75 (0.70–0.80) | 0.99 (0.97–1.01) |
| | Reference model | 0.56 (0.50–0.62) | 0.20 | 0.69 (0.66–0.73) | 0.37 (0.29–0.45) | 0.67 (0.53–0.80) | 0.50 (0.44–0.56) | 0.34 (0.26–0.42) | 0.75 (0.70–0.80) | 0.99 (0.97–1.01) |

NLR = negative likelihood ratio; NPV = negative predictive value; PLR = positive likelihood ratio; PPV = positive predictive value.
muscle mass in predicting outcomes. Further investigation is needed to establish outcome-specific cutoffs.

Machine learning algorithms are increasingly being used in spine surgery research in lieu of traditional statistical methods as these tools can be used to analyze the underlying complexity of clinical data to improve prediction accuracy. In our analysis, decision tree classification yielded more robust prediction performance than logistic regression due to its ability to process complex nonlinear relationships among predictors. Despite the superior prediction ability of machine learning approaches compared with conventional approaches, their performance was still not optimal. These findings support cautious interpretation of machine learning algorithms given the potential consequences of over- or underestimating the results for clinical decision-making purposes. Flawed enthusiasm for imperfect calculators may have clinical consequences, as patients who could benefit from surgical intervention may be dissuaded based on such data. This analysis demonstrates that though accuracy can be improved with machine learning methodologies, it does not account for the limitation that the variables entered may not include what needs to be measured. Given the broad and varied definitions of frailty used in oncology and geriatric medicine, further study is needed to determine the specific frailty phenotype for spinal metastases and incorporate other factors not yet accounted for in these models.

Limitations

Although databases such as NIS and ACS NSQIP have voluminous data, they are limited by potential errors and inconsistencies in coding, reporting, and data entry. Our study suggests that what is being measured may not be generalizable to other data sets as criteria for cutoffs and coding of variables are unclear. Furthermore, the MSTFI includes treatment variables (e.g., type of admission and surgical approaches) that can impact outcomes but are not components of physiological frailty. In addition, validation was limited by the variables available in ACS NSQIP, which are not identical to those in the NIS. The variables in these data sets may not necessarily be pertinent to spinal metastases and do not include laboratory and oncological factors considered in clinical practice. For patients with cancer, metastatic disease itself, and not necessarily age or medical comorbidities, may be the most significant physiological stressor. In fact, some of the most extensively used tools in determining frailty, such as the comprehensive geriatric assessment, include not only medical, but also cognitive, functional, mental health, and psychosocial domains. In the MSTFI, malnutrition is the only parameter that represents the overall composite health status of the patient. The NIS data set from which the MSTFI was developed was from 2002 to 2011, which may also reflect time bias since cancer therapies affecting outcomes at that time may not be reflective of outcomes today.

With regard to our institutional data, there are limitations that are inherent to retrospective data analysis spanning 10 years at a single institution. The secular trend over this period was not measured. During this period, advancements in surgical techniques, changes in practice patterns, and developments in radiation and systemic therapies likely affected the observed outcomes. Clustering of data from a single tertiary referral institution may also lead to unintended practice bias and may not necessarily be representative of other healthcare settings. Confounding from other unmeasured variables is a limitation in this analysis, as well as the construction of the MSTFI. Further study with prospective data collection across multiple centers in diverse geographic and practice environments, to account
for disparities based on biological sex, race, ethnicity, and socioeconomic status, will facilitate further development of this concept of frailty for spinal metastases, which to date has no universally accepted definition and criteria.

Conclusions
The findings in this study underscore the importance of externally validating prediction models developed from national databases, and recognizing inherent methodological limitations with prognostic model development, as significant associations between MSTSFI classes and outcomes were not observed. Continued efforts to identify and define frailty will help optimize surgical risk assessments for patients with spinal metastases.

Acknowledgments
This work was conducted with support from Harvard Catalyst/ The Harvard Clinical and Translational Science Center (National Center for Advancing Translational Sciences, National Institutes of Health Award UL 1TR002541) and financial contributions from Harvard University and its affiliated academic healthcare centers. The content is solely the responsibility of the authors and does not necessarily represent the official views of Harvard Catalyst, Harvard University and its affiliated academic healthcare centers, or the National Institutes of Health.

