

The relationship between six-month progression-free survival and 12-month overall survival end points for phase II trials in patients with glioblastoma multiforme

Karla V. Ballman,¹ Jan C. Buckner, Paul D. Brown, Caterina Giannini, Patrick J. Flynn, Betsy R. LaPlant, and Kurt A. Jaeckle

Division of Biostatistics (K.V.B., B.R.L.), Division of Medical Oncology (J.C.B.), Department of Radiation Oncology (P.D.B.), and Department of Pathology (C.G.), Mayo Clinic College of Medicine, Rochester, MN; Metro-Minnesota Community Clinical Oncology Program (P.J.F.), St. Louis Park, MN; and Department of Neurology (K.A.J.), Mayo Clinic, Jacksonville, FL; USA

Common end points for phase II trials in patients with glioblastoma multiforme (GBM) are six-month progression-free survival (PFS6) and 12-month overall survival (OS12). OS12 can be accurately measured but may be confounded with subsequent therapies upon progression, whereas the converse is true for PFS6. Our goal was to assess the relationship between these end points separately for phase II trials in patients with newly diagnosed GBM and patients with recurrent GBM. Data were pooled from 11 North Central Cancer Treatment Group trials for patients with newly diagnosed GBM (n = 1348). All patients received radiotherapy and pharmaceutical therapy (before, during, or after radiotherapy). Data were pooled from 16 trials that used various pharmaceuticals in treating patients for recurrent GBM (n = 345). All trial regimens were declared nonefficacious by predefined criteria. Overall per-patient concordance

was estimated with a kappa statistic. The relationship between OS12 and PFS6 across study arms was assessed by weighted linear regression and Pearson's correlation. Simulation was used to determine the agreement of study outcomes when using PFS6 versus OS12 end points. Cox models with progression status as a time-dependent variable and Kaplan-Meier estimators were used to ascertain the association between progression-free survival status and overall survival. At present, 97% of the patients with newly diagnosed GBM and 95% of those with recurrent GBM have died. The PFS6 and OS12 were 43% and 41%, respectively, for patients with newly diagnosed disease and 9% and 14% for patients with recurrent disease. There was only moderate concordance between the end points on both the patient level and the study level. For the simulation studies, we established phase II efficacy criteria for each end point by using the pooled estimates of OS12 (PFS6) as historical controls. The study decisions made using PFS6 and OS12 were in agreement 88% and 90% of the time for the trials of newly diagnosed and recurrent disease, respectively. Finally, there was a strong association between progression-free survival status and overall survival. PFS6 seems to be a reasonable end point for phase II trials in patients with recurrent glioblastoma. *Neuro-Oncology* 9, 29–38, 2007 (Posted to *Neuro-Oncology* [serial online], Doc. D06-00036, November 15, 2007. URL www.dukeupress.edu/neuro-oncology; DOI: 10.1215/15228517-2006-025)

Received June 1, 2006; accepted September 1, 2006.

¹ Address correspondence to Karla Ballman, Ph.D., Mayo Clinic, Division of Biostatistics, 200 First Street SW, Rochester, MN 55905, USA (ballman@mayo.edu).

² Abbreviations used are as follows: GBM, glioblastoma multiforme; NCCTG, North Central Cancer Treatment Group; OS12, 12-month overall survival; PFS6, six-month progression-free survival; PS, performance score.

Keywords: end points, glioblastoma multiforme, phase II clinical trials

Two commonly used end points for phase II trials in patients with glioblastoma multiforme (GBM)² are the proportion of patients who are alive and progression free at six months (six-month progression-free survival [PFS6]) and the proportion of patients alive at 12 months (12-month overall survival [OS12]). The purpose of phase II trials is to assess treatment regimens as a means of selecting those that most warrant phase III evaluation in a time-efficient manner, in other words, to determine as effectively as possible when new treatments are likely to be potentially efficacious. These commonly used phase II end points differ from the definitive phase III end point of overall survival. Assuming overall survival is the true end point of interest (i.e., the gold standard), treatment assessment based on PFS6 or OS12 will never supersede overall survival as the ideal end point (Begg and Leung, 2000).

