NEURO-ONCOLOGY

New validated prognostic models and prognostic calculators in patients with low-grade gliomas diagnosed by central pathology review: a pooled analysis of EORTC/RTOG/NCCTG phase III clinical trials

Thierry Gorlia, Wenting Wu, Meihua Wang, Brigitta G. Baumert, Minesh Mehta, Jan C. Buckner, Edward Shaw, Paul Brown, Roger Stupp, Evanthia Galanis, Denis Lacombe, and Martin J. van den Bent

EORTC Headquarters, Brussels, Belgium (T.G., D.L.); North Central Cancer Treatment Group, Rochester, Minnesota (W.W., J.C.B., P.B., E.G.); Radiation Therapy Oncology Group, Philadelphia, Pennsylvania (M.W., M.M., E.S.); Radiation Oncology (MAASTRO), GROW, Maastricht University Medical Centre, Maastricht, the Netherlands (B.G.B.); MediClin Robert Janker Klinik, Bonn, Germany (B.G.B.); Centre Hospitalier Universitaire Vaudois (CHUV) and University of Lausanne, Vaud, Switzerland (R.S.); Department of Neuro-Oncology, Dr. Daniel den Hoed Cancer Center, AE Rotterdam, the Netherlands (M.J.v.d.B.)

Background. In a previous study, the European Organisation for Research and Treatment of Cancer (EORTC) reported a scoring system to predict survival of patients with low-grade gliomas (LGGs). A major issue in the diagnosis of brain tumors is the lack of agreement among pathologists. New models in patients with LGGs diagnosed by central pathology review are needed.
Methods. Data from 339 EORTC patients with LGGs diagnosed by central pathology review were used to develop new prognostic models for progression-free survival (PFS) and overall survival (OS). Data from 450 patients with centrally diagnosed LGGs recruited into 2 large studies conducted by North American cooperative groups were used to validate the models.
Results. Both PFS and OS were negatively influenced by the presence of baseline neurological deficits, a shorter time since first symptoms (<30 wk), an astrocytic tumor type, and tumors larger than 5 cm in diameter. Early irradiation improved PFS but not OS. Three risk groups have been identified (low, intermediate, and high) and validated.

Conclusions. We have developed new prognostic models in a more homogeneous LGG population diagnosed by central pathology review. This population better fits with modern practice, where patients are enrolled in clinical trials based on central or panel pathology review. We could validate the models in a large, external, and independent dataset. The models can divide LGG patients into 3 risk groups and provide reliable individual survival predictions. Inclusion of other clinical and molecular factors might still improve models' predictions.

Keywords: low-grade glioma, predictive accuracy, prognostic factors.

Low-grade glioma (LGG) is a heterogeneous group of primary, diffuse, and slowly growing glial brain tumors. These tumors often remain clinically stable for many years, and patients are commonly only followed clinically without specific antitumor therapy. Based on retrospective studies suggesting an improved survival with early, extensive, and maximal tumor resections, radical surgery is often advocated. Prospective controlled studies evaluating the role of surgery are lacking, and a large part of the benefit presumed from extensive resection may be due to patient selection. If tumor location makes the surgery difficult or even impossible, a biopsy is

performed to ascertain the nature of the tumor and establish a pathological diagnosis.

Immediate (postoperative) radiotherapy has not been shown to offer an advantage in overall survival (OS) over deferred radiotherapy; although progression-free survival (PFS) is lengthened, the optimal timing remains debatable. There is no apparent effect of dose; 2 randomized studies by the European Organisation for Research and Treatment of Cancer (EORTC) and of the North Central Cancer Treatment Group (NCCTG)/Radiation Therapy Oncology Group (RTOG)/Eastern Cooperative Oncology Group (ECOG) Intergroup showed no significant difference in survival when researchers compared lower and higher irradiation doses (45 vs 59.4 Gy and 50.4 vs 64.8 Gy, respectively).[1,2,8] The role of postoperative chemotherapy alone or in combination with radiation therapy remains investigational.[3]

Individual prognosis of patients is highly variable. In order to choose the best strategy for a patient among the various treatment options, prognostic models and scores can be useful. A major limitation in addressing prognostic models for LGG is the considerable interobserver variability in both the grading and the typing of these tumors.[4,5] The widely used EORTC prognostic scoring model for LGG was based on 2 prospective randomized clinical trials. However, patient inclusion into these trials relied upon a diagnosis made by the local pathologist, which was often not confirmed by central pathology review.[6] The external validity of the EORTC scoring system was recently evaluated in a dataset of LGG patients treated in a North American Intergroup trial (NCCTG 86-72-51).[7] In that dataset, the distinction between the low-risk and the high-risk group was predominantly determined by the prognostic impact of histology and tumor size; other factors, like age and extent of surgery, did not contribute significantly. A major difference between the US and European trials was the mandatory central pathology review prior to inclusion in the American trials. Thus, we reanalyzed the pooled data from the 2 EORTC studies, restricting the analysis to patients with LGG whose histology had been confirmed upon central pathology review. Patients with histologies other than grade II glioma and patients from whom no tumor tissue was available for central review were excluded. We subsequently assessed the external validity of the EORTC studies with the individual patient data from large studies conducted by 2 US cooperative groups (RTOG and NCCTG).[8,9] Based on this analysis, we developed prognostic calculators for PFS and OS that provide estimates for both median and fixed time probabilities of survival.

Materials and Methods

Patient Selection

Three hundred seventy-nine and 311 patients with LGG at first diagnosis were randomized in EORTC trials 22844 and 22845, respectively. Central pathology review was available for 428 patients, 182 (53%) in 22844 and 246 (81%) in 22845, out of 648 eligible patients.[6] Table 1 describes how patients were selected for analysis.

Candidate Prognostic Factors

Factors screened for their prognostic value were patient's age, gender, postoperative neurological signs and symptoms (history of seizures and/or headaches, presence of mental and/or motor disturbance), time since first tumor-associated symptoms, postoperative World Health Organization (WHO) performance status and Medical Research Council (MRC) neurological score (see Supplemental Table S1), extent of resection assessed by the surgeon, time since surgery, baseline administration of steroids and/or anticonvulsants, histological type (astrocytic vs oligodendroglial), predominant tumor location, tumor crossing midline, and largest tumor diameter (details are provided in Supplemental Table S1).

