

External validation of a model to predict the survival of patients presenting with a spinal epidural metastasis

Ronald H. M. A. Bartels; Ton Feuth; Dirk Rades; Rune Hedlund; Carlos Villas; Yvette van der Linden; Wolfgang Börm; Arnoud Kappelle; Richard W. M. van der Maazen; J. André Grotenhuis; André L. M. Verbeek

R. H. M. A. Bartels & J. A. Grotenhuis

Department of Neurosurgery, Radboud University Nijmegen Medical Centre,
R. Postlaan 4, 6500 HB, Nijmegen, The Netherlands

e-mail: r.bartels@nch.umcn.nl

T. Feuth & A. L. M. Verbeek

Department of Epidemiology, Biostatistics, and Health Technology Assessment,
Radboud University, Nijmegen Medical Centre, Nijmegen, The Netherlands

D. Rades

Department of Radiation Oncology, University of Luebeck, Luebeck, Germany

R. Hedlund

Department of Orthopedics, Sahlgrenska University Hospital, Gothenburg, Sweden

C. Villas

Department of Orthopaedic Surgery and Traumatology, University of Navarra,
Pamplona, Spain

Y. van der Linden

Radiotherapeutic Institute Friesland, Leeuwarden, The Netherlands

W. Börm

Diakonissenanstalt, Neurosurgical Clinic, Flensburg, Germany

A. Kappelle

Department of Neurology, Radboud University Nijmegen Medical Centre,
Nijmegen, The Netherlands

R. W. M. van der Maazen

Department of Radiation Oncology, Radboud University, Nijmegen Medical Centre,
Nijmegen, The Netherlands

ABSTRACT

The surgical treatment of spinal metastases is evolving. The major problem is the selection of patients who may benefit from surgical treatment. One of the criteria is an expected survival of at least 3 months. A prediction model has been previously developed. The present study has been performed in order to validate externally the model and to demonstrate that this model can be generalized to other institutions and other countries than the Netherlands. Data of 356 patients from five centers in Germany, Spain, Sweden, and the Netherlands who were treated for metastatic epidural spinal cord compression were collected. Hazard ratios in the test population corresponded with those of the developmental population. However, the observed and the expected survival were different. Analysis revealed that the baseline hazard function was significantly different. This tempted us to combine the data and develop a new prediction model. Estimating iteratively, a baseline hazard was composed. An adapted prediction model is presented. External validation of a prediction model revealed a difference in expected survival, although the relative contribution of the specific hazard ratios was the same as in the developmental population. This study emphasized the need to check the baseline hazard function in external validation. A new model has been developed using an estimated baseline hazard.

KEYWORDS

Spinal metastases; Epidural compression; Surgical treatment; Prediction model

1. INTRODUCTION

The surgical treatment of spinal surgery is rapidly evolving. New insights contributed to this development. To be a surgical candidate, an expected survival of at least 3 months apart from other factors is mandatory [1].

Recently, a model to predict the survival of patients presenting with a spinal epidural metastasis has been developed [2]. The main purpose was to evaluate whether patients could be surgical candidates. Although the model was internally validated, external validation is recommended to demonstrate adequate, predictive performance in new patients in other hospitals, and also other countries [3, 4]. Therefore, this external, clinical validation including centers from several European countries was performed.

1.1 PATIENTS AND METHODS

Since the main interest of the study was testing the performance in groups of patients other than the study population, the model was geographically validated [5]. Several centers dealing with spine pathology in Europe were addressed. The centers were asked to provide data from a consecutive series of patients treated for spinal extradural metastasis within the same period as the original study population was recruited from. A computer-based form was offered. The following items were requested: date of birth,

gender, site of primary cancer, date of diagnosis of primary, curative intention of treatment of primary, date of diagnosis of spinal metastasis, location of spinal metastasis-producing symptoms, metastasis confirmed by biopsy, Karnofsky Performance Score (KPS), instituted treatment, and date of death.

