Title:

A novel risk calculator to predict outcome after surgery for symptomatic spinal metastases; use of a large prospective patient database to personalize surgical management.

Authors

David Choi[a] FRCS, Menelaos Pavlou[b] PhD, Rumana Omar[b] PhD, Mark Arts[c] MD, Laurent Balabaud[d] MD, Jacob Maciej Buchowski[e] MD, Cody Bunger[f] MD, Chun Kee Chung[g] PhD, Maarten Hubert Coppes[h] MD, Bart Depreitere[i] MD, Michael George Fehlings[j] MD, Norio Kawahara[k] MD, Chong-Suh Lee[l] PhD, YeeLing Leung[m] FRCS, Juan Antonio Martin-Benlloch[n] MD, Eric Maurice Massicotte[j] MD, Christian Mazel MD[o], Bernhard Meyer[p] MD, Fetullah Cumhur Oner[q] MD, Wilco Peul[r] MD, Nasir Quraishi[s] FRCS, Yasuaki Tokuhashi[t] MD, Katsuro Tomita[u] MD, Christian Ulbricht[v] FRCS, Jorrit-Jan Verlaan[q] MD, Michael Wang[w] MD, Hugh Alan Crockard[a] FRCS.

a. Department of Neurosurgery, The National Hospital for Neurology and Neurosurgery, University College London, London, UK
b. Department of Statistical Science, University College London, London UK.
c. Department of Neurosurgery, Medical Center Haaglanden, Haaglanden, The Netherlands
d. Orthopaedics and Traumatology Centre, Clinique Mutualiste de la Porte de L'Orient, Lorient, France
e. Departments of Orthopedic and Neurological Surgery, Washington University, Missouri, USA
f. Department of Orthopedic Surgery, University Hospital of Aarhus, Aarhus, Denmark
g. Department of Neurosurgery, Seoul National University Hospital, Seoul National University, Seoul, Republic of Korea
h. Department of Neurosurgery, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands
i. Division of Neurosurgery, University Hospital Leuven, Leuven, Belgium
j. Division of Neurosurgery and Spinal Program, University Hospital of Toronto and Toronto Western Hospital, Toronto, Canada

k. Department of Orthopedic Surgery, Kanazawa Medical University Hospital, Kanazawa, Japan
l. Department of Orthopedic Surgery, Samsung Medical Centre, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea
m. Department of Orthopaedics, Musgrove Park Hospital, Taunton, UK
n. Spinal Unit, University Hospital Dr Peset, Valencia, Spain
o. Department of Orthopedic surgery, L'Institut Mutualiste Montsouris, Paris, France
p. TUM School of Medicine, Technische Universitat Munchen, Munich, Germany
q. Department of Orthopedic Surgery, University Medical Center Utrecht, Utrecht, The Netherlands
r. Department of Neurosurgery, Leiden University Medical Centre, Leiden, The Netherlands
s. Centre for Spine Studies and Surgery, Queens Medical Centre, Nottingham, UK
t. Department of Orthopaedic Surgery, Nihon University School of Medicine, Tokyo, Japan
u. Department of Orthopedic Surgery, Kanazawa University, Kanazawa, Japan
v. Department of Neurosurgery, Charing Cross Hospital, London, UK
w. Department of Neurosurgery, University of Miami Hospital, Miami, USA

Corresponding author:

Mr. David Choi

The National Hospital for Neurology and Neurosurgery, Queen Square, London WC1N 3BG

david.choi@nhs.net

+44 (0)845 155 5000

Acknowledgements:

This work was carried out in part at University College London Hospitals, which received funding from the Department of Health (UK) National Institute for Health Research, Biomedical Research Centre. Funding was also provided by the Global Spine Tumour Study Group (Registered charity 1134934, Charities Commission of England and Wales) and DePuy Spine (Johnson and Johnson).

Keywords:

Spine surgery, risk, survival, metastasis, tumor, outcome

ABSTRACT

AIM

Surgery for spinal metastases can improve symptoms, but sometimes complications can negate the benefits. Operations may have different indications, complexities and risks, and the choice for an individual is a tailor-made personalized decision. Previous prognostic scoring systems are becoming out of date and inaccurate. We designed a risk calculator to estimate survival after surgery, to inform clinicians and patients when making management decisions.

