RJNMSR
Review Journal of Neurological & Medical Sciences Review

Volume 3, Issue 8, 2025

VALIDATION OF THE TOKUHASHI SCORE AND TOMITA SCORE IN PREDICTING SURVIVAL IN PATIENTS WITH SPINAL METASTASES: A RETROSPECTIVE COHORT STUDY IN AN LMIC POPULATION

Ihsan Ullah[1], Saif Ullah Khalid[2], Syed Shayan Shah[3], Muhammad Nawaz Khan[*4], Syed Jawad Ahmad[5], Sardar Ali[6]

[*1,4]*Assistant Professor, Department of Neurosurgery, Lady Reading Hospital, Peshawar, Pakistan*
[2,3,5,6]*Postgraduate Residents Neurosurgery, Lady Reading Hospital, Peshawar, Pakistan*

[1]ihsan.fatih@lrh.edu.pk, [2]saifullah215@gmail.com, [3]shayanshah93@gmail.com, [*4]muhammad.nawaz@lrh.edu.pk, [5]jawadahmed235@yahoo.com, [6]sardarali8053@gmail.com

Corresponding Author: *
Muhammad Nawaz Khan
DOI: https://doi.org/10.5281/zenodo.19694335

Received	Accepted	Published
27 October 2025	11 December 2025	25 December 2025

ABSTRACT

Background: *Proper preoperative prognostication is critical in making surgical decisions and resource allocation in patients with spinal metastases. The Tokuhashi and Tomita scores have been revised, but their effectiveness in low- and middle-income country (LMIC) settings with delayed presentation and limited adjuvant therapies remains poorly assessed.*

Methods: *This retrospective cohort study included adult patients (aged 18 years and older) who underwent surgery for spinal metastasis at a tertiary neurosurgical unit between January 1, 2020, and December 31, 2025. Standardized parameters were used to calculate preoperative revised Tokuhashi and Tomita scores. The main outcome was overall survival (OS) after surgery. Harrell measured discrimination using the C-index and 6- and 12-month time-dependent area under the receiver operating characteristic curve (AUC) of survival. The DeLong test was used to assess comparative discrimination. Decision curve analysis (DCA) measured net clinical benefit at threshold probabilities. Independent predictors were determined using multivariable Cox regression.*

Results: *Among 265 surgical patients, 228 were eligible (mean age 56.8 ± 13.1 years; 61% male). Median OS was 8.4 months (IQR 3.2–15.6); 12-month survival was 42.5%. The revised Tokuhashi score was better discriminatory (C-index 0.73, 95% CI 0.68 0.78) than the Tomita score (C-index 0.68, 95% CI 0.62 0.73). Tokuhashi was statistically better on DeLong testing (12-month survival) (AUC 0.71 vs 0.67; 0.04 difference, p=0.041). DCA showed a higher net benefit of the Tokuhashi score with a 20-50% threshold probability of 12-month mortality. A subgroup with high risk (Tokuhashi ≤8 and untreatable visceral metastases, n=62) had a median OS of 3.9 months (6-month survival 19.4%). The independent predictors were Tokuhashi score (adjusted HR 0.82 per point increase, 95% CI 0.760.88, p<0.001), rapid-growing primary tumours, visceral metastases, and postoperative complications.*

Conclusion: *In this LMIC cohort, the revised Tokuhashi score shows slightly better discrimination and clinical utility than the Tomita score. Preoperative use is supported by routine use to help select surgical versus palliative care, especially in resource-constrained settings, based on evidence.*

Keywords: *spinal metastases; Tokuhashi score; Tomita score; survival prediction; decision curve analysis, LMIC, prognostic validation*

Introduction

One of the most significant effects of advanced cancer is spinal metastatic disease. Skeletal metastasis most frequently occurs in the spine, and imaging and clinical series indicate that spinal disease is experienced in a significant percentage of patients with disseminated malignancy. It is not only significant in frequency, but in the morbidity it causes: chronic axial pain, pathological fracture, mechanical instability, epidural spinal cord compression, neurological deficit, loss of ambulation, and a rapid decline in quality of life. [1-3]

