Prediction of life expectancy in patients with spinal epidural metastasis

Ronald H.M.A. Bartels, Godard de Ruiter, Ton Feuth, and Mark P. Arts

Department of Neurosurgery, Radboud University Medical Center, Nijmegen, Netherlands (R.H.M.A.B.); Department of Neurosurgery, The Haaglanden Medical Center, The Hague, Netherlands (G.d.R., M.P.A.); Department for Health Evidence, Radboud University Medical Center, Nijmegen, Netherlands (T.F.)

Corresponding Author: Ronald H.M.A. Bartels, MD, PhD, Radboud university medical center, Department of Neurosurgery, Geert Grote Plein-Zuid 10, 6525 GA Nijmegen, The Netherlands (ronald.bartels@radboudumc.nl).

Background. The treatment of spinal epidural metastasis is multidisciplinary and usually involves a team of medical oncologists, radiologists, radiotherapists, and spinal surgeons. Life expectancy is one of the factors considered when deciding whether surgery is warranted. Because expert estimates of life expectancy are generally not reliable, a prediction model is needed. Here, we temporally validated a model that was previously validated geographically.

Methods. The records of 110 consecutive patients who were referred with a spinal epidural metastasis were collected prospectively from 2009 to 2013 in order to validate the model, which was published in 2011. The actual and estimated life expectancies were represented graphically, and calibration and discrimination were determined. The calibration slope, Harrell’s c-index, D, and $R^2_D$ were calculated. Hazard ratios in the derivation set of 2011 were compared with the validation set. Misspecification was determined using the joint test for $b^*$.

Results. The calibration slope was $0.64 \pm 0.15$ (95% CI: 0.34–0.94), Harrell’s c-index was 0.72, D was 1.08, and $R^2_D$ was 0.22, indicating slightly worse discrimination in the derivation set. The joint test for $b^* = 0$ was statistically significant and indicated misspecification; however, this misspecification was attributed entirely to the surgical group.

Conclusions. We validated a prediction model for surgical decision making, showing that the model’s overall performance is good. Based on these results, this model will help clinicians to decide whether to offer surgery to patients with spinal epidural metastasis.

Keywords: life expectancy, prediction, spinal epidural metastasis, treatment.
who were admitted to the neurosurgical department of the Radboud University Medical Center and Canisius Wilhelmina Hospital were collected prospectively (87 patients); in addition, data from all patients with spinal epidural metastases who were referred in 2012 to the neurosurgical department at The Haaglanden Medical Center in The Hague were collected prospectively (23 patients).

Patients with missing data were excluded from the study. In addition to the patients’ baseline characteristics, the following features were documented: Karnofsky performance score (in case of a sudden [<24 h] onset of severe spinal cord lesion [paraparesis], the score was estimated just before the deterioration); the curative intention to treat the primary tumor; the nature of the metastasis; and the spinal level. The original model did not include the instituted therapy as a predictor; however, this was recorded as well. Patients were treated using either radiotherapy or surgery (decompression of the neural structures and spinal stabilization) followed by radiotherapy. These data were used to validate the prediction model published in 2011.10 The data were truncated at 10 months, as this contributed to maintenance of the proportional hazards assumption and was also done in the original model.

The Cox model was validated using the test population as described by Royston and Altman. First, predictive ability was evaluated graphically by plotting the actual survival as a Kaplan–Meier curve and predicted survival as the mean of the predicted Cox survival curves. Next, predictive ability was calculated and expressed as Harrell’s c-index (discrimination) and the Royston–Sauerbrei D-statistic, R^2_D (calibration). The calibration slope—expressed as the regression coefficient on the prognostic index (PI)—was also calculated. The distributions of the PI in the derivation set and in the current validation set were compared and statistically tested using the Mann–Whitney U-test. Model misspecification/fit was checked using a joint test of β^* = 0. The values of β^* are differences between the regression coefficients estimated in the derivation set and those estimated in the model fitted to the validation set. The hazard ratios for the various predictors were compared against the hazard ratios of the derivation set. The predictive ability was also shown by Kaplan–Meier curves of 2 other groups of patients: those with a predicted median survival of 3 months or more, and those with a predicted median survival of less than 3 months. Survival is expressed as the median in months (range: minimum–maximum). Other values are expressed as the mean ± standard error, with 95% confidence intervals. Differences were considered to be statistically significant if P < .05.

Results

All 110 consecutive patients in the indicated time period had complete records and were included. Therefore, none of the patients had missing data. At the time of the analysis (October 2014), 90 patients were deceased, and the median survival time was 5.7 months (range: 0.3–68.3 mo); 18.1% of the patients had censored survival times.

