J BONE ONCOL 57 (2026) 100739

Contents lists available at ScienceDirect

Journal of Bone Oncology

journal homepage: www.elsevier.com/locate/jbo

Research Paper

Prospective external validation of a three-predictor frailty model for 90-day survival and complications following spinal metastasis surgery

Pedro Reggiani Anzuategui [a,b,*], Glauco José Pauka Mello [a], Ana Valéria Brunetti Rigolino [a], Lucas Emanuel Sauer Larocca [a], Cássio Zini [c,d], Carmen Australia Paredes Marcondes Ribas [c,d]

[a] Orthopedic Oncology Service, Erasto Gaertner Hospital, Curitiba, Brazil
[b] School of Medicine, Department of Surgery, Federal University of Paraná, Curitiba, Brazil
[c] School of Medicine, Mackenzie Evangelical College of Paraná, Curitiba, Brazil
[d] Graduate Program in Principles of Surgery, Mackenzie Evangelical College of Paraná, Curitiba, Brazil

HIGHLIGHTS

- Three-predictor model showed external validity for spinal metastasis surgery.
- AUC was 0.78 for 90-day survival and 0.68 for complications.
- Model reached 70% accuracy and 85% specificity for frailty detection.
- Frailty-based tool supports urgent surgical decisions at bedside.

ARTICLE INFO

Keywords:
Spinal neoplasms
Surgical decision-making
Prognostic model
Frailty index
Prognosis
Perioperative morbidity
Receiver operating characteristic
Risk assessment
Validation study

ABSTRACT

Background context: Surgical decision-making in patients with spinal metastases remains complex due to the need to balance potential surgical benefits with limited survival and common frailty. Predictive models can assist in this process, but their clinical utility is often limited by complexity and lack of validation.
Purpose: To externally validate a simple three-predictor frailty model for 90-day survival and complications, and to compare its performance with other commonly used tools.
Study design/setting: Prospective external validation study conducted at a single tertiary cancer center.
Patient sample: A consecutive cohort of 126 patients who underwent open posterior surgery with instrumentation for spinal metastases from solid tumors between 2018 and 2024.
Outcome measures: Primary outcomes were 90-day survival and the occurrence of postoperative complications. Secondary outcomes included 30-day, 180-day and overall survival. Model performance was evaluated through discrimination (AUC), risk stratification, accuracy for surgical indication and calibration.
Methods: The Anzuategui model (three predictors: tumor growth rate, comorbidities, and lymphocyte count) was applied preoperatively, along with four other three-predictor models (Tomita, Modified Bauer, Van der Linden, and Sioutos). Discrimination was assessed using ROC curves. Risk stratification was evaluated using predefined low-, moderate-, and high-risk categories, analyzed through Kaplan–Meier curves and complication rates. Model accuracy for surgical indication was calculated using a 90-day survival threshold as the reference. Calibration for both 90-day survival and postoperative complications was performed by comparing category-specific predicted probabilities derived from the development cohort with observed event rates in the validation cohort.
Results: The Anzuategui model demonstrated predictive performance for the primary outcomes comparable to the other models under evaluation. It achieved an AUC of 0.78 (95% CI: 0.70–0.85) for 90-day survival and 0.68 (95% CI: 0.59–0.76) for postoperative complications. Risk stratification showed clear separation between survival curves across the three predefined categories. Accuracy for predicting appropriate surgical indication was 70% (95% CI: 61–78), with a sensitivity of 64% and specificity of 85%. Tomita and Modified Bauer models showed comparable accuracy (75% and 74%, respectively) but lower specificity. Calibration indicated overestimation of 90-day mortality (intercept –1.75; slope 2.05) and modest miscalibration for postoperative complications (intercept –0.40; slope 0.67).

* Corresponding author at: Rua Ari José Valle, 980, Casa 16, 82030-025 Curitiba, PR, Brazil.
E-mail address: pedroanzuategui@hotmail.com (P.R. Anzuategui).

https://doi.org/10.1016/j.jbo.2026.100739
Received 22 August 2025; Received in revised form 7 December 2025; Accepted 6 January 2026
Available online 12 January 2026

Conclusions: The Anzuategui model demonstrated acceptable external performance, with greater validity for predicting 90-day survival than for postoperative complications. Its simplicity and frailty-centered structure make it a practical bedside tool, particularly in urgent or resource-limited settings. Integrating this approach with established prognostic models may support more balanced decision-making across diverse clinical scenarios.

1. Introduction

Surgical decision-making in patients with spinal metastases is particularly challenging. The potential benefit of surgery often conflicts with the patient's frail health and limited life expectancy [1]. This raises the critical question: which patients should undergo surgery [2]?

Several multivariable predictive models have been developed to estimate outcomes following surgery for spinal metastases and to support clinical decision-making [3,4]. These models have proven useful by offering prognostic classifications across multiple risk categories—such as low, moderate, and high risk. Others provide binary classifications, supporting dichotomous decisions such as whether or not to operate. More advanced models yield probabilistic predictions, offering numerical estimates ranging from 0 % to 100 % for a given outcome [5].

A particularly useful outcome in this context is 90-day survival, which is often considered a key threshold in surgical decision-making [6,7]. Additional models have been proposed to predict survival at other time points (e.g., 30 [8,9], 45 [10], 180 days [11,12], and 1 year [13]), as well as to estimate the risk of complications such as surgical site infection [14,15], massive bleeding [16], neurological deterioration [17], quality of life [18], and overall morbidity [19,20].

Simplified predictive models rely on up to three clinical predictors capable of anticipating favorable or unfavorable outcomes. Among the most widely known are the Tomita [21] and Modified Bauer [22,23] scores. Their ease of memorization makes them especially useful in bedside evaluations and urgent hospital settings, such as in cases of metastatic spinal cord compression.

The pursuit of higher accuracy has led to the development of machine learning algorithms that incorporate dozens of predictors, including advanced imaging, non-routine laboratory tests, and diverse clinical features [24]. While these approaches have increased discriminative performance to approximately 75–85 %, they introduce a degree of complexity that may hinder their clinical adoption—particularly when they delay decision-making until all tests are completed and interpreted [25].

In the era of artificial intelligence, the practicality of three-predictor models for surgical decision-making in spinal metastases remains uncertain [26]. This study aims to externally validate a three-predictor frailty model proposed by Anzuategui et al. in 2019 [27], and to compare its performance with other widely used prognostic tools.

The relevance of this study lies in its potential to simplify and enhance surgical decision-making by providing a practical, user-friendly tool for managing patients with spinal metastases, with a specific focus on frailty assessment.

2. Materials and methods

2.1. Study design

This study followed the methodological guidelines outlined in the TRIPOD Statement [5], (Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis). The completed TRIPOD checklist is provided in the Supplementary Material S1. All procedures complied with ethical standards for human research and were approved by the local Institutional Review Board.

The target population consisted of a prospective cohort of consecutive patients treated at a single tertiary cancer center who underwent surgery for spinal metastases between 2018 and 2024.

2.2. Study population

All consecutive patients undergoing open surgical treatment for spinal metastases from solid tumors were prospectively enrolled. To ensure clinical and biological homogeneity, patients with hematologic malignancies (e.g., multiple myeloma, lymphoma, leukemia) were excluded. This decision was prespecified, given the distinct pathophysiology, metastatic patterns, and surgical indications of hematologic neoplasms compared with solid tumors.

Exclusion criteria were applied to preserve cohort homogeneity and to align with the study's primary endpoint. Patients with postoperative follow-up shorter than 90 days were excluded (n = 2), as 90-day survival represents the principal outcome of interest and a clinically meaningful benchmark in this setting. Five privately treated patients were also excluded because, unlike the standardized and integrated public healthcare system of our institution, private-system care is frequently fragmented across multiple facilities and oncology protocols, introducing heterogeneity in both surgical and nonsurgical aspects of care.

Two additional cases were excluded because no spinal fixation/stabilization was performed—one due to intraoperative complications leading to early termination of the procedure and another involving a predominantly sacral lesion for which stabilization was not feasible. Anterior-only approaches were prespecified as an exclusion criteria, although none occurred during the study period.

Because the cohort comprises all surgically treated patients over the study period, no selection was made regarding tumor histology, metastatic burden, comorbidities, or functional status. Therefore, the heterogeneity observed in the sample reflects the real-world case mix of a tertiary cancer center and represents the population for whom predictive models for postoperative outcomes are intended in clinical practice.

2.3. Data collection

Data were collected prospectively as clinical events occurred, including surgery, perioperative care, outpatient follow-up, and hospital readmissions. All prediction models were applied preoperatively by an investigator blinded to postoperative outcomes.

The primary outcomes were 90-day survival and the occurrence of postoperative complications. Secondary outcomes included 30-day and 180-day survival and overall survival. Survivors were followed until the last available clinical assessment or until death or censoring.

Postoperative complications occurring within 30 days of surgery were classified as systemic or local; infectious or non-infectious; respiratory or non-respiratory; and graded according to the classification proposed by Rampersaud et al. [28]. All collected data were reviewed by the principal investigator at the end of the study to ensure consistency and accuracy.

2.4. Comorbidity assessment

Comorbidities were identified following the operational criteria used in the original development study, based on Charlson [29] and Elixhauser [30] domains. The principal investigator personally evaluated all patients and verified each condition during the preoperative assessment. A comorbidity was recorded when supported by at least one of the following: (i) documented prior diagnosis, particularly when chronic pharmacologic therapy was in place; (ii) inpatient laboratory abnormalities consistent with the condition; (iii) assessment by the hospitalist team; or (iv) confirmatory findings from echocardiography,

electrocardiography, spirometry, or other ancillary tests when available. The comorbidities systematically assessed were: diabetes mellitus, chronic pulmonary disease, prior myocardial infarction, congestive heart failure, cardiac arrhythmia, pulmonary circulation disorder, peripheral vascular disease, cerebrovascular disease, dementia, renal insufficiency, hepatic insufficiency, connective tissue disease, coagulopathy, prior paralysis, peptic ulcer disease, and acquired immunodeficiency syndrome.