1.2 STATISTICAL ANALYSIS

Since in the development of the original model survival data were truncated at 10 months to maintain the proportional hazard assumption, and the model is used to predict survival at 3 months, survival data are again truncated at 10 months. To evaluate predictive discrimination, the concordance (C) index was calculated [6]. The generalized $R^2_{\text{Nagelkerke}}$ as a measure of predictive ability [6] that can range from 0 (poor) to 1 (perfect) was used.

Missing data were managed in two ways. First, the cases with missing data were not included in the analysis, performing a complete case analysis. Second, multiple imputation was used for missing values. Logistic regression on the event indicator was used. Both sets of data were used for analysis, and presented. Heterogeneity between centers considering survival was tested using a log-rank test.

Validation of the original model was done in several ways: fitting a Cox model on the test population, using the same predictors as did the original model, and then comparing the new estimates of the regression coefficients and associated hazard ratios (HRs) to those of the original model. Also, comparison of the new and original baseline hazard functions took place, using a method previously described [7, 8].

The observed survivals and the predicted survivals are graphically presented. The baseline survival function was checked for correctness [7]. The survival for patients predicted by the original model in the test population was compared to the observed survival. For this purpose, the test population was divided in quartiles considering the predicted linear part of the regression equation using the original model. In each quartile, the actual Kaplan–Meier estimate of the survival curve was compared with the averaged Cox-predicted survival curves. Ideally, these should be identical.

To update the model, a new Cox proportional hazards model was constructed combining the test population and the original population using the same predictors. As already mentioned, missing data were imputed, and the Nelson–Aalen estimator of the cumulative hazard $H_0(T)$ was used as were other covariates for imputing a particular binary predictor that was fully missing in one of the test sites [9]. $H_0(T)$ was estimated iteratively in two cycles as has been described previously [9].

Imputation was performed five times, and by pooling the results, averaged estimates of the regression coefficients along with a variance–covariance matrix of the averaged estimates, adjusted for variability due to imputation, were obtained [10]. To assess the performance of this updated model, the median $R^2_{\text{Nagelkerke}}$ and median concordance index c of the five models obtained after imputation were calculated. To be able to calculate and draw survival curves and associated 95% confidence bands, the survival curves resulting from the different imputations were averaged.

1.3 SAMPLE SIZE

Although recommended for the development for a new model [10], 20 patients at least per factor were included totaling at least 140 patients.

2. RESULTS

Five international centers responded to the invitation to participate. These were located in the Netherlands, Sweden, Germany (two centers), and Spain. The patients were treated between January 1996 and December 2008. Sixteen patients were diagnosed with spinal metastasis in 2008.

A total number of 356 patients were included. Males outnumbered females: 217 versus 139. Two hundred eighty-eight (80.9%) patients were dead. When the database was closed in April 2009, from eight patients was not known whether they were alive or not. The other patients were either dead or alive at least 3 months after the diagnosis of spinal metastasis. The median survival of the patients who died within the study period was 6.1 ± 17.7 months. The location of the primary tumors was shown in Table 1. Thirty-four patients underwent surgery prior to radiation therapy. Sixteen patients underwent a combined treatment of surgery, chemotherapy, and radiation therapy. The remaining 306 patients received radiation therapy solely. In 124 (34.8%) patients, the diagnosis was confirmed histologically by a biopsy of the vertebral lesion.

Of one center, it was not known whether the primary cancer was treated with curative intention. Therefore, in 60 patients, the data on curative intention of the treatment of the primary were missing. For these, multiple imputation was used.

In Table 2, the estimated coefficient in the final proportional hazards model is represented based on the data of the test population including the data that have been multiple imputed. Since only two patients were included with a KPS of 10–20, the estimated hazard ratio was extremely high and did not contribute to the estimation of the expected survival time. Furthermore, the 95% CI was very wide. Therefore, these patients were excluded from further analysis in the test population.

Comparing observed survival time with the predicted survival time resulted in differences (Fig. 1). Analysis of the correctness of the baseline survival function showed a significant difference between original population and the validation population (Fig. 2).

The hazards of the two populations were disproportionate. This is a remarkable and unexpected finding. The hazard ratios using the data of the test population were comparable with the hazard ratios from the original model. Also, their relative contribution within the respective models was nearly equal.

