Radiotherapy and Oncology 123 (2017) 189–194

Contents lists available at ScienceDirect

Radiotherapy and Oncology

journal homepage: www.thegreenjournal.com

Brain metastases

Individualized early death and long-term survival prediction after stereotactic radiosurgery for brain metastases of non-small cell lung cancer: Two externally validated nomograms

Jaap D. Zindler [a,*], Arthur Jochems [a], Frank J. Lagerwaard [b], Rosemarijne Beumer [a], Esther G.C. Troost [a,c], Daniëlle B.P. Eekers [a], Inge Compter [a], Peter-Paul van der Toorn [d], Marion Essers [e], Bing Oei [e], Coen W. Hurkmans [d], Anna M.E. Bruynzeel [b], Geert Bosmans [a], Ans Swinnen [a], Ralph T.H. Leijenaar [a], Philippe Lambin [a]

[a] Department of Radiation Oncology (MAASTRO Clinic), GROW School for Oncology and Developmental Biology, Maastricht University Medical Centre; [b] Department of Radiation Oncology, VU University Medical Center, Amsterdam, The Netherlands; [c] Institute of Radiooncology, Helmholtz-Zentrum Dresden-Rossendorf, Germany; [d] Department of Radiation Oncology, Catharina Hospital Eindhoven; and [e] Department of Radiation Oncology, Verbeeten Institute, Tilburg, The Netherlands

ARTICLE INFO

Article history:
Received 25 August 2016
Received in revised form 24 December 2016
Accepted 5 February 2017
Available online 23 February 2017

Keywords:
Individualized brain metastases
Stereotactic radiosurgery
Prognostic models

ABSTRACT

Introduction: Commonly used clinical models for survival prediction after stereotactic radiosurgery (SRS) for brain metastases (BMs) are limited by the lack of individual risk scores and disproportionate prognostic groups. In this study, two nomograms were developed to overcome these limitations.

Methods: 495 patients with BMs of NSCLC treated with SRS for a limited number of BMs in four Dutch radiation oncology centers were identified and divided in a training cohort ($n = 214$, patients treated in one hospital) and an external validation cohort $n = 281$, patients treated in three other hospitals). Using the training cohort, nomograms were developed for prediction of early death (<3 months) and long-term survival (>12 months) with prognostic factors for survival. Accuracy of prediction was defined as the area under the curve (AUC) by receiver operating characteristics analysis for prediction of early death and long term survival. The accuracy of the nomograms was also tested in the external validation cohort.

Results: Prognostic factors for survival were: WHO performance status, presence of extracranial metastases, age, GTV largest BM, and gender. Number of brain metastases and primary tumor control were not prognostic factors for survival. In the external validation cohort, the nomogram predicted early death statistically significantly better ($p < 0.05$) than the unfavorable groups of the RPA, DS-GPA, GGS, SIR, and Rades 2015 (AUC = 0.70 *versus* range AUCs = 0.51–0.60 respectively). With an AUC of 0.67, the other nomogram predicted 1 year survival statistically significantly better ($p < 0.05$) than the favorable groups of four models (range AUCs = 0.57–0.61), except for the SIR (AUC = 0.64, $p = 0.34$). The models are available on www.predictcancer.org.

Conclusion: The nomograms predicted early death and long-term survival more accurately than commonly used prognostic scores after SRS for a limited number of BMs of NSCLC. Moreover these nomograms enable individualized probability assessment and are easy into use in routine clinical practice.

Stereotactic Radiosurgery (SRS) is an established treatment for a limited number of brain metastases (BMs) with a maximum diameter up to 4 cm [1]. To predict survival in BM patients, several prognostic models have been published in the past decades [2–4]. The most commonly used is the Recursive Partitioning Analysis (RPA), which is a relatively simple scoring system, initially developed in patients who were treated with whole brain radiotherapy (WBRT), and subsequently validated for other treatment modalities [5]. RPA classification takes into account age, presence of extracranial metastases, primary tumor control, and performance status. The RPA divides the patient cohort into three prognostic categories; however, a major disadvantage of the RPA is that approximately two-third of patients suitable for SRS will fall in the intermediate prognostic class, and probabilities for both short and long-term survival are group-based and not individualized [2]. Lack of individualized survival probability and

* Corresponding author at: MAASTRO Clinic, P.O. Box 3035, 6202 NA Maastricht, The Netherlands.
E-mail address: jaap.zindler@maastro.nl (J.D. Zindler).

http://dx.doi.org/10.1016/j.radonc.2017.02.006
0167-8140/

disproportional size of prognostic groups were also observed in other more recently published prognostic models for survival, such as the Golden Grading System (GGS), Disease-Specific Graded Prognostic Assessment (DS-GPA), Score Index for Radiosurgery in brain metastases (SIR), and Rades 2015 [2,6–12]. With nomograms, however, it is possible to assess individualized probabilities for endpoints, and relevant prognostic factors can be evaluated. In this study, two validated nomograms were developed for the prediction of early death (<3 months) and long-term (>1 year) survival of patients treated with SRS for a maximum of four BMs of NSCLC. The rationales for these endpoints were that (1) accurate prediction of early death can be relevant for SRS patient selection, and (2) accurate prediction of long-term survival can be particularly useful for the choice of either radical or palliative treatment of extracranial disease [13,14].