In urgent or emergent cases, confirmatory testing, especially spirometry for chronic pulmonary disease, was often not feasible; therefore, diagnoses could rely on clinical history, imaging features, and treating-team impressions. This reflects routine real-world practice in oncologic spine surgery. Inter-rater reliability was not formally assessed because comorbidity classification was performed by a single experienced evaluator.

2.5. Perioperative management

The predictive models were not used as the sole criteria for surgical decision-making. Preoperative evaluation included restaging with computed tomography scans of the head, chest, abdomen, and pelvis, as well as laboratory testing performed within three days prior to surgery. Surgical risk assessment followed institutional protocols and involved cardiology and anesthesiology consultations. In complex clinical scenarios, the hospitalist service was engaged to optimize the management of comorbidities.

A multidisciplinary team routinely provided nutritional, physical therapy, and nursing support. Psychological and social work services were offered selectively, based on individual patient needs. Postoperative care typically included intensive monitoring, wound management, and early rehabilitation.

2.6. Surgical technique

The standard procedure involved direct neural decompression combined with pedicle screw fixation for spinal stabilization, as described by Patchell et al. [31]. Patients were positioned prone for a posterior-only approach. A midline incision centered over the affected vertebra was made, followed by posterior element exposure. In thoracic lesions, the standard construct included fixation of two levels above and two below the lesion. Most decompressions were performed using a partial transpedicular corpectomy combined with multilevel laminectomy (two or three levels). Sutures were removed after three weeks, and all patients received adjuvant radiotherapy.

2.7. Three-predictor models

For external validation, the Anzuategui prediction model was compared with other clinical tools of similar structure, each limited to three predictors, as listed in Table 1. The model was operationalized using three risk categories (Low, Moderate, High), rather than the four originally defined in the development cohort. This intentional simplification aimed to improve clinical usability while preserving the conceptual framework of the original stratification.

Because all prognostic models were applied to the same patient cohort, no between-group baseline differences existed, and adjustment for histologic distribution or other clinical characteristics was unnecessary.

The Anzuategui model was compared with widely accepted benchmark models proposed by Tomita et al. [21], Bauer et al. [22] (as modified by Leithner et al. [23]), Van der Linden et al. [32] and Sioutos et al. [33], using the following performance metrics:

1. **Discriminative ability**: assessed using ROC curve analysis and the area under the curve (AUC);
2. **Prognostic stratification**: risk groups (low, moderate, high) were defined, and Kaplan-Meier survival curves and complication rates per group were analyzed;
3. **Surgical indication accuracy**: appropriate surgical indication (i.e., "true" indication) was defined as a predicted surgery with actual survival exceeding 90 days. This allowed for construction of a confusion matrix including:
   - True positives (predicted for surgery and survived $>$ 90 days),

Table 1
Summary of prognostic models evaluated in this study.

Model	Predictors	Scoring system	Risk classes	Surgical recommendation
Anzuategui	Tumor growth rate Comorbidities Peripheral blood lymphocyte count	1 point if non-slow tumor progression, significant comorbidities, and lymphocyte count $< 1 \times 10^3/\mu L$	Low: 0 points Moderate: 1 point High: 2–3 points	0 or 1 point
Tomita	Tumor growth rate Visceral metastases Bone metastases	4 points for rapid, 2 for moderate, and 1 for slow progression; 4 points for untreatable visceral metastasis, 2 if treatable; 2 points for multiple bone metastases, 1 if solitary	Low: 2–3 points Moderate: 4–7 points High: 8–10 points	2 to 7 points
Modified Bauer	Histologic type Visceral metastases Bone metastases	1 point for non-pulmonary tumor; 1 point if originated from breast, kidney, lymphoma, or multiple myeloma; 1 point if no visceral metastasis; 1 point if bone metastasis is solitary	Low: 3–4 points Moderate: 2 points High: 0–1 point	2 to 4 points
Van der Linden	Histologic type Visceral metastases Karnofsky Performance Status (KPS)	3 points for breast tumor, 2 for prostate, 1 for lung, 0 for others; 1 point if visceral metastasis is present; 2 points if KPS 80–100, 1 if 50–70, 0 if 20–40	Low: 6 points Moderate: 4–5 points High: 0–3 points	4 to 6 points
Sioutos	Histologic type Vertebral metastases Preoperative muscle strength	1 point if tumor originated from lung or colon, multiple vertebral metastases, and muscle strength grade 0 to 3	Low: 0 point Moderate: 1 point High: 2–3 points	0 or 1 point

Notes: Risk classes were adapted to three categories for standardization. **Abbreviations:** KPS, Karnofsky Performance Status.

- True negatives (predicted for conservative treatment and died ≤ 90 days),
- False positives (predicted for surgery but died ≤ 90 days),
- False negatives (predicted for conservative treatment but survived > 90 days).

A 90-day postoperative survival threshold was used to define appropriate surgical indication. This interval represents a clinically meaningful minimum period during which patients are expected to derive benefit from major spine surgery, including pain relief, mechanical stabilization, and the opportunity to receive adjuvant oncologic treatments. Survival below this threshold is generally associated with limited utility of extensive surgical intervention; thus, the 90-day cutoff provides a pragmatic and widely used benchmark for decision-oriented analyses in metastatic spine disease.

2.8. Statistical analysis

Statistical tests were selected based on data distribution and study objectives: Fisher's exact test, Student's *t*-test, McNemar's test, chi-square test, and the Mann-Whitney *U* test were applied as appropriate. Cochran–Armitage trend test was used to test for risk stratification effects on complication incidence. Continuous variables were dichotomized when relevant.

The Kaplan–Meier method was used to estimate survival times and generate survival curves.

To compare the discriminative performance of the prognostic models, pairwise differences between AUC values were assessed using the DeLong test for correlated ROC curves.

Missing data occurred only in laboratory variables that were not used as predictors in any of the prognostic models under evaluation. Because C-reactive protein, International Normalized Ratio (INR), and serum albumin served exclusively for descriptive characterization of the cohort, no imputation procedure was performed. A complete-case approach was used for all variables required by the prediction models, all of which had complete data. Therefore, missingness did not affect model calculation, discrimination, risk assessment or calibration.

All analyses adopted a 95 % confidence interval. Statistical analyses and graph generation were performed using R (version 4.4.3) and MedCalc (version 23.2.8, 64-bit).

2.9. Calibration analysis

For calibration analyses, 90-day survival was transformed to its complementary outcome (90-day mortality) so that both primary endpoints—mortality and postoperative complications—could be evaluated uniformly as adverse events. This allowed the calibration intercept, slope, and graphical patterns to be interpreted in the same direction across outcomes.

Calibration was assessed separately for 90-day mortality and postoperative complications using the model's predefined three-category structure (Low, Moderate, High Risk). Each patient in the validation cohort was assigned the category-specific predicted probability derived from the development dataset.

Logistic recalibration was performed by regressing each observed outcome on the logit of its assigned predicted probability. The resulting calibration intercept (ideal value: 0) reflects overall under- or overestimation of risk, and the calibration slope (ideal value: 1) represents the degree of risk separation relative to the development cohort. Overall accuracy was quantified with the Brier score, calculated using the original category-specific predicted probabilities.

Graphical calibration was displayed by plotting, for each risk group, predicted probabilities against observed event rates, with the 45° line representing perfect agreement.

2.10. Open science and transparency

An anonymized patient-level dataset is available as Supplementary Table S2. Although no protocol was preregistered, raw data and full statistical outputs are available from the corresponding author upon reasonable request.

This investigator-initiated study received no external funding, and the authors report no relevant conflicts of interest. No patients were involved in the design, conduct, or reporting of this research.

3. Results

3.1. Sample: composition, outcomes, and characteristics

The final sample of this study comprised 126 unique and consecutive patients (Fig. 1). The mean duration of surgery was 147 min, and the mean estimated blood loss was 456 mL. Table 2 compares the original development cohort of the Anzuategui model with the present temporal external validation cohort. The clinical characteristics, predictive variables, and outcomes of the validation cohort are detailed in Table 3.

The most common histological type was breast cancer (n = 37), followed by prostate (n = 21), lung (n = 11), colorectal (n = 10), renal (n = 10), other solid tumors (n = 10), uterine (n = 8), pharyngeal/laryngeal (n = 5), melanoma (n = 3), sarcoma (n = 3), unknown primary (n = 3), esophageal (n = 2), bladder (n = 2), and thyroid (n = 1) cancers.

The median overall survival was estimated at 228 days (95 % CI: 156 to 327), and the mean overall survival was 575 days (95 % CI: 443 to 707). A total of 21 % of the sample (n = 26) were censored.

Postoperative complications occurred in 41 patients (32 %), and eight of these developed a second complication, all of which were systemic. In total, 49 adverse events were recorded. These are categorized and described in Table 4.