Surgery is still a pillar of treatment to selected patients. Operative intervention aims are palliative yet clinically significant: neural decompression, restoration or preservation of stability, alleviation of mechanical pain, functional maintenance, and support of subsequent systemic therapy or radiotherapy. Direct decompressive surgery with radiotherapy has been shown to be better than radiotherapy alone in metastatic epidural spinal cord compression in terms of ambulatory outcomes, whereas prospective multicentre data have demonstrated that surgery may also provide long-term benefits in terms of pain, neurological functioning and health-related quality of life when patients are selected appropriately. [3-5]

Since the anticipated survival determines whether a patient is likely to gain benefit through decompression and stabilisation, through more limited palliation, or through non-operative treatment, a number of prognostic systems have been created to aid in decision-making. The Tokuhashi score has been revised to give a total of 0 to 15 points depending on the performance status, neurological status, number of extraspinal bone metastases, number of involved vertebrae, primary tumour type, and visceral metastases. The Tomita score is a smaller 2 to 10-point scale, based on primary tumour growth rate, visceral metastases, and skeletal metastatic burden, and was created to guide the desired aggressiveness of surgery.[6-8] These systems have had long-term clinical impact, although even subsequent analyses of prognostic modelling in spinal metastases have acknowledged that legacy scores tend to offer broad survival grouping, rather than high-fidelity individual prediction, and that no historical system has demonstrated consistently high concordance between predicted and observed survival across environments. [8]

The need to reappraise these models has become more pressing in contemporary oncology. The survival of metastatic cancer has improved significantly for specific tumour types due to stereotactic radiotherapy, surgical resection plans, targeted agents, immune checkpoint blockade, bone-modifying therapy, and more efficient systemic sequencing. Recent literature has thus cast doubt on whether the conventional spinal metastasis survival scores are becoming obsolete, under-calibrated, or not responsive enough to tumour-specific therapeutic improvements. [9,10] Meanwhile, the majority of validation cohorts mentioned are based on high-income or selected East Asian centres, but the extrapolability of these scores to resource-constrained systems is much less assured. Inequality in referral timing, molecular work-up, access to systemic therapy, and perioperative optimisation can change both case-mix and observed prognosis in LMIC settings.[11-13] We thus externally validated the revised Tokuhashi and Tomita scores, compared their survival discrimination with Harrell C-index and time-dependent AUCs, their clinical utility with decision curve analysis, and independent predictors of postoperative survival in a retrospective cohort study of surgically treated adults with spinal metastases in a tertiary referral centre in Pakistan.

Materials and Methods

Study Design and Setting

This single-center retrospective cohort study was conducted at the Department of Neurosurgery, Lady Reading Hospital, Peshawar, Pakistan. The study period was January 1, 2020, to December 31, 2025. Ethical approval was obtained from the institutional review board (LRH-IRB/2024-112). Informed consent was waived due to the retrospective nature.

Participants I

nclusion criteria: adult patients (≥18 years) undergoing open decompression with or without instrumentation for histologically confirmed or radiologically/clinically evident spinal metastases from a known primary malignancy. Exclusion criteria: incomplete

records precluding score calculation, primary spinal tumors, or non-operative management.

Sample Size

Sample size was determined using the rule of thumb for multivariable Cox regression (minimum 10 events per predictor variable). With 8 key predictors and an anticipated 12-month mortality of approximately 58%, a minimum of 138 patients (80 events) was required for >80% power to detect moderate associations (HR ≥1.5) at α=0.05. The final cohort of 228 patients exceeded this threshold.

Data Collection

Variables included demographics, Karnofsky performance status, Frankel grade, primary tumor histology (classified as slow-, moderate-, or rapid-growing per score criteria), number of metastases, visceral metastases status, calculated Tokuhashi and Tomita scores, postoperative complications, and survival data (hospital records supplemented by telephonic follow-up where available). Survival was censored at last contact.