To meet the goal of a prediction model that could be used in many centers, a new model has been developed combining the data of the original and test population. All five imputed prediction models were pooled.

The model was built on the data of 567 patients. The results are represented in Table 3. Examples of the prediction model are shown in Fig. 3 and Table 4. Because of the pooling, $R^2_{\text{Nagelkerke}}$ and C index cannot be determined in a standard fashion. However, based on five imputations using the Nelson–Aalen estimator, median $R^2_{\text{Nagelkerke}}$ is 0.318 (minimum 0.314, maximum 0.320), and median C index is 0.719 (minimum 0.718, maximum 0.721). The model that can be consulted on the web (www.nccn.nl) has been adapted accordingly.

3. DISCUSSION

One of the most important factors that has to be considered by the physician in order to tailor a palliative treatment to the individual patient with spinal metastasis is the patient's survival prognosis. It has been shown that an estimation based on clinical grounds is inaccurate [11, 12]. Recently, this inaccuracy has been demonstrated in a phase II trial. One of the inclusion criteria was a life expectancy of at least 2 years determined by the investigator. More than half of the patients did not survive as long as 12 months [13]. Decision support is one of the major purposes of a prediction model [6].

Therefore, models to estimate the survival of patients presenting with a spinal metastasis have been developed. The prediction for survival is, for example, important to choose between long—or short—course radiotherapy [14]. The predicted survival apart from other factors also determines whether a patient will be a surgical candidate [1].

A model to predict survival of patients in order to decide for surgery was published previously [2]. Although internally validated, transportability to other cohorts of patients was not proven. Therefore, an external validation is warranted. During this procedure, the model should be kept as simple as it was, and the urge to improve the model by adding predictive factors should be repressed [15].

Whether the survival of patients can be influenced by surgery is still matter of research. In the study by Patchell et al., in selected cases, only a small difference was found [16]. The major advantage of surgery was improvement of the quality of life. It can be argued whether survival can be improved by removing a metastatic lesion in the spine. Firstly, it cannot be removed with adequate margins. Secondly, a metastatic lesion of the spine is an expression of systemic disease that will not be influenced by removing one small part of it. Therefore, a correction for surgical intervention was not done.

During the external validation, the hazard ratios were nearly in the same range as in the original model. Also, the internal relationship of the hazard ratios did not differ significantly.

However, major differences occurred comparing the observed survival time and the expected survival time. The expected survival was in all cases too restrictive. Evaluation of the possible cause learned that the baseline hazard of the validation population was completely different. Many explanations are possible. The genetic constitution and therefore the influence of a malignancy could be different in the two populations. Race was not incorporated in the predictors. We do not believe this is of influence since all

patients are living in Western Europe. Another possibility was that treatments were different. We doubt this is the right explanation since most of the patients were only treated by radiation therapy. As was proven by Patchell [16], surgery did not contribute to a major difference in survival. The selection of not all relevant predictors in the original set of data is also a possibility. However, selection was done according to commonly accepted procedures. Bootstrapping confirmed the internal validity of the model [2].

After interpretation of the results, two options were considered. Firstly, the model should not be changed, since we demonstrated that the predictors in the original model were adequate, but the baseline hazard function was dissimilar. However, this would interfere with generalizability, and our underlying goal. The model was not meant to be used only in Nijmegen.

Secondly, the model could be adapted combining the data of the two populations. In our opinion, this was appropriate since the relative contribution of the predictors was nearly the same in both populations. Only the baseline hazard was different. Therefore, the combined data were used to develop a new model. The major advantage was that it was constructed from a larger and different population. The relative contribution of the hazard ratios was again nearly the same as in the original model. However, the 95% CI are more narrow as can be expected since the number of patients was larger. Especially, considering $R^2_{\text{Nagelkerke}}$ and C index, we are convinced that the current model more accurately predicts the survival in different geographical settings.