Graphically, the estimated survival corresponded well with the patients’ actual survival (Fig. 1). The calibration slope was 0.64 ± 0.15 (95% CI: 0.34–0.94), indicating poorer discrimination in the validation set than in the derivation set. The c-index, D-statistic, and R^2_D values from the derivation and validation sets are summarized in Table 1. The joint test for β^* was 0 (P = .0003), indicating significant misspecification.

Next, the validation group was divided into patients who underwent surgery followed by radiotherapy (58 patients) and patients who received radiotherapy only (52 patients). With these 2
subgroups, the joint test for $\beta^* = 0$ was significant for the surgery group ($P = .001$) but nonsignificant for the nonsurgery group ($P = .52$). Thus, misspecification did not occur in the group that did not undergo surgery (Figs 2 and 3). This phenomenon also occurred when evaluating discrimination (Table 1). Only 6 patients died within 3 months of presentation in the surgery group. The hazard ratios are summarized in Table 2. The distributions of the PI in the derivation set and in the validation set are represented in Table 3. Statistical difference was not present ($P = .58$). The Kaplan–Meier curves after dichotomizing for estimated median survival ($<3$ vs $\geq3$ mon) are shown in Fig. 4. The model predicted accurately for those with an estimated median survival of 3 months or more. The other group performed slightly less well: predicted survival was worse than observed survival.

**Discussion**

Estimating the survival of patients with epidural metastasis is essential for identifying patients' individual treatment options. Aside from other factors, in the decision to perform surgery, an estimated survival of $\geq3$ months is generally considered to be acceptable. Given that the opinion of experts is generally unreliable, a validated prediction model would support the decision to recommend surgery.1

The current model has been developed as an adjunct in the decision to offer surgery or not. Therefore, the threshold of 3 months was important. In the original model, the data were truncated at 10 months, since the proportional hazards assumption had to be met.9 This was also done in this validation process. After previous external validation, the model was slightly adapted.10 This final version developed on data of more than 500 patients is now being validated.

Other models have been described. Tokuhashi and colleagues12 described a model that required information about (i) general condition, (ii) number of extraspinal bone metastases, (iii) number of metastases in the vertebral body, (iv) metastases to the major internal organs (lungs, liver, kidneys, and brain), (v) primary site of the cancer, and (vi) severity of spinal cord palsy. It was merely used to estimate which kind of surgery should be performed.10 The revised version13 seemed to perform modestly.14

Another well-known model was presented by Tomita et al.15 The authors used data from surgically treated patients, which introduced a selection of patients. It was constructed merely to define the type of surgery and not the survival. A search for other metastases was also needed. Furthermore, one of the predicting factors was the malignancy of the primary, which was related to the growth speed, which could be slow, moderate, or rapid. This model was not appropriate for estimating survival for a patient presenting with an epidural metastasis.

In 2005 van der Linden et al16 published a prediction model. The model was built based on data from a strictly defined population. Patients who had metastasis in the cervical spine, a pathologic fracture or compression of the cord, a renal cell carcinoma, or melanoma were excluded. This restricted its use in

---

**Table 2.** Summary of hazard ratios (HR) of the derivation (2011) and validation (2014) sets

<table>
<thead>
<tr>
<th>Predictor</th>
<th>2011 HR (95% CI)</th>
<th>2014 HR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender (female vs male)</td>
<td>0.62 (0.49–0.79)</td>
<td>0.90 (0.51–1.58)</td>
</tr>
<tr>
<td>Lung carcinoma</td>
<td>1.89 (1.4–2.56)</td>
<td>1.23 (0.52–2.95)</td>
</tr>
<tr>
<td>Kidney carcinoma</td>
<td>2.52 (1.64–3.87)</td>
<td>0.31 (0.04–2.47)</td>
</tr>
<tr>
<td>Other carcinoma</td>
<td>1.76 (1.31–2.26)</td>
<td>0.75 (0.36–1.57)</td>
</tr>
<tr>
<td>Curative treatment of primary</td>
<td>0.69 (0.54–0.89)</td>
<td>0.41 (0.21–0.79)</td>
</tr>
<tr>
<td>Cervical location of metastasis</td>
<td>2.32 (1.68–3.19)</td>
<td>1.47 (0.79–2.74)</td>
</tr>
<tr>
<td>KPS: 10–20</td>
<td>80.92 (33.26–196.77)</td>
<td>–</td>
</tr>
<tr>
<td>KPS: 30–40</td>
<td>10.12 (5.32–19.25)</td>
<td>8.36 (3.16–22.07)</td>
</tr>
<tr>
<td>KPS: 50–70</td>
<td>5.23 (2.83–9.67)</td>
<td>1.82 (0.73–4.59)</td>
</tr>
<tr>
<td>KPS: 80</td>
<td>3.84 (1.95–7.53)</td>
<td>1.30 (0.51–3.33)</td>
</tr>
</tbody>
</table>