Materials and methods

Data

This multicenter cohort study was approved by the local institutional review board of MAASTRO clinic and registered at ClinicalTrials.gov (NCT02265549). Clinical data were collected from all patients with newly diagnosed BMs treated with linear accelerator-based SRS between December 2002 and March 2015 in four participating Dutch Radiation Oncology centers: MAASTRO clinic in Maastricht (MC), VU University medical center (VUmc) in Amsterdam, Verbeeten Institute in Tilburg (VT), and Catharina Hospital in Eindhoven (CZE). Patients were generally eligible for SRS if they had a maximum of three BMs, with a maximum diameter of 4 cm, on diagnostic magnetic resonance imaging (MRI) performed by the referring hospital. Prior to treatment, a contrast enhanced high-resolution MRI serving radiation planning purposes was performed with three-dimensional distortion correction. If a fourth BM was identified on this planning-MRI, three of the four participating centers also treated these patients with SRS as the single treatment modality. The gross tumor volume (GTV) was defined as the contrast enhancement on the planning-MRI. An isotropic margin of 1–3 mm was used to generate the planning target volume (PTV) [15]. SRS dose was prescribed at the PTV in the range of 15–24 Gy in one to three fractions. Treatment planning in VUmc and CZE have been described previously [2,15]. MC used iPlan (Brainlab AG, Feldkirchen, Germany) and Eclipse (Varian, Palo Alto) software, and treatment planning was performed with non-coplanar dynamic conformal arcs or coplanar volumetric modulated arc therapy (VMAT). At VT, the XiO software (Elekta, Stockholm, Sweden) was used for treatment planning, which was accomplished with a non-coplanar static arcs technique or VMAT. During follow-up, MRI scans were acquired every three months; an outpatient visit was planned if both the physical and mental conditions of the patient allowed it.

Variable selection

A database was available of all patients treated with SRS for newly diagnosed brain metastases of several primary tumors ($n = 929$) in four Dutch hospitals. For this study, patients with BM of NSCLC from whom the date of death was known, or patients with BM of NSCLC who had a follow-up of at least of 1 year were selected ($n = 495$). In the training cohort ($n = 214$) Kaplan–Meier analysis including multivariate Cox regression analysis was performed on the baseline characteristics to identify significant prognostic factors for survival. Dependent prognostic factors were excluded from the multivariate analysis: PTV largest BM is dependent on GTV largest BM; cumulative GTV is dependent on GTV largest metastasis; and dose is dependent on GTV largest BM. In the training cohort, the following baseline characteristics were statistically significant prognostic factors for survival in multivariate cox regression analysis: WHO performance status ($p < 0.01$, beta regression coefficient (β) = 0.41, odds ratio (OR) = 1.50, 95% confidence interval (95% CI) = 1.20–1.88), presence of extracranial metastases ($p < 0.01$, $\beta = 0.73$, OR = 2.08, 95% CI = 1.44–3.00), age ($p < 0.01$, $\beta = 0.03$, OR = 1.03, 95% CI = 1.02–1.05), GTV largest BM ($p = 0.01$, $\beta = 0.03$, OR = 1.03, 95% CI = 1.01–1.06), and gender ($p = 0.04$, $\beta = -0.35$, OR = 0.70, 95% CI 0.51–0.98); Other baseline characteristics were not prognostic for survival: primary tumor control ($p = 0.98$), and number of treated BM ($p = 0.18$).

Nomograms

The patient cohort treated in the VUmc ($n = 214$) was used as the training cohort for development of the two nomograms. The other patient cohort ($n = 281$, patients treated in MC, VT, and CZE) was used as an external validation cohort in which the two developed nomograms were tested independently from the training cohort. Prognostic factors for survival identified with Cox multivariate analysis in the training cohort of patients ($n = 214$) were used to develop the nomograms for the prediction of early death (<3 months) and long-term survival (>1 year), respectively. Nomograms were made based on logistic regression analysis and learned on the VUmc cohort. The primary endpoint of this study was the area under the curve (AUC) obtained using receiver operating characteristics (ROC) analysis for early death and long-term survival prediction. In the training and validation cohorts, the AUCs of the developed nomogram models were compared with the AUCs of the RPA, DS-GPA, GGS, SIR, and Rades 2015 prognostic models. Comparison of ROC curves was done using DeLong's test for correlated ROC curves. Statistical analyses were performed using SPSS (version 23, IBM, New York), using R (version 3.1.3, R Foundation for Statistical Computing, Vienna, Austria) using the rms, PredictABEL, and pROC packages. Validation was performed according to established methods [16]. Calculating AUC confidence intervals and calibration R2 values (predicted *versus* observed risk) was done according previously described methods [17,18].

Results

Median survival of the total cohort of patients ($n = 495$) was 6.8 months. Baseline characteristics of the training ($n = 214$) and validation ($n = 281$) cohorts are shown in Table 1.

The first developed nomogram specific for the prediction of early death is shown in Fig. 1 containing the previously identified prognostic factors for survival. With an AUC of 0.77, the nomogram predicted early death statistically significantly better than the unfavorable groups of the RPA, DS-GPA, GGS, SIR, and Rades 2015 (range AUC = 0.52–0.59). Similar results were observed in the external validation cohort with an AUC = 0.70 of the nomogram *versus* range AUCs = 0.51–0.60 with the other prognostic models, Table 2). For the ROC curves of the nomogram, see Supplementary materials 1. Calibration curves (predicted *versus* observed probability) of the nomogram are shown in Supplementary materials 2 with R2 values of 0.98 and 0.82 in respectively the training and validation cohort.

The independently developed second nomogram is specific for the prediction of long-term survival and shown in Fig. 2 containing the same prognostic factors for survival, but otherwise ranked in the nomogram. With an AUC = 0.77, this nomogram predicted 1 year survival statistically significantly better than the favorable groups of the RPA, DS-GPA, GGS, SIR, and Rades 2015 in the training cohort (range AUCs = 0.55–0.68, Table 2). In the external validation cohort comparable results were observed with AUC = 0.67 of the nomogram *versus* range AUCs = 0.57–0.61, $p < 0.05$ of four