SPINE An International Journal for the study of the spine, Publish Ahead of Print

DOI: 10.1097/BRS.0000000000002576

Prediction accuracy of common prognostic scoring systems for metastatic spine disease: results of a prospective international multicentre study of 1469 patients.

David Choi, PhD[1], Federico Ricciardi, PhD[2], Mark Arts, MD[3], Jacob M. Buchowski, MD[4], Cody Bunger, MD[5], Chun Kee Chung, PhD[6], Maarten Coppes, MD[7], Bart Depreitere, MD[8], Michael Fehlings, PhD[9], Norio Kawahara, PhD[10], Yee Leung, FRCS[11], Antonio Martin-Benlloch, MD[12],Eric Massicotte, MD[9], Christian Mazel, MD[13], Bernhard Meyer, MD[14], CumhurOner, MD[15], Wilco Peul, MD[16], Nasir Quraishi, FRCS[17], Yasuaki Tokuhashi, MD[18], Katsuro Tomita, MD[10], Christian Ulbricht, FRCS[19], Jorrit-JanVerlaan, MD[15], Mike Wang, MD[20], Alan Crockard, DSc[21]

[1]Department of Neurosurgery, National Hospital for Neurology and Neurosurgery, University College London, London, United Kingdom

[2]Department of Statistical Science, University College London, London, United Kingdom

[3]Department of Neurosurgery, Medical Center Haaglanden, Haaglanden, the Netherlands

[4]Departments of Orthopaedic Surgery, Washington University, St. Louis, Missouri, United States of America

[5]Department of Orthopedic Surgery, University Hospital of Aarhus, Aarhus, Denmark

[6]Department of Neurosurgery, Seoul National University Hospital, Seoul, Republic of Korea

[7]Department of Neurosurgery, University Medical Center Groningen, Groningen, the Netherlands

[8]Division of Neurosurgery, University Hospital Leuven, Leuven, Belgium

[9]Division of Neurosurgery and Spinal Program, University of Toronto and Toronto Western Hospital, Toronto, Ontario, Canada

[10]Department of Orthopedic Surgery, Kanazawa Medical University Hospital, Kanazawa, Japan

[11]Department of Orthopaedics, Musgrove Park Hospital, Taunton, United Kingdom

[12]Spinal Unit, University Hospital Dr Peset, Valencia, Spain

[13]Department of Orthopedic Surgery, InstitutMutualisteMontsouris, Paris, France

[14]TUM School of Medicine, Technische Universität München, Munich, Germany

[15]Department of Orthopedic Surgery, University Medical Center Utrecht, Utrecht, the Netherlands

[16]Department of Neurosurgery, Leiden University Medical Centre, Leiden, the Netherlands

[17]Centre for Spine Studies and Surgery, Queens Medical Centre, Nottingham, United Kingdom

[18]Department of Orthopaedic Surgery, Nihon University School of Medicine, Tokyo, Japan

[19]Department of Neurosurgery, Charing Cross Hospital, London, United Kingdom

[20]Department of Neurosurgery, University of Miami Hospital, Miami, Florida, United States of America

[21]Department of Neurosurgery, National Hospital for Neurology and Neurosurgery, University College London, London, United Kingdom

Corresponding author

Mr. David Choi

Department of Neurosurgery

The National Hospital for Neurology and Neurosurgery

Queen Square, London WC1N 3BG, UK

Fax number +44 203 448 3340

Telephone +44 203 448 3395

e-mail david.choi@uclh.nhs.uk

The manuscript submitted does not contain information about medical device(s)/drug(s).

This work was carried out by the Global Spine Tumour Study Group, which receives funding from DePuy Synthes and charitable donations from patients. Part of this work was performed at University College London Hospitals, which receives funding from the UK National Institute for Health Research Biomedical Research Centre.

Relevant financial activities outside the submitted work:board membership, consultancy, royalties, payment for lecture, grants.

ABSTRACT

Study Design: A prospective multicenter cohort study.

Objective: To assess the clinical accuracy of six commonly cited prognostic scoring systems for patients with spinal metastases.

Summary of Background Data: There are presently several available methods for the estimation of prognosis in metastatic spinal disease, but none are universally accepted by surgeons for clinical use. These scoring systems have not been rigorously tested and validated in large datasets to see if they are reliable enough to inform day-to-day patient management decisions. We tested these scoring systems in a large cohort of patients. A total of 1,469 patients were recruited into a secure internet database, and prospectively collected data were analysed to assess the accuracy of published prognostic scoring systems.

Methods: We assessed six prognostic scoring systems, described by the first authors Tomita, Tokuhashi, Bauer, Van der Linden, Rades, and Bollen. Kaplan Meier survival estimates were created for different patient subgroups as described in the original publications. Harrell's C-statistic was calculated for the survival estimates, to assess the concordance between estimated and actual survival.

Results: All the prognostic scoring systems tested were able to categorize patients into separate prognostic groups with different overall survivals. However none of the scores were able to achieve "good concordance" as assessed by Harrell's C-statistic. The score of Bollen and colleagues was found to be the most accurate, with a Harrell's C-statistic of 0.66.