Table 1. Comparison of tumor grade by local and central pathology review

	Central by Local Grade			Total (*n* = 648)
	Tumor Grade by Local Pathology Review			
	Missing (*n* = 16) *n* (%)	Grade I (*n* = 47) *n* (%)	Grade II (*n* = 585) *n* (%)	*n* (%)
Tumor grade by central pathology review				
Missing	2 (12.5)	23 (48.9)	195 (33.3)	220 (34.0)
Grade II	12 (75.0)	19 (40.4)	308 (52.6)	339 (52.3)
HGG	2 (12.5)	2 (4.3)	65 (11.1)	69 (10.6)
Grade I	0 (0.0)	3 (6.4)	6 (1.0)	9 (1.4)
Other pathology	0 (0.0)	0 (0.0)	11 (1.9)	11 (1.7)

In EORTC patients, 585 were locally diagnosed with grade II (WHO 1979); 390 were centrally reviewed (gray area). Grade II was confirmed in 308 patients. Of the 648 eligible patients, 339 were diagnosed with grade II by the central pathologist (black frame). These patients were considered for the prognostic factor analyses.

Patient Outcome Measurements

Computed tomography (CT) scans were used pre- and postoperatively for diagnosis and for evaluation of disease progression. PFS was computed as the time from randomization until signs of clinical or radiological progression or death, whichever occurred first. OS was calculated as the time from randomization until death regardless of cause. In the absence of events, PFS and OS were censored at the last follow-up date. For descriptive purposes, PFS and OS from the date of first LGG symptoms were also computed.

Statistical Considerations

Model Development

Categorical data were tabulated with frequencies and percentages. Medians and ranges (minimum-maximum) were used to summarize continuous variables. Spearman rank correlation coefficient (SCC) was computed pairwise for all factors. The significance of the association between categorical (nominal) factors was assessed by the Fisher exact test. For the association between continuous covariates or scores and categorical (nominal) factors, the Wilcoxon rank sum test was used. *P*-values <1% and SCC superior or equal to 0.40 were reported. Immediate versus delayed irradiation (radiotherapy) was entered as a factor in PFS analyses. Since treatment effect did not impact OS, it was not entered in the OS models but was used as a stratification factor. For each factor, Kaplan–Meier curves and log-rank tests were computed. All factors were considered for Cox multivariate analyses; that is, no systematic screening by univariate analysis was performed. The number of factors was lower than the number of PFS or OS events divided by 10, which is generally considered to provide sufficient power in multivariate analyses.[10] Proportional hazards assumptions were tested by examining the plot of the log of negative log estimates over the log survival time and by interpreting the Schoenfeld residual plots.[11] The stepwise backward method was used for factor selection. For factors whose missing value rate was more than 5%, the missing value was considered as a dummy category in the Cox analyses. Model internal validity was assessed by the bootstrap method. Factors with an importance (percentage of bootstrap samples with factor selected in multivariate analysis) lower than 60% were not included in the final models.[12] A significance level of 5% was used in multivariate analyses. Harrell's C-index corrected for optimism by bootstrap resampling was used to assess the model's discrimination.[10] Calibration plots and Schemper's percentage of explained variation (PEV) were also computed.[13,14] A PEV of at least 20% was considered a minimum requirement for a model to provide sufficiently precise individual survival predictions.[15] From the final models, prognostic calculators were developed, and predictions for median PFS, OS, 3-year PFS (PFS3y), and 5-year OS (OS5y) were derived. Individual prognostic scores were computed. Based on their scores, patients were classified into 3 distinct risk groups: low, intermediate,

and high. In the absence of predefined cutoffs, groups were taken with equal size.

For all statistical analyses, SAS v9.2 was used except for the computation of the C-index and calibration plots, which were obtained from the R "Design" and "Hmisc" packages. The PEV was computed using the SAS macro RELIMPCR.[12] The reflected method was used to estimate median survival with a 95% confidence interval (CI).[16] The log-log transformation was used for the 95% CI of PFS3y and OS5y.

Also computed were the model's sensitivity, specificity, and positive and negative predictive value (PPV and NPV, respectively).

In this study, PPV measured the ability of the model to identify patients at high risk for progression or death at year 3 (PFS3y) or of death at year 5 (OS5y). The capacity of the model to identify patients at low risk for these events is measured by the NPV. See Supplemental Table S4 for more definitions.

Model Validation

Data from patients diagnosed with grade II glioma in RTOG 98-02 (WHO classification 1993) and NCCTG 86-72-51 (Kernohan classification) were pooled for model validation. PFS and OS curves in US data were split according to the EORTC risk groups and compared between EORTC and RTOG/NCCTG cooperative groups. EORTC Cox models were fit on US data to determine which factors kept their prognostic influence. Model sensitivity, specificity, NPV, and PPV were computed on US data.

Results

Comparison of EORTC Patient Characteristics and Outcomes

Among the 585 EORTC patients locally diagnosed with grade II (WHO classification 1979), 390 were centrally reviewed and 308 (79%) were confirmed as grade II, 65 patients (16.7%) having high-grade gliomas (HGGs; WHO grades III–IV). Six (1.5%) had grade I and another pathology was diagnosed in 11 patients (2.8%). The central pathologist identified 339 grade II gliomas (WHO classification 1979, 79%) and 69 HGGs (16%) (see Table 1). HGG patients were older (median 43 vs 39 y, *P* = .008), had a worse performance status (*P* = .007), more often underwent resection (89.9% vs 64.6%, *P* < .0001), had less frequent astrocytoma (50.7% vs 68.4%, *P* = .02), and had worse PFS (*P* = .01) and OS (*P* = .03). Table 2 compares patient and disease characteristics among various subgroups. There were no significantly different characteristics and outcomes between patients with and without central pathology review (PFS, *P* = .08, OS, *P* = .92).

