A systematic review and meta-analysis of the predictive accuracy of preoperative scoring systems for postoperative survival in patients with metastatic bone disease

Ahmed A.A. Mansour, Mohamed A. El Mowafi, Sherif I. Azmy

Department of Orthopedic, Ain Shams University, Cairo, Egypt

Correspondence to Sherif I. Azmy, MSc, Department of Orthopedic Surgery Ain Shams University, Villa 67, Banafseg 11, First Settlement, Cairo, Egypt
Tel: +20 120 275 0006;
e-mail: drsherifishak@yahoo.com

Received: 09-Jan-2024
Revised: 25-Feb-2024
Accepted: 26-Feb-2024
Published: 24-Jul-2024

The Egyptian Orthopaedic Journal 2024, 59:103–117

One significant clinical challenge is meeting the needs of patients with metastatic bone disease (MBD) who are susceptible to pathological fractures. Patients with cancer who undergo surgical procedures are more vulnerable to thromboembolic and infectious complications. The expected survival rate plays a major role in deciding the best action. The current systematic review and meta-analysis were conducted to evaluate the validity and reliability of various preoperative survival scores in surgery for MBD. The aim of the review was to examine the preoperative survival evaluations used in surgery for MBD. A comprehensive exploration of medical literature was conducted using MEDLINE (accessed through Ovid), EMBASE (accessed through Ovid), and PubMed. A systematic analysis of studies examining prognostic scores that gauged survival rates in individuals with bone metastases was performed. Additionally, a meta-analysis encompassed studies assessing the sensitivity and specificity of the Tokuhashi and Tomita scores in predicting 6-month survival rates for spinal metastases. Incorporating a total of 68 studies, with 35 included in the meta-analysis, the Tokuhashi score demonstrated sensitivities ranging from 27 to 92%, and the Tomita score exhibited sensitivities from 76 to 99%. Specificities for the Tokuhashi score ranged from 44 to 96%, while the Tomita score specificities varied from 1 to 44%. The pooled diagnostic odds ratio was 6.04 (95% confidence interval, 3.96–9.21; Tau-squared=0.90; I^2=86%) for the Tokuhashi score and 1.34 (95% confidence interval, 0.67–2.67; Tau-squared=1.02; I^2=85%) for the Tomita score. The SORG Nomogram, developed in a substantial surgical cohort, exhibited robust discrimination for 3-month and 1-year survival, reliable calibration, and outperformed counterparts with low risk of bias and applicability concerns. PATHFx 3.0, 2013-SPRING, and potentially Optimodel emerged as superior models for predicting survival in extremity metastasis surgery. The Tokuhashi score showed high sensitivity and specificity, with an overall higher diagnostic value compared to the Tomita score. The SORG Nomogram demonstrated robust performance in predicting 3-month and 1-year survival, surpassing other models in terms of reliability and applicability. Additionally, PATHFx 3.0, 2013-SPRING, and potentially Optimodel emerged as promising models for predicting survival in extremity metastasis surgery.

Keywords:
meta-analysis, metastatic bone disease, scoring system, survival, systematic review

Egypt Orthop J 2024, 59:103–117

1110-1148

Introduction

Bone metastases are frequently caused by myeloma, thyroid, breast, or prostate cancer. With the exception of prostate cancer, most metastases are either mixed or lytic, which puts patients at risk for pathological fractures. According to the Mirels scale, radiographs reveal bone metastases and make it easier to assess the risk of pathological fracture [1]. Increased cancer prevalence and improved diagnosis techniques have been linked to an 18% increase in cancer incidence over the previous 10 years, according to the Scandinavian Skeletal Metastasis registry. Bone metastasis causes severe morbidity in those who are affected, greatly lowering their quality of life [2].

Bone is the third most commonly affected site in metastatic cancer, after the liver and lung. The most common cancers to metastasize to the bone are those of the breast and prostate (65–75%). The most frequently affected areas of the body by metastasis are the spine and pelvis, although long bones such as the femur and humerus are also frequently affected [3].

The bulk of pathological fractures in metastatic bone disease (MBD) necessitating surgical intervention arise in the femur. Surgical management for femoral metastatic disease is palliative rather than curative. Surgical goals encompass attaining structural stability, reinstating function, alleviating pain, enhancing life quality, and mitigating the risk of revision surgery [4]. The surgical options include prosthetic reconstruction, intramedullary nail fixation, and plate fixation. Intramedullary nail fixation-related complications typically manifest over 1-year posttreatment, unlike prosthetic reconstruction-related complications, which emerge earlier [5].

It is generally accepted that any surgical intervention is contraindicated if the expected survival is less than 4–6 weeks. Most authors agree that a 3–12 months survival period is a reasonable cutoff point for considering less invasive surgical reconstruction techniques that do not require extended rehabilitation. 'Long' survivors survive for 12 months or longer; in these cases, more invasive resection and reconstruction are necessary despite the lengthier recovery period, given the possibility of local recurrence [6].

Numerous trials have been conducted recently to establish an accurate scoring system for predicting survival after surgeries for MBD, including PATHFx, modified Tomita, Bauer, Van der Linden, Rades, SORG nomogram, Optimodel, SPRING13, Brier score, Katagiri, and Tokuhashi [3,7]. Our present study aims to compare and assess different preoperative survival scores in MBD surgery, evaluating their validity and reliability in clinical practice.

Materials and methods

This study adheres to the recommendations outlined in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines.

Data source and searches

Up to the end of February 2023, MEDLINE, EMBASE, and PubMed were searched to find clinical studies on prognostic scores, which are used to determine the survival rates of patients treated for bone metastases. The search utilized keywords: bone metastasis, metastasis, metastatic, survival, extremity, spine, spinal, vertebral, prognosis, prognostic score, prognostic scoring system, Tomita, Tokuhashi. Additionally, relevant publications were identified through the reference lists of the chosen papers. To ensure the study quality, papers not in the English language, case studies, and reviews were excluded.

Inclusion criteria

Studies with the following criteria were included: external validation of prognostic scores for survival, utilization of independent surgical or mixed cohorts, inclusion of patients with extremity or spinal metastases regardless of primary tumor type, and availability of full-text articles in the English language.

Studies evaluating the Tokuhashi and Tomita ratings sensitivity and specificity in predicting the 6-month survival rates for spinal metastases were included in the meta-analysis. An overall survival of at least 6 months has been guaranteed by prior investigations, which established a threshold for surgical intervention at a Tokuhashi rating of more than or equal to 9 or a Tomita rating of less than or equal to 7. The standard test was used based on whether the patient was still alive or had passed away six months after the metastasis diagnosis.

Exclusion criteria

Excluded studies represented those assessed scoring systems not intended for forecasting survival, examining them without furnishing performance metrics for a straightforward quantitative contrast. Primary bone tumors, non-MBD, and alternative primary intervention techniques such as ablation, radiotherapy, chemotherapy, and bisphosphonates were all excluded. Moreover, duplicate publications were excluded.

Study selection

Two researchers performed data retrieval and quality evaluation independently. Initial screening of all articles occurred based on their titles and abstracts (n=3163), and any duplicate entries were eliminated (n=2647). In cases where the title indicated relevance to the review, the abstract underwent assessment to determine eligibility for inclusion, excluding non-English studies, case reports, reviews, letters, technical notes, and studies not evaluating scores (n=155). Articles deemed potentially eligible were then scrutinized in their entirety excluding papers where data was insufficient to make a 2 by 2 table, or there was no description of survival or evaluation of other scores or uncertain or cut-off value or duplicating enrolled cases. Only articles meeting all inclusion criteria were incorporated for subsequent analysis (n=68).

Data extraction

For the meta-analysis, the following data were extracted: actual overall survival and the numbers of TP, FP, FN, and TN. For systematic review, the following data were extracted and summarized in the summary table: first author, year of publication, prognostic scores, study design, study duration, number of patients, intervention, and survival.

Data synthesis and analysis

Predictive values were assessed for each study, including sensitivity, specificity, diagnostic odds ratio (DOR), and a 95% confidence interval (CI), to perform a meta-analysis for the diagnostic accuracy of the available studies. To avoid estimated values of 0 or infinity and to make SE computation easier, 0.5 was added to each cell in groups that had no events (a 'zero cell' in the 2 by 2 table). Summary receiver operating characteristic plots and pooled estimates were produced using the meta-disc software. The area under the curve (AUC) of ROC plots was used to compare diagnostic accuracy of the two scores. To create survival curves, Kaplan–Meier life table analysis was employed. By evaluating the model's overall performance, discriminative power, and calibration accuracy for the patients under study, the validation of the model was assessed in the systematic review.

Evaluation of bias and review strategy

The second edition of the QUADAS-2 tool was used to evaluate the quality of the studies that were part of the meta-analysis. The reference standard bias was carefully examined, and studies with a low risk of bias had a minimum 6-month gap between the diagnosis of metastases and the end of the study, either by death or survival. Because of their low accuracy, studies without a description of how to start predicting survival were deemed to have an unclear risk of bias. Research that had follow-up times shorter than 6 months was deemed to be highly biased. To assess the risk of bias and applicability concerns, the Prediction model study Risk of Bias Assessment Tool (PROBAST) was utilized. The primary purpose of this tool is to evaluate primary studies in systematic reviews.

Results

Study selection results

After a database search, 3163 citations were retrieved, which were then screened for duplicates, resulting in 2647 results. After screening the titles and abstracts, 115 studies were screened in full text. Ultimately, 68 studies met the inclusion criteria and were included in the evidence synthesis. Of them, 35 studies were included in the meta-analysis, and 33 in the systematic review as depicted in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram (Fig. 1).