METHODS

A prospective cohort study was performed, including 1430 patients with spinal metastases who underwent surgery. 1264 patients from 20 centers were used for model development using a Cox frailty model. Calibration slope, D-statistic and C-index were used for model validation based on 166 patients. Follow-up was to death or minimum of 2 years after surgery. Pre-operative indices (examination findings, pain, Karnofsky physical functioning score, and radiology) were assessed.

RESULTS

An algorithm to predict survival was constructed including the tumor type, ambulatory status, analgesic use, American Society of Anesthesiologists score, number of spinal metastases, previous radiotherapy or chemotherapy, presence of visceral metastases, cervical or thoracic spine involvement, as predictors. An internet-based Risk Calculator was developed based on this algorithm, with similar or improved accuracy compared to other validated prognostic scoring systems (C-index 0·68, 95% CI 0·63-0·73, and calibration slope 1·00, 95% CI 0·68-1·32).

CONCLUSION

A large prospective surgical series of patients with symptomatic spinal metastases was used to create a validated risk calculator that can help clinicians to inform patients about the most appropriate treatment plan. The calculator is available at www.spinemet.com

INTRODUCTION

Spinal surgery can improve quality of life for patients with symptomatic spinal metastases,[1-3] but may sometimes cause complications.[4-6] Choosing the right operation for an individual is therefore paramount; increasing the complexity of surgery can improve outcome but may also increase the likelihood of complications.[6,7] Management decisions are usually guided by prognosis, but this can be difficult to estimate;[8,9] systems designed to assess prognosis have often been based on small retrospective datasets,[10,11] have poor accuracy,[12,13] or were based on non-contemporary data. To address these deficiencies, we sought to develop a risk calculator similar to those for stroke[14] and cardiovascular disease[15] that is individualized for the patient. The risk algorithm may be updated as new treatments become available later, allowing for a more accurate and contemporary risk calculator compared to commonly cited prognostic scoring systems.[13]

The Global Spine Tumour Study Group is a collaboration of surgeons studying the outcome of surgery for spine tumors (registered Charities Commission England and Wales, #1134934).[16]

METHODS

We performed a prospective cohort study of consecutive patients who were admitted for surgical treatment of symptomatic spinal metastases at 20 specialist spinal centers in Belgium, Canada, China, Denmark, France, Germany, Japan, the Netherlands, Spain, South Korea, the United Kingdom, and the United States of America. Patients were included if they provided consent for their anonymous data to be included in the secure database but were excluded if they were under 18 or unable to consent due to learning disabilities, unconsciousness, mental illness, young age, or dementia. Data were prospectively entered into an encrypted internet database and validated by surgeons at the spine centers, anonymized at point of entry. Ethical regulatory approval was granted for all centers (in the UK, NRES registration 08/H0714/44).

Patients were recruited between 1st January 2000 and September 2016. Follow-up was scheduled in the participating centers as close to three, six, 12 and 24-months after surgery, or to the date of death.

Data collection included the following:

Pre-operative tumor type, Frankel score, sphincter control, Karnofsky status, visual analogue pain score (VAS), American Society of Anesthesiologists (ASA) score, number of spinal levels affected by tumor, number of visceral metastases, extra-spinal bone metastases, ambulatory status, EQ-5D questionnaire.

Surgical data; type of operation, number of levels of tumor and fixation, intra-operative complications.

Follow-up data; post-operative complications, neurological status, sphincter control, Karnofsky score, EQ-5D, and survival.

STATISTICAL ANALYSIS AND DATA HANDLING

Statistical analyses were performed with Stata 14 software (StataCorp LP, Texas USA) and R version 3·2 (R Foundation, Vienna Austria)

Model Development

A Cox proportional hazards model was used to develop the risk prediction model.[17] The follow-up time for each patient was taken from the date of surgery to death, end of study period or last follow-up. Patients who were alive at the end of study period or who were lost to follow-up were treated as censored. Patient data were obtained from 20 centers and therefore unmeasured center characteristics may independently influence outcome. Thus, a Cox frailty model (with random effects for the center) was used to account for potential clustering by center. The random effects (frailty terms) were assumed to have a gamma distribution, in essence assigning a different baseline hazard to each center. The proportional hazards assumption was tested using Schoenfeld residuals. Pre-specified predictors were considered in the model and a backwards-elimination procedure with a 15% significance level used.