Our study has some limitations. Although the study population is retrieved from 1996 till 2008, development of new medicines, alternative radiation schemes, or more aggressive surgical intervention may have contributed to an improved survival. However, the study by Patchell did not show a major contribution of surgery to survival [16]. Multiple imputation was used for missing values. This enabled us to use data on other predictors. It has been shown that the used method of multiple imputation did provide a lower bias to the null than older ones as, for example, the multiple imputation based on log T [9]. Growing evidence is available supporting the use of multiple imputation instead of complete case analysis [17]. However, more complete data would have been preferable.

Currently, prospectively data are collected to validate further the prediction model externally. These will be separately presented. This study proofed that external validation was obligatory before using a prediction model clinically. In our opinion, this was the first time that a difference in baseline hazards had been shown during an external validation process. Future investigations should pay explicitly attention to the baseline hazards of the original and test population.

In conclusion, an externally validated and adapted model is presented to predict the survival of patients presenting with spinal metastasis. The model is especially appealing since knowledge of only five clinical parameters is necessary. Furthermore, this model can be used in multiple countries. Incorporation of this model in daily practice will certainly contribute to an improvement of the quality of the care for the patients with spinal metastatic disease

Disclaimer: None

REFERENCES

1. Bartels, R. H. M. A., van der Linden, Y. M., & van der Graaf, W. T. A. (2008). Spinal extradural metastasis: review of current treatment options. *CA: A Cancer Journal for Clinicians*, 58, 245–259.
2. Bartels, R. H. M. A., Feuth, T., van der Maazen, R., et al. (2007). A model to predict the life expectancy of patients with spinal epidural metastasis. *Cancer*, 110, 2042–2049.
3. Moons, K. G. M., Royston, P., Vergouwe, Y., Grobbee, D. E., & Altman, D. G. (2009). Research methods & reporting. Prognosis and prognostic research: what, why, and how? *BMJ*, 338, b375.
4. Altman, D. G., & Royston, P. (2000). What do we mean by validating a prognostic model? *Statistics in Medicine*, 19, 453–473.
5. Steyerberg, E. W. (2009). Validation of prediction models. 1, 299–311.
6. Steyerberg, E. W., Vickers, A. J., Cook, N. R., et al. (2010). Assessing the performance of prediction models. A framework for traditional and novel measures. *Epidemiology*, 21, 128–138.
7. van Houwelingen, H. C. (2010). Validation, calibration, revision and combination of prognostic survival models. *Statistics in Medicine*, 19, 3401–3415.
8. van Houwelingen, H. C., & Thorogood, J. (1995). Construction, validation and updating of a prognostic model for kidney graft survival. *Statistics in Medicine*, 14, 1999–2008.
9. White, I. R., & Royston, P. (2009). Imputing missing covariate values for the Cox model. *Statistics in Medicine*, 28, 1982–1998.
10. Harrell Jr, F. E. (2001). *Regression modeling strategies. With applications to linear models, logistic regression, and survival analysis*. 1,
11. Chow, E., Davis, L., Panzarella, T., et al. (2005). Accuracy of survival prediction by palliative radiation oncologists. *International Journal of Radiation Oncology, Biology, Physics*, 61, 870–873.
12. Chow, E., Harth, T., Hruby, G., Finkelstein, J., Wu, J., & Danjoux, C. (2001). How accurate are physicians' clinical prediction of survival and the available prognostic tools in estimating survival times in terminally ill cancer patients? A systematic review. *Clinical Oncology*, 13, 209–218.
13. Amdur, R. J., Bennett, J., Olivier, K., et al. (2009). A prospective, phase II study demonstrating the potential value and limitation of radiosurgery for spine metastases. *American Journal of Clinical Oncology*, 32, 515–520.
14. Rades, D., Dunst, J., & Schild, S. E. (2008). The first score predicting overall survival in patients with metastatic spinal cord compression. *Cancer*, 112, 157–161.
15. van Houwelingen, H. C. (2000). Validation, calibration, revision and combination of prognostic survival models. *Statistics in Medicine*, 19, 3401–3415.
16. Patchell, R. A., Tibbs, P. A., Regine, W. F., et al. (2005). Direct decompressive surgical resection in the treatment of spinal cord compression caused by metastatic cancer: a randomised trial. *Lancet*, 366, 643–648.
17. Steyerberg, E. W., Mushkudiani, N., Perel, P., et al. (2008). Predicting outcome after traumatic brain injury: development and international validation of prognostic scores based on admission characteristics. *PLoS Medicine*, 5, e165. doi:10.1371/journal.pmed.0050165-