3.2. Comparative model performance

The distribution of patients according to risk categories defined by the predictive models was as follows:

- **Anzuategui**: 17 % (n = 22) low risk, 34 % (n = 43) moderate risk, 48 % (n = 61) high risk;
- **Tomita**: 31 % (n = 39) low risk, 42 % (n = 53) moderate risk, 27 % (n = 34) high risk;

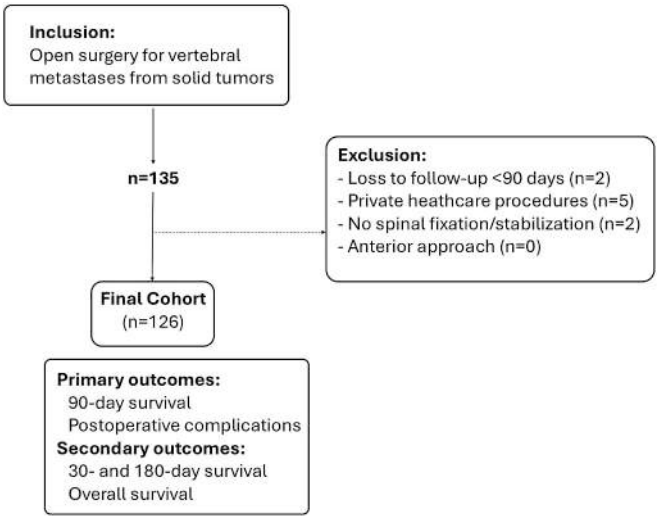


Fig. 1. Flowchart of patient selection for the prospective study cohort, detailing inclusion and exclusion criteria and resulting in 126 unique and consecutive surgeries.

Table 2
Characteristics, Predictors, and Outcomes of the Development Cohort (Retrospective Analysis, n = 205) and the Temporal External Validation Cohort (Prospective Analysis, n = 126) for the Anzuategui Predictive Model.

Variable	Development Cohortn (%) / median (IQR)	External validation Cohortn (%) / Median (IQR)	p-value
Clinical Characteristics			
Age	59 (51–69)	58 (48–68)	0.50
Male sex	114 (55 %)	58 (46 %)	0.11
Surgical approach			
Cervical or Cervicothoracic	11 (5 %)	8 (6 %)	0.71
Thoracic	70 (34 %)	53 (42 %)	0.15
Thoracolumbar	71 (35 %)	49 (39 %)	0.49
Lumbar or Lumbosacral	49 (24 %)	16 (16 %)	0.08
Multiple	4 (2 %)	0 (0 %)	0.11
Primary tumor histology			
Slow-growing	124 (60 %)	53 (42 %)	0.001
Intermediate	49 (24 %)	25 (20 %)	0.39
Rapid	32 (16 %)	48 (38 %)	<0.0001
Predictors			
One or more comorbidities	65 (32 %)	50 (40 %)	0.17
Non-slow progression tumor	81 (40 %)	75 (59 %)	<0.0001
Lymphocytes < 1 ($x10^3$ / μL)	51 (25 %)	59 (47 %)	<0.0001
Outcomes			
Survival > 90 days	117 (57 %)	93 (74 %)	0.002
One or more complications	64 (31 %)	41 (32 %)	0.92

Notes: The development cohort included patients with hematologic malignancies, which explains the higher proportion of slow-growing tumors. The development cohort also included grade III–IV complications according to Rampersaud et al., whereas the external validation cohort included grade II–IV complications. **Abbreviations:** IQR, interquartile range.

- **Bauer (modified)**: 25 % (n = 32) low risk, 44 % (n = 55) moderate risk, 31 % (n = 39) high risk;
- **Van der Linden**: 9 % (n = 12) low risk, 27 % (n = 34) moderate risk, 63 % (n = 80) high risk;
- **Sioutos**: 8 % (n = 10) low risk, 51 % (n = 64) moderate risk, 41 % (n = 52) high risk.

The predictive performance of all five models in relation to postoperative survival and complications is illustrated in Figs. 2–9 and summarized in Tables 5 and 6.

3.3. Calibration results

Calibration analysis for 90-day survival demonstrated that the model systematically overestimated short-term mortality in the validation cohort. Logistic recalibration yielded a calibration intercept of −1.75 (95 % CI −3.01 to −1.23), indicating lower-than-expected event rates, and a calibration slope of 2.05 (95 % CI 1.31 to 3.57), reflecting greater risk separation in the development cohort compared with the external sample. The Brier score for 90-day mortality prediction was 0.225.

For postoperative complications, the calibration pattern was more modest. The calibration intercept was −0.40 (95 % CI −0.84 to 0.04), suggesting slight overestimation of complication risk, while the calibration slope was 0.67 (95 % CI 0.16 to 1.27), indicating attenuated discrimination across risk categories in the validation cohort. The Brier score for complications was 0.213.

Graphical assessment of calibration for both primary outcomes was presented in Fig. 10.

Table 3
Other characteristics, predictors, and outcomes of the temporal external validation Cohort, N = 126.

Variable	n (%) / Median (IQR)
Age ≥ 70 years	24 (19 %)
Age ≥ 65 years	47 (37 %)
Comorbidities	
Diabetes	24 (19 %)
Chronic pulmonary disease	15 (12 %)
AIDS	4 (3 %)
Previous paralysis	4 (3 %)
Renal failure	4 (3 %)
ASIA impairment (A to D)	89 (71 %)
ECOG performance status	
0–2	92 (73 %)
3–4	34 (27 %)
Known visceral metastases	
Any site	69 (55 %)
Lung	58 (46 %)
Liver	28 (22 %)
Brain	13 (10 %)
Known lymph node metastases	74 (59 %)
Known vertebral metastases	
Solitary or isolated	20 (16 %)
Three or more	97 (77 %)
Prior systemic therapy	89 (71 %)
Hemoglobin (g/dL)	11.9 (11–13.1)
Platelets ($\times 10^3$/μL)	270 (203–339)
White blood cells ($\times 10^3$/μL)	10.5 (7.1–13.3)
Lymphocytes ($\times 10^3$/μL)	1.10 (0.67–1.76)
Neutrophils ($\times 10^3$/μL)	8.46 (5.24–11.73)
Neutrophil-to-lymphocyte ratio	7.82 (3.51–14.71)
Platelet-to-lymphocyte ratio	262 (150–399)
Albumin (g/dL)	3.5 (3.2–3.9)
Creatinine (mg/dL)	0.7 (0.5–0.9)
INR	1.02 (1.00–1.12)
C-reactive protein (mg/dL)	2.40 (1.50–4.70)
Overall survival	
30 days	115 (91 %)
90 days	93 (74 %)
180 days	69 (55 %)

Abbreviations.
IQR, interquartile range; AIDS, acquired immunodeficiency syndrome; ASIA, American Spinal Injury Association scale; ECOG, Eastern Cooperative Oncology Group performance status; INR, international normalized ratio.

3.4. Missing data

C-reactive protein values were missing in 7 % of patients, INR in 5 %, and serum albumin in 9 %. None of the predictor variables required by the prognostic models had missing data.

4. Discussion

The present study provides prospective external evidence supporting the clinical usefulness of the Anzuategui model after surgery for spinal metastases. Although its discriminatory ability for 90-day survival (AUC 78 %) was comparable to that of traditional prognostic tools, including the Tomita, Modified Bauer, Van der Linden, and Sioutos models, these findings should be interpreted as confirmation of *acceptable* rather than superior or strong performance. Similar effect sizes have been reported in prior comparative studies, reinforcing that most three-predictor frameworks converge toward moderate discrimination for early postoperative survival [4,11,24].

Regarding the prediction of postoperative complications, the model demonstrated limited predictive accuracy (AUC 68 %), similar to the other four evaluated models, which showed acceptable yet suboptimal performance, with AUC values ranging from 60 to 70 %. These findings are consistent with prior literature indicating that even tools specifically developed to assess frailty or predict postoperative complications exhibit limited discriminative ability [8,14]. Ramos et al. [34] similarly

Table 4
Postoperative complications.

Complication Type	n = 49
Systemic	42 (33 %)
Pneumonia	9 (7.1 %)
Non-infectious respiratory failure	4 (3.2 %)
Urinary tract infection	4 (3.2 %)
Infected pressure ulcer	3 (2.4 %)
Sepsis of unknown origin	3 (2.4 %)
Seizure	3 (2.4 %)
Acute abdomen	2 (1.6 %)
Renal failure	2 (1.6 %)
Venous thrombosis	2 (1.6 %)
Death from unknown cause	2 (1.6 %)
Other	8 (6.3 %)
Local	7 (5.5 %)
Wound infection	3 (2.4 %)
Screw loosening	1 (0.8 %)
Disease progression with paralysis	1 (0.8 %)
Wound dehiscence	1 (0.8 %)
Excessive bleeding	1 (0.8 %)
Infectious	22 (17 %)
Non-infectious	27 (21 %)
Respiratory	13 (10 %)
Non-respiratory	36 (28 %)
Severity grade	
II	9 (7.1 %)
III	21 (17 %)
IV	19 (15 %)

Notes: Complications were categorized as systemic or local and further subclassified as infectious or non-infectious, respiratory or non-respiratory, and by severity grade according to the Rampersaud classification.

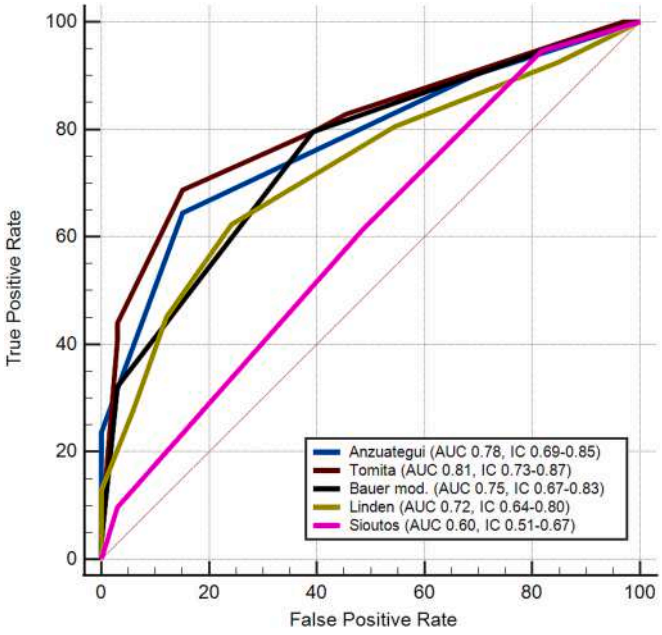


Fig. 2. Discriminative performance of predictive models for 90-day post-operative survival based on ROC curve analysis. Abbreviations: AUC, area under the curve; CI, confidence interval.

reported this limitation in a comparative validation of the New England Spinal Metastasis Score (NESMS) proposed by Ghori et al. [8,35], the Metastatic Spinal Tumor Frailty Index (MSTFI) proposed by Ramos et al. [1], and the Anzuategui models, underscoring the inherent complexity of predicting adverse outcomes in oncologic patients undergoing surgery.