Since the objective was to develop a model that can predict the survival of patients generally, we mainly focused on marginal predictions, *i.e.* using only the estimates of the fixed regression coefficients from the frailty model and ignoring the estimates of the frailty terms.[18]

The marginal predictions for the risk of death at t years were obtained by the equation:

$$P(death\ at\ t\ years) = 1 - (S_0(t))^{\exp(risk\ score)}$$

The risk score is defined as the sum of the products of the predictors and their estimated regression coefficients and $S_0(t)$ is the baseline survival at time t.

However, we also calculated the conditional predictions, i.e. predictions which use the estimates of the frailty terms additionally from the development sample of centers. These predictions are typically applicable for risk prediction for patients in the centers in the development sample.[19]

Sample size

At least 10 events are required per estimated regression coefficient in a model to estimate the regression coefficients with adequate precision.[17] Estimation of the variance for the frailty terms requires an additional 10 events. Data from 1264 patients from 20 centers in Europe, Asia and North America were included in the development sample. A total of 491 deaths were observed within the two year follow-up period, which is adequate for the estimation of 48 regression coefficients and the variance of the frailty terms.

Model Validation

The predictive ability of the risk model was assessed at both three months and two years after surgery, based on measures of calibration and discrimination. Both internal and external validations of the model were performed. Internal validation was performed using bootstrapping with 200 bootstrap datasets. External validation was performed using the data from the London center, which was not used for model development.

Sample size for external validation

There were 100 events available for external validation at the 2-year time point, which should allow reliable estimation of the performance measures.[20] However, at the 3-month time point there were only 31 events observed, thus performance measures from the external validation of the model at 3 months should be interpreted with caution.

Measures of predictive performance

Calibration was assessed using the calibration slope (and calibration plot) and discrimination was assessed using Uno's C-index and the D-statistic.

A Calibration slope of 1 suggests perfect agreement between the observed and predicted risks, while a value less than 1 is indicative of model overfitting. If the calibration slope was less than 0·9, the regression coefficients were multiplied by a linear shrinkage factor (obtained from the internal validation) to alleviate model overfitting. The D-statistic assesses the observed separation between subjects with low and high predicted risks as predicted by the model and can be interpreted as the log hazard ratio comparing patients in the upper and lower half of the risk distribution.

Missing data

Logistic regression was used to identify the predictors of missingness and data were assumed to be missing at random. Missing predictor values were imputed using Multiple Imputation by Chained Equations (ICE)[21] using appropriate imputation models. The outcome, the Nelson-Aalen estimate of the cumulative hazard, all pre-specified potential predictors for the outcome, center (as fixed effects) and predictors of missingness were included in the imputation model. The imputation procedure was performed separately for the development and validation samples and 10 imputed datasets were produced. A 'stacked approach' was used to perform variable selection in the multiply imputed datasets.[22] The coefficient estimates for the final model were combined from the imputed datasets using Rubin's rules.[23] The performance measures were estimated in each imputed validation dataset; overall measures (optimism-adjusted for the bootstrap validation) were calculated by combining the estimates using Rubin's rules.[23]

RESULTS

A total of 491 patients died within the two year follow-up period (total follow-up time was 923 person-years), 168 patients were still alive at the end of the two years, and the remaining patients were lost to follow-up with their last status recorded as being alive (table 1).

Cause of death was directly due to the tumor in 87.6% of patients, due to surgical complications in 0.9%, other treatment complications in 0.2%, unrelated causes in 1.3% and unknown causes in 9.9% of patients.

Factors that were not found to be independent predictors of survival at the 15% level of significance were gender, age, Karnofsky score, Frankel score (since mobility was the stronger predictor), previous surgery, pain quality or intensity (which was less influential than analgesic requirement), the presence of brain metastases, or spinal metastases at the craniocervical, cervicothoracic, thoracolumbar or lumbosacral junctional levels and calendar period. Prior surgery was not included in the final model because very few patients had had previous surgery.

Table 1

We have previously analyzed and described factors which predict survival (i.e. tumor type, number of spinal metastases and visceral metastases, and Karnofsky functional status) or post-operative quality of life (i.e. Karnofsky and EQ-5D score) in an earlier cohort.[24]

The variables in Table 1 were observed to be significantly associated with survival in the Cox proportional hazard's model. Since mobility is easier to assess but has a strong association with Karnofsky status, we focused on mobility in the regression model. The estimated regression coefficients with hazards ratios and corresponding confidence intervals are shown in Table 2. The proportional hazards assumption was satisfied in each of the imputed datasets. In internal validation, the calibration slope of 0·78 at two years was indicative of moderate model overfitting and thus all regression coefficients in the prediction equation were multiplied by a shrinkage factor of 0·78. The model demonstrated reasonable discrimination with a C-index of 0·68 (95% CI 0·66-0·71) and 0·71 (95% CI 0·67-0·75) for two years and three months respectively. In external validation, the predictive ability of the model for the risk of death at two years was consistent with the results from internal validation with a C-index of 0·68 (95% CI 0·63-0·73) and a calibration slope of 1·00 (95% CI 0·68-1·31) which indicated that the linear shrinkage estimated from internal validation was effective in improving calibration.