Conclusions: No prognostic scoring system was found to have a good predictive value. The scores of Bollen and Tomita were the most effective with Harrell's C-statistic of 0.66 and 0.65 respectively. Prognostic scoring systems are calculated using data from previous years, and are subject to inaccuracies as treatments advance in the interim. We suggest that other methods of assessing prognosis should be explored, such as prognostic risk calculation.

Key Words: Metastasis; Spine; Surgery; Prediction; Prognosis; Survival; Score

Level of Evidence:3

INTRODUCTION

Spinal metastases are common in patients with cancer, and may be found in 70% of patients with cancer. In addition, up to 10% of patients with cancer present with cord compression and neurological symptoms.(1)

Symptomatic spinal metastases may be treated with surgery, radiotherapy, or combination of treatments to minimise pain, instability or neurological impairment. These treatments, however, are associated with side effects or complications, and therefore it is important to match the level of intervention with the prognosis of the patient to avoid exposing the patient to unnecessary risk.

The Global Spine Tumour Study Group is a group of spine surgeons from around the world who are together studying the outcome of surgery for spinal metastases by using a secure internet registry of consecutive surgeries for patients with symptomatic spinal metastases.(2) One of the principal aims of the registry database was to assess the outcome of surgery for spinal metastases and assess factors which might predict the outcome after treatment, to better inform clinicians and allow patient choice(3). Previously there have been very few reports of quality of life after surgery in this vulnerable patient group (4)and therefore these data will benefit decision making in symptomatic spinal metastases.

Predicting who is likely to benefit from surgery may be calculated mathematically if we analyse patient cohorts and identify those pre-operative factors that are associated with outcome measures such as survival and quality of life(5). However, these mathematical models will never be precise and will be subject to “noise” due to unknown predictors or random variation.

Individual prognostic scores will not necessarily be generalisable since they will depend on the selection bias of the cohort being studied.

Despite a degree of random variation, scientific prediction is still useful todetermine factors that might help a clinician and patient decide which treatment is the most appropriate. Performing a technically challenging, risky and painful operation on a patient who is unlikely to survive to appreciate the benefits of the surgery would be inappropriate. Conversely, avoiding surgery in a patient who would clearly benefit would be equally incorrect.

Several prognostic scoring systems have been published to help predict the survival of patients with metastatic spine disease, undergoing treatment with either surgery, radiotherapy, or combination of treatments. The most cited prognostic scoring systems are those of Tomita(6) and Tokuhashi(7), but other published systems include those of Bauer(8), van der Linden(9), Rades(10) and Bollen(11).

Tomita et al. developed a scoring system including primary tumour type, presence of visceral metastases and number of bone metastases including the spine, based on data from 67 patients. This produced a total prognostic score between 2 to 10, and Tomita et al. divided patients into four distinct prognostic groups (2-3, 4-5, 6-7, 8-10), which determined the treatment goal. (6)

Tokuhashi et al. originally published a scoring system in 1990, (7)which was later revised to include more primary tumour subgroups, (12)as well as the number of spinal, visceral, extraspinal bone metastases, Karnofsky performance status, Frankel neurological grade and the ability to walk. This generated a prognostic score between 0-15, and Tokuhashi et al considered patients in three management categories, scoring 0-8, 9-11, 12-15.

Bauer and Wedin(8) developed a scoring system based on data from 153 patients, assessing the primary tumour type, skeletal and visceral metastases, and the presence of pathological vertebral fracture, generating patient subgroups with 0-1, 2-3, or 4-5 significant criteria.

Bollen et al.(11) were the first group to utilise larger datasets to generate a prognostic score, in 1043 patients, and divided patients into groups A-D depending on the primary tumour type, presence of visceral metastases, and Karnofsky performance status. Only 46 of these patients, however, were treated with surgery.

Rades et al. (13) reviewed the outcome of 1128 patients who underwent primary radiotherapy, of whom 564 were used to calculate the prognostic scoring system. Similar to other systems, they incorporated primary tumour type, number of bone and visceral metastases, into a score, but uniquely also inputed the time from cancer diagnosis to the development of cord compression, and the rapidity of motor deficit.

Van der Linden et al (9) also looked at a registry of patients who underwent radiotherapy, and reviewed 342 patients, and developed a scoring system based on the primary tumour type, Karnofsky performance status and presence of visceral metastases, forming three groups: group A (score 0-3); group B (score 4-5); and group C (score 6).

These scoring systems have not been assessed prospectively in a large group of patients, and some would question the validity of scoring systems. (14, 15)

We aimed to test the accuracy of these well-cited prognostic scoring systems by using a large database of patients which is independent of other published patient groups(16) to validate the scoring systems themselves and the work of other research groups. (16)Unlike other published studies (16), our series has the advantage of being prospectively collected data, and therefore did

not suffer from retrospective reporting bias and potential estimation of parameters using surrogate variables.