References
1. Dea N, Versteeg A, Fisher C, et al. Adverse events in emergency oncological spine surgery: a prospective analysis. J Neurosurg Spine. 2014;21(5):698–703.
2. Nater A, Tetreault LA, Kopjar B, et al. Predictive factors of survival in a surgical series of metastatic epidural spinal cord compression and complete external validation of 8 multivariable models of survival in a prospective North American multicenter study. Cancer. 2018;124(17):3536–3550.
3. Rothrock RJ, Barzilai O, Reiner AS, et al. Survival trends after surgery for spinal metastatic tumors: 20-year cancer center experience. Neurosurgery. 2021;88(2):402–412.
4. Ahangar P, Aziz M, Rosenzweig DH, Weber MH. Advances in personalized treatment of metastatic spine disease. Ann Transl Med. 2019;7(10):223.
5. Lafer I, Jorgulescu JB, Chapman T, et al. Local disease control for spinal metastases following “separation surgery” and adjuvant hypofractionated or high-dose single-fraction stereotactic radiosurgery: outcome analysis in 186 patients. J Neurosurg Spine. 2013;18(3):207–214.
6. Barzilai O, Amato MK, McLaughlin L, et al. Hybrid surgery-radiosurgery therapy for metastatic epidural spinal cord compression: a prospective evaluation using patient-reported outcomes. Neurooncol Pract. 2018;5(2):104–113.
7. Lau D, Leach MR, Than KD, et al. Independent predictors of complication following surgery for spinal metastasis. Eur Spine J. 2013;22(6):1402–1407.
8. Patchell RA, Tibbs PA, Regine WF, et al. Direct decompressive surgical resection in the treatment of spinal cord compression caused by metastatic cancer: a randomized trial. Lancet. 2005;366(9486):643–648.
9. Newman WC, Lafer I, Bilsky MH. Neurologic, oncologic, mechanical, and systemic and other decision frameworks for spinal disease. Neurosurg Clin N Am. 2020;31(2):151–166.
10. Lafer I, Rubin DG, Lis E, et al. The NOMS framework: approach to the treatment of spinal metastatic tumors. Oncologist. 2013;18(6):744–751.
11. Hussain I, Barzilai O, Reiner AS, et al. Patient-reported outcomes after surgical stabilization of spinal tumors: symptom-based validation of the Spinal Instability Neoplastic Score (SINS) and surgery. Spine J. 2018;18(2):261–267.
12. Hussain I, Barzilai O, Reiner AS, et al. Spinal Instability Neoplastic Score component validation using patient-reported outcomes. J Neurosurg Spine. 2019;30(4):432–438.
13. Lakomkin N, Zuckerman SL, Stannard B, et al. Preoperative risk stratification in spine tumor surgery: a comparison of the modified Charlson Index, Frailty Index, and ASA score. Spine (Phila Pa 1976). 2019;44(13):E782–E787.
14. Massaad E, Hadzipasic M, Alvarez-Brekkenridge C, et al. Predicting tumor-specific survival in patients with spinal metastatic renal cell carcinoma: which scoring system is most accurate? J Neurosurg Spine. 2020;33(4):529–539.
15. Massaad E, Fatima N, Hadzipasic M, et al. Predictive analytics in spine oncology research: first steps, limitations, and future directions. Neurospine. 2019;16(4):669–677.
16. Choi D, Ricciardi F, Arts M, et al. Prediction accuracy of common prognostic scoring systems for metastatic spine disease: results of a prospective international multicenter study of 1469 patients. Spine (Phila Pa 1976). 2018;43(23):1678–1684.
17. Fedarko NS. The biology of aging and frailty. Clin Geriatr Med. 2011;27(1):27–37.
18. Morley JE, Vellas B, van Kan GA, et al. Frailty consensus: a call to action. J Am Med Dir Assoc. 2013;14(6):392–397.
19. Chi JH, Gokaslan Z, McCormick P, et al. Selecting treatment for patients with malignant epidural spinal cord compression—does age matter?: results from a randomized clinical trial. Spine (Phila Pa 1976). 2009;34(5):431–435.
20. Amelot A, Balabaud L, Choi D, et al. Surgery for metastatic spine tumors in the elderly. Advanced age is not a contraindication to surgery! Spine J. 2017;17(6):759–767.
21. Ishayek E, Candanedo C, Fraifeld S, et al. Ambulation and survival following surgery in elderly patients with metastatic epidural spinal cord compression. Spine J. 2018;18(7):1211–1221.
22. Rockwood K, Mitnitski A. Frailty in relation to the accumulation of deficits. J Gerontol A Biol Sci Med Sci. 2007;62(7):722–727.
23. Makary MA, Segev DL, Pronovost PJ, et al. Frailty as a predictor of surgical outcomes in older patients. J Am Coll Surg. 2010;210(6):901–908.
24. Bullock AF, Greenley SL, McKenzie GAG, et al. Relationship between markers of malnutrition and clinical outcomes in older adults with cancer: systematic review, narrative synthesis and meta-analysis. Eur J Clin Nutr. 2020;74(11):1519–1535.
25. Cadwell JB, Afonso AM, Shahroki A. Prognostic nutritional index (PNI), independent of frailty is associated with six-month postoperative mortality. J Geriatr Oncol. 2020;11(5):880–884.
26. Shehadi JA, Scibbha DM, Suk I, et al. Surgical treatment strategies and outcome in patients with breast cancer metastatic to the spine: a review of 87 patients. Eur Spine J. 2007;16(8):1179–1192.
27. Goodwin CR, Khattab MH, Sankey EW, et al. Factors associated with life expectancy in patients with metastatic spine disease from adenocarcinoma of the lung. Global Spine J. 2015;5(5):417–424.
28. Ju DG, Zadnik PL, Groves ML, et al. Factors associated with improved outcomes following decompressive surgery for prostate cancer metastatic to the spine. Neurosurgery. 2013;73(4):657–666.
29. De la Garza Ramos R, Goodwin CR, Jain A, et al. Development of a metastatic spinal tumor frailty index (MSTFI) using a nationwide database and its association with inpatient morbidity, mortality, and length of stay after spine surgery. World Neurosurg. 2016;95:548–555.e4.
30. Bourassa-Moreau É, Versteeg A, Moskven E, et al. Sarco...
34. Ethun CG, Bilen MA, Jani AB, et al. Frailty and cancer: Implications for oncology surgery, medical oncology, and radiation oncology. *CA Cancer J Clin*. 2017;67(5):362–377.