Statistical Analysis

Continuous data were summarized as mean ± SD or median (IQR); categorical data as frequencies (%). Survival was estimated by the Kaplan-Meier method and compared with the log-rank test. Discrimination was assessed by Harrell's C-index (with 95% CI) and time-dependent AUC for 6- and 12-month survival. Comparative discrimination between scores was performed using the DeLong test for paired ROC curves, reporting ΔAUC and p-value.

Decision curve analysis (DCA) evaluated clinical utility by plotting the net benefit of each score across threshold probabilities (10–60%) for 12-month mortality, using "treat-all" and "treat-none" as references.

Univariable Cox regression screened predictors ($p<0.10$ entry threshold). Multivariable Cox models with backward stepwise elimination yielded adjusted hazard ratios (HR) with 95% CI. The proportional hazards assumption was verified using Schoenfeld residuals and log-minus-log plots. A two-tailed $p<0.05$ was considered statistically significant. Analyses were performed in IBM SPSS Statistics version 26.0 and R (version 4.3.1) for DCA and DeLong testing.

Results

Patient Characteristics

A total of 265 patients underwent surgery; 228 were analyzed after excluding 37 with incomplete records (Figure 1). Baseline characteristics are presented in Table 1.

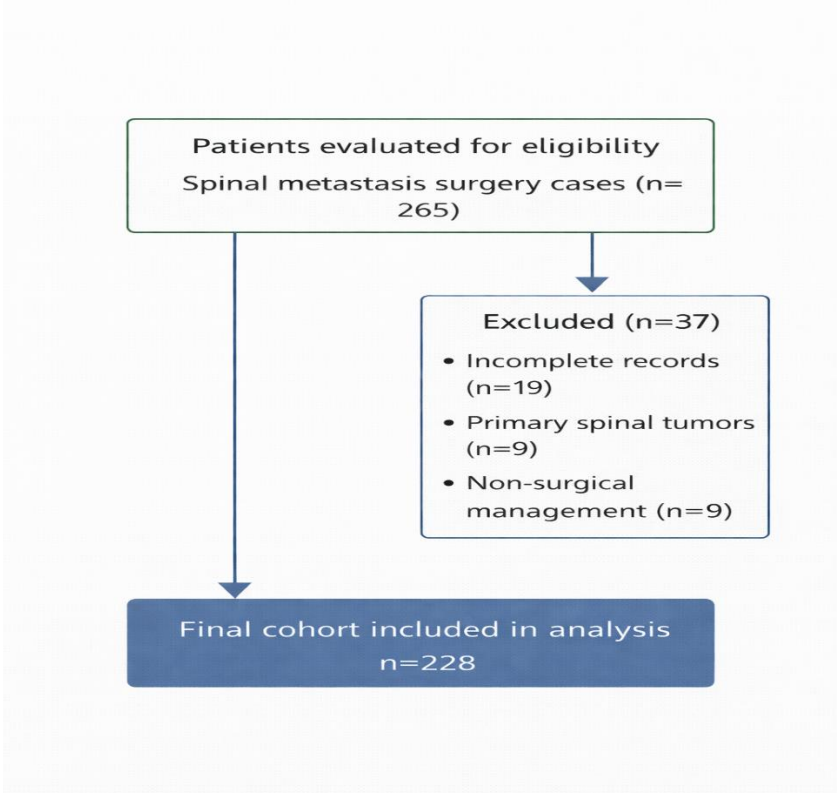

Figure 1: Flow diagram of patient inclusion.

Table 1: Baseline demographic and clinical characteristics (N=228)

Variable	Value
Age (years), mean ± SD	56.8 ± 13.1
Male, n (%)	139 (61.0)
Karnofsky performance status, median (IQR)	60 (40–70)
Frankel grade A–C, n (%)	98 (43.0)
Rapid-growing primary tumor, n (%)	82 (36.0)
Untreatable visceral metastases, n (%)	112 (49.1)
Revised Tokuhashi score, mean ± SD	9.1 ± 2.9
Tokuhashi 0–8 (poor), n (%)	96 (42.1)
Tokuhashi 9–11 (intermediate), n (%)	82 (36.0)
Tokuhashi 12–15 (good), n (%)	50 (21.9)
Tomita score, mean ± SD	6.4 ± 1.8
Postoperative complications, n (%)	51 (22.4)

Outcomes

Median overall survival was 8.4 months (95% CI 6.9–9.9). Six-month survival was 61.4%; 12-month survival was 42.5%. Kaplan-Meier curves showed significant stratification by both scores (log-rank $p<0.001$; Figure 2).