Phase II studies are limited because they lack a control arm and rely on a historical value as a comparison for treatment efficacy. It is therefore crucial that patients enrolled on a phase II trial are representative of the patient population from which the historical comparison value was obtained, because both end points would otherwise suffer from patient selection bias. Beyond this, there are advantages and limitations for both PFS6 and OS12. OS12 is objectively ascertained. One drawback of using OS12 is that it could potentially be influenced by life-prolonging subsequent therapy administered after patients leave the study (typically, upon progression); additionally, investigators must wait at least 12 months from the time of last patient enrollment before the study results are known. On the other hand, PFS6 directly measures the efficacy of initial therapy, unaffected by treatment at progression. When PFS6 is the end point, study results can be obtained six months sooner than when OS12 is used. The drawback is that PFS6 is based on clinical and/or imaging criteria, both which have an element of subjectivity (e.g., progression erroneously declared) and may be influenced by prior therapies (e.g., surgery, radiotherapy, intratumoral therapy, and corticosteroids), imaging technique, and observer subjectivity.

In this study, our primary goal was to determine the relationship between PFS6 and OS12 as end points in phase II GBM trials. A secondary goal was to determine the relationship between PFS6 and the true end point of overall survival (the assumed gold standard). We were also interested in whether the relationships assessed in our primary and secondary goals were similar for patients with newly diagnosed GBM and those with recurrent disease. Knowing these relationships will enable investigators to make more informed choices between OS12 and PFS6 as outcomes in phase II trials for GBM patients. We aggregated all GBM patients that were treated on a North Central Cancer Treatment Group (NCCTG) protocol and stratified our analysis by patients with newly diagnosed GBM and by those with recurrent disease.

Methods

Two separate groups of patients were obtained from trials conducted by the NCCTG by pooling patients across all trials of newly diagnosed GBM and all trials of recurrent GBM. Each trial was multicenter, had appropriate institutional review board approval, and obtained informed consent from patients or their designated representative prior to enrollment into the trial. Our retrospective investigation was approved by the Mayo Clinic institutional review board and was covered by the patients' consent obtained at the time of their enrollment on the clinical trial.

General characteristics of the 11 trials that enrolled patients with newly diagnosed GBM and the 16 trials that enrolled patients with recurrent GBM are reported in Table 1. Results for all trials were deemed negative by the primary end point decision criterion; that is, no improvement over the historical experience was found in phase II trials, and no significant difference in the primary end points was observed among the treatment arms in phase III trials.

Patient Eligibility

Newly Diagnosed GBM. Patient eligibility was similar across the trials that enrolled patients with newly diagnosed GBM. All patients underwent a biopsy and/or tumor resection prior to study enrollment, and central pathology review was performed in all cases. Only patients with histologically identified GBM were used in our analysis. Patients were not allowed to enroll until they recovered from their biopsy or surgery but were required to enroll within six to eight weeks of their biopsy or surgery. For most trials, patients had to be adults (≥ 18 years old) and had to have an Eastern Cooperative Oncology Group performance score (PS) of 2 or less. Patients were not eligible if they had prior chemotherapy for a brain tumor. All patients received radiotherapy, and all trials involved chemotherapy.

Recurrent GBM. Likewise, patient eligibility criteria were similar across the protocols that enrolled patients with recurrent GBM. All trials required evidence of recurrence after prior radiotherapy. For most studies, the age criterion (≥ 18 years old) and PS criterion ($PS \leq 2$) were the same as for trials of newly diagnosed patients. Prior chemotherapy regimens were allowable. Patient enrollment was allowed if it had been 8–12 weeks or longer since the completion of radiotherapy and four to six weeks or longer since the completion of chemotherapy for GBM. Only patients with histologically identified GBM, either at initial diagnosis or at recurrence as confirmed by central pathology review, were eligible for this study.

Treatment Evaluation and Event-Monitoring Schedules

All NCCTG trials of patients with newly diagnosed or recurrent GBM had equivalent evaluation schedules. Generally, patients were evaluated every two months

Table 1. Brief description of the trials from which patients were pooled for these analyses