Development of Prognostic Models

Supplemental Table S2 presents the factors with correlation coefficients ≥0.40. Presence of mental ($\rho = 0.44$, *P* < .0001) or motor disturbances ($\rho = 0.59$, *P* < .0001)

Table 2. Comparison of patients' characteristics and outcomes between subgroups

	No Central Review (1) (n = 220) n (%)	Central Review (2) (n = 428) n (%)	P(2)/(1)	EORTC Centrally Reviewed LGG (3) (n = 339) n (%)	EORTC Centrally Reviewed HGG (4) (n = 99) n (%)	P(4)/(3)	RTOG/NCCTG Centrally Reviewed LGG (5) (n = 450) n (%)	P(5)/(3)
Age (y)								
Median	39.2	39.3	.60	39.0	43.0	.008	40.0	.07
Range	16.0–66.4	17.2–65.7		17.2–65.3	18.1–65.7		18.0–82.0	
Gender								
Male	122 (55.5)	264 (61.7)	.13	214 (63.1)	41 (59.4)	.59	259 (57.6)	.12
Female	98 (44.5)	164 (38.3)		125 (36.9)	28 (40.6)		191 (42.4)	
History of seizure								
No	143 (65.0)	304 (71.0)	.10	235 (69.3)	54 (78.3)	.19	101 (22.4)	<.0001
Yes	77 (35.0)	121 (28.3)		101 (29.8)	15 (21.7)		98 (21.8)	
Missing	0 (0.0)	3 (0.7)		3 (0.9)	0 (0.0)		251 (55.8)	
Headache								
No	181 (82.3)	334 (78.0)	.30	267 (78.8)	53 (76.8)	.63	143 (31.8)	.06
Yes	39 (17.7)	91 (21.3)		69 (20.4)	16 (23.2)		56 (12.4)	
Missing	0 (0.0)	3 (0.7)		3 (0.9)	0 (0.0)		251 (55.8)	
Mental disturbance								
No	184 (83.6)	347 (81.1)	.51	277 (81.7)	52 (75.4)	.18	178 (39.6)	.03
Yes	35 (15.9)	78 (18.2)		59 (17.4)	17 (24.6)		21 (4.7)	
Missing	1 (0.5)	3 (0.7)		3 (0.9)	0 (0.0)		251 (55.8)	
Motor disturbance								
No	186 (84.5)	312 (72.9)	.05	245 (72.3)	51 (73.9)	1.0	175 (38.9)	.003
Yes	34 (15.5)	90 (21.0)		72 (21.2)	15 (21.7)		24 (5.3)	
Missing	0 (0.0)	26 (6.1)		22 (6.5)	3 (4.3)		251 (55.8)	
Performance status								
0	81 (36.8)	149 (34.8)	.81	127 (37.5)	15 (21.7)	.007	109 (24.2)	.40
1	98 (44.5)	216 (50.5)		164 (48.4)	42 (60.9)		154 (34.2)	
>1	41 (18.6)	62 (14.5)		47 (13.9)	12 (17.4)		25 (5.6)	
Missing	0 (0.0)	1 (0.2)		0 (0.0)	0 (0.0)		162 (36.0)	
MRC score								
No	139 (63.2)	246 (57.5)	.33	194 (57.2)	42 (60.9)	.10	201 (44.7)	.009
Some	49 (22.3)	129 (30.1)		101 (29.8)	21 (30.4)		183 (40.7)	
Moderate/major	32 (14.5)	53 (12.4)		44 (13.0)	6 (8.7)		56 (12.4)	

Continued

Table 2. *Continued*

	No Central Review (1) (n = 220) n (%)	Central Review (2) (n = 428) n (%)	P(2)/(1)	EORTC Centrally Reviewed LGG (3) (n = 339) n (%)	EORTC Centrally Reviewed HGG (4) (n = 99) n (%)	P(4)/(3)	RTOG/NCCTG Centrally Reviewed LGG (5) (n = 450) n (%)	P(5)/(3)
Missing	0 (0.0)	0 (0.0)		0 (0.0)	0 (0.0)		10 (2.2)	
Extent of resection by neurosurgeon								
Biopsy	84 (38.2)	127 (29.7)	.04	116 (34.2)	4 (5.8)	<.0001	211 (46.9)	<.0001
Resection	135 (61.4)	294 (68.7)		219 (64.6)	62 (89.9)		239 (53.1)	
Missing	1 (0.5)	7 (1.6)		4 (1.2)	3 (4.3)		0 (0.0)	
Baseline steroids								
No	102 (46.4)	201 (47.0)	.003	160 (47.2)	30 (43.5)	1.0	219 (48.7)	<.0001
Yes	86 (39.1)	94 (22.0)		77 (22.7)	14 (20.3)		228 (50.7)	
Missing	32 (14.5)	133 (31.1)		102 (30.1)	25 (36.2)		3 (0.7)	
Baseline anticonvulsants								
No	23 (10.5)	50 (11.7)	.15	36 (10.6)	13 (18.8)	.03	51 (11.3)	.12
Yes	165 (75.0)	236 (55.1)		193 (56.9)	29 (42.0)		396 (88.0)	
Missing	32 (14.5)	142 (33.2)		110 (32.4)	27 (39.1)		3 (0.7)	
Local diagnosis								
Astrocytoma	154 (70.0)	279 (65.2)	.36	232 (68.4)	35 (50.7)	.02	104 (23.1)	<.0001
Mixed oligoastrocytoma	21 (9.5)	42 (9.8)		31 (9.1)	11 (15.9)		133 (29.6)	
Oligodendroglioma	44 (20.0)	107 (25.0)		76 (22.4)	23 (33.3)		213 (47.3)	
Missing	1 (0.5)	0 (0.0)		0 (0.0)	0 (0.0)		0 (0.0)	
Frontal location								
No	137 (62.3)	215 (50.2)	.004	172 (50.7)	30 (43.5)	.29	176 (39.1)	.001
Yes	83 (37.7)	213 (49.8)		167 (49.3)	39 (56.5)		274 (60.9)	
Temporal location								
No	145 (65.9)	316 (73.8)	.04	251 (74.0)	51 (73.9)	1.0	289 (64.2)	.007
Yes	75 (34.1)	112 (26.2)		88 (26.0)	18 (26.1)		157 (34.9)	
Missing	0 (0.0)	0 (0.0)					4 (0.9)	
Parietal location								
No	181 (82.3)	356 (83.2)	.82	277 (81.7)	63 (91.3)	.05	296 (65.8)	<.0001
Yes	39 (17.7)	72 (16.8)		62 (18.3)	6 (8.7)		151 (33.6)	
Missing	0 (0.0)	0 (0.0)					3 (0.7)	
Occipital location								
No	213 (96.8)	424 (99.1)	.05	336 (99.1)	69 (100.0)	1.0	426 (94.7)	.002
Yes	7 (3.2)	4 (0.9)		3 (0.9)	0 (0.0)		20 (4.4)	