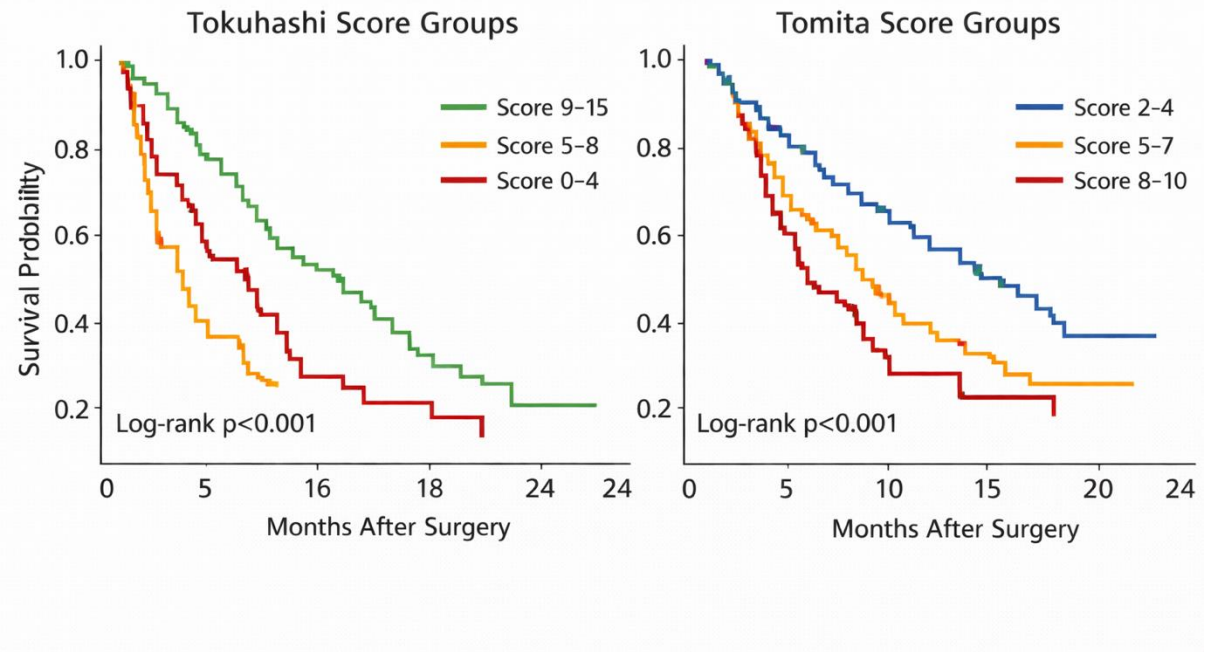

Figure 2: Kaplan-Meier survival curves stratified by Tokuhashi and Tomita score groups (log-rank p<0.001 for both).

Univariate Cox regression results are summarized in Table 2.

Table 2: Univariate Cox regression for overall survival

Predictor	HR (95% CI)	p-value
Tokuhashi score (per point)	0.79 (0.74–0.85)	<0.001
Tomita score (per point)	1.42 (1.28–1.57)	<0.001
Rapid-growing primary	2.14 (1.58–2.90)	<0.001
Untreatable visceral metastases	2.67 (1.98–3.60)	<0.001
Karnofsky <50	2.31 (1.71–3.12)	<0.001
Postoperative complications	1.78 (1.29–2.46)	0.001

Multivariable Analysis and Advanced Metrics

In multivariable Cox models, the revised Tokuhashi score remained independently associated with survival (adjusted HR 0.82 per point increase, 95% CI 0.76–0.88, p<0.001), along with rapid-growing primaries, visceral metastases, and complications (Table 3). Similar independent associations were observed in the Tomita model.