MATERIALS AND METHODS

A total of 2,148 consecutive patients who underwent surgery for symptomatic spinal metastases were prospectively recruited, and clinical data were input into the secure cloud database of the Global Spine Tumour Study Group (17). Previous data have been published on earlier patient cohorts. (3) Of these 2,148 patients, 1,469 were used to assess the prediction models of Tomita, Tokuhashi, Bauer, Bollen, van Linden, and Rades. The majority of patients were excluded from survival calculations due to inaccuracies of data entry: 587 patients had prospectively collected data which was retrospectively uploaded (at least one month after date of surgery) and were excluded due to missing or inaccurate data; and 92 patients had missing data for date of surgery or details of surgical technique.

To calculate the Bauer score, one of the criteria used is the presence or absence of pathological vertebral fracture. This was not possible to determine accurately in our data, and therefore the Bauer score was calculated with sensitivity analyses as estimate I (conservative estimate, assuming all patients had a pathological fracture) and estimate II (optimistic estimate, assuming no patient had a pathological fracture).

Statistical analysis

Descriptive statistical summary measures were used to describe relevant variables. Binary and categorical variables are summarized with frequency and percentage (table 1).

Kaplan-Meier survival estimators were fitted and curves were displayed.Hazard Ratios (HR), associated 95% confidence intervals (95%CI) and p-values testing their difference with the reference category were estimated by fitting appropriate Cox proportional hazards models. When statistical tests were performed, $p<0.05$ was taken as the significance threshold.

The probabilities of concordance between predicted and observed responses were estimated using Harrell's C-statistic, similar to previous methods in the literature (18). Based on its values, it is possible to formulate a judgment about the utility of the model:

- A value below 0.5 indicates a very poor model;
- A value of 0.5 means that the model is no better than predicting an outcome than random chance;
- Values over 0.7 indicate a good model;
- Values over 0.8 indicate a strong model;
- A value of 1 means that the model perfectly predicts those group members who will experience a certain outcome and those who will not.

Data analysis was performed using Stata 13 software (StataCorp LLC, Texas USA).

RESULTS

Of 2,148 patients in the GSTSG database on 1st September 2016, 1,469 patients were analysed: 840 males (57%) and 629 females (43%). Mean patient age was 60.6 years (standard deviation 12.7 years).

136 patients presented with symptomatic cervical spine metastases (9.5%), 678 patients had thoracic spine metastases (48.0%), 233 patients had lumbar spine metastases (16.5%), 367 patients had mixed spinal levels (26.0%).

661 patients (45.1%) of patients presented for surgery with involvement of one vertebral level, 286 patients (19.5%) had two vertebral level involvement, and 519 patients (35.4%) had three or more levels involved.

636 patients (44.7%) had no other bone metastases apart from the spine metastases, 442 patients (31.1%) had one to two extraspinal bone metastases, and 344 patients (24.2%) had three or more.

Visceral metastases were relatively common:while 262 patients (17.9%) had no visceral metastases, 665 patients (45.4%) had one visceral metastasis, and 539 (36.8%) had two or more visceral metastases.

Karnofsky performance status, neurological (Frankel) status, and primary tumour types are outlined in table 1.

Tokuhashi score(12)

Analysis of patients using the Tokuhashi scoring system demonstrated statistically significant differences in the actual survival for different Tokuhashi categories ($p<0.01$, figure 1, table 2), taken from the original scoring system.(12)

Bauer score(8)

The Bauer score was calculated using a conservative calculation (estimate I, in which all patients are assumed to have a pathological fracture) and optimistic calculation (estimate II, all patients

do not have a fracture). For both estimates, the score was able to discriminate between different prognostic categories ($p<0.01$, figure 2AB, table 2).

Tomita score (6)

Assessment of the Tomita scoring system revealed statistically significant differences for survival between the calculated categories ($p<0.01$, figure 3, table 2).

Van der Linden scoring system (9)

Assessment of the Van der Linden score revealed significant differences in the calculated survival of category B with respect to reference category A (, $p<0.01$, table 2). However, Van der Linden category C patients did not demonstrate a significant difference in survival, possibly due to the small numbers of patients from our dataset falling into category C (figure 4, table 2).

Rades scoring system(13)

There was a significant difference in calculated survival between the two groups according to the Rades classification ($p<0.01$, figure 5, table 2).

Bollen criteria (11)

There was a significant difference between the survival estimates for different Bollen categories ($p<0.01$, figure 6, table 2).

Harrell's C-statistic

The accuracy of prediction was determined using Harrell's C-statistic (table 3).

The two most cited scoring systems of Tokuhashi(7) and Tomita(6) were similarly accurate, although the Bollen score (11) performed slightly better with a C-statistic of 0.66 (table 3). However no scoring system achieved a good predictive value above the threshold of 0.7. The scoring systems of Van der Linden(9) and Rades(19) did not perform well using our surgical data, most likely because these systems were devised using data from radiotherapy databases, which are not generalisable to a surgically selected series.

DISCUSSION

All prognostic scoring systems, to some extent, will be subject to inaccuracies and non-generalisability depending on how the scoring system was designed. Databases are not necessarily complete population registries, but more commonly come from groups of patients who have been pre-selected for specific interventionss. This selection bias of the cohort is likely to create an over-optimistic prognostic scoring system if calculated from a surgical series, but a potentially more pessimistic outlook for when a radiotherapy cohort has been used as the baseline population.