35. Revenig LM, Canter DJ, Henderson MA, et al. Preoperative quantification of perceptions of surgical frailty. *J Surg Res*. 2015;193(2):583–589.

36. Karimi D, Morgen SS, Frueggaard S, et al. Long-term survival after surgical treatment of spinal metastasis—the predictive role of sex. *Anticancer Res*. 2020;40(3):1475–1479.

37. Schoenfeld AJ, Ferrone ML, Passias PG, et al. Laboratory markers as useful prognostic measures for survival in patients with spinal metastases. *Spine J*. 2020;20(1):5–13.

38. Choi BD, Shankar GM, Sivaganesan A, et al. Implication of immunotherapy status on outcomes in patients with metastatic melanoma to the spine. *Spine (Phila Pa 1976)*. 2018;43(21):E1274–E1280.

39. Shankar GM, Choi BD, Grannan BL, et al. Effect of immunotherapy status on outcomes in patients with metastatic melanoma to the spine. *Spine (Phila Pa 1976)*. 2017;42(1):E721–E725.

40. Ali R, Schwalb JM, Nerenz DR, et al. Use of the modified frailty index to predict 30-day morbidity and mortality from spine surgery. *J Neurosurg Spine*. 2016;25(4):537–541.

41. Shin JI, Kothari P, Fahn K, et al. Frailty index as a predictor of adverse postoperative outcomes in patients undergoing cervical spinal fusion. *Spine (Phila Pa 1976)*. 2017;42(5):304–310.

42. Flexman AM, Charest-Morin R, Stobart L, et al. Frailty and postoperative outcomes in patients undergoing surgery for degenerative spine disease. *Spine J*. 2016;16(11):1315–1323.

43. Leven DM, Lee NJ, Kothari P, et al. Frailty index is a significant predictor of complications and mortality after surgery for adult spinal deformity. *Spine (Phila Pa 1976)*. 2016;41(23):E1394–E1401.

44. Miller EK, Neuman BJ, Jain A, et al. An assessment of frailty as a tool for risk stratification in adult spinal deformity surgery. *Neurosurg Focus*. 2017;43(6):E3.

45. Wilson JRF, Badhiwala JH, Moghaddamjou A, et al. Frailty is a better predictor than age of mortality and perioperative complications after surgery for degenerative cervical myelopathy: an analysis of 41,369 patients from the NSQIP Database 2010-2018. *J Clin Med*. 2020;9(11):E3491.

46. Mitnitski AB, Mogilner AJ, Rockwood K. Accumulation of deficits as a proxy measure of aging. *ScientificWorldJournal*. 2001;1:323–336.