In light of these strengths and limitations, the validated version of

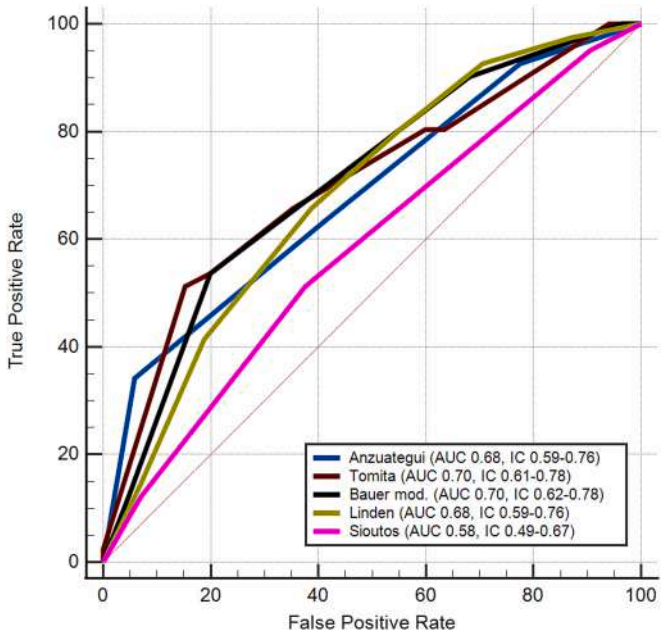
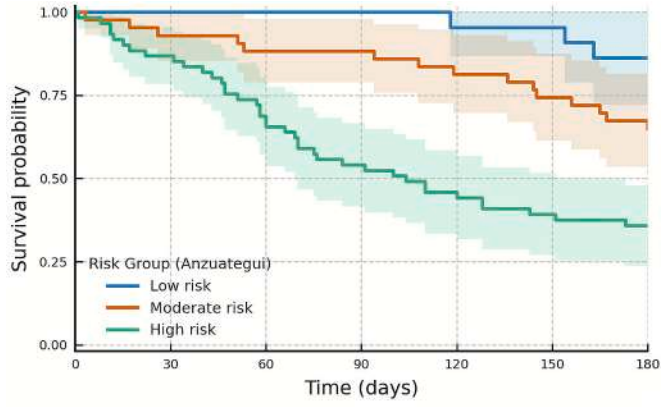


Fig. 3. Discriminative performance of predictive models for postoperative complication risk based on ROC curve analysis. Abbreviations: AUC, area under the curve; CI, confidence interval.

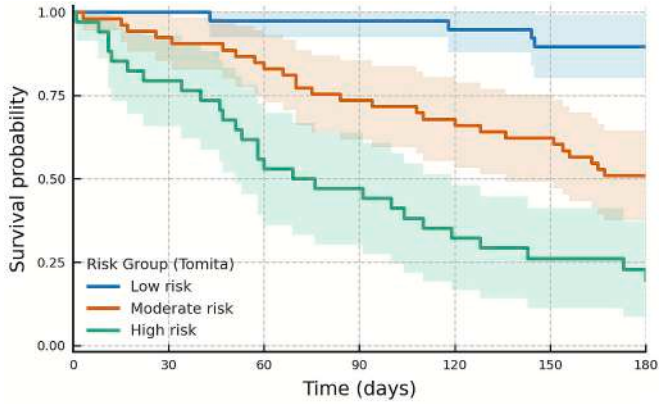


	0	90	180
Low risk	22	22	19
Moderate risk	43	38	29
High risk	61	33	21

Fig. 4. Kaplan–Meier survival curve up to 180 days according to the risk categories of the Anzuategui model. Shaded areas represent 95 % confidence intervals. The table below displays the number of patients at risk over time. P < 0.0001.

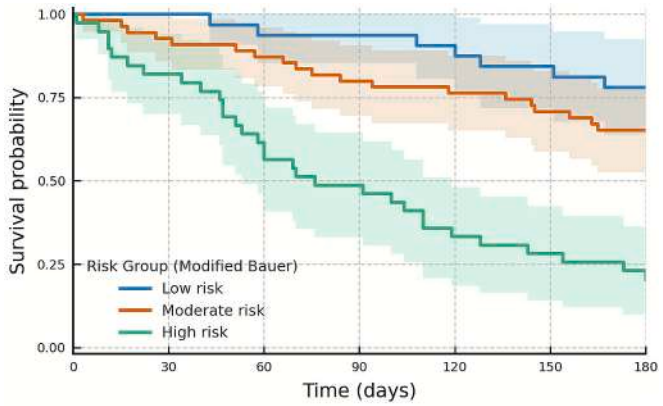
the Anzuategui model presented here in Table 7, accompanied by a streamlined decision-support structure, should be viewed as a practical aid for early perioperative risk estimation rather than a comprehensive solution to surgical prognostication. Its simplicity and frailty-centered approach may facilitate bedside applicability, but its use should be integrated with clinical judgment and complementary prognostic models, especially when decisions hinge on estimated morbidity risk.

A noteworthy and potentially novel finding of this study was the evaluation of model accuracy for surgical indication, using a tailored methodology. Considering postoperative survival beyond 90 days as a marker of appropriate surgical indication, the Anzuategui, Tomita, and Modified Bauer models each achieved approximately 70 % accuracy (Table 6). To our knowledge, no comparable analyses have been



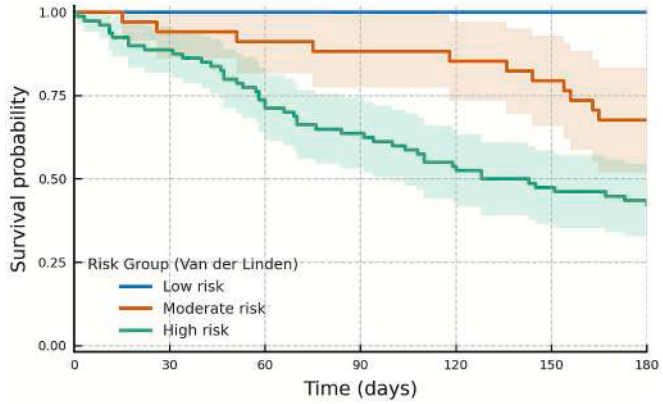
	0	90	180
Low risk	39	38	35
Moderate risk	53	39	27
High risk	34	16	7

Fig. 5. Kaplan–Meier survival curve up to 180 days according to the risk categories of the Tomita model. Shaded areas represent 95 % confidence intervals. The table below displays the number of patients at risk over time. P < 0.0001.



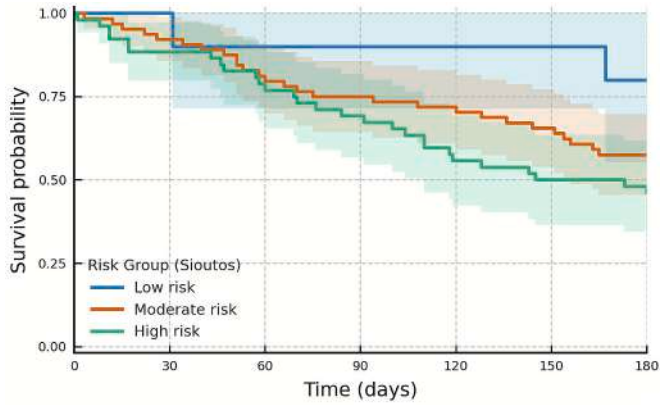
	0	90	180
Low risk	32	30	25
Moderate risk	55	44	35
High risk	39	19	9

Fig. 6. Kaplan–Meier survival curve up to 180 days according to the risk categories of the modified Bauer model. Shaded areas represent 95 % confidence intervals. The table below displays the number of patients at risk over time. P < 0.0001.



	0	90	180
Low risk	12	12	12
Moderate risk	34	30	23
High risk	80	51	34

Fig. 7. Kaplan–Meier survival curve up to 180 days according to the risk categories of the Van der Linden model. Shaded areas represent 95 % confidence intervals. The table below displays the number of patients at risk over time. P < 0.0001.



	0	90	180
Low risk	10	9	8
Moderate risk	64	48	36
High risk	52	36	25

Fig. 8. Kaplan–Meier survival curve up to 180 days according to the risk categories of the Sioutos model. Shaded areas represent 95 % confidence intervals. The table below displays the number of patients at risk over time. P = 0.05.

previously reported in the medical literature.

From a qualitative standpoint, among the three models with good accuracy (>70 %) for surgical decision-making, Tomita and Modified Bauer showed high sensitivity (>80 %), making them more effective in identifying surgical candidates. In contrast, the Anzuategui model, designed to detect patient frailty, demonstrated superior specificity (>85 %), making it particularly useful for identifying patients less likely to benefit from surgery. This divergence allows institutions to adopt a model aligned with their clinical philosophy—more conservative (emphasizing specificity) or more interventional (emphasizing sensitivity).