Our results support the findings of Bollen et al.(16), that their scoring system is the most accurate prediction model compared to the other models tested in our study. This is despite differences in the populations of patients used to test the various scoring systems: Bollen et al. used a retrospective, smaller cohort, with higher mortality rate (median survival 5.1 months) and smaller numbers (1379 patients) limited to two European centres only. (16)

Comparison with other published data

Several other papers have attempted to validate published prognostic scoring systems using limited datasets. Arana et al. assessed the Tomita and modified Bauer scoring systems using a

database of 90 patients and demonstrated good intra- and inter-observer agreement but did not validate the accuracy of survival prediction itself.(20)Balain et al. reviewed the Tokuhashi, Tomita, and modified Bauer indices in a cohort of 199 patients(21), and Bartel in 110 patients(22) recommending that these systems are reliable and useful for clinical application. However, in our larger series of patients we did not find similar levels of predictive accuracy and demonstrateda lower reliability than previously thought.

All prognostic scoring systems are weighted by the influence that primary tumour type has on survival. However, in recent years we have seen improved survival with the advent of new drug treatments, immunotherapies, and more accurate radiation techniques. Therefore prognostic scoring systems that have been devised using patient data before 2005 or even 2010 are less reliable and should be used with caution. (23, 24) In general there will always be a lag-time bias, in that new treatments are often developed during the time of data acquisition, and may affect the results at later timepoints within the study.Therefore prognostic scoring systems, in general, are likely to be outdated unless they can be regularly updated or refined by current real-time data.

The Van der Linden and Rades scoring systems were devised from patient cohorts that were preselected for more palliative management than our surgical series.(9, 13) This resulted in very uneven distribution of our patients when applied to the different prognostic groupings as described in the original papers of Van der Linden and Rades, with very few patients being designated to Van der Linden group C or Rades' palliative category. Fewer patients in these worst performing categories will inevitably bias the results, and hence Harrell's C-statistic was poor for these two scoring systems when applied to our patient population.

Many previous prognostic scoring systems were designed using small numbers of patients and retrospectively collect data. Rades et al had a large database of 2029 patients(25) but all were treated with radiotherapy, and therefore the conclusions may not be generalised to a surgical patient population. By the time a patient embarks on radiotherapy, the decision to manage the patient without surgery has been already made, and therefore prognostic scoring systems are perhaps more usefully applied before this stage, to determine whether a patient is an appropriate candidate for surgery. Bollenet al. analysedpatients before thispivotal decisionwas made: whether to operate or proceed with palliative radiotherapy. They assessed a large group of 1043 patients, but retrospectively analysed. (11)

CONCLUSION

Our study demonstrates a high variability in the accuracy of the commonly cited prognostic scoring systems, which all suffer from a low concordance between predicted and actual survival, as calculated by Harrell's statistic. The Bollen and Tomita prognostic systems appear to be the most effective, but even these scoring systems did not achieve a "good" rating for accuracy and concordance.

The scores are all useful to highlight those factors which influence survival, and should be taken into consideration when determining the most appropriate management plan for a particular patient, but the scoring systems are probably not reliable enough to use the calculated scores themselves for directly predicting survival. Using numbers calculated from scoring systems to determine thresholds for appropriate treatment for a particular patient is probably not a valid use of prognostic scoring systems. These systems should be used more qualitatively rather than

quantitively, to avoid difficulties created by zones of uncertainty, ceiling effects between categories, or the freedom of patient and clinician's choice.

The prediction of survival is useful to guide treatment plans, requirements for discharge home and after surgery, and the allocation of appropriate resources. But large variations in prediction models and non-generalizable datasets make it difficult to apply data derived from cohorts to individual patients.

Qualitative assessments can guide physicians and surgeons, and particularly in cardiovascular medicine, the use of risk calculators has revolutionised clinical decision making and perhaps is the next step in the clinical application of predictive data for spinal metastases. The Qstroke score for predicting risk of ischaemic stroke(26), the Framingham risk score for coronary heart disease(27) are some other examples of risk models, which are readily available forday-to-day practice using internet-ready software applications.

Acknowledgements

Statistical support from Dr MenelaosPavlou, Prof Rumana Omar, UCL Department of Statistical Science, London, United Kingdom

References

1. Choi D, Crockard A, Bunger C, Harms J, Kawahara N, Mazel C, et al. Review of metastatic spine tumour classification and indications for surgery: the consensus statement of the Global Spine Tumour Study Group. EurSpine J. 2010;19(2):215-22.

2. Choi D, Crockard HA. How and why should we benchmark clinical outcomes and quality of life for surgery in spinal metastases? Br J Neurosurg. 2009;23(1):3-4.

3. 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(5):698-708.

4. 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(2):e0171507.

5. Verlaan JJ, Choi D, Versteeg A, Albert T, Arts M, Balabaud L, et al. Characteristics of Patients Who Survived < 3 Months or > 2 Years After Surgery for Spinal Metastases: Can We Avoid Inappropriate Patient Selection? Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2016;34(25):3054-61.

6. 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.