		Gender		Total
		Female	Male	
Location	Prostate	0	61	61
	Breast	64	3	67
	Lung	23	66	89
	Kidney	7	17	24
	Other	44	71	115
Total		138	218	356

Table 2. Comparison of hazard ratios estimating an unfavorable outcome when using the data from the original population, the test population without imputation, and also the test population after multiple imputation for the missing data of curatively intended treatment of the primary tumor.

	Estimated HR in original model	Estimated HR in test population without imputation	Estimated HR in test population after multiple imputation
Number of patients	219	286	346
Variable			
Gender (female versus male)	0.669	0.572	0.616
Lung carcinoma ^a	2.232	2.788	2.391
Kidney carcinoma ^a	2.798	2.116	2.449
Other carcinoma ^a	2.052	1.595	1.680
Curatively intended treatment of primary tumor ^b	0.675	0.790	0.709
Cervical location of spinal metastasis ^c	1.630	1.528	1.676
KPS (10–20)	31.455	–	–
KPS (30–40)	7.687	5.599	7.275
KPS (50–70)	5.670	4.475	4.627
KPS (80)	4.446	2.143	2.041

HR hazard ratios, 95% CL 95% confidence limits, KPS Karnofsky Performance Score.

For the test population, it is assumed that the baseline hazard is the same as in the original population

^a The HR was calculated compared with patients with breast or prostate carcinoma.

^b This is a true or false expression. It is true when considering histology, location, and extension of the primary tumor (including metastases), the initial treatment is aimed at cure. When the treatment is palliative, the expression is false.

^c Although in other parts of the spine metastases can be present, the symptomatic metastases should be present in the cervical spine.

Table 3. Hazard ratios estimating an unfavorable outcome when using the combined data from the original and test population

Variable	Estimated coefficient in original proportional hazards model	HR	95% CL for the HR	P
Gender (female versus male)	-0.473	0.623	0.491–0.79	<0.0001
Lung carcinoma ^a	0.639	1.894	1.401–2.562	<0.0001
Kidney carcinoma ^a	0.924	2.520	1.642–3.868	<0.0001
Other carcinoma ^a	0.546	1.725	1.319–2.257	<0.0001
Curatively intended treatment of primary tumor ^b	-0.368	0.692	0.540–0.887	0.004
Cervical location of spinal metastasis ^c	0.841	2.318	1.684–3.189	<0.0001
KPS (10–20)	4.393	80.918	33.275–196.774	<0.0001
KPS (30–40)	2.314	10.117	5.318–19.246	<0.0001
KPS (50–70)	1.654	5.228	2.826–9.671	<0.0001
KPS (80)	1.344	3.836	1.954–7.529	<0.0001

HR hazard ratios, 95% CL 95% confidence limits, KPS Karnofsky Performance Score.

The total number of patients is 567.

^a The HR was calculated compared with patients with breast or prostate carcinoma.

^b This is a true or false expression. It is true when considering histology, location, and extension of the primary tumor (including metastases), the initial treatment is aimed at cure. When the treatment is palliative, the expression is false.

^c Although in other parts of the spine metastases can be present, the symptomatic metastases should be present in the cervical spine.

Table 4. Prediction of survival at 3 months and after 6 months after the first diagnosis of spinal metastasis using the model developed combining the original data and the data of the patients collected for external validation