	(1)	(2)	p	(3)	(4)	p	(5)	p
Missing							4 (0.9)	
Other location								
No	204 (92.7)	401 (93.7)	.62	320 (94.4)	63 (91.3)	.41	227 (50.4)	.08
Yes	16 (7.3)	27 (6.3)		19 (5.6)	6 (8.7)		24 (5.3)	
Missing							199 (44.2)	
Left lobe								
No	114 (51.8)	237 (55.4)	.45	188 (55.5)	39 (56.5)	.90	221 (49.1)	.08
Yes	105 (47.7)	191 (44.6)		151 (44.5)	30 (43.5)		229 (50.9)	
Missing	1 (0.5)	0 (0.0)					0 (0.0)	
Right lobe								
No	114 (51.8)	201 (47.0)	.24	158 (46.6)	31 (44.9)	.89	252 (56.0)	.01
Yes	105 (47.7)	227 (53.0)		181 (53.4)	38 (55.1)		198 (44.0)	
Missing	1 (0.5)	0 (0.0)					0 (0.0)	
Midline crossing								
No	163 (74.1)	310 (72.4)	.55	247 (72.9)	50 (72.5)	.87	136 (30.2)	.06
Yes	53 (24.1)	88 (20.6)		68 (20.1)	15 (21.7)		57 (12.7)	
Missing	4 (1.8)	30 (7.0)		24 (7.1)	4 (5.8)		257 (57.1)	
Tumor size, cm								
<5	84 (38.2)	157 (36.7)	.47	124 (36.6)	24 (34.8)	.67	218 (48.4)	.07
≥5	102 (46.4)	220 (51.4)		173 (51.0)	38 (55.1)		231 (51.3)	
Missing	34 (15.5)	51 (11.9)		42 (12.4)	7 (10.1)		1 (0.2)	
Time since first LGG symptoms (weeks)								
Median	27.9	30.0	.55	30.5	26.3	.40	14.5	<.0001
Range	2.9–1749.0	2.0–1542.4		2.0–1542.4	2.1–828.7		1.3–787.7	
N obs	220	426		338	68		426	
Time since surgery (wk)								
Median	3.6	2.3	<.0001	2.3	1.7	<.0001	4.1	<.0001
Range	0.0–31.7	0.3–157.6		0.3–157.6	0.4–9.7		0.3–214.6	
N obs	220	428		339	69		420	
Median PFS, mo (95% CI)	53 (44,76)	54 (47,60)	.08	55 (49,63)	41 (27,55)	.01	66 (55,75)	.01
Median OS, mo (95% CI)	80 (61,111)	84 (77,95)	.92	87 (79,99)	62 (44,89)	.03	110 (96,129)	.02

(1) EORTC patients without central pathology review. (2) EORTC patients with central pathology review. (3) EORTC patients centrally diagnosed with low grade gliomas (LGGs, grade II). (4) EORTC patients centrally diagnosed with high grade gliomas (HGGs, grade III-IV). (5) RTOG/NCCTG centrally diagnosed with LGGs (grade II).

Fisher test was used for binary or categorical factors. Wilcoxon rank sum test was used for continuous variables and scores. Log-rank test was used for outcome comparisons.

and WHO performance status ($\rho = 0.46$, $P < .0001$) were correlated with the MRC neurological scale. In order to minimize the problems linked to multicollinearity, separate multivariate models with MRC score or WHO performance were fit and their performance was compared. Supplemental Table S3 displays the results of PFS and OS univariate analyses. History of use of steroids and anticonvulsants was collected during radiotherapy and not collected in the delayed radiotherapy arm; these data did not show prognostic significance in univariate analyses in the radiotherapy arms and were not used in the multivariate analyses. Only tumor location involving the temporal lobe was significant for OS and thus considered for multivariate analyses. Information on tumor crossing midline was not systematically available in the US dataset. Models without and with this factor were fit for sensitivity. Midline crossing and tumor size were missing in 8% and 13%, respectively, of the EORTC patients. A dummy category ("missing") was used instead.

For all LGG patients ($n = 339$), median PFS was 55.3 months (95% CI = 49.4–63.4) and PFS3y was 68.0% (62.6–72.7). Median OS was 86.5 months (95% CI = 78.6–99.2) and OS5y was 65.9% (60.4–70.9).

For both PFS and OS, models had all similar discrimination power (for PFS, a C-index ranging 64%–66% and PEV ranging 10%–13%; for OS, a C-index of 67% for all models, PEV ranging 9%–15%), irrespective of the combination of covariates. For all these models, PEV was below the 20% threshold necessary to consider a model sufficiently precise for individual predictions. Extent of resection and age were not identified with significant prognostic value in any analysis. Among all tested EORTC models, final selections took into account availability of covariates and maximal sample size in the US validation datasets.

Five factors were retained in the final PFS prognostic model: immediate irradiation ($P = .0008$, hazard ratio [HR] = 0.62, 95% CI = 0.47–0.82), time since first LGG symptoms (<30 wk vs ≥30 wk, $P = .01$, HR = 0.70, 95% CI = 0.53–0.92), presence of neurological deficit ($P = .0003$, HR = 1.64, 95% CI = 1.25–2.15), independent confirmation of astrocytoma ($P = <.0001$, HR = 1.93, 95% CI = 1.47–2.54), and tumor size (<5 cm/≥5 cm, $P = .004$, HR = 1.53, 95% CI = 1.15–2.03). C-index was 0.64 and PEV was 10.1%.