Meta-analysis to assess Tokuhashi and Tomita score diagnostic accuracy

The meta-analysis included 35 articles in total, with a total of 10 125 patients; of them, 4460 patients evaluated by Tokuhashi scores and 1381 patients were evaluated by Tomita scores. Sensitivity ranged from 27 to 92% (Fig. 2) and specificity from 44 to 96% (Fig. 3). The pooled DOR was 6.04 [95% confidence interval (CI), 3.96–9.21; Tau-squared=0.90; I^2=86%] (Fig. 4), and the AUC was 0.795 (Fig. 5). The Tokuhashi score is reflected in these outcomes. Figure 6 shows the sensitivity range of 76–99% and Fig. 7 shows the specificity range of 1–44% for the Tomita score. Figure 8 shows the pooled DOR of 1.34 (95% CI, 0.67–2.67; Tau-squared=1.02; I^2=85%) and Fig. 9 shows the AUC of 0.879.

Pooled estimates for the Tokuhashi score were 58% (95% CI, 0.56–0.60) and 79% (95% CI, 0.77–0.81), respectively, for sensitivity and specificity; for the Tomita score, they were 85% (95% CI, 0.83–0.86) and 20% (95% CI, 0.17–0.22), respectively. This demonstrated high sensitivity but low specificity, with a Tomita score of less than or equal to 7, leading to an overestimation of 6-month predicted survival. Conversely, an underestimation of survival was observed with a Tokuhashi score of more than or equal to 9, which was associated with high specificity and low sensitivity. Summary table of included studies is reported in Table 1.

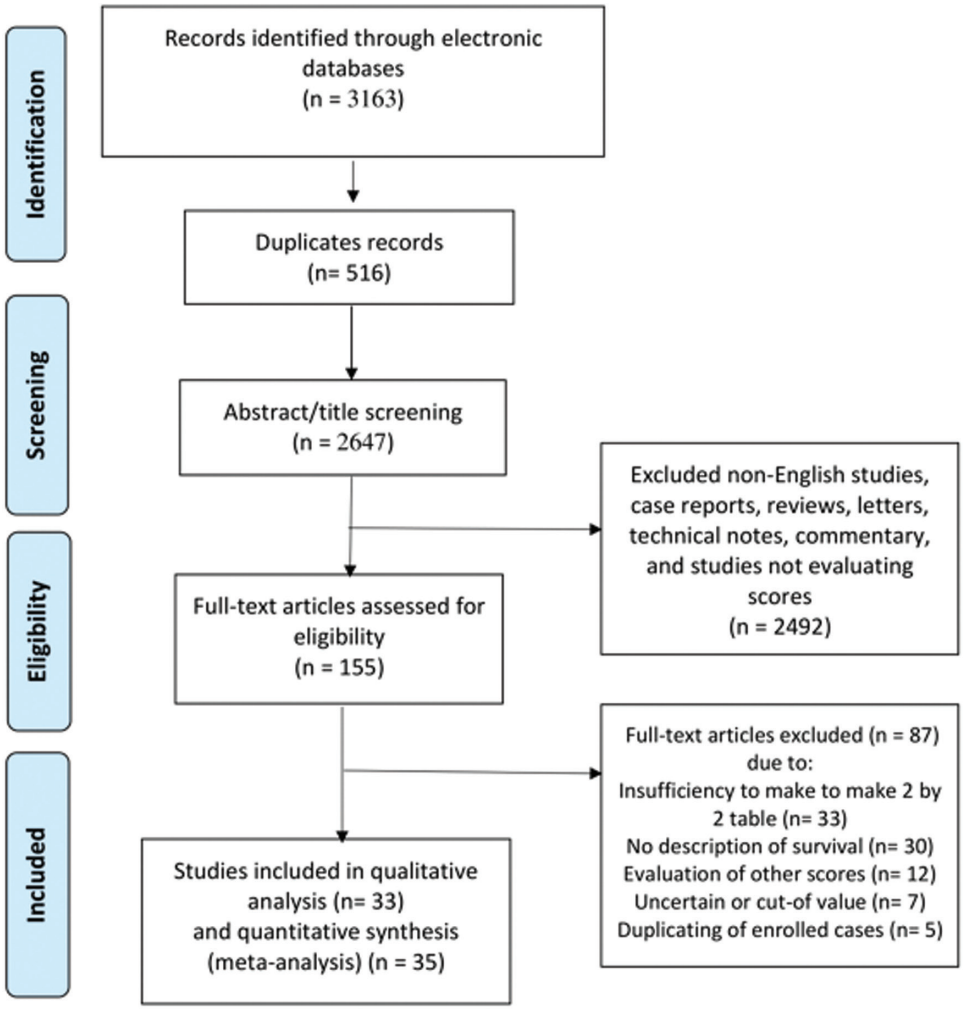
Systematic reviews of preoperative scoring systems in prediction of postoperative survival in patients with metastatic spine disease

The patient population across the 12 studies varied greatly, ranging from 61 to 1469 patients. Of the 10 databases, only two were fully prospective [8,9]. The median survival time ranged from 5.1 to 13.6 months [10,11]. Nine investigations exclusively enrolled patients who underwent surgical treatment, while the remaining three evaluated a mixed cohort, predominantly comprising nonsurgical patients. Four studies specifically excluded individuals with hematological malignancies. The prognostic scoring systems evaluated across the studies encompassed the modified (revised) Tokuhashi [12-14], (modified) Bauer (mBauer) [15,16], Tomita [17], Katagiri [18], Sioutos [19], van der Linden [20], Bartels [21], Rades [22], Oswestry Spinal Risk Index [23], Bollen [24], New England Spinal Metastasis Score (NESMS) [25], Skeletal Oncology Research Group (SORG) Classic [26], SORG Nomogram [26], and SORG machine learning algorithm (stochastic gradient boosting, SGB) [27]. The summary of those studies is reported in Table 2.

Risk of bias assessment

Of the 12 studies examined, half were found to have a high or uncertain risk/concern in the areas of bias and applicability. Both the studies by Choi and colleagues and Nater and colleagues showed a low overall risk of bias and ambiguous applicability concerns. Pereira

Figure 1



Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram of study selection.

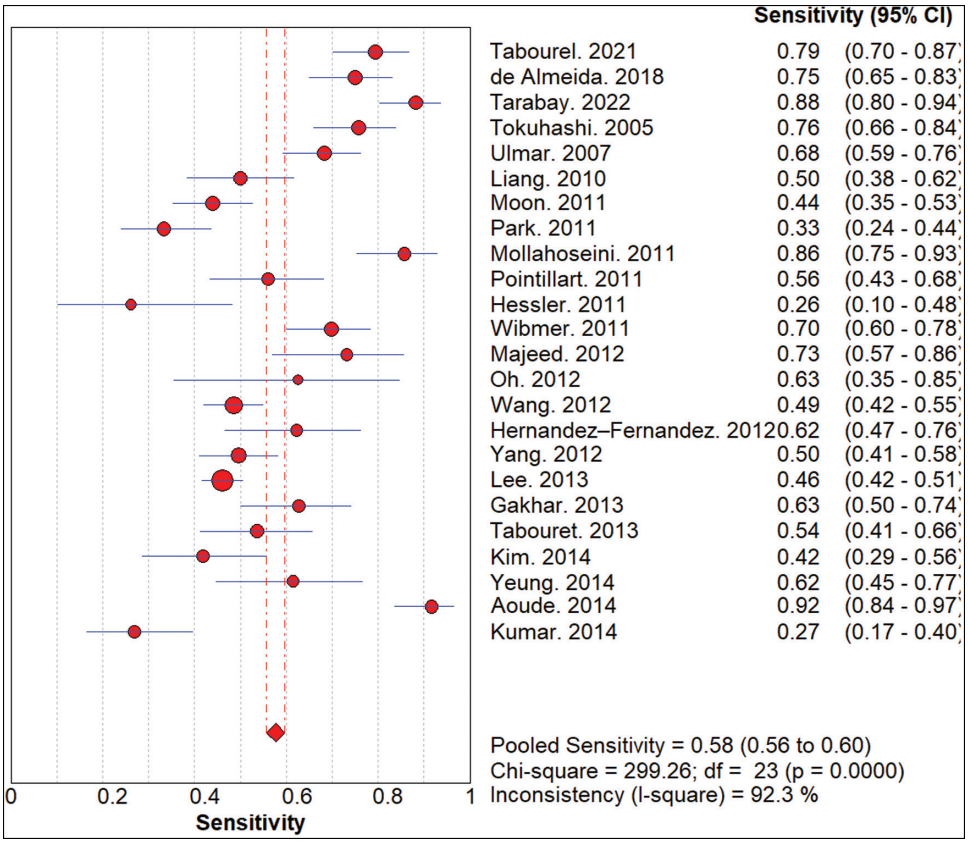
and colleagues, Ahmed and colleagues, Karhade and colleagues, and Westermann and colleagues were the four studies that showed the least concern in the bias and applicability domains.

There was little chance of bias in the research done by Karhade and colleagues and Nater and colleagues. Karhade and colleagues only looked at certain aspects of calibration, discrimination, and Brier score, whereas Nater and colleagues neglected to evaluate specific survival time points or include more recent individual risk prediction scores.