The calibration analysis demonstrated heterogeneous performance across outcomes and cohorts. For 90-day survival, the model systematically underestimated survival in the validation cohort and showed greater separation between risk strata in the development cohort. A key contributor to this discrepancy is the disproportionate representation of high-risk patients in the validation sample. This occurred, in part, because hematologic malignancies, typically associated with more favorable prognosis, were intentionally excluded from the validation cohort, shifting the case mix toward biologically more aggressive tumors.

Despite this higher-risk distribution, the validation cohort paradoxically exhibited superior 90-day survival. This pattern highlights meaningful contextual differences between cohorts, potentially driven by temporal improvements in oncologic therapies, evolving perioperative practice, shifts in surgical indications, and more refined selection of operative candidates. These observations align with broader longitudinal trends in our institution, which are being examined in a separate study. Such structural changes may alter baseline risk distributions and limit the transportability of models developed in earlier clinical eras.

A similar pattern was observed for postoperative complications, although the magnitude of miscalibration was smaller. The model modestly overestimated complication risk and demonstrated attenuated risk separation in the validation cohort. Interpretation of these findings

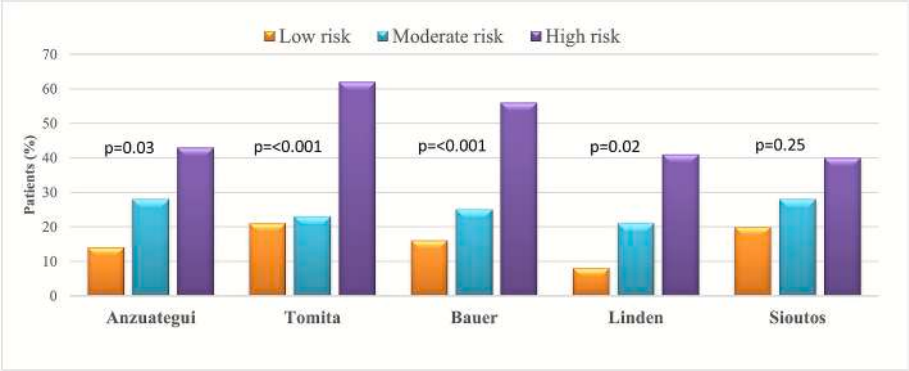


Fig. 9. Incidence of postoperative complications according to the risk categories of predictive models. P-value calculated using the Cochran–Armitage trend test.

Table 5
Discriminative ability according to the area under the curve.

Predictive model	30-day survival AUC(CI)/p	90-day survival AUC(CI)/p	180-day survival AUC(CI)/p	Complications AUC(CI)/p
Anzuategui	0.72 (0.63–0.80) Ref.	0.78 (0.70–0.85) Ref.	0.73 (0.64–0.81) Ref.	0.68(0.59–0.76) Ref.
Tomita	0.76 (0.68–0.83) 0.43	0.81 (0.73–0.87) 0.49	0.81 (0.74–0.88) 0.04	0.70(0.61–0.78) 0.67
Bauer (mod.)	0.73 (0.65–0.81) 0.78	0.75 (0.67–0.83) 0.63	0.74 (0.66–0.82) 0.83	0.70(0.62–0.78) 0.66
Linden	0.66 (0.58–0.75) 0.46	0.72 (0.64–0.80) 0.26	0.73 (0.64–0.80) 0.96	0.68(0.59–0.76) 0.99
Sioutos	0.60 (0.51–0.69) 0.24	0.60 (0.51–0.69) 0.003	0.60 (0.51–0.69) 0.02	0.58(0.49–0.67) 0.10

Notes: AUC values represent the probability that the model correctly classifies patients; a value of 0.50 indicates no predictive power and a value of 1.0 indicates maximum predictive accuracy. **Abbreviations:** AUC, area under the ROC curve; ROC, receiver operating characteristics; CI, confidence interval; Ref., reference (used in the calculation of p-values); mod., modified.

Table 6
Accuracy of surgical indication according to predictive models.

Predictive Model	Accuracy (CI) / p	Sensitivity (CI) / p	Specificity (CI) / p
Anzuategui	70 % (61–78) / Ref.	64 % (54–74) / Ref.	85 % (68–95) / Ref.
Tomita	75 % (67–83) / 0.37	83 % (74–90) / <0.001	54 % (36–72) / <0.001
Bauer (mod.)	74 % (66–82) / 0.48	80 % (70–87) / 0.005	59 % (41–76) / <0.001
Linden	56 % (47–65) / 0.02	45 % (35–56) / 0.002	88 % (72–97) / 0.49
Sioutos	59 % (50–67) / 0.07	61 % (51–71) / 0.62	51 % (33–69) / <0.001

Notes: Surgical indication was assumed when each model classified the case as low or moderate risk, according to Table 1. A correct surgical indication was defined when postoperative survival exceeded 90 days; likewise, when the model indicated high risk and observed survival was less than 90 days, this was also considered a correct prediction. In this context, sensitivity reflects the model's ability to identify individuals suitable for surgery, whereas specificity represents its ability to correctly identify frailty (high risk). **Abbreviations:** CI, confidence interval; Ref., reference (used in the calculation of p-values); mod., modified.

requires particular caution because the definitions of complications differed between cohorts: in the validation sample, grade 2 events according to Rampersaud were also captured, thereby increasing the recorded incidence, whereas the development cohort, retrospective by design, was subject to expected underreporting of events. These methodological differences alone could narrow apparent risk gradients or create the impression of overestimation.

Moreover, the same contextual factors that may explain improved 90-day survival—such as advances in oncologic care, evolving perioperative practices, and refined surgical selection—could also plausibly reduce the relative incidence of complications over time. Whether these improvements occurred to a meaningful degree cannot be fully determined here, given the differences in complication ascertainment across cohorts. Collectively, these considerations highlight the inherent challenges of calibrating complication risk and underscore the need for cautious interpretation of these results.

4.1. Clinical applicability of three-predictor models

With at least 10 validated and widely recognized therapeutic prediction models currently available for spinal metastases [36], selecting the most appropriate tool remains a frequent challenge for spine surgeons. In an era dominated by increasingly sophisticated machine learning–based tools [3,37], the question arises: do traditional models relying on only three variables still hold clinical relevance [38]?

It is important to recognize that the ultimate decision to undergo surgery lies with the patient. Subjective, ethical, and existential factors frequently influence this difficult choice, as the final goal of treatment is to improve and preserve independence and acceptable health-related quality of life [39]. Even highly accurate risk estimates produced by complex algorithms may fail to capture the personal values and priorities that shape how patients weigh the risks and potential benefits of surgery—often in the context of a progressive or terminal illness. In this setting, simpler models that classify risk into intuitive categories (low, moderate, high) may be more effective in facilitating communication and supporting shared decision-making.

By selecting a model that includes three out of the 20 currently described predictors [12], the clinician focuses on a specific biological or clinical dimension of the patient. We propose that predictors be grouped into three distinct domains: (1) expected oncologic progression, (2) functional status, and (3) patient frailty. The combination of predictors used in each model thus defines its specific clinical perspective.

The Tomita and Modified Bauer models are examples of tools that focus exclusively on oncologic progression, incorporating tumor histology and the extent of metastatic disease. In contrast, more comprehensive traditional models such as the modified Katagiri [11] include predictors from all three proposed domains, while Van der Linden and Sioutos incorporate variables related to both oncologic progression and functional status.

Models that incorporate functional performance scales—such as Eastern Cooperative Oncology Group (ECOG), Karnofsky, Frankel, or

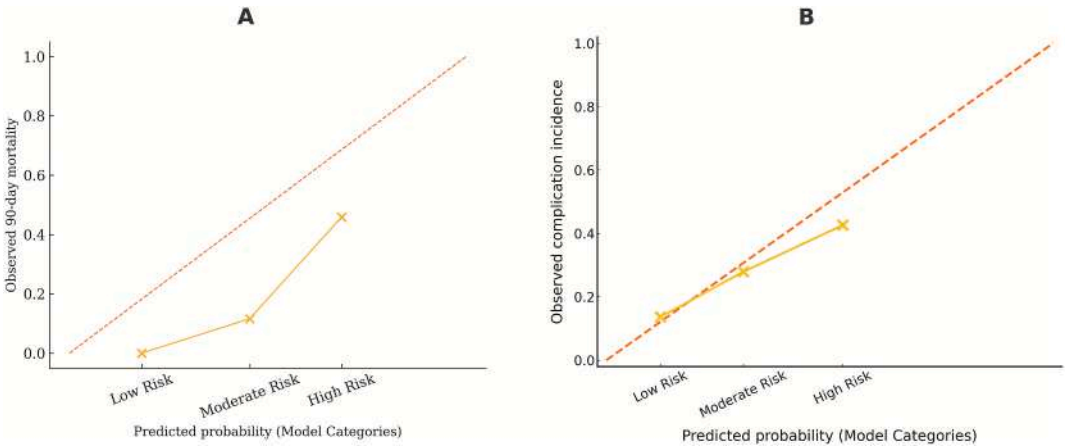


Fig. 10. Calibration plots comparing predicted and observed event rates for the primary outcomes. Panel A shows calibration for 90-day postoperative survival, and Panel B depicts calibration for the incidence of complications.

Table 7
Validated version of the Anzuategui frailty-based prediction model to support decision-making in cases of spinal metastases from solid tumors.