The estimate of the C-index was higher when conditional predictions were used; 0·73 (95% CI 0·71-0·75) and 0·76 (95% CI 0·72-0·79) at 2 years and 3 months, respectively (Table 3).

Table 2

Table 3

Observed survival and survival predicted by our risk model were calculated and shown in Figures 1 and 2. These calibration plots show good agreement between observed and predicted risks of death at two years, with some underestimation of the risk for the lowest risk group.

Figure 1

Figure 2

Prediction equation

The marginal predictions for the risk of death at two years were obtained by the following equation:

$$P(death\ at\ 2\ years) = 1 - (0 \cdot 8397648)^{\exp(\text{risk score})}$$

We compared the performance of our risk model with the two most widely cited and used prediction tools described by Tomita[10] and Tokuhashi[11] (Table 4).

Table 4

DISCUSSION

We have developed a risk calculator that provides an estimate of survival, determined by specific patient characteristics at the time of presentation for surgical treatment. This risk calculator is more accurate than previously published prognostic scoring systems, and is available as an online clinical tool. The significant variables which influenced survival and were incorporated in the risk calculator were: the type of metastatic tumor; mobility; analgesic usage; presence of bone and visceral metastases; co-morbidities (ASA); previous chemotherapy or radiotherapy; and presence of tumor at cervical or thoracic spinal levels.

Methodological considerations

Clinical prediction models may not be generalizable due to selection bias, which occurs in any database. In our study, patients were more likely to represent those patients whom clinicians had decided were candidates for surgery, but may also include patients with poor prognosis.[8,9]

The most common metastatic tumors in our series were from breast, renal, lung, and prostate carcinomas and myeloma (Table 1). The validation cohort had a lower proportion of patients with lung cancer (7·8%) compared to the development cohort (15·9%), which may give rise to the differences in predicted and actual risk.

Findings of single cohort studies may not be generalizable due to center-specific bias. Although there will never be a perfect dataset for clinical prediction modelling,[25] the GSTSG database includes data from multiple international centers, minimizing bias from clustering.

Including too many variables in the prediction model will decrease the practical use of the model, and therefore the variables chosen for the model will inevitably be determined by a combination of expert opinion, previous published relevant models, clinical relevance and practicality.

Predictions of risk

Around half the patients had normal mobility, but with general comorbidities ASA 2-3. Most patients did not have lung or liver metastases at the time of presentation, however the presence of liver and lung metastases were the most influential in predicting survival, with 54% and 37% increased risk of dying if these metastases were present, respectively (Table 2).

Similar to the four most cited and accurate prognostic scoring systems of Tomita,[10] Tokuhashi,[11] Bollen,[26] and Bauer,[27] we found that the tumor type was highly associated with survival: lung carcinoma patients had a 3.2 times greater risk of death, and gastric cancer 5 times higher risk than that for a patient with breast carcinoma. The number of bone metastases and visceral metastases were also associated with poor survival. In agreement with the Tokuhashi[11] scoring system, we found that the bed-bound patient had a two-fold increased risk of death compared to a patient who was walking unaided.

In external validation, the predictive ability of the model for the risk of death at two years was satisfactory and consistent with the results from internal validation with a C-index of 0·68 (95% CI 0·63-0·73). For predicting the risk of death at two years in the centers that were included in model development, one could use the conditional predictions which provide more accurate predictions at 2 years with a C-index of 0·73 (95% CI 0·71-0·75, Table 3) as seen in internal validation. This level of predictive accuracy is slightly better than other popular cited prognostic scoring systems including Bollen,[26] Tomita,[10] and Tokuhashi[11] scores.

The online risk calculator can be accessed via the web address www.spinemet.com (Figure 3).