Comparative discrimination analysis showed the Tokuhashi score was statistically superior for 12-month survival (AUC 0.71, 95% CI 0.64–0.78 vs 0.67, 95% CI 0.60–0.74 for Tomita; ΔAUC = 0.04, p=0.041 by DeLong test). Results were consistent for 6-month survival.

Decision curve analysis revealed greater net clinical benefit for the Tokuhashi score across threshold probabilities of 20–50% for 12-month mortality risk compared with the Tomita score and default strategies (Figure 3).

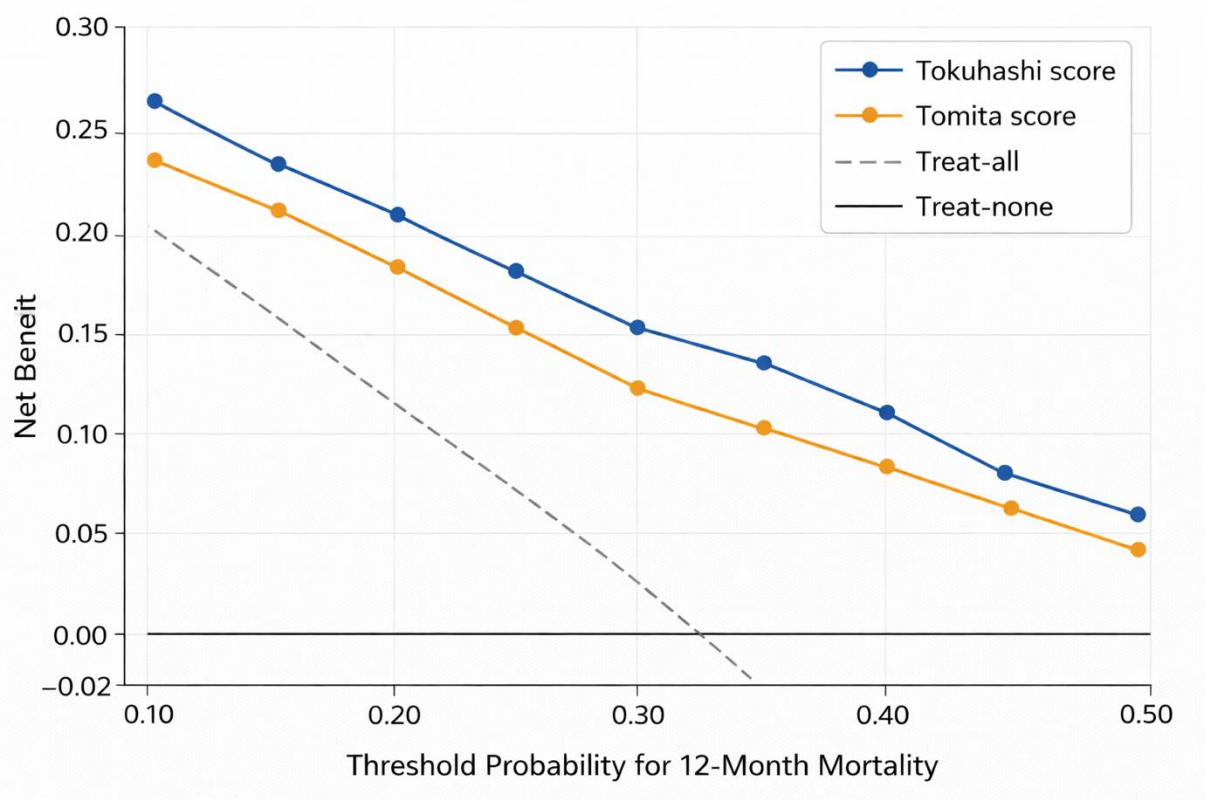

Figure 3: Decision curve analysis for 12-month mortality showing net benefit of Tokuhashi versus Tomita scores.