In the final OS model, independent prognostic factors were identified as time since first LGG symptoms ($P = .009$, HR = 0.67, 95% CI = 0.49–0.91), MRC score ($P = .0001$, HR = 1.51, 95% CI = 1.22–1.86), independent confirmation of astrocytoma ($P < .0001$, HR = 1.96, 95% CI = 1.43–2.69), and tumor size ($P = .001$, HR = 1.74, 95% CI = 1.25–2.43). C-index was 0.67 and PEV was 8.8%. Final multivariate models are presented in Table 3. Figure 1 shows PFS and OS Kaplan–Meier curves by the 3 equally sized risk groups in the EORTC data. For PFS and OS, sensitivity, specificity, and PPV were low (36%–57%). NPV was 74% for PFS and 61% for OS (see Table 3 for details). In both PFS and OS models, calibration plots did not suggest large systematic differences (biases) between predicted and observed outcomes (data not shown). Variability was nevertheless high.

Validation in RTOG/NCCTG Data

Baseline characteristics were different between EORTC and RTOG/NCCTG patients (Table 2). In particular, tumors of EORTC and RTOG/NCCTG patients were of different histological types. Oligodendrogliomas or mixed oligoastrocytomas were diagnosed in 76.9% of US patients compared with 31.5% of EORTC patients ($P < .0001$). It is beyond the scope of this paper to interpret these differences. For both PFS and OS, MRC score or presence of neurological deficit, central pathology diagnosis (nonastrocytic vs astrocytic tumor type), and tumor size but not time since first LGG symptoms had significant prognostic influence in US data (Table 3). An explanation for this could be that time since first LGG symptoms was significantly shorter in US data ($P < .0001$, median 14 vs 30 wk), which may reflect a more aggressive therapeutic approach. Compared with EORTC, the C-index and PEV were slightly lower in US data (PFS: C-index = 0.61, PEV = 5.5%; OS: C-index = 0.62, PEV = 7.1%; Table 3). Table 4 compares PFS and OS by risk groups between EORTC and US data. Overall, US patients had significantly different outcomes compared with EORTC patients (PFS: $P = .01$, OS: $P = .03$). Nevertheless, there was no significant difference in PFS and OS between EORTC and US patients within risk groups. There was no difference in either PFS or OS when they were computed from the time of first symptoms (PFS: $P = .95$, OS: $P = .92$, curves not shown). Figure 2 shows PFS and OS Kaplan–Meier curves by risk group in US data. Curves separated well between the 3 risk groups. Sensitivity, specificity, and PPV were low (<70%). NPV was 73% for PFS and 71% for OS (see Table 3).

Prognostic Calculators

Prognostic calculators based on new prognostic models have been developed. Like nomograms, these prognostic calculators provide patients with PFS3y and OS5y estimates based on their individual characteristics. Prognostic calculators are available online for physicians and patients at http://www.eortc.be/tools/lggcalculator. Individual predictions can also be extracted from Supplemental Table S5. As a disclaimer, prognostic calculators must be used cautiously—individual precision and prediction of outcome are limited. A patient's prognosis may depend on other factors than those taken into account. Any decisions concerning patient care should not be based only on the use of these calculators but should also take into account the patient's past history, other current patient and tumor characteristics, and new therapeutic development.

Discussion

This is a pooled prognostic factor analysis of 2 large EORTC trials of patients with LGG. Only patients with independently confirmed WHO grade II glioma were included, thus providing a more homogeneous dataset and increasing the precision of the prediction. A total of

Table 3. Multivariate analyses of PFS and OS

	PFS				OS			
	EORTC		RTOG/NCCTG		EORTC		RTOG/NCCTG	
	P *n* = 338/E = 235	HR (95% CI)	*P* *n* = 418/E = 293	HR (95% CI)	*P* *n* = 338/E = 183	HR (95% CI)	*P* *n* = 418/E = 239	HR (95% CI)
Treatment (delayed/immediate irradiation)*	.0008	0.62 (0.47−0.82)		N/A†		N/A†*		N/A†
Time since first symptoms (<30 /≥30 wk)	.01	0.70 (0.53−0.92)	.34	1.14 (0.87−1.47)	.009	0.67 (0.49−0.91)	.42	1.13 (0.85−1.50)
MRC score (no/some/moderate or major deficit)	NI	NI	NI	NI	.0001	1.51 (1.22−1.86)	<.0001	1.46 (1.22−1.75)
MRC score (no/at least some deficit)	.0003	1.64 (1.25−2.15)	.01	1.36 (1.07−1.71)	NI	NI	NI	NI
Central histological type (OA or OD/AA)	<.0001	1.93 (1.47−2.54)	<.0001	1.93 (1.49−2.52)	<.0001	1.96 (1.43−2.69)	<.0001	2.08 (1.56−2.76)
Tumor size, cm (<5/≥5)	.004	1.53 (1.15−2.03)	<.0001	1.78 (1.41−2.26)	.001	1.74 (1.25−2.43)	.0005	1.58 (1.22−2.05)
C-index♪		0.64		0.61		0.67		0.62
PEV, %‡		10.1		5.5		8.8		7.1
Sensitivity, %♫		36		65		38		61
Specificity, %♫		47		26		43		29
NPV, %♫		74		73		61		71
PPV, %♫		50		58		57		51