Because Pereira and colleagues created the SORG Nomogram, their study is prone to optimism bias. Similarly, the percentage of correctly stratified patients served as a calibration measure in Pereira and colleagues comparative study, which assessed AUCs at particular time points (discrimination). Because the SORG Nomogram development study was published by the same authors, it is also vulnerable to optimism bias. On the other hand, Choi and colleagues did not include an assessment of the NESMS or SORG scores, instead focusing on a sizable multicenter prospective surgical validation study. They did not take into account calibration measures or analyze individual time points; they only compared c-statistics.

Pereira and colleagues also measured the proportion of correctly stratified patients as a calibration measure. Ahmed and colleagues evaluated AUCs at particular time points (discrimination), incorporating more recent individual risk prediction scores. While presenting both calibration and discrimination measures, Westermann and colleagues did not include the more recent NESMS or SORG scores. Additionally, because they were the only ones to compare a modified revised Tokuhashi score, it was difficult to compare their findings with those of other studies. The risk of bias and applicability (PROBAST) is detailed in Table 3.

Figure 2



Sensitivity of Tokuhashi score for postoperative survival in patients with metastatic spine disease.

Discrimination and calibration performance assessment

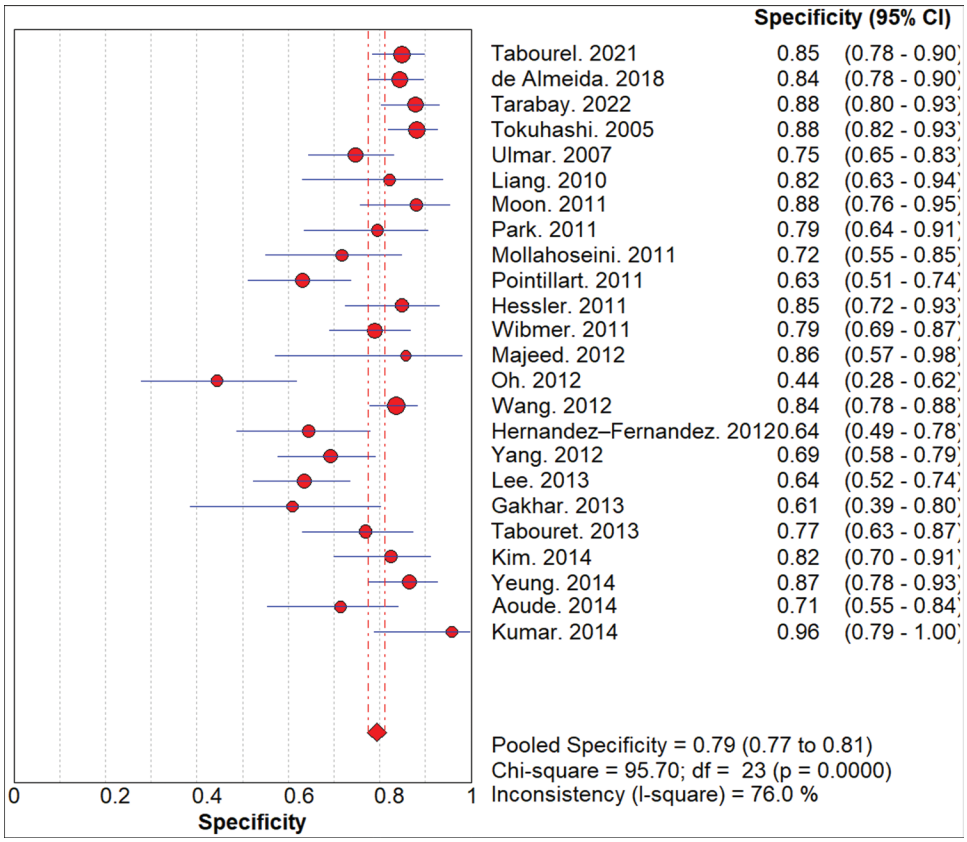
Discrimination is the ability to differentiate between two patients based on their outcomes. The SORG machine learning algorithm, NESMS, SORG Nomogram, and Katagiri score had a good discriminatory performance at 3 months, while rTokuhashi, Tokuhashi, and mBauer had a moderate performance. The SORG Nomogram, SORG machine learning algorithm, Katagiri, Tokuhashi, NESMS, Tomita, and mrTokuhashi showed good discriminatory performance at 1 year, while rTokuhashi, SORG Classic, and mBauer showed moderate performance.

Calibration is the degree to which the predicted survival probabilities match the actual outcomes. However, several studies showed poor calibration for multivariate prognostic scoring systems. The most recent scoring systems, like NESMS, SORG Nomogram, and SORG machine learning algorithms, were not included in the Bartels score, despite having the best calibration. Only the SORG Nomogram and SORG machine learning algorithm were evaluated for calibration out of the top-performing scores for discrimination. In a comparative study, the SORG Nomogram proved to be accurate in estimating survival, and in its external comparison study, the SORG machine learning algorithm demonstrated a fair calibration. While the rTokuhashi displayed moderate calibration, the Tokuhashi score had poor calibration overall. Mr Tomohashi was evaluated in only one study, but his overall calibration was good. The included studies did not evaluate the NESMS and Katagiri for calibration.

Systematic reviews of preoperative scoring systems in prediction of postoperative survival in patients with metastatic extremity bone disease

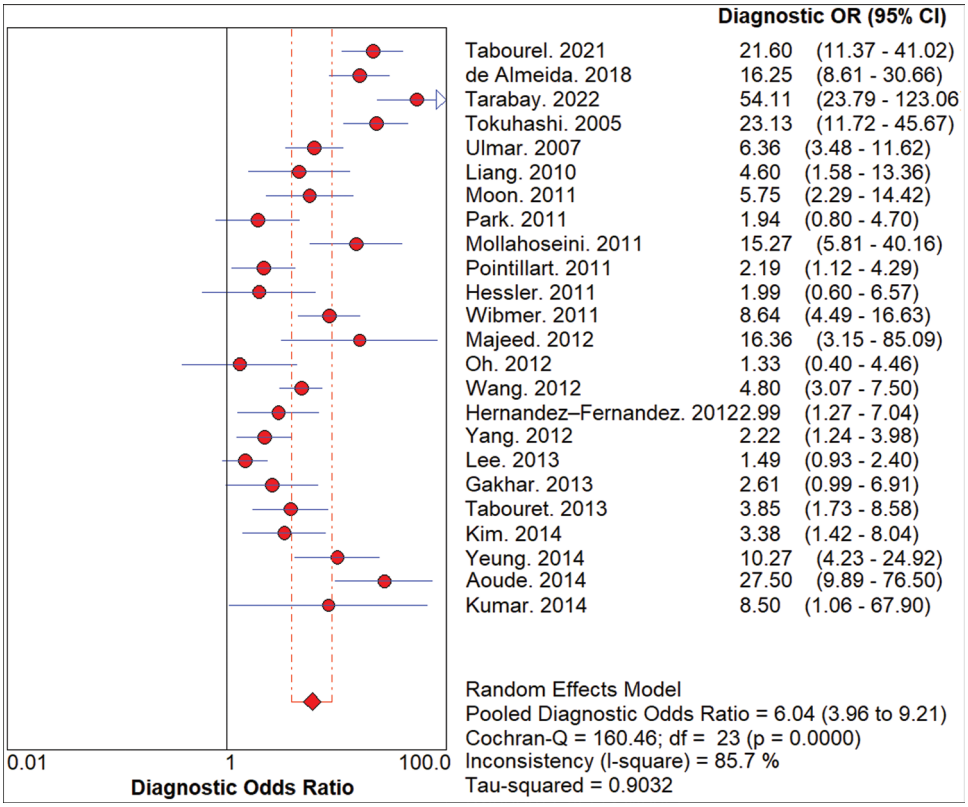
The 11 papers in this review are examined using seven different models intended for survival estimation. Four studies, which represent improvements over previous iterations, expand on two models each, while five focus on a single model. The two studies that remain take a comparative stance. The models that were analyzed included the following: the PathFx 1.0 model by Forsberg and colleagues, the SORG model by Thio and colleagues, the Janssen score by Janssen and colleagues, the Optimodel by Willeumier and colleagues, the Janssen score by Janssen and colleagues, the Sorensen and colleagues SPRING 2008 and SPRING 2013 nomograms, and the Errani and colleagues IOR score [3,5,6,28–35]. Table 4 provides a summary of the scoring systems used for postoperative survival in patients with MBD of the extremities.

Figure 3



Specificity of Tokuhashi score for postoperative survival in patients with metastatic spine disease.

Figure 4



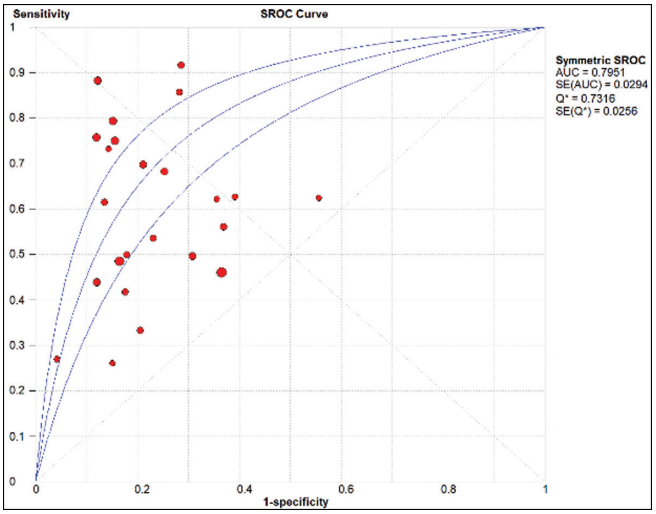
Diagnostic odds ratio of Tokuhashi score for postoperative survival in patients with metastatic spine disease.