**Table 3.** Distribution of prognostic index in derivation set and validation set

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Minimum</th>
<th>Q25&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Median</th>
<th>Q&lt;75&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Maximum</th>
<th>Skewness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Derivation N = 567</td>
<td>−0.8</td>
<td>1.2</td>
<td>1.7</td>
<td>2.3</td>
<td>5.9</td>
<td>0.07</td>
</tr>
<tr>
<td>Validation N = 110</td>
<td>−0.8</td>
<td>0.8</td>
<td>1.7</td>
<td>2.3</td>
<td>3.8</td>
<td>−0.24</td>
</tr>
</tbody>
</table>

<sup>a</sup>25% quartile.<br><sup>b</sup>75% quartile.
oncologic practice. It also required a search for visceral metastases.

Another model has recently been described by Bollen et al.6 The information needed was type of primary cancer; performance status; presence of visceral, brain, and bone metastases; number and location of spinal metastases; and neurological functioning. Performance status was assessed with KPS and neurological functioning with the Frankel scale. It resulted in 4 categories of different survival. However, none of the above-mentioned models has ever been validated like the one we report in this paper.

The major advantage of the current model is its simplicity. Only 5 factors have to be known: gender, histology (kidney cancer, breast/prostate cancer, lung cancer or other), whether the primary was curatively treated, cervical location of the symptomatic metastasis, and KPS. Extensive radiological examinations of different parts of the body are not warranted, such as CT of thorax/abdomen or nuclear scans. Therefore, an estimation can be made within minutes if the primary is known (most instances). Discussions about the experience of individual physicians will not be necessary.

The performance of the model is good, as the c and $R^2$ values were nearly the same in the validation set, with only a slight reduction, which was to be expected.11 The calibration of the model presented was also good.

Discrimination was slightly poorer in the validation set than in the derivation set. Discrimination in the nonsurgical group was good, whereas discrimination in the surgically treated group was worse; the same pattern applies to misspecification/fit. This misfit might be due to a relatively small number of patients in the separate groups in relation to number of predictors; moreover, the model was developed solely to predict survival in order to optimize the selection of surgical candidates. Therefore, the instituted therapy was not introduced as a separate indicator, given that most patients underwent radiotherapy only. Furthermore, introducing surgery would be a complicated exercise, given that the surgical options are quite diverse and can include kyphoplasty but also total vertebrectomy, and many more options as well. In practice, the type of surgical intervention is determined only after the patient is found to be a suitable candidate for surgery.

With respect to the patients who underwent surgery, the finding that actual survival was slightly better than estimated...
survival is consistent with a study by Patchell et al.\(^3\) The relatively low percentage of patients (6.9\%) who died within 3 months of surgery in the Nijmegen population justified the use of the presented prediction model. Thus, for patients referred to hospitals in the Nijmegen region, an estimated life expectancy of <3 months was an important determining factor for offering radiotherapy only. Overall, although the model predicted the minimum survival time, we must emphasize that other factors—such as the patient’s wishes, histology of the metastasis, radiosensitivity, surgical accessibility, and duration of neurological deficits—should also be considered.

The small sample size might be considered a flaw. Overfitting is a problem when the number of variables are not in balance with the sample size. It might be a cause for the difference in the effect of predictors. However, we did not have any problems with overfitting in the development of the prediction model validated in the current study. The original model was not adjusted based on the data from the validation set, and so overfitting was not an issue. Sample size can also contribute to a difference in variation between variables in different datasets. Since the distribution of the PI is comparable in the original set and in the validation set, any difference was not relevant for clinical use. The difference in the Kaplan–Meier curve and the mean of the predicted survival curves for patients with a predicted median survival of <3 months could also be attributed to the sample size for this group.

In conclusion, this model will help to optimize the treatment options for individual patients who have spinal epidural metastasis and are treated in a multidisciplinary setting. Finally, the model provides a reliable estimate of life expectancy based on easily retrievable data (www.nccn.nl/nccn-env/).

**Funding**

None declared.

**Acknowledgment**

We are indebted to Bert Keurentjes, Radboud in’to Languages, Center of Expertise for language and communication, for correcting English grammar and style.

**Conflict of interest statement.** None declared.

**References**