Gender	Primary tumor treated curatively	Primary tumor	KPS, range	Cervical location	Survival at 3 month	Survival at 6 month
F	+	Breast	50–70%	–	0.87 (0.83–0.91)	0.75 (0.69–0.81)
F	–	Breast	50–70%	–	0.81 (0.76–0.87)	0.66 (0.59–0.74)
F	+	Breast	50–70%	+	0.72 (0.63–0.82)	0.51 (0.39–0.66)
F	–	Breast	50–70%	+	0.62 (0.52–0.74)	0.38 (0.27–0.54)
F	+	Breast	80%	–	0.90 (0.86–0.94)	0.81 (0.74–0.88)
F	–	Breast	80%	–	0.86 (0.81–0.91)	0.74 (0.65–0.83)
F	+	Breast	80%	+	0.79 (0.70–0.88)	0.61 (0.49–0.77)
F	–	Breast	80%	+	0.71 (0.60–0.83)	0.49 (0.36–0.67)
M	+	Prostate	50–70%	–	0.80 (0.74–0.85)	0.63 (0.55–0.72)
M	–	Prostate	50–70%	–	0.72 (0.66–0.79)	0.51 (0.43–0.61)
M	+	Prostate	50–70%	+	0.59 (0.47–0.74)	0.34 (0.22–0.53)
M	–	Prostate	50–70%	+	0.47 (0.35–0.63)	0.21 (0.12–0.39)
M	+	Prostate	80%	–	0.85 (0.79–0.91)	0.71 (0.62–0.81)
M	–	Prostate	80%	–	0.79 (0.72–0.86)	0.61 (0.51–0.73)
M	+	Prostate	80%	+	0.68 (0.56–0.82)	0.46 (0.31–0.67)
M	–	Prostate	80%	+	0.57 (0.44–0.74)	0.32 (0.19–0.54)
M	+	Kidney	80%	–	0.66 (0.53–0.81)	0.43 (0.28–0.65)
M	–	Kidney	80%	–	0.54 (0.40–0.74)	0.29 (0.16–0.54)
M	+	Kidney	80%	+	0.38 (0.21–0.66)	0.14 (0.04–0.43)
M	–	Kidney	80%	+	0.24 (0.11–0.54)	0.06 (0.01–0.29)