7. Tokuhashi Y, Ajiro Y, Umezawa N. Outcome of treatment for spinal metastases using scoring system for preoperative evaluation of prognosis. . Spine. 2008;34(1):69-73.

8. Bauer HC, Wedin R. Survival after surgery for spinal and extremity metastases. Prognostication in 241 patients. Acta Orthop Scand. 1995;66(2):143-6.

9. van der Linden YM, Dijkstra SP, Vonk EJ, Marijnen CA, Leer JW, Dutch Bone Metastasis Study G. Prediction of survival in patients with metastases in the spinal column: results based on a randomized trial of radiotherapy. Cancer. 2005;103(2):320-8.

10. Rades D, Rudat V, Veninga T, Stalpers LJ, Basic H, Karstens JH, et al. A score predicting posttreatment ambulatory status in patients irradiated for metastatic spinal cord compression. Int J Radiat Oncol Biol Phys. 2008;72(3):905-8.

11. 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-oncology. 2014;16(7):991-8.

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(19):2186-91.

13. Rades D, Evers JN, Bajrovic A, Veninga T, Karstens JH, Schild SE. Metastatic spinal cord compression: a validated survival score for elderly patients. Strahlenther Onkol. 2014;190(10):919-24.

14. 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. European Spine Journal. 2013;22(3):565-8.

15. Hernandez-Fernandez A, Velez R, Lersundi-Artamendi A, Pellise F. External validity of the Tokuhashi score in patients with vertebral metastasis. . J Cancer Res Clin Oncol. 2012;138:1493-500.

16. 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(3):E155-62.

17. Wang MY, Parr AM. An international initiative to define the role for surgery for spinal metastases: the GSTSG. World Neurosurg. 2010;73(5):426-7.

18. Harrell FE, Jr., Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med. 1996;15(4):361-87.

19. Rades D, Conde-Moreno AJ, Garcia R, Veninga T, Schild SE. A Tool to Estimate Survival of Elderly Patients Presenting with Metastatic Epidural Spinal Cord Compression (MESCC) from Cancer of Unknown Primary. Anticancer Res. 2015;35(11):6219-22.

20. Arana E, Kovacs FM, Royuela A, Asenjo B, Perez-Ramirez U, Zamora J, et al. Agreement in the assessment of metastatic spine disease using scoring systems. Radiother Oncol. 2015;115(1):135-40.

21. 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 Journal. 2013;95-B(2):210-6.

22. Bartels RH, Feuth T, Rades D, Hedlund R, Villas C, van der Linden Y, et al. External validation of a model to predict the survival of patients presenting with a spinal epidural metastasis. Cancer Metastasis Rev. 2011;30(2):153-9.

23. Gregory TM, Coriat R, Mir O. Prognostic scoring systems for spinal metastases in the era of anti-VEGF therapies. Spine (Phila Pa 1976). 2013;38(11):965-6.

24. Choi D, Bilsky M, Fehlings M, Fisher C, Gokaslan Z. Spine Oncology-Metastatic Spine Tumors. Neurosurgery. 2017;80(3S):S131-S7.

25. Rades D, Hueppe M, Schild SE. A score to identify patients with metastatic spinal cord compression who may be candidates for best supportive care. Cancer. 2013;119(4):897-903.

26. Hippisley-Cox J, Coupland C, Brindle P. Derivation and validation of QStroke score for predicting risk of ischaemic stroke in primary care and comparison with other risk scores: a prospective open cohort study. BMJ. 2013;346:f2573.

27. D'Agostino RB, Sr., Vasan RS, Pencina MJ, Wolf PA, Cobain M, Massaro JM, et al. General cardiovascular risk profile for use in primary care: the Framingham Heart Study. Circulation. 2008;117(6):743-53.

Figure legends

Figure 1. Kaplan Meier survival curves for different Tokuhashi score categories.