Abbreviations: *n*, sample size; E, number of events; NI, not included in this model; OA, oligoastrocytoma; OD, oligodendroglioma; AA, anaplastic astrocytoma.
*In PFS analyses, treatment was considered a variable in the regression equation. In OS analyses, it was used as a stratification factor in the Cox model.
†All US patients were treated with immediate radiotherapy. ♪ C-index was corrected for optimism by bootstrap technique.
‡A PEV of at least 20% is considered a minimum requirement for a model to provide sufficiently precise individual survival predictions. ♫ see Supplemental Table S4 for definition.
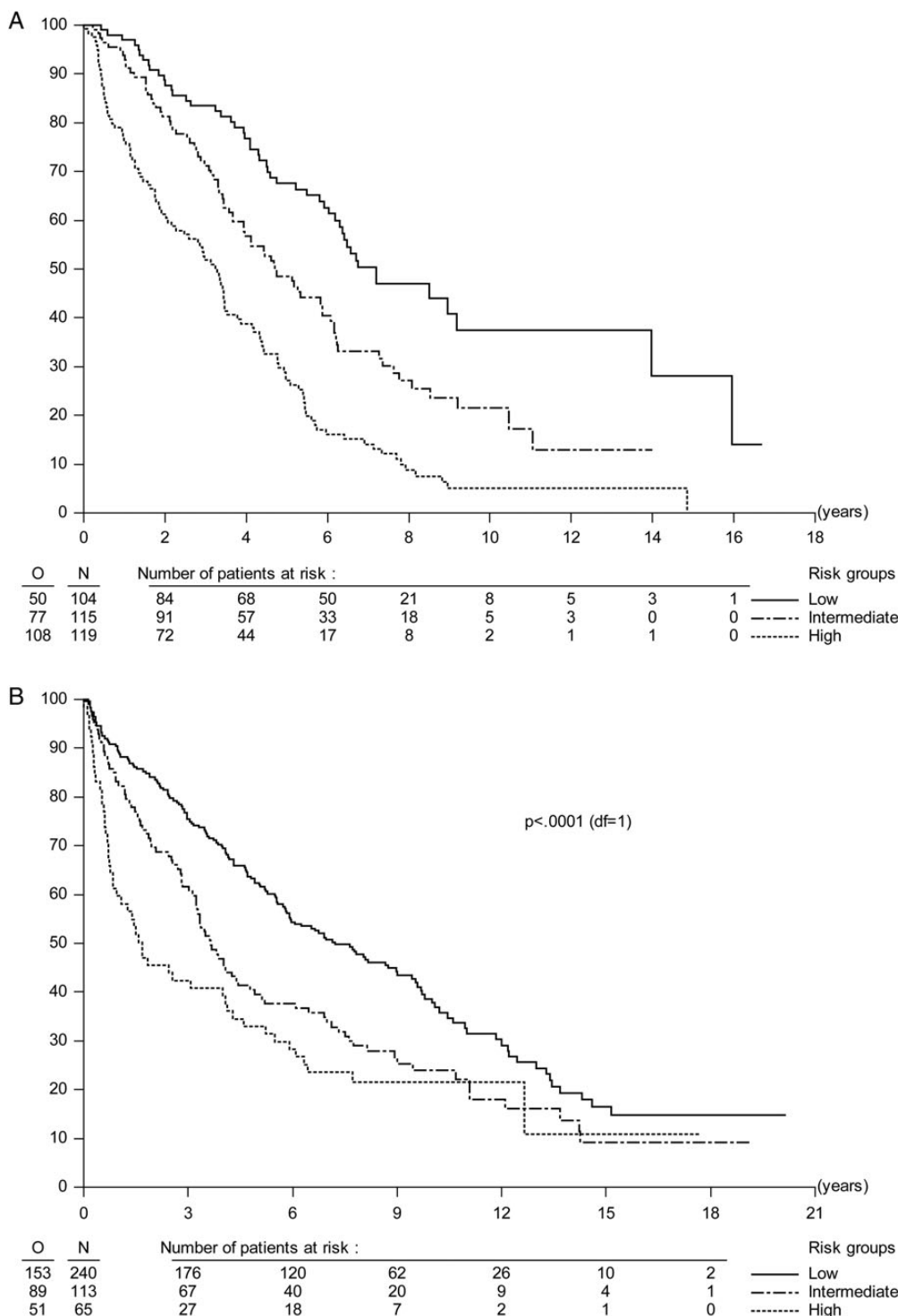
Fig. 1. (A) PFS curves split by risk group in the EORTC dataset. Compared with low-risk patients, patients with intermediate risk had PFS HR = 1.78 with 95% CI = 1.24–2.55 and patients with high risk had PFS HR = 3.32 with 95% CI = 2.36–4.67. (B) PFS curves split by prognostic risk group in the NCCTG/RTOG dataset. PFS HRs and 95% CIs were 1.56 and 1.20–2.03, P = .0009, for intermediate risk patients and 2.17 and 1.58–2.99, P < .0001, for high-risk compared with low-risk patients.

21% of cases reviewed had to be excluded due to differing central pathology review diagnosis, 17% being qualified as HGGs. Survival was substantially worse for the excluded patients retrospectively considered as having HGG. This interobserver variation is a known factor in trials

on LGG and HGG and is related to the subjectivity of the criteria used.[5]

Both PFS and OS were negatively influenced by a worse baseline neurological status (ie, presence of neurological deficits [in PFS] or MRC score [in OS]), a shorter time

Table 4. Comparison of outcomes between EORTC and RTOG/NCCTG data: all patients and by risk group

		PFS				OS			
		N patients/n events	Median, mo (95% CI)	PFS3y (95% CI)	P	N/E	Median, mo (95% CI)	OS5y	P
All patients	EORTC	339/235	55.33 (49.35, 63.38)	67.96 (62.60, 72.73)	.01	339/183	86.54 (78.55, 99.19)	65.92 (60.36, 70.90)	.025
	RTOG/NCCTG	450/313	65.97 (55.13, 74.87)	66.47 (61.86, 70.65)		450/256	110.06 (95.87, 129.05)	67.66 (63.06, 71.82)	
Low risk	EORTC	104/50	86.28 (74.28, 167.66)	83.43 (74.38, 89.51)	.48	116/41	152.02 (101.72, N)	76.88 (67.47, 83.89)	0.92
	RTOG/NCCTG	240/153	86.54 (69.62, 113.15)	75.43 (69.41, 80.43)		229/110	139.17 (120.97, 161.41)	76.91 (70.76, 81.93)	
Intermediate risk	EORTC	115/77	56.48 (44.09, 70.60)	71.18 (61.76, 78.67)	.53	106/56	90.71 (74.71, 106.97)	72.18 (62.07, 80.02)	.67
	RTOG/NCCTG	113/89	43.73 (37.39, 57.40)	61.54 (51.85, 69.85)		102/65	86.60 (62.26, 132.83)	61.78 (51.46, 70.53)	
High risk	EORTC	119/108	39.33 (25.92, 42.32)	51.85 (42.49, 60.41)	.34	116/86	57.72 (45.17, 75.27)	49.90 (40.25, 58.79)	.55
	RTOG/NCCTG	65/51	20.01 (9.99, 48.82)	42.39 (30.20, 54.05)		87/64	65.35 (31.57, 86.74)	50.04 (39.09, 60.02)	