Risk of bias assessment

Two studies with a high risk of bias and four with an uncertain risk of bias were identified among the included studies. Five studies had uncertain risk on applicability domains, while three had low risk and applicability domains. Table 5 displays the specifics of the applicability and bias risk.

In a number of studies, prognostic factors were ascertained through univariate [29,30] and multivariate Cox analyses [5,6,31,35], with machine learning models employing particular algorithms. The prognostic scores were produced by the studies using a variety of techniques, including machine learning algorithms (used in three studies [29,30,35]), nomograms (used in two studies), and traditional scoring systems like survival estimation methods (used in four studies [5,6,31,35]). The calibration scores and discrimination accuracy were used to evaluate the prognostic scores' accuracy while Ratasvuori and colleagues, Willeumier and colleagues, Sorensen and colleagues, and Anderson and colleagues used an external population, Thio and colleagues and Janssen and colleagues used a subset of the population that generated the data for external validation. Following comparative studies, six models – Optimodel, SPRING 2013, PATHfx 1.0, 7SSG, Janssen, and IOR – with varying demographics were externally validated; the SORG model was not included in these studies. Because only Errani and colleagues carried out a prospective study, the study design's inherent bias was minimized. The 12-month survival prediction served as the common endpoint for all models [3,28].

Figure 5



Summary receiver operating characteristics of Tokuhashi score for postoperative survival in patients with metastatic spine disease.

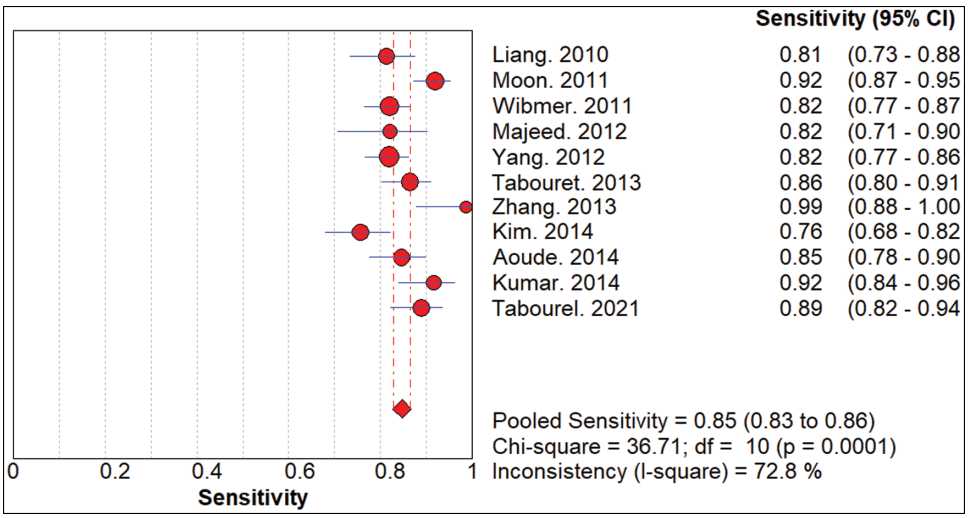
Discrimination accuracy and calibration score

In every study that used AUC as a performance measure, the externally validated cohort showed a decline in AUC. The AUCs and Brier scores for each validation set at 3-, 6-, and 12-month survivals. All models had mean AUCs between 0.57 and 0.87 and mean Brier scores between 0.13 and 0.25.

Discussion

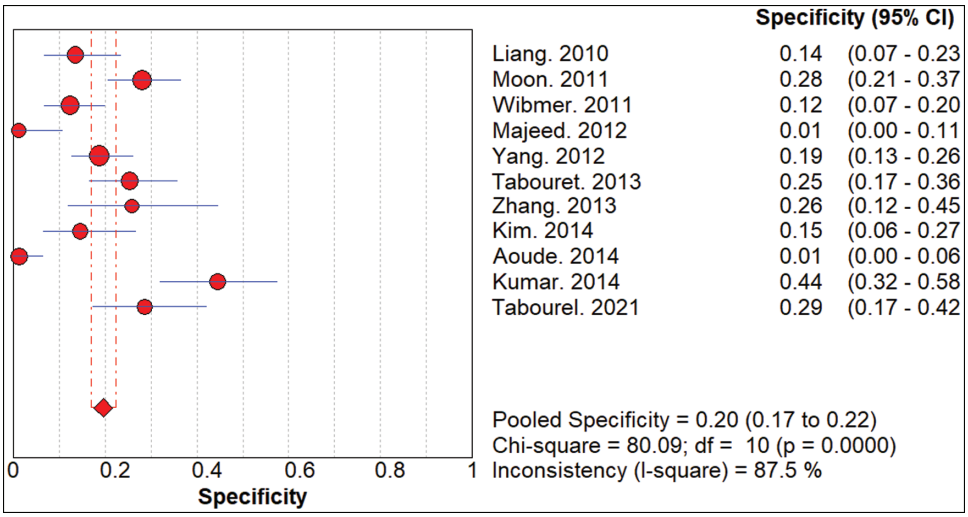
A meta-analysis comprising 35 articles was carried out, encompassing 4460 patients evaluated using Tokuhashi scores and 1381 patients evaluated using Tomita scores. A diagnostic odds ratio of 1.43 for the Tomita scores and 6.04 for the Tokuhashi scores was found in the results. Additionally, our analysis revealed that the Tokuhashi and Tomita scores had AUC estimates of 0.795 and 0.879, respectively. A sensitivity of 58% and

Figure 6



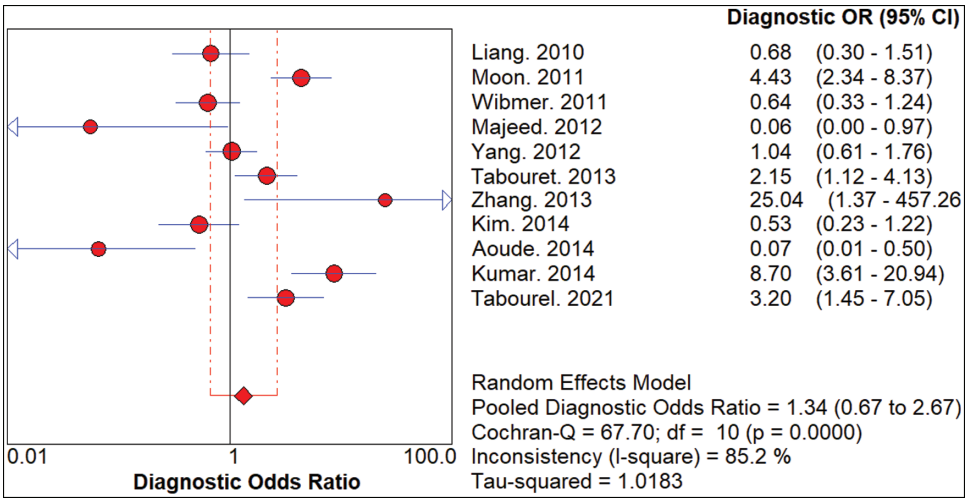
Sensitivity of Tomita score for postoperative survival in patients with metastatic spine disease.

Figure 7



Specificity of Tomita score for postoperative survival in patients with metastatic spine disease.

Figure 8



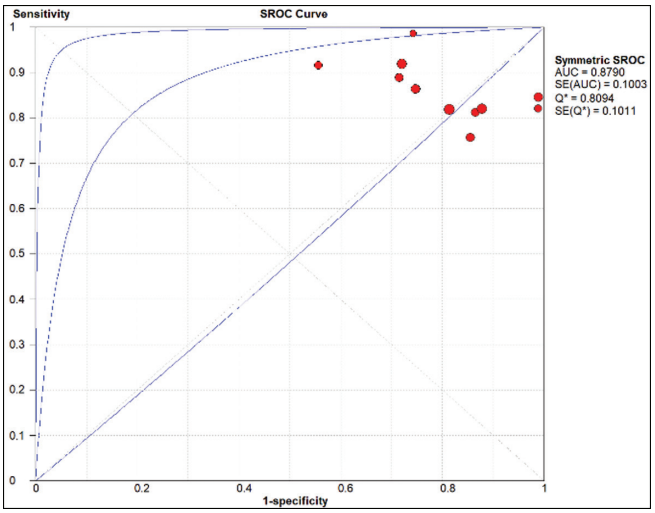
Diagnostic odds ratio of Tomita score for postoperative survival in patients with metastatic spine disease.

a specificity of 79% were found using the Tokuhashi score. Despite this, the Tomita score showed 20% specificity and 85% sensitivity.

The primary cancer site affects survival prediction in bone metastasis disease by dictating the aggressiveness of spinal metastasis and the primary disease and its response to treatment [36]. The primary cancer score in the Tomita score accurately predicted survival, according to the cumulative 6-month survival rate [37].