Trial	Phase	Trial Start Year	Treatment(s)	No. of GBMs	Reference	Trial End Point
Trials for Patients with Newly Diagnosed GBM						
797251	III	1980	Carmustine vs. dibromodulcitol	181	Elliott et al., 1997	OS
857251	III	1985	Carmustine vs. 1-(2-chloroethyl)-3(2,6 dioxo-1-piperidyl)-1-nitrosourea	248	Dinapoli et al., 1993	OS
887202	I	1989	Interferon α + carmustine + radiation therapy	7	Rajkumar et al., 1998a	MTD
887203	Pilot	1989	Accelerated hyperfractionated radiation therapy + carmustine	5	Rao et al., 2002	Toxicity
887252	III	1990	Radiation therapy + carmustine \pm interferon-alpha	292	Buckner et al., 2001b	OS
907201	I/II	1991	Radiosurgery + carmustine + radiation therapy	6	Toxicity	
917201	Pilot	1991	Carmustine + cisplatin + etoposide	12	Rajkumar et al., 1999a	MTD
927203	I	1993	Carmustine + cisplatin + etoposide	11	Rajkumar et al., 1998b	MTD
937252	III	1994	Carmustine + cisplatin vs. carmustine; accelerated hyperfractionated radiation therapy vs. radiation therapy	401	Buckner et al., 2001a	OS
987252	II	1999	Carmustine + cisplatin + etoposide	91	Moynihan et al., 2002	OS12
N0074	II	2001	Gefitinib	94	Uhm et al., 2004	OS12
Trials for Patients with Recurrent GBM						
797251	III	1980	Etoposide vs. teniposide	68	Elliott et al., 1997	Tumor response
817202	II	1985	Carmustine + dianhydrogalatitol + N-(phosphonacetyl)-L-aspartate	1		Tumor response
847251	II	1985	Fludarabine	6	Cascino et al., 1988	Tumor response
867202	II	1989	Interferon α + eflornithine	14	Buckner et al., 1998	Tumor response
867253	II	1986	Interferon α + carmustine	9	Buckner et al., 1995	Tumor response
867254	II	1987	Etoposide + cisplatin	11	Buckner et al., 1990	Tumor response
887251	II	1988	Ifosfamide + sodium 2-mercaptoethane sulfonate	8	Elliott et al., 1991	Tumor response
897251	II	1989	5-Fluorouracil + leucovorin	22	Cascino et al., 1996	Tumor response
897252	II	1991	Amonafide	15	Levitt et al., 1995	Tumor response
917251	II	1992	Nitrogen mustard + vincristine \pm procarbazine	27	Galanis et al., 1998	Tumor response
927251	II	1993	Topotecan	21	Burch et al., 2000	Tumor response
937251	II	1993	2-Chlorodeoxyadenosine	7	Rajkumar et al., 1999b	Tumor response
957253	II	1996	Dacarbazine	18	Rajkumar et al., 2000	Tumor response
967251	II	1998	Irinotecan	34	Reid et al., 2000	Tumor response
987254	I/II	2000	Pyrazoloacridine + carboplatin	27	Galanis et al., 2005b	Tumor response
N997B	II	2001	Temsirolimus	57	Galanis et al., 2005a	PFS6

Abbreviations: GBM, glioblastoma multiforme; MTD, maximum tolerated dose; OS, overall survival; OS12, overall survival at 12 months; PFS6, progression-free survival at six months.

during the active-monitoring phase (i.e., while receiving study treatment). An evaluation included a neurologic examination and an imaging study (either CT or MRI); the same imaging modality was used consistently on a patient throughout the study. When patients completed or went off study treatment (e.g., because of disease progression, refusal of further treatment, or toxicity), they entered an event-monitoring phase. Patients remained in event monitoring until death or the end of study.

Tumor progression was determined by a combination of the neurologic examination status and imaging results. An imaging progression occurred when there was a greater than 25% increase in the product of perpendicular diameters of contrast enhancement or mass (compared with pretreatment scan) for bidimensionally

measurable disease or an unequivocal increase in the size of contrast enhancement or increase in mass effect (compared with pretreatment scan), as agreed upon independently by the primary physician and quality-control physicians for evaluable disease (i.e., contrast-enhancing mass on CT/MRI that is not measurable but clearly evaluable for response to therapy). For both measurable and evaluable tumors, the appearance of new lesions signified disease progression regardless of the status of the initial tumor. The neurologic examination status was deemed better, same, or worse compared with the pretreatment examination. A patient was identified as having disease progression when there was progression by the imaging study and the neurologic examination status was determined as the same or worse. If there

was a discrepancy between the neurologic status examination and CT/MRI measurement (i.e., neurologic examination status was better but CT/MRI indicated disease progression, or neurologic examination status was worse but CT/MRI did not indicate progression), the patient continued treatment until the next evaluation. If the discrepancy remained, the patient was classified as having disease progression at the time of the subsequent evaluation.

Statistical Considerations

Summary statistics used for categorical variables were the frequency and percent. Those used for continuous variables were the mean \pm 1 SD and the median and range (minimum to maximum values). Survival curves were estimated with the Kaplan-Meier estimates (Kaplan and Meier, 1958) and compared using the log-rank test (Peto and Peto, 1972). Survival experiences were summarized with the median value and 95% CI.