47. Searle SD, Mitnitski A, Gahbauer EA, et al. A standard procedure for creating a frailty index. *BMC Geriatr*. 2008;8:24.

48. Miller EK, Ailion T, Neuman BJ, et al. Assessment of a novel adult cervical deformity frailty index as a component of preoperative risk stratification. *World Neurosurg*. 2018;109: e800–e806.

49. Passias PG, Bortz CA, Segreto FA, et al. Development of a modified cervical deformity frailty index: a streamlined clinical tool for preoperative risk stratification. *Spine (Phila Pa 1976)*. 2019;44(3):169–176.

50. Rockwood K, Song X, MacKnight C, et al. A global clinical measure of fitness and frailty in elderly people. *CMAJ*. 2005;173(5):489–495.

51. Obeid NM, Azuh O, Reddy S, et al. Predictors of critical care-related complications in colectomy patients using the National Surgical Quality Improvement Program: exploring frailty and aggressive laparoscopic approaches. *J Trauma Acute Care Surg*. 2012;72(4):878–883.

52. Pennington Z, Ehresman J, Cottrill E, et al. To operate, or not to operate? Narrative review of the role of survival predictors in patient selection for operative management of patients with metastatic spine disease. *J Neurosurg Spine*. 2021;34(1):135–149.

53. 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.

54. Schoenfeld AJ, Ferrone ML, Schwab JH, et al. Prospective validation of a clinical prediction score for survival in patients with spinal metastases: the New England Spinal Metastasis Score. *Spine J*. 2021;21(1):28–36.

55. Zakaria HM, Wilkinson BM, Pennington Z, et al. Sarcopenia as a prognostic factor for 90-day and overall mortality in patients undergoing spine surgery for metastatic tumors: a multicenter retrospective cohort study. *Neurosurgery*. 2020;87(5):1025–1036.

56. Moskven E, Bourassa-Moreau É, Charest-Morin R, et al. The impact of frailty and sarcopenia on postoperative outcomes in adult spine surgery. A systematic review of the literature. *Spine*. 2018;18(12):2354–2369.

57. Cruz-Jentoft AJ, Bahat G, Bauer J, et al. Sarcopenia: revised European consensus on definition and diagnosis. *Age Ageing*. 2019;48(1):16–31.

58. Gallacher E, McCarron P, Sudlow C, et al. The impact of frailty on postoperative complications following spinal surgery: a systematic review. *Spine*. 2018;43(2):137–142.

59. Chang M, Canseco JA, Nicholson KJ, et al. The role of machine learning in spine surgery: the future is now. *Front Surg*. 2020;7:54.

60. Bongers MER, Karhade AV, Villavieja J, et al. Does the SORG algorithm generalize to a contemporary cohort of patients with spinal metastases on external validation? *Spine J*. 2020;20(10):1646–1652.

61. Karhade AV, Thio QCBS, Oqink PT, et al. Predicting 90-day and 1-year mortality in spinal metastatic disease: development and internal validation. *Neurosurgery*. 2019;85(4):E671–E681.

62. Pazniokas J, Gandhi C, Theriault B, et al. The immense heterogeneity of frailty in neurosurgery: a systematic literature review. *Neurosurg Rev*. 2021;44(1):189–201.

63. Wildiers H, Heeren P, Puts M, et al. International Society of Geriatric Oncology consensus on geriatric assessment in older patients with cancer. *J Clin Oncol*. 2014;32(24):2595–2603.

64. Andrew MK, Mitnitski AB, Rockwood K. Social vulnerability, frailty and mortality in elderly people. *PLoS One*. 2008;3(5):e2232.

**Disclosures**

Dr. Schoenfeld: support of non-study-related clinical or research effort from NIH, OREF, and DoD; royalties from Wolters Kluwer and Springer Nature; advisory board member of the Commonwealth Fund; and editorial board of *Journal of Bone & Joint Surgery*.

**Author Contributions**

Conception and design. Shin, Massaad, Williams. Acquisition

---

**Author Contributions**

Conception and design. Shin, Massaad, Williams. Acquisition...
Supplemental Information
Online-Only Content
Supplemental material is available online.
   Supplementary Methodology. https://thejns.org/doi/suppl/10.3171/2021.2.FOCUS201113.

Correspondence
John H. Shin: Massachusetts General Hospital, Boston, MA. shin.john@mgh.harvard.edu.