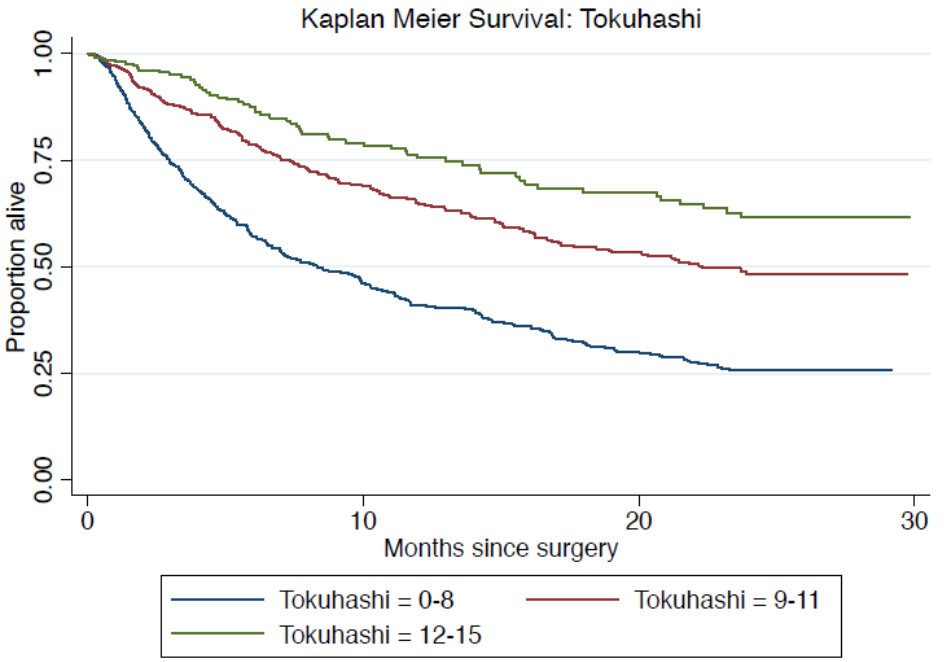

Figure 2A. Kaplan Meier survival curves for different Bauer score categories (*conservativeestimate I*).

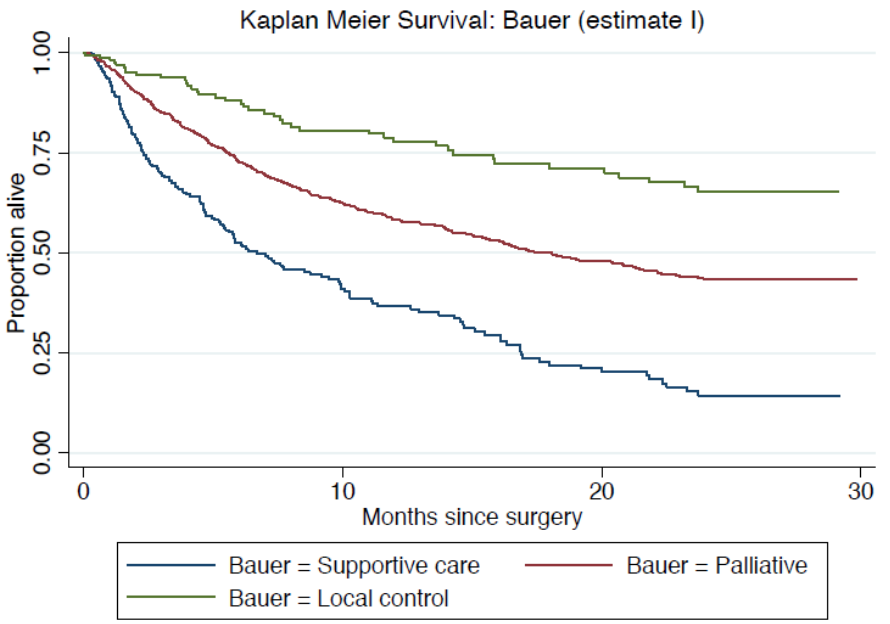

Figure 2B. Kaplan Meier survival curves for different Bauer score categories (*optimistic estimate II*).

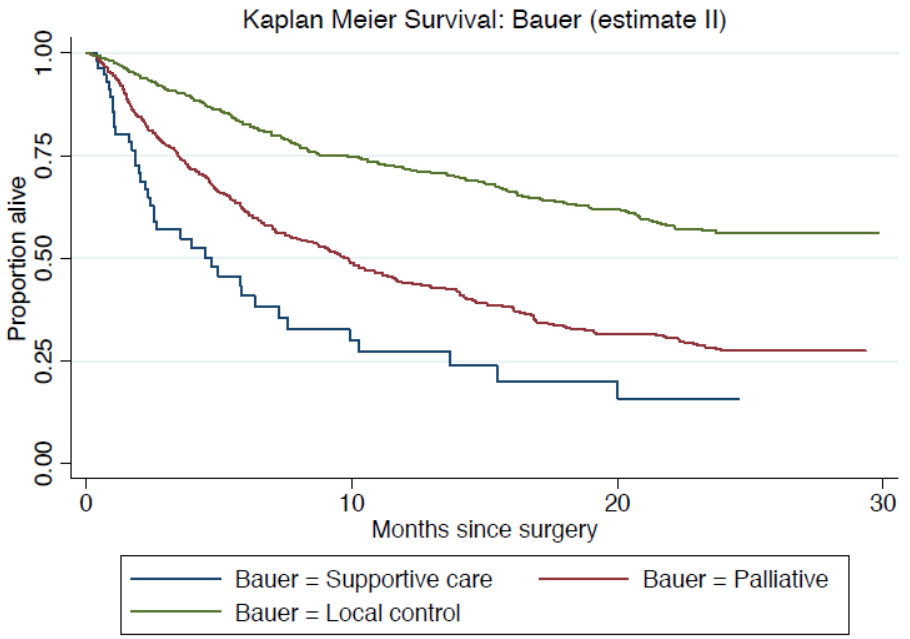

Figure 3. Kaplan Meier survival curves for different Tomita score categories.