Patient-level agreement between the PFS6 and the OS12 end points meant that a patient was progression free at six months and alive at 12 months or that a patient had disease progression by six months and was dead by 12 months. The patient-level agreement was summarized by the raw agreement, the number of patients for which the end points agreed divided by the total number of patients. Furthermore, the expected levels of agreement due to chance alone were also determined. A kappa statistic (Cohen, 1968) and 95% CI were used to summarize the amount of agreement above and beyond that expected by chance alone. One can interpret a kappa statistic as a type of correlation coefficient. It ranges from 0 (no agreement) to 1 (perfect agreement). Values observed in the 0.4–0.6 range indicate moderate agreement, those in the 0.6–0.8 range indicate substantial agreement, and those in the 0.8–1.0 range indicate strong agreement.

Study-level agreement was assessed with two methods: The first was the determination of the association between the study PFS6 proportion and OS12 proportion, and the second was a measure of agreement of the overall study decision between using a PFS6 end point and an OS12 end point. The first method used weighted linear regression in which the numbers of patients in the study were the weights, thus giving more weight to studies with more patients. The equation of the regression line is given; if there was perfect agreement between the values of the two end-point values per study, the slope of the line would be 1 and the intercept would be 0. In addition, the value of the correlation coefficient was also computed and reported. Note that the correlation coefficient is somewhat dependent on the range of observed values for the independent variable. Even in cases in which variables are strongly correlated, when the relationship is assessed over a small portion of the potential range, the correlation value will be less than when assessed over the entire range.

Because of the small number of studies and small sample sizes for some studies, we used simulation to explore the agreement at the study level between the two end points. Of particular interest was the amount

of agreement between the study decisions when using the PFS6 end point versus the OS12 end point. To assess this, we replicated our current approach for designing phase II clinical trials. We computed the historical control values of PFS6 and OS12 to be used in our phase II trial designs based on the observed outcomes of our database of 1348 patients with newly diagnosed GBM and 345 patients with recurrent GBM: PFS6 and OS12 values were 43% and 41%, respectively, for newly diagnosed cases and 9% and 14%, respectively, for recurrent cases. For each patient group, we designed a two-stage phase II study using a Simon design (Simon, 1989). The level of significance used was 0.10, and there was 0.90 power to detect a minimum increase of 0.15 above the historical control value. The required sample size for the patients with newly diagnosed GBM was $n = 83$ patients for both the PFS6 and OS12 end points, and the required sample sizes for the patients with recurrent GBM was $n = 53$ for both end points. We performed a simulation of 10,000 trials for each patient group. Specifically, for each trial, we selected the required number of patients with replacement from the pooled data. Using the observed progression-free survival and overall survival times for each patient, we determined their PFS6 status and OS12 status. On the basis of these, we determined the study final decision (including the possibility for stopping because of futility after the first stage)—that is, sufficient evidence to warrant further investigation versus no evidence of activity, meaning the regimen does not merit further investigation. In each case, we recorded whether the two study end points agreed at the end of each simulation. We then computed the percentage of trials for which there was agreement in the final study decision between the two end points. This is in the spirit of a criterion advocated by Begg and Leung (2000) for determining whether one end point is a good surrogate for another.

PFS6 is merely a potential surrogate end point for the real quantity of interest: overall survival. A good surrogate needs to be correlated with the end point it is intended to replace, although this is not a sufficient condition. To assess the relationship between PFS6 and overall survival, we used Cox proportional hazard modeling. The outcome was overall survival, and the time-dependent covariate was progression status (disease progression vs. no progression). Conceptually, this means that, at the beginning of the trial, all of the patients were in the no-progression group. When patients experienced documented disease progression, they switched from the no-progression group to the progression group. Patients who died without a documented disease progression were censored for death at the time of their last evaluation for progression; patients without documented disease progression and with documentation of a death due to a cause other than their GBM were not censored. We generated separate models for the patient groups with newly diagnosed and recurrent GBM. Furthermore, we compared the survival experience of patients who were deemed progressors at six months with the experience of those who did not have disease progression at six months. We employed a conservative approach for

this analysis and used only those patients who were alive and not censored at six months. For the purposes of this analysis, *not censored* means their follow-up (including their last tumor evaluation) was six months or longer. Again, separate analyses were performed for patients with newly diagnosed GBM and those with recurrent disease.

Results

Trial and Patient Characteristics

Eleven trials of newly diagnosed primary brain tumor enrolled 1348 GBM patients between June 1980 and August 2002 (Table 1). Overall, 34 patients (3%) were enrolled on phase I or pilot trials, 192 (14%) were enrolled on phase II trials, and 1122 (83%) were enrolled on phase III trials. The median number of patients enrolled on a study was 91 (range, 5–401). All trials had negative results, meaning that none of the treatment regimens investigated met the predefined trial criteria for efficacy. At present, almost all patients (1314 of 1348) have died, and only 134 (10%) died without documented disease progression prior to death.