Predictors	Present predictors	Risk category	Surgical strategy
Comorbidities[1]	0	Low	Surgery
Tumor with non-slow progression[2]	1	Moderate	Upfront surgery if urgent, or proceed with in-depth risk assessment.
Lymphocyte count $< 1{,}000/\mu L^3$	2 or 3	High	Conservative treatment

Notes: [1] Presence of at least one of the following comorbidities: diabetes mellitus, chronic pulmonary disease, prior myocardial infarction, congestive heart failure, cardiac arrhythmia, pulmonary circulation disorder, peripheral vascular disease, cerebrovascular disease, dementia, renal failure, hepatic failure, connective tissue disease, coagulopathy, prior paralysis, peptic ulcer disease, acquired immunodeficiency syndrome. [2] Solid tumors with slow progression: hormone-dependent breast cancer, prostate cancer, thyroid cancer, and other rare histological types with slow progression. [3] Total preoperative peripheral blood lymphocyte count.

American Spinal Injury Association (ASIA) Impairment Scale—featured in tools like Tokuhashi [40], Sioutos, Van der Linden, Katagiri, Spine Oncology Research Group (SORG) Nomogram [13], and machine learning–based models such as SORG-MLA [41] and PathFx 3.0 [42], require careful evaluation. While these variables possess predictive value, their inclusion may conflict with one of the primary objectives of spinal surgery: to preserve or restore neurologic function. It is paradoxical that models might dissuade surgery precisely in patients who may benefit most from it.

Therefore, we propose that model selection should be guided by multiple considerations, including predictive accuracy (often greater with more variables), specificity (whether the model focuses on frailty, oncologic progression, or function), practicality (ease and cost of use), and clinical utility (whether it truly supports decision-making). In many cases, adopting more than one model may be a reasonable strategy: one that is simple, specific, and immediately applicable; another that is more complex and capable of providing precise, individualized predictions.

4.2. *Illustrative case*

The clinical case presented in Fig. 11 illustrates the practical application of three-predictor models in surgical decision-making for spinal metastasis. The patient was a 64-year-old male undergoing treatment for prostate cancer, referred to the Orthopedic Oncology service due to progressive motor weakness in the right lower limb, associated with severe lumbar pain. His medical history included chronic pulmonary disease secondary to long-term tobacco use, with a smoking load estimated at 50 pack-years.

Local staging revealed an expansive neoplastic lesion compressing the spinal canal at S1 and S2. Distant staging demonstrated metastatic spread to pelvic lymph nodes and multiple skeletal sites (pelvis, thoracic spine, lumbar spine, and ribs). The patient's cardiac surgical risk was classified as low according to the Lee score [43] and anesthetic risk was ASA grade II.

In accordance with the NOMS framework [44] (Neurologic,

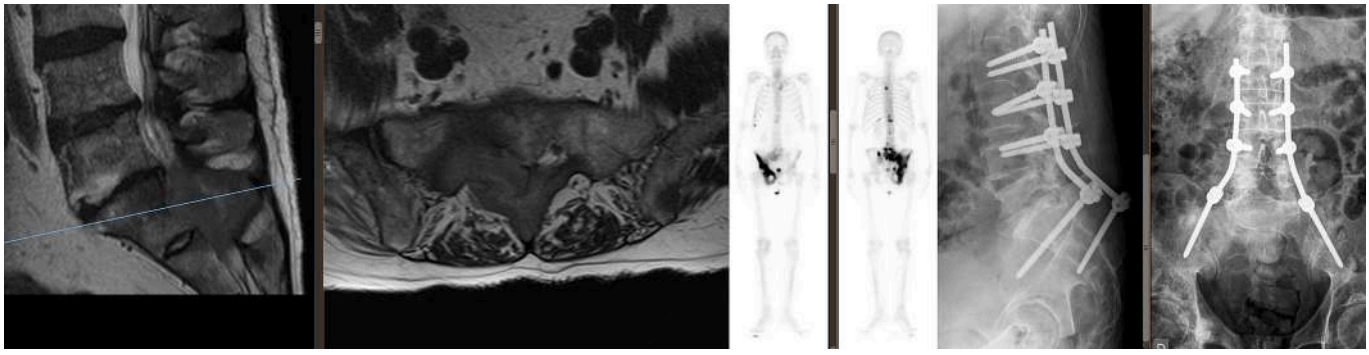


Fig. 11. Illustrative case. A 64-year-old male with metastatic prostate cancer presented with progressive right lower-limb weakness and severe low back pain. His medical history included chronic pulmonary disease related to long-term smoking (50 pack-years). *Left:* Preoperative MRI demonstrating an expansile epidural lesion at S1–S2 causing significant canal compromise. *Center:* Bone scintigraphy showing multiple osteoblastic metastases involving the right hemipelvis, ribs, sacrum, thoracic, and lumbar spine. *Right:* Postoperative lumbosacral radiograph following decompression and spinopelvic fixation for metastatic stabilization.

Oncologic, Mechanical, and Systemic considerations), surgical treatment was indicated. Subsequently, all three-predictor models analyzed in this study were applied, resulting in the following classifications: Anzuategui (moderate risk, 1 point), Tomita (moderate risk, 5 points), Modified Bauer (moderate risk, 2 points), Van der Linden (moderate risk, 4 points), and Sioutos (moderate risk, 1 point).

Given the imminent risk of cauda equina syndrome and an acceptable surgical risk profile, the team opted for decompressive surgery with spinopelvic fixation.

Postoperatively, the patient experienced significant intraoperative bleeding and required a prolonged stay in the Intensive Care Unit, which was classified as a Grade 2 complication. Cauda equina syndrome was successfully avoided. The patient reported a substantial reduction in neuropathic pain and maintained functional motor strength until his death, which occurred 664 days after surgery.

4.3. Recommendations for using the Anzuategui model

The three-predictor model proposed by our group in 2019 [27] was designed to combine one predictor related to tumor progression (non-slow-growing tumor) with two indicators of frailty (presence of comorbidities and low lymphocyte count), resulting in a simple and easy-to-remember score ranging from 0 to 3 points (one point per negative predictor).

Here, we present its temporal external validation using a prospective cohort in which the predictive variables were systematically assessed and standardized, potentially reducing various sources of bias.

A major strength of the Anzuategui model is its simplicity, which does not require advanced imaging or oncologic restaging. We believe its ideal application lies in bedside screening, as proposed in the Vertebral Metastasis Surgery Decision Tree (Fig. 12). In urgent cases involving spinal cord compression, the seemingly straightforward decision to operate may actually require several days of deliberation, during which simple and intuitive tools can offer valuable support to the clinical team.

It is worth highlighting that the total peripheral blood lymphocyte count—a predictor included in our model and associated with malnutrition, immunosuppression, and inflammation—was shown by our group to be a strong independent prognostic marker [45]. This variable is also employed in recent machine learning models such as SORG-MLA [41] and PathFx 3.0 [42]. We recommend that blood sampling be performed as close to surgery as possible, ideally within three days, due to expected variability in white blood cell differentials.

A clear limitation of the Anzuategui model lies in the heterogeneous and sometimes subjective assessment of frailty-related comorbidities.

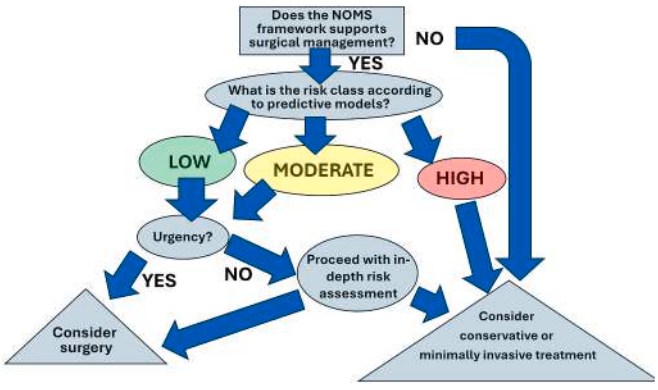


Fig. 12. Vertebral Metastasis Surgery Decision Tree. Decision-making follows the NOMS framework, which integrates neurological (N), oncological (O), mechanical (M), and systemic (S) considerations. Rectangular boxes represent decision nodes, ovals indicate chance nodes, and triangles denote terminal outcomes.

Clinical conditions often present with wide and progressive spectrums of severity. For example, newly diagnosed diabetes without end-organ damage is unlikely to significantly impact surgical outcomes. However, since the model is composed of only three predictors, the inclusion or exclusion of a single point may alter a patient's risk classification.

In our setting, chronic pulmonary disease is frequently underdiagnosed [46]. In many hospital environments, spirometry and specialist consultation are not always readily available. Therefore, we consider that a clinically assumed diagnosis of Chronic Obstructive Pulmonary Disease (COPD) is, in some cases, justified, and one point was accordingly assigned in the model. Accordingly, we recommend that experienced hospitalist clinicians participate in the evaluation of these patients and make informed clinical judgments regarding comorbidities. Only those considered clinically significant, such as diabetes and COPD [47,48] (see suggested list in Table 7) should contribute to the model score. Common conditions such as grade I or II obesity, well-controlled hypertension, mild peripheral venous insufficiency, hypothyroidism, dyslipidemia, and prediabetes are generally not considered significant comorbidities within this model.

It is also crucial to reflect on the histologic types responsible for spinal metastases and their use in predictive modeling. Understanding tumor progression requires in-depth knowledge of tumor biology, which continues to evolve—particularly with advancements in genomic, molecular, and hormonal biomarkers [49]. For instance, in 2001, Tomita et al. [21] did not have access to molecular tools necessary to accurately classify breast cancer subtypes. In contrast, Katagiri et al. [50] in 2014, incorporated molecular markers for lung cancer and considered hormonal therapy response in both breast and prostate cancers.

Rather than adopting rigid histology-based lists when developing predictive models, we advocate for evaluating tumors based on their estimated progression rate, considering available diagnostic tools—histopathologic, molecular, genetic, hormonal, or otherwise. When using the Anzuategui model, we recommend thoughtful consideration when classifying tumor aggressiveness, avoiding overreliance on Table 7 and encouraging individualized assessment whenever possible, ideally involving multidisciplinary input from pathologists and medical oncologists.