At 6 months, the Tokuhashi initial malignancy rating of four showed the largest area beneath the curve and the lowest cumulative survival percentage. This first malignancy score was consistent since it only included rectal cancer, unlike other scores. Furthermore, there was a tendency for the Tokuhashi rating to underestimate survival. Because of this, a Tokuhashi initial malignancy rating of 4 may have indicated the best accuracy in diagnosis, even though it had a higher initial malignancy rating than a lower survival rate [37]. On the other hand, a Tokuhashi initial malignancy rating of 2 showed the most encouraging cumulative survival and the least amount of AUC. The first malignancy score of 2, which was labeled 'others,' included lesions of the colon, ovary, urethra, melanoma, germinoma, liposarcoma, and leiomyosarcoma. This may help to explain why the least accurate diagnosis was found in cases where the Tokuhashi initial malignancy rating was 2 [14]. Tokuhashi and colleagues used information from fewer than three patients to classify cancers of the bladder, ovaries, colon, pancreas, and esophagus. As a result, particularly in the second score group, the initial

Figure 9



Specificity of Tomita score for postoperative survival in patients with metastatic spine disease.

malignancy group may reduce diagnostic accuracy in the Tokuhashi rating [14].

Furthermore, 12 studies were included that used various models to estimate survival in patients with metastatic spine disease. Our findings suggest that the SORG Nomogram is a good choice for predicting 3-month and 1-year survival in patients who are having surgery to preserve their quality of life but have spinal metastases. It is regarded as one of the few scoring systems that consistently exceeds the AUC threshold 0.70, demonstrating strong discriminatory power for both endpoints. These outcomes followed the Smeijers and Depreitere [38] study. Unlike scoring systems, the SORG Nomogram's performance is maintained in three external validation studies. In addition to superior results in three out of three comparative studies with low risk of bias and low applicability concerns, good calibration was noted for both endpoints in one of the studies [38]. Moreover, the Nomogram began life as a multivariate Cox model that employed relevant and readily available prognostic factors, such as the primary tumor category, the Eastern Cooperative Oncology Group, the presence of visceral metastasis, laboratory markers (hemoglobin, white blood cell count), the number of spinal metastases, and prior systemic treatment [7].

Previous systematic reviews have reported these factors as independent survivorship predictors [39–41]. The Nomogram has an easy-to-use format that can be printed for convenience. It was created utilizing a sizable retrospective surgical assembly of 649 patients treated at two tertiary facilities between 2002 and 2014 [39–41]. An additional positive rating comes from the SORG SGB machine learning algorithm, which demonstrated excellent qualities in its development study (732 surgical patients) and exceeded assessment standards [27].

As demonstrated by Karhade *et al.* [27], this score performs comparably, if not better, in discrimination and calibration than the SORG Nomogram during external authentication using the same patient as Ahmed *et al.* [42]. A net benefit is also indicated by the decision curve analysis and overall performance. Its potential has also been confirmed by two recent retrospective external validation studies in surgical cohorts, which showed good performance in calibration and discrimination for 3-month and 1-year survival. The SORG scoring systems are currently not fully prospectively validated or compared [43–45].

Due to their superior performance in development, validation, and direct comparison, individual risk prediction scores – like the SORG Nomogram and SORG SGB machine learning algorithm – are becoming more and more popular than traditional risk score tables like Tomita and Tokuhashi [45]. Therefore, adopting individual risk prediction modeling should be the main focus of future research and treatment choices. Even the top-performing models, though, can still be improved upon; 70–78% performance is expected for both calibration and discrimination [38]. It is essential to follow methodological guidelines like PROBAST and the TRIPOD statement. Prospective clinical incorporation should evaluate patient outcomes such as quality of life and give clinical utility for patient counseling and decision-making priority after validation [46–48].

Prognostic models that predict the preoperative survival of patients suffering from disease-related extremity metastases. Our findings indicated that the PATHfx 3.0 model and the SPRING 2013 nomogram had the most recent externally validated survival likelihood scores, exhibiting superior performance status in accuracy and discrimination. At the 3-month endpoint, PATHfx 3.0 showed similar performance to the SPRING 2013 model, despite the latter's marginally superior survival prediction at 6 and 12 months. Itis interesting that both models are updated iterations of earlier prognostic models. These outcomes followed a prior investigation conducted by Ben Gal *et al.* [4]. Nevertheless, using the same external validation database, Meares *et al.* [3] showed that the Optimodel performed better than SPRING 2013 in predicting survival at 12 months. At 3- and 6-month intervals, PATHfx 1.0 was demonstrated to be more accurate.

Table 1 Summary table of studies included in the meta-analysis assessing Tokuhashi and Tomita systems for postoperative survival in patients with metastatic spine disease

References	Years of publication	Country	Study period	Total	TP	FP	FN	TN
Tokuhashi scores								
Tabourel *et al.*[53]	2021	France	2014-2017	739	81	25	21	140
de Almeida *et al.*[54]	2018	Brazil	2008-2015	117	72	24	24	130
Tarabay *et al.*[55]	2022	Canada	2008–2020	94	90	14	12	101
Tokuhashi *et al* [12].	2005	Japan	-	246	72	18	23	133
Ulmar *et al.*[56]	2007	Germany	1984.09–2005	217	86	23	40	68
Liang *et al.*[57]	2010	China	1996–2009.	104	38	5	38	23
Moon *et al.*[58]	2011	South Korea	1987–2008	182	58	6	74	44
Park *et al.*[59]	2011	South Korea	2001–2008	135	32	8	64	31
Mollahoseini *et al.*[60]	2011	Iran	2007–2009	109	60	11	10	28
Pointillart *et al.*[61]	2011	France	2005–2007	142	37	28	29	48
Hessler *et al.*[62]	2011	Germany	1999–2004	76	6	8	17	45
Wibmer *et al.*[63]	2011	Austria	2003–2011	196	74	19	32	71
Majeed *et al.*[64]	2012	UK	2007–2010	55	30	2	11	12
Oh *et al.*[65]	2012	South Korea	1996–2008	52	10	20	6	16
Wang *et al.*[66]	2012	Denmark	1992–2009	448	117	34	124	173
Hernandez–Fernandez *et al.*[67]	2012	Spain	2004–2006	90	28	16	17	29
Yang *et al* [36].	2012	South Korea	2001–2009	217	69	24	70	54
Lee *et al.*[68]	2013	South Korea	2005–2010	577	227	31	265	54
Gakhar *et al.*[69]	2013	UK	2007–2010	90	42	9	25	14
Tabouret *et al.*[70]	2013	France	2004–2010	121	37	12	32	40
Kim *et al.*[71]	2014	South Korea	2006–2011	112	23	10	32	47
Yeung *et al.*[72]	2014	China	2000–2010	128	24	12	15	77
Aoude *et al.*[73]	2014	Canada	2003–2012	126	77	12	7	30
Kumar *et al.*[74]	2014	Singapore	2007–2011	87	17	1	46	23
Tomita Scores								
Tabourel *et al.*[58]	2021	France	2014–2017	120	40	15	16	126
Liang *et al.*[75]	2010	China	1996–2009	104	64	24	10	6
Aoude *et al.*[73]	2014	Canada	2003–2012	126	83	23	1	19
Moon *et al.*[58]	2011	South Korea	1987–2008	182	95	16	37	34
Zhang *et al.*[76]	2013	China	2003–2011	36	23	0.5	8	5
Wibmer *et al.*[63]	2011	Austria	2003–2011	196	93	43	13	47
Majeed *et al.*[64]	2012	UK	2007–2010	55	41	12	0.5	2
Yang *et al* [36].	2012	South Korea	2001–2009	217	113	48	26	30
Tabouret *et al.*[70]	2013	France	2004–2010	146	65	23	22	36
Kim *et al.*[71]	2014	South Korea	2006–2011	112	47	36	8	21
Kumar *et al.*[74]	2014	Singapore	2007–2011	87	35	8	28	16

FN, false negative; FP, false positive; TN, true negative; TP, true positive.

There is a lack of agreement on the variables that should be included in survival prediction models. Although primary tumor histology is a major prognostic factor that influences survival, different models differ greatly in the content and subdivision of this variable. Bollen and colleagues' division is used in some studies [5,31], but Katagiri and colleagues's division is used in others [32–34]. Many models [6,29,30,35] include multiple myeloma as the primary tumor; however, some, like Willeumier *et al.* [31], do not include it, arguing that the impact of primary hematological cancer on survival differs from that of osseous metastases from solid carcinomas. Some recent prognostic models [29,30,33,34] have found a significant impact on survival associated with impending and pathologic fractures, as identified by Bauer and Wedin [15]. However, other models [5,6,31,35] have not found the same impact. Important predictors of preoperative survival have been identified, including hemoglobin level [32–34], absolute lymphocyte [33,34], platelets, ALP, albumin [32], or CRP [5]. Nonetheless, these labs may be interchangeable due to collinearity with other clinical variables [4]. Using only three variables, Willeumier *et al.* [31] produced a clinically useful model that accepted reduced discriminatory ability compared to the many variable-based models [30,33–35]. The kitchen sink approach is used by machine learning-based models, which choose a large number of variables to improve accuracy while providing a clinician-friendly tool. The disparity in how prognostic factors are thought of compromises the quality and reliability of comparative studies in various populations and demands consideration in future research [31].