A clinically relevant high-risk subgroup—patients with revised Tokuhashi score ≤8 and untreatable visceral metastases (n=62, 27.2%)—had markedly poor outcomes (median OS 3.9 months, 95% CI 2.5–5.3; 6-month survival 19.4%). In this group, surgical intervention offered limited survival benefit, favoring individualized consideration of palliative approaches to minimize morbidity.

Discussion

Both the revised Tokuhashi and Tomita scores in this externally validated LMIC surgical cohort had retained clinically relevant and only moderate discriminatory ability of postoperative survival. The re-calculated Tokuhashi score was generally more successful, having a larger reported C-index, better 12-month AUC and greater net benefit in decision curve analysis at clinically plausible threshold probabilities. Notably, the practical significance of this observation lies not so much in the numerical magnitude of the difference in AUC as in the fact that the improved-performing model also provided more helpful information for treatment decisions. The fact that a high-risk subgroup with Tokuhashi score revised to 8 or less and untreatable visceral metastases, with a median overall survival of less than 4 months, was identified further indicates that the score can be used to identify patients in whom the burden of major surgery warrants particularly close justification [14].

The results are directionally consistent with previous direct comparisons of the two systems, but must be viewed in the context of a more critical contemporary literature. Aoude and Amiot have found that modified Tokuhashi has been shown to be more effective in a surgical cohort than Tomita, which confirms the trend here, but the wider literature has grown more sceptical of the performance of legacy models alone.[14] No widely used prognostic score has shown consistent good predictive value in a large prospective international multicentre study,[15] nor does the current study support the assertion that old-fashioned scores are still entirely good, as Li et al. found significant heterogeneity in the performance of seven models in a contemporary Chinese cohort,[16] while Bindels et al., in a 2025 external validation of 12 survival prediction models, showed poor-to-fair discrimination

overall and poor calibration across the model set.[17]

The current study also sits within a broader shift in how prognosis is modelled for spinal metastases. Park et al. recently reported that most commonly used scoring systems had low discriminative power for poor prognosis, with only more contemporary modelling approaches showing moderate performance.[18] In prospective comparison, the New England Spinal Metastasis Score differentiated survival more effectively than several legacy systems,[19] and external validation work on the SORG algorithms has shown that machine-learning models can achieve credible discrimination when tested in independent populations.[20] More recently, combined approaches, such as integration of the revised Tokuhashi and New England scores, have also been proposed to improve predictive accuracy.[21] Taken together, these studies suggest that the most defensible interpretation of the present findings is that revised Tokuhashi remains a useful benchmark, but not the final word in prognostication

One reason traditional scores may underperform is that contemporary survival is increasingly shaped by tumour-specific systemic therapy. This is particularly evident in lung cancer, renal cell carcinoma, breast cancer, and other tumour groups in which targeted therapy, immunotherapy, and modern multidisciplinary treatment have altered historical survival expectations.[9,10] In a non-surgical lung cancer spinal metastasis cohort, Yan et al. found that multiple prognostic systems were ineffective,[22] while Huang et al. showed in surgically treated non-small cell lung cancer that Tomita and revised Tokuhashi tended to underestimate survival.[23] The present cohort's median survival of 8.4 months is therefore best understood not simply as a local descriptive statistic, but as the product of tumour biology, disease burden, treatment access, and selection for surgery. In that context, the absence of molecular and treatment-response variables is not a minor omission; it is a structural limitation of legacy prognostic models.

Clinically, the relevance of these results lies in decision support rather than decision replacement. Modern spinal oncology frameworks such as NOMS already emphasise that prognosis is only one component of treatment planning and must be integrated with neurological status, oncological radiosensitivity, mechanical instability, systemic reserve, and patient goals.[13] In that setting, the revised Tokuhashi score appears better suited than Tomita to function as a pragmatic survival-estimation tool in this LMIC surgical population. The decision curve analysis strengthens that claim, because it indicates potential usefulness across threshold probabilities that are close to real-world discussions about whether to proceed with major decompression and stabilisation or to favour less invasive palliation.[24] The key message is therefore not that Tokuhashi should determine care, but that it can contribute meaningfully to multidisciplinary, shared decision-making when used with appropriate clinical judgement.