There were 16 trials for patients with recurrent or progressive brain tumors, which included 345 patients with recurrent GBM. These studies enrolled patients between June 1980 and September 2004. The median number of patients enrolled on a study was 17 (range, 1–68). Of the 345 patients, 277 (80%) were enrolled on phase II trials, and 68 (20%) were enrolled on phase III trials. Results for all these trials were also declared negative, meaning they did not meet the predefined efficacy criteria. At present, almost all patients (328 of 345) have died, and 39 (11%) died without documented disease progression prior to death. This percentage is similar to that for the patients with newly diagnosed GBM.

The average age (\pm SD) at trial entry was 57 ± 12 years for patients with newly diagnosed GBM and 53 ± 12 years for patients with recurrent GBM, and 61% of the patients in both groups were men (Table 2). The PS value was 0 in a considerably greater fraction of patients with newly diagnosed disease compared with patients with recurrent disease (30% compared with 16%, respectively), and a considerably greater fraction of patients with recurrent GBM had a PS of 2 compared with patients with newly diagnosed GBM (36% compared with 16%). The patients with recurrent GBM were more likely to have had a gross total resection upon initial diagnosis than were the patients with newly diagnosed disease (33% compared with 19%). Finally, a large majority of patients with recurrent GBM (72%) did not undergo resection or biopsy at the time of recurrence or progressive disease.

Appropriateness of PFS6 End Point

Figure 1 shows the survival and progression-free survival curves. The median survivals were 10.2 months (95% CI, 9.7–10.7 months) for patients with newly diag-

Table 2. Patient characteristics by glioblastoma multiforme (GBM) tumor type: newly diagnosed and recurrent

Characteristic	Newly Diagnosed GBM	Recurrent GBM
Number of patients	1348	345
Age (years)		
Mean \pm SD	57 \pm 12	53 \pm 12
Median (min., max.)	58 (15, 84)	54 (19, 79)
Gender, n (%)		
Female	531 (39)	135 (39)
Male	817 (61)	210 (61)
ECOG performance status, n (%)		
0	408 (30)	53 (16)
1	627 (47)	142 (42)
2	218 (16)	121 (36)
>2	88 (7)	24 (7)
Missing	7	5
Extent of initial resection, n (%)		
None	0 (0)	2 (1)
Biopsy only	257 (20)	53 (17)
Subtotal	764 (60)	156 (50)
Gross total	245 (19)	104 (33)
Extent of recurrent resection, n (%)		
None	—	198 (72)
Biopsy only	—	19 (7)
Subtotal	—	44 (16)
Gross total	—	15 (5)

Abbreviation: ECOG, Eastern Cooperative Oncology Group.

nosed GBM and 5.0 months (95% CI, 4.6–5.4 months) for patients with recurrent GBM. Median progression-free survival times were 5.3 months (95% CI, 5.0–5.6 months) for patients with newly diagnosed GBM and 1.8 months (1.7–2.0 months) for patients with recurrent disease.

A relevant question is whether six months is a reasonable time point for measuring progression-free survival or whether another time point might be more relevant when considering an alternative for a 12-month overall survival end point. Figure 2 shows the progression rate (number of progressions per 100 patients) as a function of time from study enrollment for patients with newly diagnosed GBM and those with recurrent disease, separately. For the patients with recurrent disease, there was a significant drop in the disease progression rate after six months. The median time to progression was 1.8 months, so a large fraction of patients had progression by six months. The median survival time after patients had disease progression was 2.3 months (95% CI, 1.7–2.7 months). These data together seem to indicate that a majority of patients had disease progression and died before six months. As a consequence, six months seems to be a reasonable time point to assess progression-free survival for patients with recurrent disease. The picture is not quite as clear for the patients with newly diagnosed GBM. The disease progression rate for the