Table 2 Summary table of studies assessing scoring systems for postoperative survival in patients with metastatic spine disease

References	Year of publication	Prognostic scores	Study design	Study duration	Number of patients	Intervention	Survival
Leithner *et al* [16].	2008	Tokuhashi, Bauer, Tomita, Sioutos, Van der Linden	Prospective and retrospective	1998–2006	69	Surgery	21
Wibmer *et al.*[63]	2011	Tokuhashi, Bauer, Tomita, Sioutos, Van der Linden	Retrospective	1998–2006	254	Systemic/RT 76% Surgery 24%	10.6
Dardic *et al.*[77]	2015	Tokuhashi, Bauer, Tomita, Van der Linden	Retrospective	2005–2010	196	Conservative 65% Surgery 35%	7, 14
Bolen *et al.*[10]	2016	Tokuhashi, Bauer, Tomita, Van der Linden, Rades	Retrospective	2000–2010	1379	RT 83% Palliation 7% Surgery 10%	5.1
Hibberd *et al.*[78]	2017	Tokuhashi, Bauer, Tomita, Sioutos, Van der Linden	Retrospective	2010–2013	61	Surgery	NR
Pereira *et al.*[79]	2017	Tokuhashi, Bauer, Tomita	Retrospective	2014	100	Surgery	9
Pollner *et al.*[80]	2018	Tokuhashi, Bauer, Tomita, Van der Linden	Retrospective	2007–2015	329	Surgery	7.4
Choi *et al* [8].	2018	Tokuhashi, Bauer, Tomita, Van der Linden, Rades	Prospective	2003–2016	1469	Surgery	NR
Ahmed *et al* [42].	2018	Tokuhashi, Bauer, Tomita, Katagiri, Van der Linden	Retrospective	2008–2012	176	Surgery	9.4
Nater *et al* [9].	2018	Tokuhashi, Bauer, Tomita, Van der Linden, Bartels	Prospective	2000–2016	142	Surgery	7.5
Karhade *et al* [27].	2019	Tokuhashi, Bauer, Tomita, Katagiri, Van der Linden	Retrospective	2008–2015	732	Surgery	3
Westermann *et al* [11].	2020	Tokuhashi, Bauer, Tomita, Van der Linden	Retrospective	-	233	Surgery	13.6

NR, not reported; RT, radiotherapy.

Table 3 Risk of bias and applicability (Prediction model study Risk of Bias Assessment Tool) of studies assessing scoring systems for postoperative survival in patients with metastatic spine disease

References	Risk of bias				Concern regarding applicability			Overall	
	Participants	Predictors	Outcomes	Analysis	Participants	Predictors	Outcome	ROB	CRA
Leithner *et al* [16].	–	*	*	**	–	*	–	**	–
Wibmer *et al.*[63]	–	*	*	**	**	*	–	**	**
Dardic *et al.*[77]	**	*	*	**	**	*	–	**	–
Bolen *et al.*[10]	–	*	*	–	**	*	–	–	**
Hibberd *et al.*[78]	–	*	*	**	*	*	–	**	–
Pereira *et al.*[79]	*	*	*	*	*	*	*	*	*
Pollner *et al.*[80]	–	*	*	**	*	**	–	**	**
Choi *et al.*[8]	*	*	*	*	*	*	–	*	–
Ahmed *et al* [42].	*	*	*	*	*	*	*	*	*
Nater *et al* [9].	*	*	*	*	*	*	–	*	–
Karhade *et al* [27].	*	*	*	*	*	*	*	*	*
Westermann *et al* [11].	*	*	*	*	*	*	*	*	*

CRA, concern regarding applicability; ROB, risk of bias.
–, unclear risk.
*, low risk.
**, high risk.

The model's clinical representation ought to be dynamic, encompassing not only the rate of tumor growth but also the efficacy of treatment progression and the emergence of different primary tumor subtypes. For this goal, every model in our evaluation used clinical profiles. To ensure accuracy and applicability, Anderson and colleagues and Sorensen and colleagues improved their original models by developing new iterations based on larger and more recent databases [29,30,33,34]. The model applicability was dependent on how similar the treated populations were, as the patients in these studies received different treatments and had different characteristics [4]. Approach variations could result in less-than-ideal performance. As documented in the literature [49], external validation cohorts performed worse than internal validation cohorts for all survival endpoints in models where performance criteria were defined as AUCs. Whereas Pathfx 1.0 was externally validated across several patient cohorts, the Janssen score, OptiModel, and SPRING 2013 nomogram

Table 4 Summary table assessing scoring systems for postoperative survival in patients with metastatic extremity bone disease

References	Year of publication	Prognostic score	Study design	Study duration	Number of patients	Intervention	Survival in months
Foresberg *et al.*[81]	2011	PATHfx. 1.0	Retrospective	1999–2009	815	Surgery	3, 12
Anderson *et al.*[34]	2019	PATHfx. 3.0	Retrospective	1999–2003	397	Surgery	3, 6, 12,
				2015–2018			18, 24
Sorensen *et al.*[29]	2016	SPRING 2008	Retrospective	2003–2008	121	EPR	3, 6, 12
Sorensen *et al.*[30]	2018	SPRING 2013	Retrospective	2003–2013	270	EPR	3, 6, 12
Ratasvuori *et al.*[82]	2013	7SSG	Retrospective	1999–2009	833	Surgery	6, 12
Willeumier *et al* [31].	2018	OPTIModel	Retrospective	2000–2013	1520	RT or Surgery	3, 6, 12
Thio *et al.*[32]	2019	SORG	Retrospective	1999–2017	1090	Surgery	1, 12
Janssen *et al* [35].	2015	Nomogram	Retrospective	2009–2013	927	Surgery	1, 3, 12
Errani *et al* [5].	2021	IOR	Prospective	2015–2018	159	Surgery	12
Meares *et al* [3].	2019	Revised Katagiri model, PathFx model, SSG score, Janssen nomogram, OPTModel, SPRING 13	Retrospective	2003-2014	114	Surgery	36
Alfaro *et al* [28].	2021	Revised Katagiri, PathFx, Optimodel and IOR score	Retrospective	2016-2019	136	Surgery	12

EPR, endoprosthetic replacement; IOR, instituto Orthopedico Rizzoli; RT, radiotherapy; SSG, Scandinavian Sarcoma Group.

Table 5 Risk of bias and applicability (Prediction model study Risk of Bias Assessment Tool) of studies assessing scoring systems for postoperative survival in patients with metastatic extremity bone disease

References	Risk of bias				Concern regarding applicability			Overall	
	Participants	Predictors	Outcomes	Analysis	Participants	Predictors	Outcome	ROB	CRA
Foresberg *et al.*[81]	–	*	*	–	*	*	*	–	–
Anderson *et al.*[34]	*	*	*	–	*	*	–	–	–
Sorensen *et al.*[29]	–	*	*	–	**	*	*	–	**
Sorensen *et al.*[30]	*	*	*	*	**	*	–	–	**
Ratasvuori *et al.*[82]	–	*	*	*	*	*	–	–	–
Willeumier *et al* [31].	*	*	*	*	*	*	*	*	*
Thio *et al.*[32]	*	*	*	*	*	**	–	**	–
Janssen *et al* [35].	*	*	*	*	*	*	–	*	*
Errani *et al* [5].	*	*	*	*	*	*	*	*	*
Meares *et al* [3].	*	*	*	*	*	*	*	*	*
Alfaro *et al* [28].	*	*	*	*	*	*	*	*	*

CRA, concern regarding applicability; ROB: risk of bias.
–, unclear risk.
*, low risk.
**, high risk.

were externally validated twice [16,17,37–39]. When compared to other studies, the SORG model performed better in our investigation at the 3- and 12-month endpoints in its internal validation. However, since external validation was not carried out, this may not accurately reflect its performance and applicability in diverse populations.

While the projected 1-year survival rate is seen as a yardstick for prolonged survival, roughly corresponding to the median survival of surgical cohorts, the anticipated 3-month survival rate is still considered a threshold for surgical decision making. When choosing the best surgical strategy to achieve long-term results, this metric can be crucial in calculating the risk of local recurrence in long-term survivors, for example [8,50]. Prognostic systems are widely acknowledged in the literature as essential instruments for treating metastatic disease to the spine and extremities [51,52].

Our research highlights the challenges that recent models of survival prediction for symptomatic MBD patients have faced with regard to their applicability and performance in a variety of populations. This

highlights the need for multidisciplinary cooperation between radiologists, orthopedic surgeons, and oncologists to determine the best course of treatment instead of depending exclusively on risk prediction models. Additionally, it recommends that future prospective multicenter studies employ standardized treatment regimens to investigate and compare the most promising models [4].

Limitations

Our study was subject to certain limitations, such as inter-study heterogeneity, treatment effects on overall survival and cause of death not being taken into account, and limited availability of individual data from enrolled studies. Heterogeneity has been created by using different statistical methods to assess performance, incorporating different scoring systems, and using different time periods, all of which have presented certain difficulties. However, these limitations were assessed by focusing on similar metrics and presenting the results in a comprehensive manner. Most comparative studies have focused primarily on evaluating discrimination, with insufficient attention paid to calibration. In current research, it is challenging to lessen the influence of primary tumor categories, and there is currently not enough evidence to support the use of scoring systems in cancer-specific prognostications. All external validation studies were not included in individual scores; instead, validation studies were only included for comparison to obtain better results. However, it is unlikely that any other scoring system would perform better than the ones listed above.

Conclusion

The Tokuhashi score exhibited elevated sensitivity and specificity, along with a generally superior diagnostic value when compared with the Tomita score. The SORG Nomogram displayed strong predictive performance for both 3-month and 1-year survival, outperforming other models regarding reliability and practicality. Furthermore, promising outcomes were observed with PATHFx 3.0, 2013-SPRING, and potentially Optimodel as models for forecasting survival in surgeries related to extremity metastasis.

Orthopedic surgeons can make informed decisions using preoperative prognostic models for patients with MBD. Their discriminative ability and calibration precision have, however, only slightly improved. It is challenging to recommend a particular model as the gold standard for predicting survival in patients with BMD.

Financial support and sponsorship

Nil.

Conflicts of interest

The authors declare they have no conflict of interest in preparing this paper.