The finding that postoperative complications were independently associated with poorer survival is also important. This reinforces two related points: first, that metastatic spine surgery is performed in a medically fragile population; and second, that any prognostic discussion that considers surgery only in terms of tumour burden and score-based survival risks underestimating perioperative vulnerability. Prospective outcome studies have shown that surgery can produce rapid and sustained improvement in pain and quality of life, but they also underscore non-trivial adverse-event burdens and the need for careful selection and optimisation.[25,26] In practical terms, a model that is only moderately discriminative may still be clinically valuable if it helps avoid high-morbidity surgery in patients whose systemic condition makes meaningful postoperative recovery unlikely.

Several limitations should be acknowledged explicitly. This was a retrospective single-centre surgical cohort, and the inevitable selection bias means the results cannot be extrapolated to all patients with spinal metastases, especially those managed non-operatively. The study evaluates discrimination and decision utility, but does not fully address calibration, which is now recognised as a critical component of external validation. It also lacks granular molecular data, contemporary systemic therapy variables, and tumour-subtype detail, all of which are increasingly relevant to survival estimation in metastatic oncology. In addition, although primary tumour histology was collected, the current presentation would be strengthened by explicit reporting of site-specific

primary tumour distribution rather than relying mainly on score-derived growth categories. Future work should therefore focus on recalibration using recent local data, incorporation of tumour-specific and treatment-specific variables, and comparison with contemporary machine-learning or hybrid models that may capture heterogeneity more effectively than legacy scores alone.

Conclusion

The revised Tokuhashi score has slightly better discriminative and clinical utility for predicting survival following spinal metastasis surgery than the Tomita score. The preoperative use of these validated instruments can support evidence-based risk stratification, shared decision-making between surgical intervention and palliative care, and efficient resource allocation in high-burden, resource-constrained health care systems.

REFERENCES

Coleman RE. Metastatic bone disease: clinical features, pathophysiology and treatment strategies. **Cancer Treat Rev.** 2001;27(3):165–176. PMID: 11417967.

Shah LM, Salzman KL. Imaging of spinal metastatic disease. **Int J Surg Oncol.** 2011;2011:769753. PMID: 22312523.

Klimo P Jr, Schmidt MH. Treatment of metastatic spinal epidural disease: a review of the literature. **Neurosurg Focus.** 2003;15(5):E1. PMID: 15323458.

Patchell RA, Tibbs PA, Regine WF, Payne R, Saris S, Kryscio RJ, et al. Direct decompressive surgical resection in the treatment of spinal cord compression caused by metastatic cancer: a randomized trial. **Lancet.** 2005;366(9486):643–648. PMID: 16112300.

Fehlings MG, Nater A, Tetreault L, Kopjar B, Arnold PM, Massicotte EM, et al. Survival and clinical outcomes in surgically treated patients with metastatic epidural spinal cord compression: results of the prospective multicenter AOSpine study. **J Clin Oncol.** 2016;34(3):268–276. PMID: 26598751.

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(19):2186–2191. PMID: 16205345.

Tomita K, Kawahara N, Kobayashi T, Yoshida A, Murakami H, Akamaru T. Surgical strategy for spinal metastases. **Spine (Phila Pa 1976).** 2001;26(3):298–306. PMID: 11224867.

Tokuhashi Y, Uei H, Oshima M, Ajiro Y. Scoring system for prediction of metastatic spine tumor prognosis. **World J Orthop.** 2014;5(3):262–271. PMID: 25035829.

Tabouret E, Terrier LM, Jalinier C, Montava M, Barlesi F, Fuentes S, et al. Are spine metastases survival scoring systems outdated and irrelevant in today's practice? **J Neurosurg Spine.** 2021;35(4):527–534. PMID: 34298515.

Fomchenko EI, Byvalets JV, Alvarez-Breckenridge CA, Rhines LD, Tatsui CE. Spinal metastases and the evolving role of molecular targeted therapy, chemotherapy, and immunotherapy. **Neurosurgery.** 2022;90(4):978–993. PMID: 36597635.