HRs were not included in any comparisons. They cannot be interpreted when curves are crossing over, which was the case for some of them.

since first symptoms (<30 wk), an astrocytic histology, and a tumor size of >5 cm in diameter. Of note, presence of neurological deficits and WHO performance status measure were interrelated. Treatment by immediate irradiation improved PFS but not OS.[1] Contrary to earlier report, age no longer showed a prognostic importance in the now more homogeneous dataset of histologically confirmed low-grade tumors.[6] Elderly HGG patients were removed from this dataset. In this analysis, debulking surgery or complete tumor resection (as reported by the operating neurosurgeon without confirmation by imaging) did not significantly improve either PFS or OS (although tumor size was inversely correlated with extent of resection).[6] In everyday clinical practice, histological diagnosis is based on the skills and expertise of the local pathologist—independent central expert review is rarely routine practice. Thus general applicability of our data may be confounded by a higher variation in histological subtypes and grades seen in clinical practice. Our models had moderate discrimination measured by C-index (max, 0.67). Their percentage of explained variation in survival times was limited (PEV < 20%), leading to large CIs for outcome estimates. Sensitivity, specificity, and PPV were low for both PFS and OS in all datasets, as was NPV for OS in EORTC data. Our models had moderate NPV in both EORTC and US datasets for PFS (~74%). They could nevertheless separate patients into 3 distinct risk groups—low, intermediate, and high—in both EORTC (development) and US (validation) datasets. A major limitation of our study is the absence of molecular data in EORTC trials designed in the mid-1980s, without tissue collection, as well as the estimation of tumor size based on CT as opposed to MRI. The prognostic value of new biomarkers relevant for gliomas could therefore not be assessed in our dataset. In particular, 1p/19q codeletion has since been identified as a favorable prognostic factor for oligodendroglial tumors, associated with more indolent disease, prolonged natural history, and increased responsiveness to therapy.[17] Results of randomized trials must further distinguish between prognostic and predictive information related to 1p/19q status. Similarly, *IDH* mutations are of major prognostic significance in diffuse gliomas, although their value in grade II tumors is disputed.[18] Furthermore, not all trials collected the same clinical data. As an example, the Mini Mental State Examination (MMSE) scores were not collected in the EORTC patients. Previous reports have shown that the presence of an abnormal baseline MMSE score is a strong predictor of poorer PFS and OS.[19] The addition of these factors to our prognostic models might significantly improve their predictive accuracy and precision.

Conclusions

In our previous report, all patients diagnosed by local pathologists were used in the prognostic modeling.[6] In this study, patients were selected based on the LGG diagnosis of a central pathology reviewer. This population is more homogeneous because fewer patients with higher-grade gliomas were included. It better fits with modern practice,
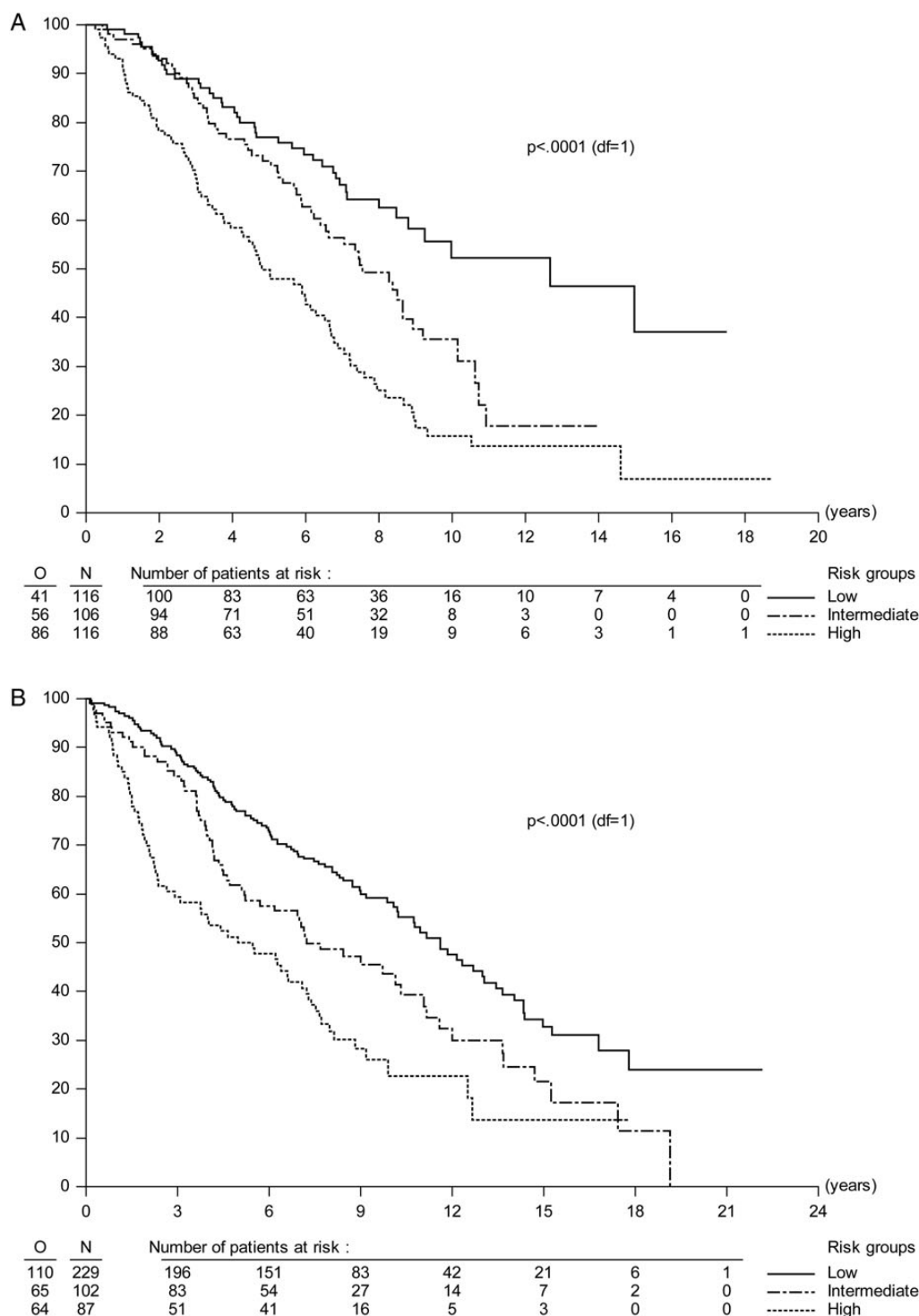
Fig. 2. (A) OS curves split by prognostic risk group in the EORTC dataset. Compared with low-risk patients, patients with intermediate risk had OS HR = 1.67 with 95% CI = 1.12–2.51 and patients with high risk had OS HR = 2.90 with 95% CI = 2.00–4.22. (B) OS curves split by prognostic risk group in the NCCTG/RTOG dataset. OS HRs and 95% CIs were 1.58 and 1.16–2.14, $P = 0.004$, for intermediate risk patients and 2.47 and 1.81–3.38, $P < .0001$, for high-risk compared with low-risk patients.