References

1 Guzik G. Oncological and functional results after surgical treatment ofbone metastases at the proximal femur. BMC Surg 2018;18:5.

2 Soeharno H, Povegliano L, Choong PF. Multimodal treatment of bone metastasis—a surgical perspective. Front Endocrinol 2018;9:518.

3 Meares C, Badran A, Dewar D. Prediction of survival after surgical management of femoral metastatic bone disease – a comparison of prognostic models. J Bone Oncol 2019;15:100225.

4 Ben Gal O, *et al*. The prediction of survival after surgical management of bone metastases of the extremities – a comparison of prognostic models. Curr Oncol 2022;29:4703-4716.

5 Errani C, Cosentino M, Ciani G, Ferra L, Alfaro PA, Bordini B, Donati DM. C-reactive protein and tumour diagnosis predict survival in patients treated surgically for long bone metastases. Int Orthop 2021;45:1337-1346.

6 Ratasvuori M, Wedin R, Hansen BH, Keller J, Trovik C, Zaikova O, *et al*. Prognostic role of en-bloc resection and late onset of bone metastasis in patients with bone-seeking carcinomas of the kidney, breast, lung, and prostate: SSG study on 672 operated skeletal metastases. J Surg Oncol 2014;110:360-365.

7 Paulino Pereira NR, Janssen SJ, Van Dijk E, Harris MB, Hornicek FJ, Ferrone ML, *et al*. Development of a prognostic survival algorithm for patients with metastatic spine disease. Spine J 2016;16:S318.

8 Choi D, Ricciardi F, Arts M, Buchowski JM, Bunger C, Chung CK, *et al*. Prediction accuracy of common prognostic scoring systems for metastatic spine disease: results of a prospective international multicentre study of 1469 patients. Spine (Phila Pa 1976) 2018;43:1678-1684.

9 Nater A, Tetreault L, Kopjar B, Arnold P, Dekutoski M, Finkelstein J, *et al*. Predictive factors of survival in a surgical series of metastatic epidural spinal cord compression and complete external validation of 8 multivariate models of survival in a prospective North American multicenter study. Cancer 2018;124:3536-3550.

10 Bollen L, Wibmer C, Van der Linden YM, Pondaag W, Fiocco M, Peul WC, *et al*. Predictive value of six prognostic scoring systems for spinal bone metastases: an analysis based on 1379 patients. Spine (Phila Pa 1976) 2016;41:E155-E162.

11 Westermann L, Olivier AC, Samel C, Eysel P, Herren C, Sircar K, *et al*. Analysis of seven prognostic scores in patients with surgically treated epidural metastatic spine disease. Acta Neurochir (Wien) 2020;162:109-119.

12 Tokuhashi Y, Matsuzaki H, Oda H, Oshima M, Ryu J. A revised scoring system for preoperative evaluation of metastatic spine tumor prognosis. Spine (Phila Pa 1976) 2005;30:2186-2191.

13 Tokuhashi Y, Matsuzaki H, Toriyama S, Kawano H, Ohsaka S. Scoring system for the preoperative evaluation of metastatic spine tumor prognosis. Spine (Phila Pa 1976) 1990;15:1110-1113.

14 Morgen SS, *et al*. A revision of the Tokuhashi revised score improves the prognostic ability in patients with metastatic spinal cord compression. J Cancer Res Clin Oncol 2018;144:33-38.

15 Bauer HC, Wedin R. Survival after surgery for spinal and extremity metastases. Prognostication in 241 patients. Acta Orthop Scand 1995;66:143-146.

16 Leithner A, Radl R, Gruber G, Hochegger M, Leithner K, Welkerling H, *et al*. Predictive value of seven preoperative prognostic scoring systems for spinal metastases. Eur Spine J 2008;17:1488-1495.

17 Tomita K, Kawahara N, Kobayashi T, Yoshida A, Murakami H, Akamaru T. Surgical strategy for spinal metastases. Spine (Phila Pa 1976) 2001;26:298-306.

18 Katagiri H, Takahashi M, Wakai K, Sugiura H, Kataoka T, Nakanishi K. Prognostic factors and a scoring system for patients with skeletal metastasis. J Bone Joint Surg Br 2005;87:698-703.

19 Sioutos PJ, Arbit E, Meshulam CF, Galicich JH. Spinal metastases from solid tumors. Analysis of factors affecting survival. Cancer 1995;76:1453-1459.

20 Van der Linden YM, Dijkstra SP, Vonk EJ, Marijnen CA, Leer, *et al*. Prediction of survival in patients with metastases in the spinal column: results based on a randomized trial of radiotherapy. Cancer 2005; 103:320-328.

21 Bartels RH, *et al*. Development of a model with which to predict the life expectancy of patients with spinal epidural metastasis. Cancer 2007;110:2042-2049.

22 Rades D, Dunst J, Schild SE. The first score predicting overall survival in patients with metastatic spinal cord compression. Cancer 2008;112:157-161.

23 Balain B, Jaiswal A, Trivedi JM, Eisenstein SM, Kuiper JH, Jaffray DC. The Oswestry risk index: an aid in the treatment of metastatic disease of the spine. Bone Joint J 2013;95:210-216.

24 Bollen L, van der Linden YM, Pondaag W, Fiocco M, Pattynama BP, Marijnen CA, *et al*. Prognostic factors associated with survival in patients with symptomatic spinal bone metastases: a retrospective cohort study of 1,043 patients. Neuro Oncol 2014;16:991-998.

25 Ghori AK, Leonard DA, Schoenfeld AJ, Saadat E, Scott N, Ferrone ML, *et al*. Modeling 1-year survival after surgery on the metastatic spine. Spine J 2015;15:2345-2350.

26 Paulino Pereira N, Janssen SJ, Van Dijk E, Harris MB, Hornicek FJ, Ferrone ML, *et al*. Development of a prognostic survival algorithm for patients with metastatic spine disease. J Bone Jt Surg 2016;98:1767-1776.

27 Karhade AV, Thio QC, Ogink PT, Bono CM, Ferrone ML, Oh KS, *et al*. Predicting 90-day and 1-year mortality in spinal metastatic disease: development and internal validation. Neurosurgery 2019;85: E671-E681.

28 Alfaro PA, Delgado J, Dumas A, Mesa C, Wevar O, Herrera C, *et al*. Comparison between different prognostic models to be used for metastatic bone disease on appendicular skeleton in a Chilean population. Eur J Orthop Surg Traumatol 2021;31:1657-1662.

29 Sørensen MS, Gerds TA, Hindsø K, Petersen MM. Prediction of survival after surgery due to skeletal metastases in the extremities. Bone Joint J 2016;98-b:271-277.

30 Sørensen MS, Gerds TA, Hindsø K, Petersen MM. External validation and optimization of the SPRING model for prediction of survival after surgical treatment of bone metastases of the extremities. Clin Orthop Relat Res 2018;476:1591-1599.

31 Willeumier JJ, Van der Linden YM, Van der Wal CWPG, Jutte PC, van der Velden, *et al*. An easy-to-use prognostic model for survival estimation for patients with symptomatic long bone metastases. J Bone Joint Surg Am 2018;100:196-204.

32 Thio QC, Karhade AV, Ogink PT, Bramer JA, Ferrone ML, Calderón SL, *et al*. Development and internal validation of machine learning algorithms for preoperative survival prediction of extremity metastatic disease. Clin Orthop Relat Res 2020;478:322-333.

33 Forsberg JA, Wedin R, Bauer HC, Hansen BH, Laitinen M, Trovik CS, *et al*. External validation of the Bayesian Estimated Tools for Survival (BETS) models in patients with surgically treated skeletal metastases. BMC Cancer 2012;12:493.

34 Anderson AB, *et al*. External validation of PATHFx version 3.0 in patients treated surgically and nonsurgically for symptomatic skeletal metastases. Clin Orthop Relat Res 2020;478:808-818.

35 Janssen SJ, van der Heijden AS, van Dijke M, Ready JE, Raskin KA, Ferrone ML, *et al*. 2015 Marshall Urist Young Investigator Award: prognostication in patients with long bone metastases: does a boosting algorithm improve survival estimates? Clin Orthop Relat Res 2015;473:3112-3121.

36 Yang SB, Cho W, Chang U-K. Analysis of prognostic factors relating to postoperative survival in spinal metastases. J Korean Neurosurg Soc 2012;51:127-134.

37 Lee CH, *et al*. Which one is a valuable surrogate for predicting survival between Tomita and Tokuhashi scores in patients with spinal metastases? A meta-analysis for diagnostic test accuracy and individual participant data analysis. J Neurooncol 2015;123:267-275.

38 Smeijers S, Depreitere B. Prognostic scores for survival as decisional support for surgery in spinal metastases: a performance assessment systematic review. Eur Spine J 2021;30:2800-2824.

39 Nater A, Martin AR, Sahgal A, Choi D, Fehlings MG. Symptomatic spinal metastasis: a systematic literature review of the preoperative prognostic factors for survival, neurological, functional and quality of life in surgically treated patients and methodological recommendations for prognostic studies. PLoS One 2017;12:e0171507.

40 Luksanapruksa P, Buchowski JM, Hotchkiss W, Tongsai S, Wilartratsami S, Chotivichit A. Prognostic factors in patients with spinal metastasis: a systematic review and meta-analysis. Spine J 2017;17:689-708.

41 Bollen L, Jacobs WCH, Van der Linden YM, Van der Hel O, Taal W, Dijkstra PDS. A systematic review of prognostic factors predicting survival in patients with spinal bone metastases. Eur Spine J 2018;27:799-805.