Figure 3

Use of the risk calculator

This risk calculator is intended to be used as a guide to clinicians, to provide an objective estimate of survival for clinical decision-making and considering surgical options for an individual. With the advent of immunomodulatory drug therapies, previous prognostic scoring systems will not accurately predict survival for renal cell carcinoma metastases,[28] whereas risk calculators can be updated to maintain usefulness, in a similar fashion to the Framingham stroke risk calculator.[15,29]

It is important to understand the limitations of the risk calculator, which should not be generalized to non-surgical cancer patients, and due to standard error may be inaccurate for some patients. If the data is presented too explicitly to patients, without simultaneous expert opinion, this may cause alarm. The European Association of Palliative Care recommends that physicians should always take into account the preferences and expectations of patients, as well as considering prognosis.[30]

CONCLUSION

In this modern era of personalized medicine, it is useful to predict the risk for an individual patient. The use of prognostic scoring systems which place patients into different categories that influence the choice of treatment is becoming outdated. We have developed an internet-calculator to estimate the risk of death in patients who are selected for surgical management of symptomatic spinal metastases, for use by experienced doctors and medical staff as an aid to patient management.

TABLES

Table 1: Summary data for the development and validation cohorts (risk factors that were included in the final model)

		Development			Validation	
Secondary Tumor type		**N**	**%**		**N**	**%**
Breast		204	16·4		32	19·2
Renal		151	12·2		20	12·1
Lung (any)		197	15·9		13	7·8
Prostate		143	11·5		24	14·5
Myeloma		110	8·9		14	8·4
Gastric		24	1·9		3	1·8
Sarcoma		30	2·4		12	7·2
Other specified		306	24·6		37	22·3
Other/unknown		77	6·2		11	6·6
Total		1,242	100		166	100
Missing		*22*	*1·7*		*0*	*0*
Mobility						
Walking normally		655	53·2		75	45·2
Walking with 1 stick/crutch		69	5·6		17	10·2
Walking with 2 sticks/crutches		135	11·0		16	9·6
Wheelchair-bound		165	13·4		13	7·8
Bed-bound		207	16·8		45	27·1
Total		1231	100		166	100
Missing		*33*	*2·6*		*0*	*0*

Bone Metastases						
0 sites		509	41·7		11	66·9
1-2 sites		400	32·8		38	22·9
>2 sites		312	25·6		17	10·2
Total		1221	100		166	100
Missing		*43*	*3·4*		*0*	*0*
Use of analgesic						
Strong opioids		516	42·0		81	48·8
Weak opioids		273	22·2		31	18·7
No regular analgesia		140	11·4		18	10·8
Non-opioid analgesia		301	24·5		36	21·7
Total		1230	100		166	100
Missing		*34*	*2·7*		*0*	*0*
ASA						
1		131	10·7		17	10·3
2		554	45·1		77	46·7
3		508	41·4		66	40·0
4-5		35	2·9		5	3·0
Total		1228	100		165	100
Missing		*36*	*2·9*		*1*	*0·6*
Other visceral metastases						
0 metastases		467	37·9		67	40·4
1 metastasis		543	44·1		67	40·4
2 or more metastases		221	18·0		32	19·3

Total		1231	100		166	100
Missing		*33*	*2·6*		*0*	*0*
Radiotherapy						
No		981	81·2		133	80·1
Yes		217	18·1		33	19·9
Total		1198	100		166	100
Missing		*66*	*5·2*		*0*	*0*
Chemotherapy						
No		996	81·6		135	81·8
Yes		225	18·4		30	18·2
Total		1221	100		165	100
Missing		*43*	*3·4*		*1*	*0·6*
Lung Metastases						
No		1022	83·0		125	75·3
Yes		209	17·0		41	24·7
Total		1231	100		166	100
Missing		*33*	*2·6*		*0*	*0*
Liver Metastases						
No		1104	89·7		145	87·3
Yes		127	10·3		21	12·7
Total		1231	100		166	100
Missing		*33*	*2·6*		*0*	*0*

Cervical spine involvement						
No		966	78·6		131	78·9
Yes		263	21·4		35	21·1
Total		1229	100		166	100
Missing		*35*	*2·8*		*0*	*0*
Lumbar spine involvement						
No		343	27·9		40	24·1
Yes		886	72·1		126	75·9
Total		1229	100		166	100
Missing		*35*	*2·8*		*0*	*0*

Table 2: Model development cohort. Calculation of Hazards ratios and confidence intervals for the different pre-operative predictors.