Farmer P, Frenk J, Knaul FM, Shulman LN, Alleyne G, Armstrong L, et al. Expansion of cancer care and control in countries of low and middle income: a call to action. **Lancet.** 2010;376(9747):1186–1193. PMID: 20709386.

Sullivan R, Alatise OI, Anderson BO, Audisio R, Autier P, Aggarwal A, et al. Global cancer surgery: delivering safe, affordable, and timely cancer surgery. **Lancet Oncol.** 2015;16(11):1193–1224. PMID: 26427363.

Laufer I, Rubin DG, Lis E, Cox BW, Stubblefield MD, Yamada Y, et al. The NOMS framework: approach to the treatment of spinal metastatic tumors. **Oncologist.** 2013;18(6):744–751. PMID: 23709750.

Aoude A, Amiot LP. A comparison of the modified Tokuhashi and Tomita scores in determining prognosis for patients afflicted with spinal metastasis. **Can J Surg.** 2014;57(3):188–193. PMID: 24869611.

Choi D, Ricciardi F, Arts M, Buchowski JM, Bunger C, Fehlings MG, et al. Prediction accuracy of common prognostic scoring systems for metastatic spine disease: results of a prospective international multicenter study of 1469 patients. **Spine (Phila Pa 1976).** 2016;41(5):338–346. PMID: 30422958.

Li Z, Guo S, Zhang Z, Wang Y, Chen X, Liu H, et al. Evaluation of different scoring systems for spinal metastases based on a Chinese cohort. **Cancer Med.** 2023;12(4):4125–4136. PMID: 36128836.

Bindea BJ, Kuijten RH, Groot OQ, Vleggeert-Lankamp CLA, Verlaan JJ, Oner FC, et al. External validation of 12 existing survival prediction models for patients with spinal metastases. **Spine J.** 2025;25(7):1347–1359. PMID: 39894281.

Park SJ, Park JS, Kang DH, Lee CS. Which prognostic model best predicts poor prognosis in patients with spinal metastases? A comparative analysis of 8 scoring systems. **World Neurosurg.** 2025;193:553–566. PMID: 39414132.

Schoenfeld AJ, Ferrone ML, Blucher JA, et al. Prospective comparison of the accuracy of the New England Spinal Metastasis Score to legacy scoring systems in prognosticating outcomes following treatment of spinal metastases. Spine J. 2022;22(1):39-48. PMID: 33741509.

Shah AA, Ogink PT, Harris MB, 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(10):1679-1686. PMID: 33798728.

Mavritsakis D, Amiot LP. A novel prognostic scoring system combining the revised Tokuhashi score and the New England spinal metastasis score for preoperative evaluation of spinal metastases. Front Surg. 2024;11:1349586. PMID: 38505407.

Yan Y, Zhong G, Lai H, Huang C, Yao M, Zhou M, et al. Comparing the accuracy of seven scoring systems in predicting survival of lung cancer patients with spinal metastases: an external validation from two centers. Spine (Phila Pa 1976). 2023;48(14):1009-1016. PMID: 37350732.

Huang AP, Yang CY, Xiao F, Yang SH, Chen CM, Lai DM, et al. Spinal metastases from non-small cell lung cancer: is surgical extent enough by following suggestions of the Tomita and Tokuhashi scores? Asian J Surg. 2024;47(8):3448-3456. PMID: 37684123.

Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Making. 2006;26(6):565-574. PMID: 17099194.

Choi D, Fox Z, Albert T, et al. Rapid improvements in pain and quality of life are sustained after surgery for spinal metastases in a large prospective cohort. Br J Neurosurg. 2016;30(3):337-344. PMID: 26901574.

Barzilai O, Sahgal A, Rhines LD, et al. Patient-reported and clinical outcomes of surgically treated patients with symptomatic spinal metastases: results from epidemiology, process, and outcomes of spine oncology, a prospective, multi-institutional and international study. Neurosurgery. 2024;95(5):1148-1157. PMID: 38832791.