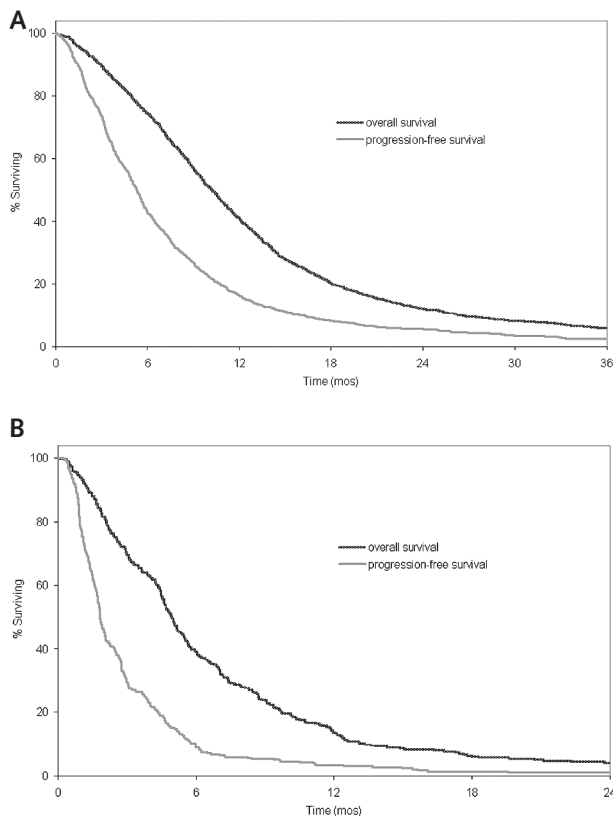


Fig. 1. Overall survival and progression-free survival. (A) Patients with newly diagnosed GBM. (B) Patients with recurrent GBM.

patients with newly diagnosed GBM was relatively flat between 4 and 12 months. It peaked at six months, but there was not a dramatic drop after six months, or at any point, as there was for the patients with recurrence. The median time to progression was 5.3 months, implying the majority of patients had progression prior to six months. The corresponding median survival time after progression was 3.1 months (95% CI, 2.8–3.5 months). Hence, the majority of patients had disease progression prior to six months and, of these, a majority died prior

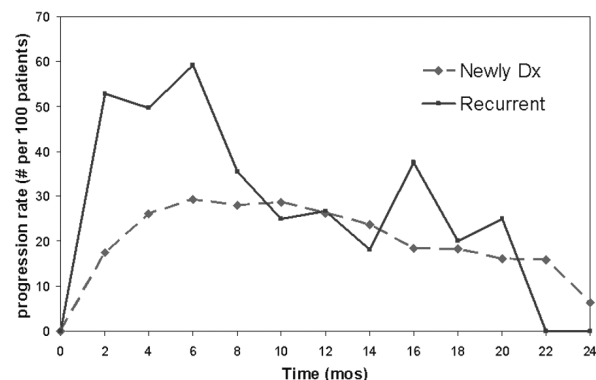


Fig. 2. Progression rates as a function of time for patients with newly diagnosed and recurrent GBM.

to 12 months. Therefore, with regard to a time point, six months seems reasonable to use for progression-free survival in the context of a 12-month overall survival alternative end point.

Patient-Level Agreement of Outcomes

Patient-level agreement between the PFS6 and OS12 end points for both newly diagnosed and recurrent GBM are summarized in Fig. 3. For the patients with newly diagnosed GBM, the overall agreement between the end points was 75%. By chance alone, we would expect an agreement level of 51.2%. The kappa statistic estimate for the level of agreement, beyond chance alone, is 0.48 (95% CI, 0.44–0.53). The level of agreement is at most moderate. The overall agreement between end points for the patients with recurrent disease was 88%. We would expect an agreement level due to chance alone of 74.7%. The kappa statistic estimate was 0.52 (95% CI, 0.39–0.65); again, this is a moderate level of agreement. In general, for both patients with newly diagnosed GBM and those with recurrent disease, there was only a moderate level of agreement between PFS6 and OS12 end points.

Study-Level Agreement

Weighted linear regression was used to assess the strength of the relationship between the observed proportion of PFS6 patients (explanatory variable) and OS12 (outcome variable) for the 11 trials of patients with newly diagnosed GBM and the 16 trials of patients with recurrent GBM. The regression line for the patients with newly diagnosed GBM was $OS12 = 0.24 + 0.40$

A

		OS12	
		Alive at 12 months	Dead at 12 months
PFS6	No progression at six months	397 (237.9)	185 (344.1)
	Progression at six months	154 (313.1)	612 (452.9)

B

		OS12	
		Alive at 12 months	Dead at 12 months
PFS6	No progression at six months	30 (7.3)	12 (34.7)
	Progression at six months	30 (52.7)	273 (250.3)

Fig. 3. Patient-level agreement between PFS6 and OS12 end points. In each cell the top number is the number of patients observed; bottom number in parentheses is the expected number of patients for the cell due to chance. (A) Patients with newly diagnosed GBM. (B) Patients with recurrent GBM.