Risk Factor	**Coefficient**	**95% CI**		**Hazard Ratio**	**95% CI (for HR)**		**p-value**
Secondary tumor type							
Breast				1·00			<0·001
Renal	0·54	0·13	0·96	1·72	1·14	2·60	
Lung (any)	1·17	0·80	1·53	3·21	2·23	4·61	
Prostate	0·76	0·38	1·14	2·14	1·46	3·14	

Myeloma	-0·29	-0·82	0·25	0·75	0·44	1·28	
Gastric	1·61	0·98	2·23	4·98	2·66	9·30	
Sarcoma	0·90	0·28	1·52	2·46	1·32	4·57	
Other specified	0·60	0·26	0·95	1·83	1·29	2·58	
Other/unknown	0·68	0·21	1·16	1·98	1·23	3·19	
Mobility							
Walking normally				1·00			
Walking with 1 stick/crutch	0·51	0·01	1·01	1·66	1·01	2·73	<0·001
Walking with 2 sticks/crutches	0·37	0·05	0·70	1·45	1·05	2·00	
Wheelchair-bound	0·30	-0·00	0·61	1·36	1·00	1·84	
Bed-bound	0·68	0·42	0·94	1·98	1·52	2·57	
				.			
Bone Metastases							
0 sites				1·00			
1-2 sites	0·35	0·09	0·61	1·41	1·09	1·83	0·032
>2 sites	0·21	-0·05	0·48	1·23	0·95	1·61	
Use of analgesic							
No regular analgesia				1·00			
Non-opioid analgesia	0·39	-0·01	0·79	1·48	0·99	2·20	0·015
Weak Opioids	0·60	0·20	0·99	1·82	1·22	2·70	
Strong Opioids	0·58	0·20	0·96	1·79	1·22	2·62	
ASA							
1				1·000			0·003
2	0·66	0·22	1·10	1·94	1·25	3·01	

3	0·76	0·31	1·21	2·15	1·37	3·37	
4 and 5	1·15	0·48	1·83	3·16	1·61	6·21	
Clinical Status							
0 metastases				1·00			
1 metastasis	0·30	0·05	0·55	1·35	1·05	1·73	0·062
2 or more metastases	0·34	-0·03	0·72	1·41	0·97	2·05	
Previous spine radiotherapy	0·21	-0·04	0·46	1·24	0·96	1·59	0·100
Previous chemotherapy	0·28	0·03	0·53	1·32	1·03	1·69	0·028
Presence of liver metastases	0·43	0·11	0·75	1·54	1·11	2·13	0·010
Presence of lung metastases	0·32	0·03	0·60	1·37	1·03	1·83	0·029
Cervical spine involvement	0·34	0·10	0·58	1·41	1·11	1·78	0·005
Lumbar spine involvement	0·21	-0·04	0·46	1·23	0·96	1·59	0·099

Table 3: Results from internal validation.

Internal validation (Bootstrap) Marginal predictions at 2 years		
C-index	0·68 (95% CI 0·66-0·71)	
Calibration Slope	0·78 (95% CI 0·67-0·88)	
D-statistic	1·11 (95% CI -1·20-3·41)	

Internal validation (Bootstrap) Marginal predictions at 3 months		
C-index	0·71 (95% CI 0·67-0·75)	
Calibration Slope	0·86 (95% CI 0·69-1·03)	
D-statistic	1·21 (95% CI -1·09-3·51)	
Internal validation (Bootstrap) Conditional predictions at 2 years		
C-index	0·73 (95% CI 0·71-0·75)	
Calibration Slope	0·87 (95% CI 0·78-0·97)	
D-statistic	1·40 (95% CI -1·68-4·48)	
Internal Validation (Bootstrap) Conditional predictions at 3 months		
C-index	0·76 (95% CI 0·72-0·79)	
Calibration Slope	0·92 (95% CI 0·77-1·06)	
D-statistic	1·50 (95% CI -1·70-4·70)	

Table 4: External validation (using the London center), survival at 2 years. 95% confidence intervals are represented in brackets.

(A shrinkage factor of 0.78 obtained from the internal validation of the model has been applied to the regression coefficients)

External validation (London center) - 2 years	**Our Score**	**Tokuhashi**	**Tomita**

C-index (95% CI)	0·68 (0·63-0·73)	0·66 (0·60-0·71)	0·52 (0·45-0·60)
Calibration Slope (95% CI)	1·01 (0·68-1·31)		
D-statistic (95%CI)	1·06 (0·72-1·40)	0·87 (0·55-1·21)	0·16 (-0·19-0·51)

FIGURE LEGENDS

Figure 1:

Groups are defined according to the predicted risk at 2 years, using the coefficients estimated in Cox regression and the Nelson Aalen estimate of the baseline survival at two years. The observed risk at 2 years is then obtained from the Kaplan-Meier survival curve restricted to the patients in each risk group. Number of patients per risk group at 2 years: 25, 33, 36, 72 for groups 1, 2, 3 and 4 respectively.