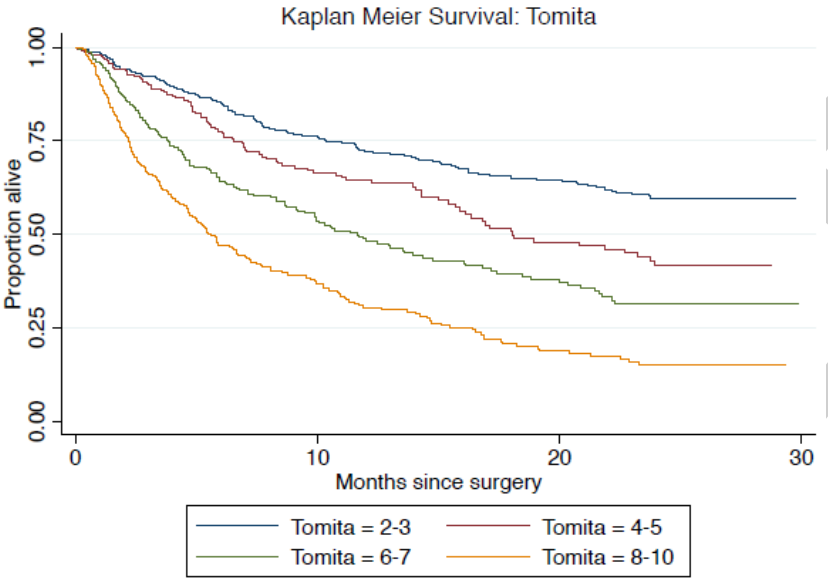

Figure 4. Kaplan Meier survival curves for different van der Linden score categories.

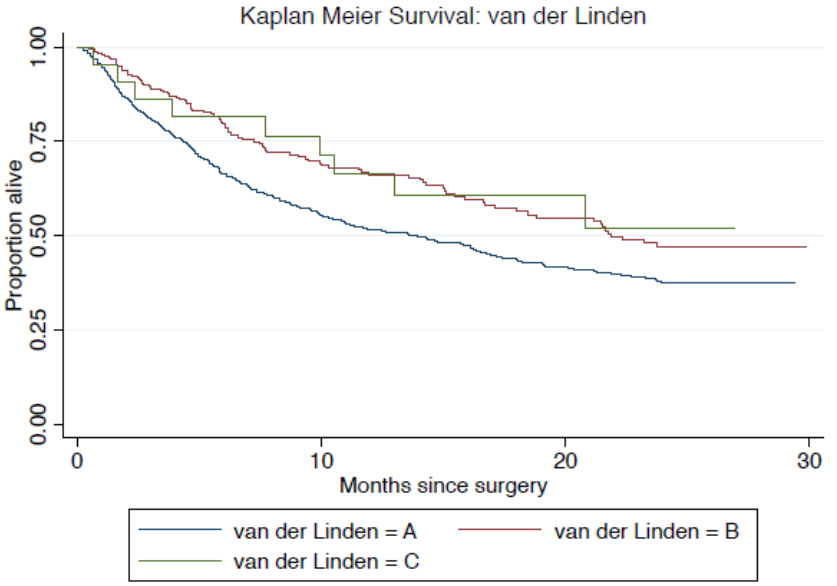

Figure 5. Kaplan Meier survival curves for different Rades score categories.