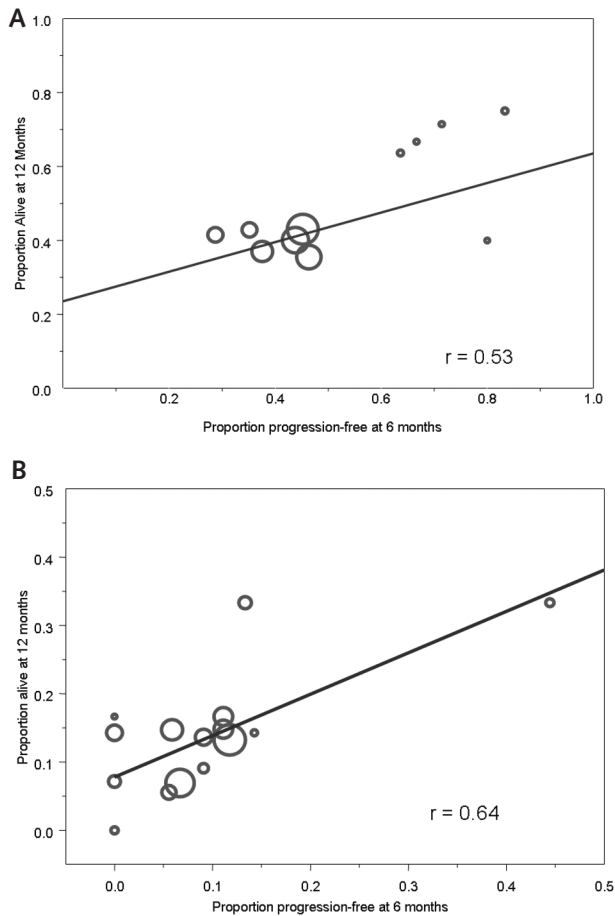


Fig. 4. Scatterplot and weighted linear regression line and corresponding correlation (r) for the relationship between the PFS6 and OS12 rates on a study level. There is one point per study, and the size of the points is proportional to the number of patients on the study. (A) Patients with newly diagnosed GBM. (B) Patients with recurrent GBM.

\times PFS6 (Fig. 4A). There was a positive linear relationship between the end points, although it did not quite achieve statistical significance (P for the slope = 0.09). The correlation coefficient value was $r = 0.53$. Overall, the observed relationship was only moderately strong. In particular, the slope of the line was 0.40, which differed significantly from 1 ($P = 0.02$), and the correlation was only 0.53, which indicates that only 28% of the variation in the observed OS12 rates was explained by the PFS6 rates. A partial explanation for the poor correlation is that most of the larger studies span only a relatively small range of the observed progression-free survival values (between 0.25 and 0.50). Specifically, low correlation values are often observed when the values on the horizontal axis span only a small range of potential values.

For the patients with recurrence, the equation of the weighted regression line was $OS12 = 0.08 + 0.61 \times PFS6$, and the corresponding correlation coefficient was 0.64. There was a significant positive relationship between the two end points (P for the slope = 0.01),

although there was some evidence that the slope differed from 1 ($P = 0.07$). As for the patients with newly diagnosed GBM, the strength of the relationship was at most moderate. The observed values were slightly higher for the patients with recurrent versus newly diagnosed disease, with a slope closer to 1.0 and a higher correlation coefficient. As for the patients with newly diagnosed GBM, the observed PFS6 rates span only a small fraction of the potential values for the majority of the studies (i.e., 0–0.15). There is one potentially influential study (a PFS6 rate of approximately 0.45) that could have artificially inflated the observed correlation value. The correlation value with this study removed was 0.47.

Overall, the relationships between PFS6 and OS12 for patient trials of newly diagnosed GBM and of recurrent GBM were positive, which is as expected. However, the relationships were of moderate strength, meaning that if we know the PFS6 value for a trial, we would not be able to predict accurately what its OS12 value would be. This is a little surprising but likely explained by the observed range of values for PFS6 being relatively limited.

Study Decision Agreement

To assess the number of times the study decision based on a PFS6 end point agrees with the study decision based on an OS12 end point, we performed the simulation studies described in the Statistical Considerations section. For the trials of newly diagnosed disease, there was agreement in 88% of the simulated trials between the PFS6 and OS12 end points. For the trials of recurrent GBM, there was 90% agreement between studies designed with the OS12 and PFS6 end points. In both patient groups, the simulations resulted in a substantially high level of agreement in study decisions between these two end points.