Figure 2:

Groups are defined according to the predicted risk at 3 months, using the coefficients estimated in Cox regression and the Nelson Aalen estimate of the baseline survival at two years. Number of patients per risk group at 3 months: 20, 58, 63, 25 for groups 1, 2, 3 and 4 respectively.

Figure 3:

Snapshot of the online risk calculator

CONFLICT OF INTEREST STATEMENT

Dr Arts holds stock in Nuvasive, Stryker, Galapagos and Pharming. He holds a consulting role with Amedica, Zimmer Biomet, Silony and EIT.

Dr. Buchowski reports grants and non-financial support from AO Spine North America, grants and non-financial support from OMeGA, other support from Globus Medical, other support from Medtronic, other support from Globus Medical, other support from Orthofix, outside the submitted work. In addition, Dr. Buchowski has a patent Globus Medical issued, a patent K2M issued, and a patent Wolters Kluwer Health with royalties paid.

Dr Choi reports other funding from Global Spine Tumour Study Group and DePuy Spine during the conduct of the study; other funding from Department of Health, UK, outside of submitted work.

Dr. Fehlings reports a Consulting or Advisory Role with Pfizer, Zimmer Biomet and InVivo Therapeutics.

Dr. Massicotte reports personal fees from AO Spine North America, outside the submitted work.

Dr. Mazel reports other support from DePuy Spine, other support from Medtronic, other support from Zimmer, other support from Clariance, outside the submitted work. In addition, Dr. Mazel has a patent CHD issued and Stock ownership is held with Amplitude; a consulting role is held with Ethicon.

Dr. Meyer reports grants, personal fees and non-financial support from Medtronic; personal fees and non-financial support from Depuy; grants, personal fees and non-financial support from Ulrich Medical; grants, personal fees and non-financial support from Relievant; grants, personal fees and non-financial support from Icotec; personal fees and non-financial support from Brainlab; non-financial support and other from Nexstim; personal fees, non-financial support and other from Spineart, outside the submitted work.

Dr. Quraishi reports personal fees from AO Spine, personal fees from Medtronic and personal fees from DePuy Synthes, outside the submitted work.

Dr. Verlaan reports grants from DePuy Synthes, outside the submitted work; and Dr Verlaan holds an advisory role with DePuy Synthes.

Dr. Wang reports grants from Department of Defence, other support from Depuy Spine, Aesculap Spine, JoiMax, K2M, other support from Innovative Surgical Devices, Spinicity, other support from Children's Hospital of Los Angeles, Depuy Spine, Springer Publishing, Quality Medical Publishing, outside the submitted work; In addition, Dr. Wang has a patent Childrens Hospital of LA issued, a patent DePuy Spine issued, a patent Springer Publishing issued, and a patent Quality Medical Publishing issued and Dr Wang holds stocks in Innovative Surgical Devices and Spinicity. He also holds a consulting role with DePuy Spine, AesculapSpine, Jointax and K2M.

All other authors have nothing to disclose.

REFERENCES

1 Patchell RA, Tibbs PA, Regine WF, et al. Direct decompressive surgical resection in the treatment of spinal cord compression caused by metastatic cancer: a randomised trial. *Lancet* 2005; **366**: 643–8.

2 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**: 337–44.

3 Fehlings MG, Nater A, Tetreault L, et al. Survival and Clinical Outcomes in Surgically Treated Patients With Metastatic Epidural Spinal Cord Compression: Results of the Prospective Multicenter AOSpine Study. Journal of clinical oncology : official journal of the American Society of Clinical Oncology 2016; 34(3): 268-76.

4 Dea N, Versteeg A, Fisher C, et al. Adverse events in emergency oncological spine surgery: a prospective analysis. *J Neurosurg Spine* 2014; **21**: 698–703.

5 Jansson KA, Bauer HC. Survival, complications and outcome in 282 patients operated for neurological deficit due to thoracic or lumbar spinal metastases. *Eur Spine J* 2006; **15**: 196–202.