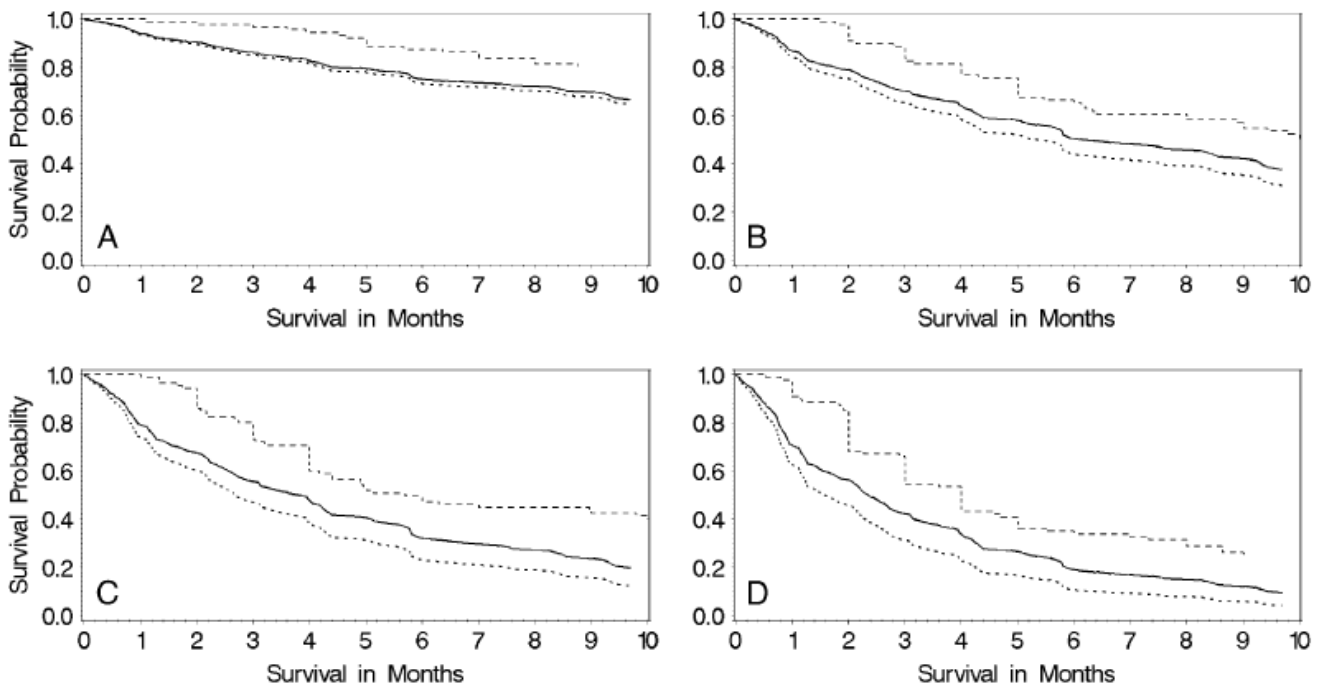


Figure 1. A difference is shown between the observed (dashed line) and the expected survival without shrinkage (dotted line) and with shrinkage (solid line). The data were divided in quartiles based on the Predictive Index. Those with the best prognosis are represented in **a** and subsequently in **b**, **c**, and **d**

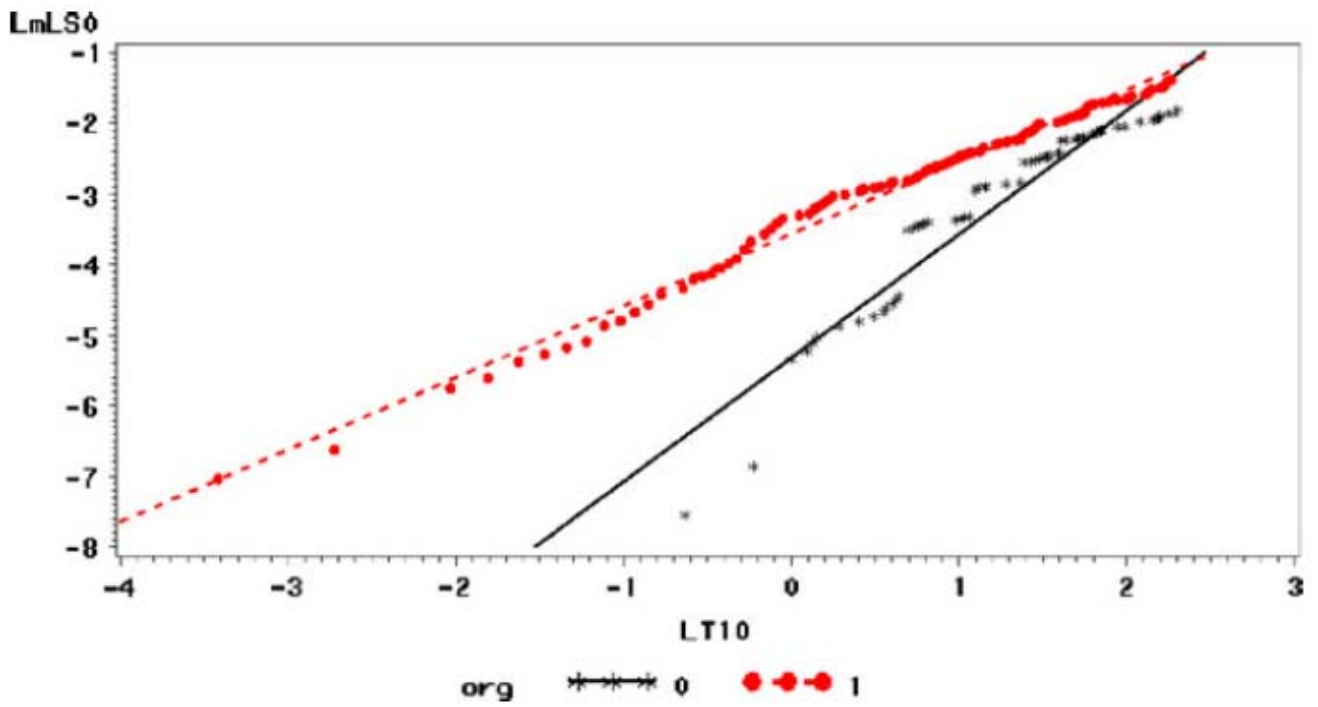


Figure 2. Direct comparison of log cumulative baseline hazard function disclosing a significant difference in slope ($P < 0.0001$).