42 Ahmed AK, Goodwin CR, Heravi A, Kim R, Abu-Bonsrah N, Sankey E, *et al*. Predicting survival for metastatic spine disease: a comparison of nine scoring systems. Spine J 2018;18:1804-1814.

43 Karhade AV, Ahmed AK, Pennington Z, Chara A, Schilling A, Thio QC, *et al*. External validation of the SORG 90-day and 1-year machine learning algorithms for survival in spinal metastatic disease. Spine J 2020;20:14-21.

44 Bongers ME, Karhade AV, Villavieja J, Groot OQ, Bilsky MH, Laufer I, *et al*. Does the SORG algorithm generalize to acontemporary cohort of patients with spinal metastases on external validation? Spine J 2020;20:1646-1652.

45 Shah AA, Karhade AV, Park HY, Sheppard WL, Macyszyn LJ, Everson RG, *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 J 2021;21:1679-1686.

46 Wolff RF, Moons KG, Riley RD, Whiting PF, Westwood M, Collins GS, *et al*. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. Ann Intern Med 2019;170:51-58.

47 Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD). Ann Intern Med 2015;162:735-736.

48 Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, *et al*. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med 2015;162:W1-W73.

49 Siontis GC, Tzoulaki I, Castaldi PJ, Ioannidis JP. External validation of new risk prediction models is infrequent and reveals worse prognostic discrimination. J Clin Epidemiol 2015;68:25-34.

50 Depreitere B, Ricciardi F, Arts M, Balabaud L, Buchowski JM, Bunger C, *et al*. Loss of local tumor control after index surgery for spinal metastases: A prospective cohort study. World Neurosurg 2018;117:e8-e16.

51 Barzilai O, Fisher CG, Bilsky MH. State of the art treatment of spinal metastatic disease. Neurosurgery 2018;82:757-769.

52 Choi D, Fox Z, Albert T, Arts M, Balabaud L, Bunger C, *et al*. Prediction of quality of life and survival after surgery for symptomatic spinal metastases: a multicenter cohort study to determine suitability for surgical treatment. Neurosurgery 2015;77:698-708.

53 Tabourel G, Terrier LM, Dubory A, Cristini J, Le Nail LR, Cook AR, *et al*. Are spine metastasis survival scoring systems outdated and do they underestimate life expectancy? Caution in surgical recommendation guidance. J Neurosurg: Spine 2021;35:527-534.

54 Alves CR, Faria DDP, Carneiro CDG, Garcez AT, Gutierrez VP, das Neves W, *et al*. 18F-Fluoride PET/CT and 99mTc-MDP SPECT/CT can detect bone cancer at early stage in rodents. Life Sciences 2018;206:29-34.

55 Serratrice N, Faddoul J, Tarabay B, Attieh C, Chalah MA, Ayache SS, Abi Lahoud GN. Ten years after SINS: Role of surgery and radiotherapy in the management of patients with vertebral metastases. Front Oncol 2022;12:802595.

56 Ulmar B, Naumann U, Catalkaya S, Muche R, Cakir B, Schmidt R, *et al*. Prognosis scores of Tokuhashi and Tomita for patients with spinal metastases of renal cancer. Ann Surg Oncol 2007;14:998-1004.

57 Oh CD, Maity SN, Lu JF, Zhang J, Liang S, Coustry F, *et al*. Identification of SOX9 interaction sites in the genome of chondrocytes. PloS One 2010;5:e10113.

58 Moon KY, Chung CK, Jahng TA, *et al*. Postoperative survival and ambulatory outcome in metastatic 7 spinal tumors: prognostic factor analysis. J Korean Neurosurg Soc 2011;50:216-23.

59 Park JH, Rhim SC, Jeon SR, *et al*. Efficacy of decompression and fixation for metastatic spinal cord compression: analysis of factors prognostic for survival and postoperative ambulation. J Korean Neurosurg Soc 2011;50:434.

60 Mollahoseini R, Farhan F, Jouibari MAM, Gholipour F. Is Tokuhashi score suitable for evaluation of life expectancy before surgery in Iranian patients with spinal metastases? J Res Med Sci: Off J Isfahan Univ Med Sci 2011;16:1183.

61 Pointillart V, Vital JM, Salmi R, Diallo A, Quan GM. Survival prognostic factors and clinical outcomes in patients with spinal metastases. J Cancer Res Clin Oncol 2011;137:849-856.

62 Hessler C, Vettorazzi E, Madert J, Bokemeyer C, Panse J. Actual and predicted survival time of patients with spinal metastases of lung cancer: evaluation of the robustness of the Tokuhashi score. Spine 2011;36:983-989.

63 Wibmer C, Leithner A, Hofmann G, Clar H, Kapitan M, Berghold A, Windhager R. Survival analysis of 254 patients after manifestation of spinal metastases: evaluation of seven preoperative scoring systems. Spine 2011;36:1977-1986.

64 Majeed H, Kumar S, Bommireddy R, Klezl Z, Calthorpe D. Accuracy of prognostic scores in decision making and predicting outcomes in metastatic spine disease. Ann Royal Coll Surg Engl 2012;94:28-33.

65 Oh I-S, Kim S-I, Ha K-I. Significant predictive values for the life expectancy in patients with spinal metastasis following surgical treatment. Eur J Orthop Surg Traumatol 2012;22:17-24.

66 Wang XS, Rhines LD, Shiu AS, Yang JN, Selek U, Gning I, *et al*. Stereotactic body radiation therapy for management of spinal metastases in patients without spinal cord compression: a phase 1–2 trial. Lancet Oncol 2012;13:395-402.

67 Hernandez-Fernandez A, Vélez R, Lersundi-Artamendi A, Pellisé F. External validity of the Tokuhashi score in patients with vertebral metastasis. J Cancer Res Clin Oncol 2012;138:1493-1500.

68 Lee J, Rhee WJ, Chang JS, Chang SK, Koom WS. Evaluation of predictive factors of vertebral compression fracture after conventional palliative radiotherapy for spinal metastasis from colorectal cancer. J Neurosurg Spine 2017;28:333-340.

69 Gakhar H, Swamy GN, Bommireddy R, Calthorpe D, Klezl Z. A study investigating the validity of modified Tokuhashi score to decide surgical intervention in patients with metastatic spinal cancer. Eur Spine J 2013;22:565-568.

70 Tabouret E, Cauvin C, Fuentes S, *et al*. Reassessment of scoring systems and prognostic factors for metastatic spinal cord compression. Spine J 2015;15:944-950. doi:10.1016/j.spinee.2013.06.036.

71 Kim J, Lee SH, Park SJ, *et al*. Analysis of the predictive role and new proposal for surgical strategies based on the modified Tomita and Tokuhashi scoring systems for spinal metastasis. World J Surg Oncol 2014;12:245. doi:10.1186/1477-7819-12-245

72 Yeung YN, Cheung KK, Lam TC, Cheng HO, Chow YY. A study of the predictive value of the modified Tokuhashi score in metastatic spinal tumour causing cord compression in a southern Chinese population. J Orthopaed Trauma Rehab 2014;18:15-21.

73 Aoude A, Louis-Philippe A. A comparison of the modified Tokuhashi and Tomita scores in determining prognosis for patients afflicted with spinal metastasis. Canadian J Surg 2014;57:188.

74 Kumar N, Tan JJ, Zaw AS, Lim JL, Wai KL, Malhotra R, *et al*. Evaluation of scoring systems and prognostic factors in patients with spinal metastases from nasopharyngeal carcinoma. Spine J 2014;14: 2946-2953.

75 Liang TZ, Wan Y, Long GH, Zou XN, Peng XS, Zheng ZM. Predictive value of three surgical scoring systems for estimation of life expectancy in patients with extradural spinal metastasis. Zhonghua zhong liu za zhi (Chin J Oncol) 2010;32:875-879.

76 Zhang D, Xu W, Liu T, Yin H, Yang X, Wu Z, Xiao J. Surgery and prognostic factors of patients with epidural spinal cord compression caused by hepatocellular carcinoma metastases: retrospective study of 36 patients in a single center. Spine 2013;38:E1090-E1095.

77 Dardic M, Wibmer C, Berghold A, *et al*. Evaluation of prognostic scoring systems for spinal metastases in 196 patients treated during 2005-2010. Eur Spine J 2015;24:2133-2141.

78 Hibberd CS, Quan GMY. Accuracy of preoperative scoring systems for the prognostication and treatment of patients with spinal metastases. Int Scholar Res Notice 2017;1320684:e1-e9.

79 Paulino Pereira NR, Mclaughlin L, Janssen SJ, van Dijk CN, Bramer JA, Laufer I, *et al*. The SORG nomogram accurately predicts 3- and 12-months survival for operable spine metastatic disease: External validation. J Surg Oncol 2017;115:1019-1027.

80 Czigléczki G, Mezei T, Pollner P, Horváth A, Banczerowski P. Prognostic factors of surgical complications and overall survival of patients with metastatic spinal tumor. World Neurosurg 2018;113:e20-e28.

81 Forsberg JA, Eberhardt J, Boland PJ, Wedin R, Healey JH. Estimating survival in patients with operable skeletal metastases: an application of a bayesian belief network. PloS One 2011;6:e19956.

82 Ratasvuori M, Wedin R, Keller J, Nottrott M, Zaikova O, Bergh P, *et al*. Insight opinion to surgically treated metastatic bone disease: Scandinavian sarcoma group skeletal metastasis registry report of 1195 operated skeletal metastasis. Surg Oncol 2013;22:132-138.