6 Lee BH, Park JO, Kim HS, Park YC, Lee HM, Moon SH. Perioperative complication and surgical outcome in patients with spine metastases: retrospective 200-case series in a single institute. *Clin Neurol Neurosurg* 2014; **122**: 80–6.

7 Boriani S, Gasbarrini A, Bandiera S, Ghermandi R, Lador R. En Bloc Resections in the Spine: The Experience of 220 Patients During 25 Years. *World Neurosurg* 2017; **98**: 217–29.

8 Glare P, Virik K, Jones M, et al. A systematic review of physicians' survival predictions in terminally ill cancer patients. *BMJ* 2003; **327**: 195–8.

9 Chow E, Davis L, Panzarella T, et al. Accuracy of survival prediction by palliative radiation oncologists. *Int J Radiat Oncol Biol Phys* 2005; **61**: 870–3.

10 Tomita K, Kawahara N, Kobayashi T, Yoshida A, Murakami H, Akamaru T. Surgical strategy for spinal metastases. *Spine* 2001; **26**: 298–306.

11 Tokuhashi Y, Matsuzaki H, Oda H, Oshima M, Ryu J. A revised scoring system for preoperative evaluation of metastatic spine tumor prognosis. *Spine* 2005; **30**: 2186–91.

12 Bollen L, Wibmer C, Van der Linden YM, et al. Predictive Value of Six Prognostic Scoring Systems for Spinal Bone Metastases: An Analysis Based on 1379 Patients. *Spine* 2016; **41**: E155–62.

13 Choi D, Ricciardi F, Arts M, et al. Prediction accuracy of common prognostic scoring systems for metastatic spine disease: results of a prospective international multicenter study of 1469 patients. **Spine** (in press)

14 Hippisley-Cox J, Coupland C, Vinogradova Y, et al. Predicting cardiovascular risk in England and Wales: prospective derivation and validation of QRISK2. *BMJ* 2008; **336**: 1475–82.

15 D'Agostino RB, Sr., Vasan RS, Pencina MJ, et al. General cardiovascular risk profile for use in primary care: the Framingham Heart Study. *Circulation* 2008; **117**: 743–53.

16 Choi D, Crockard A, Bunger 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**: 215–22.

17 Peduzzi P, Concato J, Feinstein AR, Holford TR. Importance of events per independent variable in proportional hazards regression analysis. II. Accuracy and precision of regression estimates. *J Clin Epidemiol* 1995; **48**: 1503–10.

18 Skrondal A, Rabe-Hesketh S. Multilevel modelling. SAGE: Los Angeles; 2010.

19 Skrondal A, Rabe-Hesketh S. Prediction in multilevel generalized linear models. *J R Stat Soc*; **172**; 659–87.

20 Collins GS, Ogundimu EO, Altman DG. Sample size considerations for the external validation of a multivariable prognostic model: a resampling study. *Stat Med* 2016; **35**: 214–26.

21 van Buuren S, Boshuizen HC, Knook DL. Multiple imputation of missing blood pressure covariates in survival analysis. *Stat Med* 1999; **18**: 681–94.

22 Wood AM, White IR, Royston P. How should variable selection be performed with multiply imputed data? *Stat med* 2008; **27**: 3227–46.

23 Rubin DB. Multiple Imputation for Nonresponse in Surveys. John Wiley & Sons, Inc.: New York; 2008.

24 Choi D, Fox Z, Albert T, 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. *Neurosurg* 2015; **77**: 698–708.

25 Lee YH, Bang H, Kim DJ. How to Establish Clinical Prediction Models. *Endocrinol metab* 2016; **31**: 38–44.

26 Bollen L, van der Linden YM, Pondaag W, 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**: 991–8.

27 Bauer HC, Wedin R. Survival after surgery for spinal and extremity metastases. Prognostication in 241 patients. *Acta Orthop Scand* 1995; **66**: 143–6.

28 Gregory TM, Coriat R, Mir O. Prognostic scoring systems for spinal metastases in the era of anti-VEGF therapies. *Spine* 2013; **38**: 965–6.

29 Dufouil C, Beiser A, McLure LA, et al. Revised Framingham Stroke Risk Profile to Reflect Temporal Trends. *Circulation* 2017; **135**: 1145–59.

30 Maltoni M, Caraceni A, Brunelli C, et al. Prognostic factors in advanced cancer patients: evidence-based clinical recommendations--a study by the Steering Committee of the European Association for Palliative Care. *JCO* 2005; **23**: 6240–8.