5. Future directions

Future studies should aim to validate the Anzuategui model in non-surgical or demographically diverse populations to assess its applicability in different clinical and epidemiological settings. Additionally, incorporating new evidence into the model may enhance its performance and clinical utility as a decision-support tool.

To ensure appropriate application of the validated model, future researchers should focus on accurate identification of comorbidities and timely collection of laboratory data, particularly lymphocyte counts close to the time of surgery. Therefore, we recommend that future validations be conducted through prospective designs, as retrospective studies frequently fail to control for key variables.

6. Limitations

This study has several limitations. The lack of complete blinding may have introduced selection, performance, measurement, and confirmation biases. Additionally, the generalizability of the results is limited, as the study was conducted in a single institution with a characteristically heterogeneous sample. Clinical interpretation of complex cases is inherently subject to judgment errors. Furthermore, differences in access to therapeutic resources may lead to variable surgical outcomes for spinal metastases. Cancer incidence and socioeconomic factors [51,52] also vary across countries and may influence the applicability of these findings. A further limitation is that surgical approaches vary substantially across institutions, as highlighted in recent comparative studies [53,54], and the technique predominantly used in our center may not

reflect the procedures most commonly performed worldwide. Although our cohort is sizeable, the number of complications was not large enough to support more granular inferences about these events. Moreover, many complications in oncologic patients may arise independently of surgery itself, which requires caution when interpreting associations and evaluating the predictive performance of any model.

7. Conclusions

The Anzuategui model demonstrated acceptable external performance, with greater validity for predicting 90-day survival than for postoperative complications. Its simplicity and frailty-centered structure make it a practical bedside tool, particularly in urgent or resource-limited settings. Integrating this approach with established prognostic models may support more balanced decision-making across diverse clinical scenarios.

CRediT authorship contribution statement

Pedro Reggiani Anzuatégui: Writing – review & editing, Writing – original draft, Project administration, Methodology, Formal analysis, Data curation, Conceptualization. **Glauco José Pauka Mello:** Writing – review & editing, Supervision, Data curation, Conceptualization. **Ana Valéria Brunetti Rigolino:** Data curation. **Lucas Emanuel Sauer Larocca:** Data curation. **Cássio Zini:** Writing – review & editing, Supervision. **Carmen Australia Paredes Marcondes Ribas:** Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jbo.2026.100739.

References

[1] R.R. De la Garza, C.R. Goodwin, A. Jain, N. Abu-Bonsrah, C.G. Fisher, C. Bettegowda, 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. 95 (2016) 548–555.e4, https://doi.org/10.1016/j.wneu.2016.08.029.

[2] B. Knapp, A. Govindan, S.S. Patel, K. Pepin, N. Wu, S. Devarakonda, et al., Outcomes in patients with spinal metastases managed with surgical intervention, Cancers (Basel) 16 (2024), https://doi.org/10.3390/cancers16020438.

[3] A.A. Shah, J.H. Schwab, Predictive modeling for spinal metastatic disease, Diagnostics 14 (2024), https://doi.org/10.3390/diagnostics14090962.

[4] A.K. Ahmed, C.R. Goodwin, A. Heravi, R. Kim, N. Abu-Bonsrah, E. Sankey, et al., Predicting survival for metastatic spine disease: a comparison of nine scoring systems, Spine Journal 18 (2018) 1804–1814, https://doi.org/10.1016/j.spinee.2018.03.011.

[5] G.S. Collins, J.B. Reitsma, D.G. Altman, K.G.M. Moons, Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) the TRIPOD statement, Circulation 131 (2015) 211–219, https://doi.org/10.1161/CIRCULATIONAHA.114.014508.

[6] S.J. Park, C.H. Ma, C.S. Lee, C.Y. Jeon, T.S. Shin, J.S. Park, Survival and functional outcomes after surgical treatment for spinal metastasis in patients with a short life expectancy, J. Clin. Med. 12 (2023), https://doi.org/10.3390/jcm12010046.

[7] T. Uto, S. Kato, N. Yokogawa, T. Shimizu, S. Demura, Y. Shiratani, et al., Risk factors for early 3-month postoperative mortality in metastatic spinal tumor surgery. Spine (Phila Pa 1976) 2025. https://doi.org/10.1097/BRS.0000000000005359.

[8] A.J. Schoenfeld, M.L. Ferrone, J.H. Schwab, J.A. Blucher, L.B. Barton, D.G. Tobert, et al., Prospective validation of a clinical prediction score for survival in patients with spinal metastases: the New England Spinal Metastasis score, Spine Journal 21 (2021) 28–36, https://doi.org/10.1016/j.spinee.2020.02.009.

[9] J. Ryvlin, M.K. Hamad, B. Wang, J. Xavier, R.R. De la Garza, S.G. Murthy, et al., Performance assessment of the american college of surgeons risk calculator in metastatic spinal tumor surgery, Spine (Phila Pa 1976) 48 (2023) 825–831, https://doi.org/10.1097/BRS.0000000000004644.

[10] A.V. Karhade, B. Fenn, O.Q. Groot, A.A. Shah, H.-K. Yen, M.H. Bilsky, et al., Development and external validation of predictive algorithms for six-week mortality in spinal metastasis using 4,304 patients from five institutions, Spine J. 22 (2022) 2033–2041, https://doi.org/10.1016/j.spinee.2022.07.089.

[11] Z. Li, L. Guo, B. Guo, P. Zhang, J. Wang, X. Wang, et al., Evaluation of different scoring systems for spinal metastases based on a chinese cohort, Cancer Med. 12 (2023) 4125–4136, https://doi.org/10.1002/cam4.5272.

[12] B. Santipas, K. Veerakanjana, P. Ittichaiwong, P. Chavalparit, S. Wilartratsami, P. Luksanapruksa, Development and internal validation of machine-learning models for predicting survival in patients who underwent surgery for spinal metastases, Asian Spine J 18 (2024) 325–335. https://doi.org/10.31616/asj.2023.0314.

[13] N.R.P. Pereira, S.J. Janssen, E. Van Dijk, M.B. Harris, F.J. Hornicek, M.L. Ferrone, et al., Development of a prognostic survival algorithm for patients with metastatic spine disease, J. Bone nd Joint Surg. – Am. 98 (2016) 1767–1776, https://doi.org/10.2106/JBJS.15.00975.

[14] L.C.M. Ndjonko, A. Chakraborty, F. Petri, S.M.A. Alavi, T. Matsuo, F. Borgonovo, et al., Evaluating predictive performance and generalizability of traditional and artificial intelligence models in predicting surgical site infections post-spinal surgery: a systematic review, Spine J. (2025), https://doi.org/10.1016/j.spinee.2025.07.032.

[15] J. Ryvlin, S.W. Kim, R.R. De la Garza, M. Hamad, A. Stock, E. Owolo, et al., External validation of an online wound infection and wound reoperation risk calculator after metastatic spinal tumor surgery, World Neurosurg. 185 (2024) e351–e356, https://doi.org/10.1016/j.wneu.2024.02.005.

[16] D.C. de Reus, R.H. Kuijten, P. Saha, D.A.A. Lastoria, A. Warr-Esser, C.F.C. Taylor, et al., External validation of a machine learning prediction model for massive blood loss during surgery for spinal metastases: a multi-institutional study using 880 patients, Spine Journal 25 (2025) 1386–1399, https://doi.org/10.1016/j.spinee.2025.03.018.

[17] S. Maki, Y. Shiratani, S.Orita, A. Suzuki, K. Tamai, T. Shimizu, et al., Predicting postoperative neurological outcomes in metastatic spinal tumor surgery using machine learning. Spine (Phila Pa 1976) 2025. https://doi.org/10.1097/BRS.0000000000005322.

[18] R.H. Kuijten, B. Bindels, O. Groot, E. Huele, R. Gal, M. Groot, et al., Predicting quality of life of patients after treatment for spinal metastatic disease: development and internal evaluation, Spine J. 25 (2025) 1371–1385, https://doi.org/10.1016/j.spinee.2025.03.016.

[19] R.R. De la Garza, J.A. Benton, Y. Gelfand, M. Echt, M.K. Hamad, M.D. Kinon, et al., A novel clinical scoring system for perioperative morbidity in metastatic spinal tumor surgery, Spine (Phila Pa) 46 (46) (2021) E161–E166, https://doi.org/10.1097/BRS.0000000000003733.

[20] W. Jiang, J. Zhang, W. Shi, X. Cao, X. Zhao, B. Zhang, et al., An artificial intelligence platform for predicting postoperative complications in metastatic spinal surgery: development and validation study, J Big Data 12 (2025), https://doi.org/10.1186/s40537-025-01155-0.

[21] K. Tomita, N. Kawahara, T. Kobayashi, A. Yoshida, H. Murakami, T. Akamaru, Surgical strategy for spinal Metastases, Spine (Phila Pa 1976) 26 (26) (2001) 298–306, https://doi.org/10.1097/00007632-200102010-00016.

[22] H.C.F. Bauer, R. Wedin, Survival after surgery for spinal and extremity metastases: prognostication in 241 patients, Acta Orthop. 66 (1995) 143–146, https://doi.org/10.3109/17453679508995508.

[23] A. Leithner, R. Radl, G. Gruber, M. Hochegger, K. Leithner, H. Welkerling, et al., Predictive value of seven preoperative prognostic scoring systems for spinal metastases, Eur. Spine J. 17 (2008) 1488–1495, https://doi.org/10.1007/s00586-008-0763-1.

[24] B.J.J. Bindels, R.H. Kuijten, O.Q. Groot, E.H. Huele, R. Gal, M.C.H. de Groot, et al., External validation of 12 existing survival prediction models for patients with spinal metastases, Spine Journal 25 (2025) 1347–1359, https://doi.org/10.1016/j.spinee.2025.01.014.