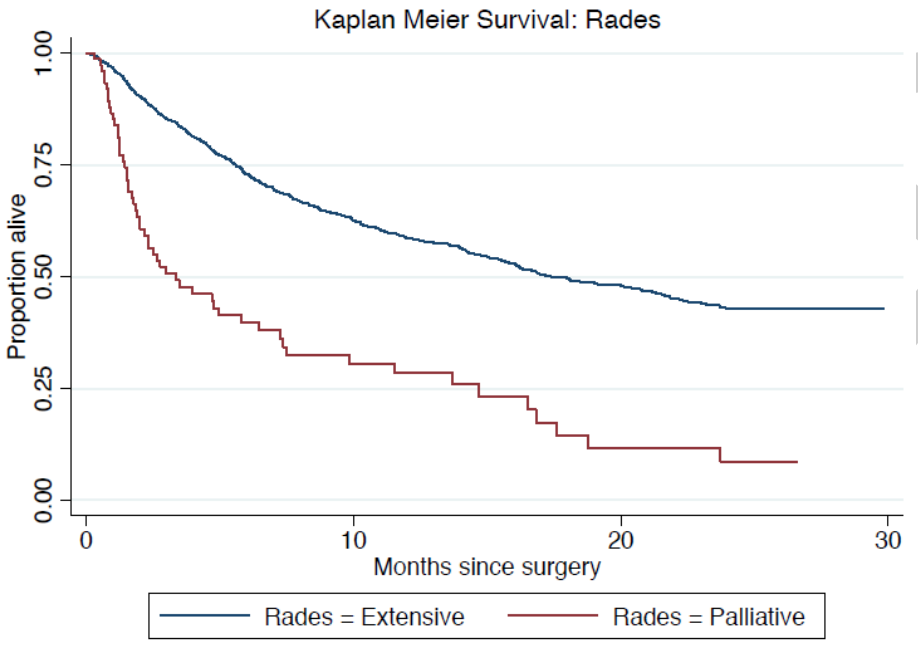

Figure 6. Kaplan Meier survival curves for different Bollen categories

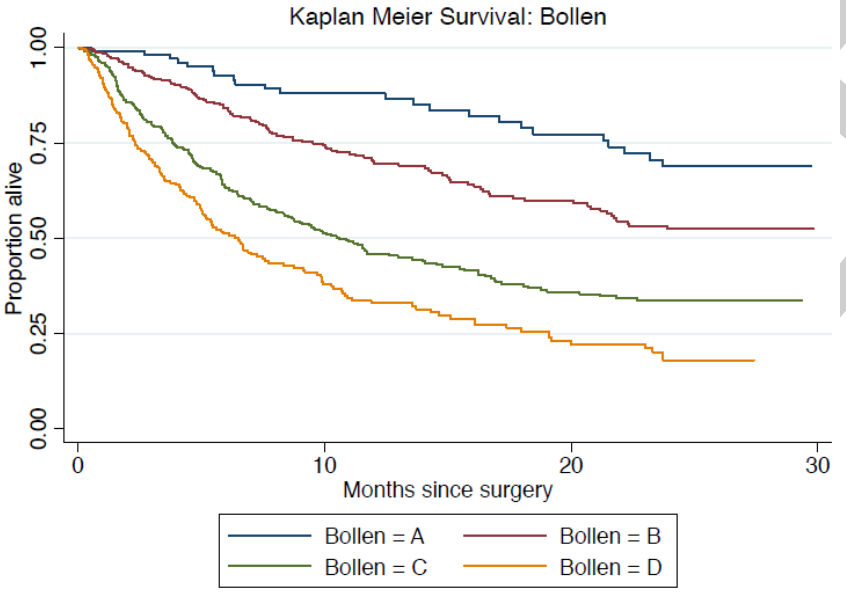

Table 1: Pre-operative clinical status and primary tumour aetiology

		Number of patients	Percentage of patients
Karnofsky	80-100	547	38.0
	50-70	678	47.2
	0-40	213	14.8
Frankel	A	23	1.6
	B	58	4.0
	C	272	18.9
	D	449	31.2
	E	638	44.3
Tumour type	Biliary	10	0.7
	Bladder	22	1.6
	Breast	238	17.1
	Cervical	14	1.0
	Colorectal	56	4.0
	Gastric	27	1.9
	Liver	48	3.5
	Lung	202	14.5

	Lymphoma	33	2.4
	Melanoma	28	2.0
	Myeloma	1	0.1
	Other	213	15.3
	Prostate	165	11.9
	Renal	169	12.2
	Sarcoma	38	2.7
	Thyroid	33	2.4
	Unknown	92	6.6

Table 2: Results from the Cox proportional hazards model for each scoring system. * denotes the reference category.

	Number of patients	Percentage of patients	HR	95%CI	p-value
Tokuhashi score category					
0-8	620	44.8	3.21	2.40; 4.24	<0.01
9-11	523	37.8	1.58	1.17; 2.14	<0.01
12-15 *	240	17.4			
Bauer score category: Estimate I (conservative)					
Supportive	273	20.2	4.41	3.12; 6.22	<0.01
Palliative	889	65.7	2.09	1.51; 2.90	<0.01
Local Control *	192	14.2			
Bauer score category: Estimate II (optimistic)					
Supportive	61	4.5	4.15	2.92; 5.90	<0.01
Palliative	625	46.2	2.38	1.98; 2.86	<0.01
Local Control *	668	49.3			
Tomita score					
2-3 *	514	36.8			
4-5	280	20.0	1.51	1.16; 1.97	<0.01
6-7	306	21.9	2.30	1.81; 2.91	<0.01
8-10	297	21.3	3.88	3.11; 4.85	<0.01

Van der Linden category					
A *	845	73.5			
B	278	24.2	0.67	0.53; 0.84	<0.01
C	27	2.3	0.64	0.32; 1.23	0.18
Rades category					
Extensive *	1268	93.8			
Palliative	84	6.2	2.95	2.23; 3.89	<0.01
Bollen category					
A *	139	11.7			
B	351	29.6	1.94	1.23; 3.03	<0.01
C	401	33.8	3.84	2.49; 5.92	<0.01
D	295	24.9	5.92	3.82; 9.18	<0.01

Table 3

Harrell's C-statistic for each prognostic scoring system and estimated survival. At 4 and 6 months survival data was calculated for the worst prognostic categories of each scoring system, and at 12 and 24 months, survival data was calculated for the patients in the best prognostic categories.

Prognostic scoring system	**Harrell's C-statistic**	**%Decease dwithin 4 months (worst prognostic groups)**	**% Deceased within 6 months (worst prognostic groups)**	**% Survival at12 months (best prognostic groups)**	**% Survival at 24 months (best prognostic groups)**
Tokuhashi	0.62	26.3	34.5	83.0	76.8
Bauer (estimate 1 all patients with fracture)	0.60	29.9	38.9	84.7	79.1
Bauer (estimate 2 no patients with fracture)	0.62	41.9	50.0	80.0	72.9
Tomita	0.65	35.1	45.0	80.7	74.8
Van der Linden	0.55	14.8	14.8	64.2	57.9
Rades	0.54	45.4	50.0	69.6	62.5

Bollen	0.66	29.6	38.5	92.1	83.5