***Original population.

●●●Test population

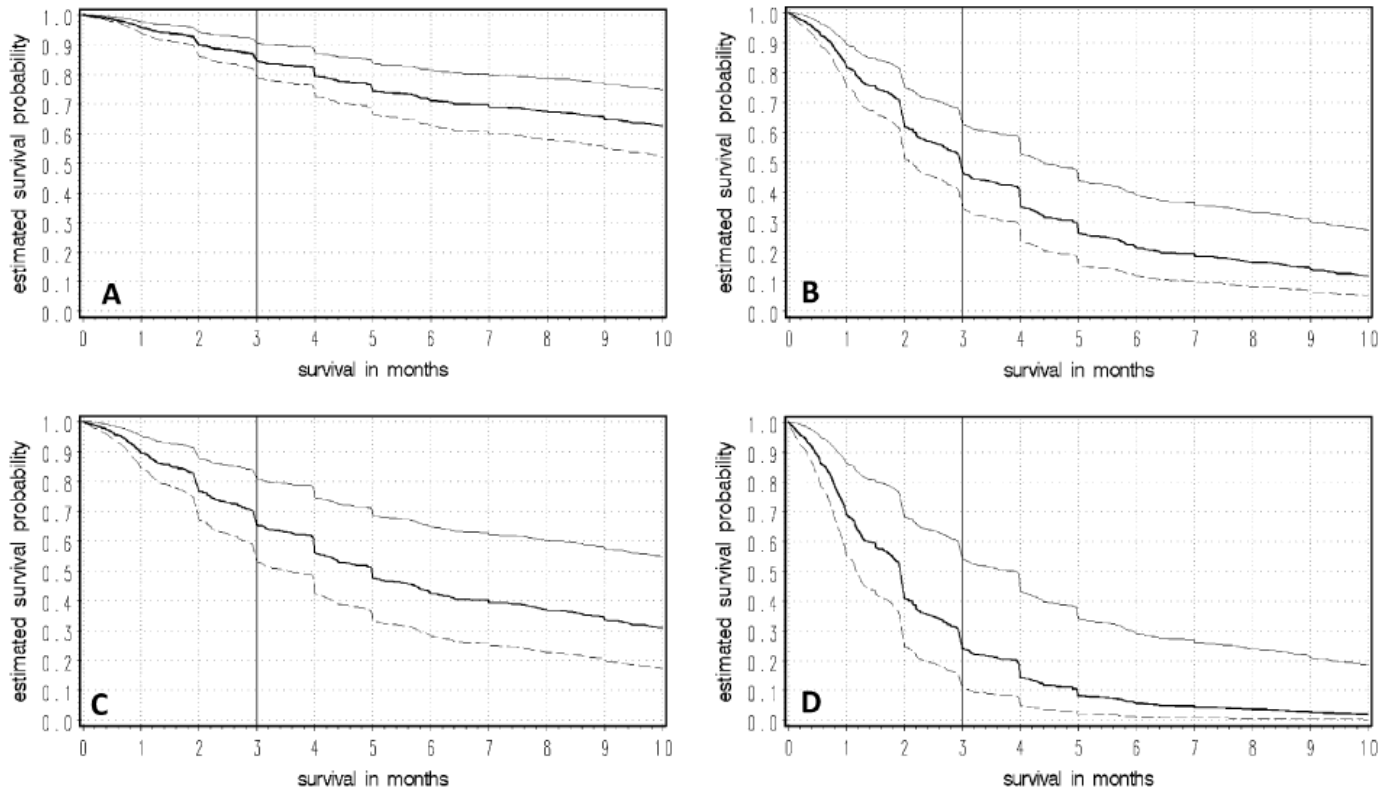


Figure 3. Predicted survivals of patients with the same characteristics as previous published [2]. **a** Prostate carcinoma, KPS 80%, no cervical location, and curative intention of primary. **b** As a except KPS 50–70% and cervical metastasis. **c** Renal cell carcinoma, KPS 80%, noncervical, curative intention treatment of primary. **d** Same as c except cervical location and noncurative intention treatment primary.