where patients are enrolled in clinical trials based on central or panel pathology review. With more similar patients, the new prognostic models provide more reliable and precise predictions. With their limitations correctly understood, they can help physicians to classify patients into 3 risk groups and propose the most adapted therapeutic strategy, including patients' participation in clinical trials. They can be used to discuss disease prognosis with patients and families. Characteristics of patients and how they were managed were different between

EORTC and RTOG/NCCTG patients, but discrimination and predictive accuracy were comparable, making these prognostic models useful for both European and American patients.

Supplementary Material

Supplementary material is available online at *Neuro-Oncology* (http://neuro-oncology.oxfordjournals.org/).

Funding

References

1. van den Bent MJ, Afra D, de Witte O, et al. Long-term efficacy of early versus delayed radiotherapy for low-grade astrocytoma and oligodendroglioma in adults: the EORTC 22845 randomised trial. *Lancet*. 2005;366(9490):985–990.

2. Kiebert GM, Curran D, Aaronson NK, et al. Quality of life after radiation therapy of cerebral low-grade gliomas of the adult: results of a randomised phase III trial on dose response (EORTC trial 22844). EORTC Radiotherapy Co-operative Group. *Eur J Cancer*. 1998;34(12):1902–1909.

3. van den Bent MJ. Can chemotherapy replace radiotherapy in low-grade gliomas? Time for randomized studies. *Nat Rev Neurol*. 2010;6(12):695–701.

4. Kros JM, Gorlia T, Kouwenhoven MC, et al. Panel review of anaplastic oligodendroglioma from European Organisation For Research and Treatment of Cancer Trial 26951: assessment of consensus in diagnosis, influence of 1p/19q loss, and correlations with outcome. *J Neuropathol Exp Neurol*. 2007;66(6):545–551.

5. van den Bent MJ. Interobserver variation of the histopathological diagnosis in clinical trials on glioma: a clinician's perspective. *Acta Neuropathol*. 2010;120(3):297–304.

6. Pignatti F, van den Bent M, Curran D, et al. Prognostic factors for survival in adult patients with cerebral low-grade glioma. *J Clin Oncol*. 2002;20(8):2076–2084.

7. Daniels TB, Brown PD, Felten SJ, et al. Validation of EORTC prognostic factors for adults with low-grade glioma: a report using intergroup 86–72–51. *Int J Radiat Oncol Biol Phys*. 2011;81(1):218–224.

8. Shaw E, Arusell R, Scheithauer B, et al. Prospective randomized trial of low- versus high-dose radiation therapy in adults with supratentorial low-grade glioma: initial report of a North Central Cancer Treatment Group/Radiation Therapy Oncology Group/Eastern Cooperative Oncology Group study. *J Clin Oncol*. 2002;20:2267–2227.

9. Shaw E, Wang M, Coons S, et al. Randomized trial of radiation therapy plus procarbazine, lomustine, and vincristine chemotherapy for supratentorial adult low-grade glioma: initial results of RTOG 9802. *J Clin Oncol*. 2012;30(25):3065–3070.

10. Harrell FE, Jr, Lee KL. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*. 1996;15(4):361–387.

11. Schoenfeld D. Partial residuals for the proportional hazards regression model. *Biometrika*. 1982;69(1):239–241.

12. Sauerbrei W, Schumacher M. A bootstrap resampling procedure for model building: application to the Cox regression model. *Stat Med*. 1992;11(16):2093–2109.

13. Schemper M, Henderson R. Predictive accuracy and explained variation in Cox regression. *Biometrics*. 2000;56(1):249–255.

14. Heinze G, Schemper M. Comparing the importance of prognostic factors in Cox and logistic regression using SAS. *Comput Methods Programs Biomed*. 2003;71(2):155–163.

15. Dunkler D, Michiels S. Gene expression profiling: does it add predictive accuracy to clinical characteristics in cancer prognosis? *Eur J Cancer*. 2007;43(4):745–751.

16. Slud EV, Byar DPA. Comparison of reflected versus test-based confidence intervals for the median survival time, based on censored data. *Biometrics*. 1984;40(3):587–600.

17. Kouwenhoven MC, Kros JM, French PJ, et al. 1p/19q loss within oligodendroglioma is predictive for response to first line temozolomide but not to salvage treatment. *Eur J Cancer*. 2006;42(15):2499–2450.

18. van den Bent MJ, Dubbink HJ, Marie Y, et al. IDH1 and IDH2 mutations are prognostic but not predictive for outcome in anaplastic oligodendroglial tumors: a report of the European Organisation for Research and Treatment of Cancer Brain Tumor Group. *Clin Cancer Res*. 2010;16(5):1597–1604.

19. Brown PD, Buckner JC, O'Fallon JR, et al. Importance of baseline mini-mental state examination as a prognostic factor for patients with low-grade glioma. *Int J Radiat Oncol Biol Phys*. 2004;59(1):117–125.