[25] M. Ghanem, A.K. Ghaith, V.G. El-Hajj, A. Bhandarkar, A. de Giorgio, A. Elmi-Terander, et al., Limitations in evaluating machine learning models for imbalanced binary outcome classification in spine surgery: a systematic review, Brain Sci. 13 (2023), https://doi.org/10.3390/brainsci13121723.

[26] G. Tabourel, L.M. Terrier, A. Dubory, J. Cristini, L.R. Le Nail, A.R. Cook, et al., Are spine metastasis survival scoring systems outdated and do they underestimate life expectancy? caution in surgical recommendation guidance, J. Neurosurg. Spine 35 (2021) 527–534, https://doi.org/10.3171/2020.12.SPINE201741.

[27] P.R. Anzuatégui, C.L. Da, G.J.P. Mello, E. Stieven Filho, X.S. Graells, Spinal metastasis surgery: a proposal for a predictive model of morbidity and mortality, Rev Bras Ortop (Sao Paulo) 54 (2019) 665–672, https://doi.org/10.1055/s-0039-1697018.

[28] Y.R. Rampersaud, E.R.P. Moro, M.A. Neary, K. White, S.J. Lewis, E.M. Massicotte, et al., Intraoperative adverse events and related postoperative complications in spine surgery: implications for enhancing patient safety founded on evidence-based protocols, Spine (Phila PA 1976) 31 (2006) 1503–1510, https://doi.org/10.1097/01.brs.0000220652.39970.c2.

[29] M.E. Charlson, P. Pompei, K.L. Ales, C.R. MacKenzie, A new method of classifying prognostic comorbidity in longitudinal studies: development and validation, J. Chronic Dis. 40 (1987) 373–383, https://doi.org/10.1016/0021-9681(87)90171-8.

[30] A. Elixhauser, C. Steiner, D.R. Harris, R.M. Coffey, Comorbidity measures for use with administrative data, Med. Care 36 (1998) 8–27, https://doi.org/10.1097/00005650-199801000-00004.

[31] R.A. Patchell, P.A. Tibbs, W.F. Regine, R. Payne, S. Saris, R.J. Kryscio, et al., Direct decompressive surgical resection in the treatment of spinal cord compression caused by metastatic cancer: a randomised trial, Lancet 366 (2005) 643–648, https://doi.org/10.1016/S0140-6736(05)66954-1.
[32] Y.M. van der Linden, S.P.D.S. Dijkstra, E.J.A. Vonk, C.A.M. Marijnen, J.W.H. Leer, Dutch Bone Metastasis Study Group. Prediction of survival in patients with metastases in the spinal column: results based on a randomized trial of radiotherapy, Cancer 103 (2) (2005) 320–328.
[33] P.J. Sioutos, E. Arbit, C.F. Meshulam, J.H. Galicich, Spinal metastases from solid tumors. Analysis of factors affecting survival, Cancer 76 (1995) 1453–1459, https://doi.org/10.1002/1097-0142(19951015)76:8<1453::AID-CNCR2820760824>3.0.CO;2-T.
[34] R.R. De la Garza, I. Naidu, J.H. Choi, Z. Pennington, C.R. Goodwin, D.M. Sciubba, et al., Comparison of three predictive scoring systems for morbidity in oncological spine surgery, J. Clin. Neurosci. 94 (2021) 13–17, https://doi.org/10.1016/j.jocn.2021.09.031.
[35] A.K. Ghori, D.A. Leonard, A.J. Schoenfeld, E. Saadat, N. Scott, M.L. Ferrone, et al., Modeling 1-year survival after surgery on the metastatic spine, Spine Journal 15 (2015) 2345–2350, https://doi.org/10.1016/j.spinee.2015.06.061.
[36] S. Smeijers, B. Depreitere, Prognostic scores for survival as decisional support for surgery in spinal metastases: a performance assessment systematic review, Eur. Spine J. 30 (2021) 2800–2824, https://doi.org/10.1007/s00586-021-06954-6.
[37] S.B. Wilson, J. Ward, V. Munjal, C.S.A. Lam, M. Patel, P. Zhang, et al., Machine learning in spine oncology: a narrative review, Global Spine J 15 (2025) 210–227, https://doi.org/10.1177/21925682241261342.
[38] D.-H. Kang, J.-S. Park, M. Kang, K. Jung, C.-S. Lee, S.-J. Park, Which scoring system best predicts long-term survival in patients with spinal metastasis in the era of targeted systemic treatment? A comparative study of eight prognostic models. Spine (Phila Pa 1976) 2024. https://doi.org/10.1097/BRS.0000000000005205.
[39] S. Terzi, C. Griffoni, S. Rosa, C. Cini, E. Asunis, C. Alcherigi, et al., Health- related quality of life after surgery for spinal metastases, J. Bone Oncol. 52 (2025), https://doi.org/10.1016/j.jbo.2025.100675.
[40] Y. Tokuhashi, H. Matsuzaki, H. Oda, M. Oshima, J. Ryu, A revised scoring system for preoperative evaluation of metastatic spine tumor prognosis, Spine (Phila Pa 1976) 30 (30) (2005) 2186–2191, https://doi.org/10.1097/01.brs.0000180401.06919.a5.
[41] A.A. Shah, A.V. Karhade, H.Y. Park, W.L. Sheppard, L.J. Macyszyn, R.G. Everson, et al., Updated external validation of the SORG machine learning algorithms for prediction of ninety-day and one-year mortality after surgery for spinal metastasis, Spine Journal 21 (2021) 1679–1686, https://doi.org/10.1016/j.spinee.2021.03.026.
[42] C. Carrwik, P. Tsagkozis, R. Wedin, Y. Robinson, Predicting survival of patients with spinal metastatic disease using PathFx 3.0 – a validation study of 668 patients in Sweden, Brain and Spine 2 (2022), https://doi.org/10.1016/j.bas.2022.101669.
[43] T.H. Lee, E.R. Marcantonio, C.M. Mangione, E.J. Thomas, C.A. Polanczyk, E. F. Cook, et al., Derivation and prospective validation of a simple index for prediction of cardiac risk of major noncardiac surgery, Circulation 100 (1999) 1043–1049, https://doi.org/10.1161/01.CIR.100.10.1043.
[44] I. Laufer, D.G. Rubin, E. Lis, B.W. Cox, M.D. Stubblefield, Y. Yamada, et al., The NOMS framework: approach to the treatment of spinal metastatic tumors, Oncologist 18 (2013) 744–751, https://doi.org/10.1634/theoncologist.2012-0293.
[45] P.R. Anzuatégui, G.J.P. Mello, A.V.B. Rigolino, Lymphopenia predicts 30-day morbidity and mortality following spinal metastasis surgery, North Am. Spine Soc. J. 6 (2021), https://doi.org/10.1016/j.xnsj.2021.100062.
[46] G.L. Moreira, B.M. Manzano, M.R. Gazzotti, O.A. Nascimento, R. Perez-Padilla, A. M.B. Menezes, et al., PLATINO, a nine-year follow-up study of COPD in the city of São Paulo, Brazil: the problem of underdiagnosis, J. Bras. Pneumol. 40 (2014) 30–37, https://doi.org/10.1590/S1806-37132014000100005.
[47] M. Pojskić, B. Saß, M.H.A. Bopp, S. Wilke, C. Nimsky, Determinants of overall and readmission-free survival in patients with metastatic epidural spinal cord compression †, Cancers (Basel) 16 (2024) https://doi.org/10.3390/cancers16244248.
[48] L.-C. Wu, Y.-Y. Hsieh, C. Ic, C.-J. Chiang, Life-threatening perioperative complications among older adults with spinal metastases: an analysis based on a nationwide inpatient sample of the US, J Geriatr Oncol 14 (2023) 101597, https://doi.org/10.1016/j.jgo.2023.101597.
[49] S. Das, M.K. Dey, R. Devireddy, M.R. Gartia, Biomarkers in cancer detection, diagnosis, and prognosis, Sensors 24 (2024), https://doi.org/10.3390/s24010037.
[50] H. Katagiri, R. Okada, T. Takagi, M. Takahashi, H. Murata, H. Harada, et al., New prognostic factors and scoring system for patients with skeletal metastasis, Can. Med. 3 (2014) 1359–1367, https://doi.org/10.1002/cam4.292.
[51] B.T. van Munster, J.J.G.J. Amelink, O.Q. Groot, T.M. de Groot, J.-J. Verlaan, D. G. Tobert, The influence of social determinants on clinical presentation in patients surgically treated for metastatic spinal disease, Spine J. (2025), https://doi.org/10.1016/j.spinee.2025.07.001.
[52] D.G. Rowe, J. Woo, S. Yoo, J.M. Emerson, E. O'Callaghan, M. Goodin, et al., Associations between problem list items and clinically significant distress in patients with metastatic spine disease, Spine (Phila PA 1976) 50 (2025) 816–822, https://doi.org/10.1097/BRS.0000000000005343.
[53] J.J.G.J. Amelink, B.T. van Munster, B.J.J. Bindels, R.J.B. Pierik, J. van Tiel, O. Q. Groot, et al., Surgical management of spinal metastases: a cross-continental study in the United States and the Netherlands, J Bone Oncol 52 (2025), https://doi.org/10.1016/j.jbo.2025.100676.
[54] F. Cofano, G. Di Perna, A. Alberti, B.M. Baldassarre, M. Ajello, N. Marengo, et al., Neurological outcomes after surgery for spinal metastases in symptomatic patients: does the type of decompression play a role? a comparison between different strategies in a 10-year experience, J Bone Oncol 26 (2021), https://doi.org/10.1016/j.jbo.2020.100340.