Relationship Between Progression and Overall Survival

Another important aspect of interest is the relationship between progression and the true end point of interest: overall survival. This was assessed in two ways: using progression as a time-dependent variable, and comparing the survival experiences between patients who were alive at six months and progression free and those who were alive at six months and had disease progression prior to then. For patients with newly diagnosed GBM, the hazard ratio for progression treated as a time-dependent variable was 16.2 (95% CI, 13.2–19.8). This indicates quite a strong relationship between progression and survival. For patients who were alive at six months ($n = 1002$), we compared the survival experience between those patients who had disease progression by six months and those patients who were progression free at six months. The estimate of the hazard ratio was 2.1 (95% CI, 1.8–2.4), indicating that patients who had disease progression by six months were approximately twice as likely to die within a given time period as those patients who had not had progression by six months. The survival curves for the two groups differed significantly (Fig. 5A): median survival (measured from six

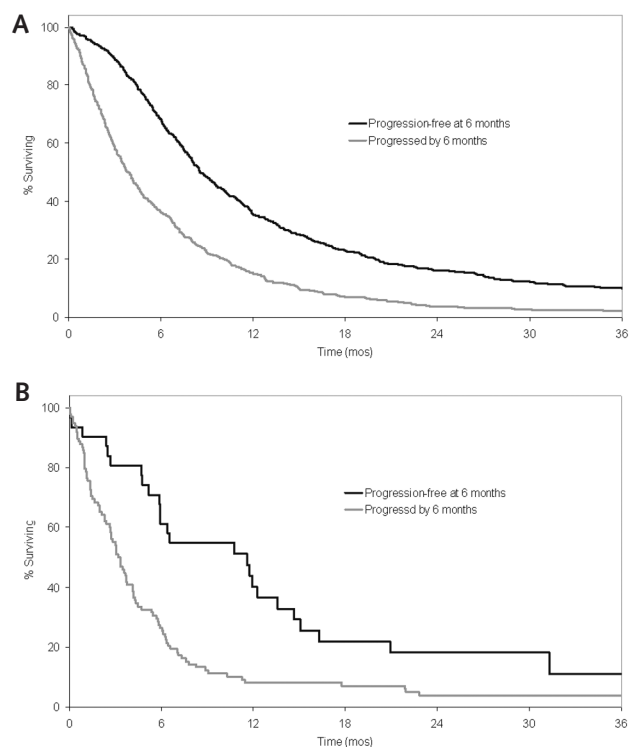


Fig. 5. Overall survival by six-month progression status. Time is measured from six months after study entry (i.e., time 0 corresponds to six months after study entry) (A) Patients with newly diagnosed GBM. (B) Patients with recurrent GBM.

months after study entry) was 8.6 months for the non-progressors and 3.8 months for progressors ($P < 0.001$). The results were similar for the patients with recurrent GBM. The hazard ratio for progression treated as a time-dependent variable was 8.5 (95% CI, 5.7–12.8). Only 129 patients were alive at six months. The hazard ratio for those who had disease progression by then and those who had not was 2.4 (95% CI, 1.6–3.8). Again, the likelihood of death for individuals who had disease progression by six months was more than twice that for patients without progression. The median remaining survival time for patients who had progression by six months was 3.3 months compared with 11.6 months for those without progression by six months ($P < 0.001$; Fig. 5B).

Discussion

Phase II studies often employ non-survival-based efficacy end points such as tumor response rate and progression-free survival, which can be obtained in a much shorter interval than overall survival, the end point for phase III studies in GBM. Few phase III investigations have shown a benefit for experimental treatment compared with a standard therapy or placebo, often because the phase II end points are weak surrogates for the phase III end point of overall survival (Fazzari and Heller, 2000). Before the adoption of a non-survival-

based phase II end point, it is important to understand its relationship to a survival-based end point in order to know whether it will provide a reliable indication of the potential benefit of an agent. Along these lines, Hess et al. (1999) examined the relationship between response to chemotherapy and subsequent progression and survival in patients with recurrent malignant glioma. They showed a substantial difference in progression-free survival and overall survival between responders and non-responders.

Few publications have compared the relationship between PFS6 and OS12 in glioma trials. Investigators advocate the use of a progression-free end point in phase II trials rather than a response end point when evaluating the cytostatic therapeutics (Batchelor et al., 2001; Brada and Yung, 2000; Mick et al., 2000). However, it is recognized that the ultimate goal is to increase overall survival. Some comparisons have been made between progression-free (or disease-free)-based and survival-based end points in other solid tumors. Sekine et al. (1999) found that the rate of progressive disease was significantly correlated with the median survival time; the correlation was higher than that between the response rate and median survival time. Sargent et al. (2005) completed a formal evaluation of a disease-free primary end point versus an overall survival end point. They concluded that, in adjuvant colon cancer trials of fluorouracil-based regimens, disease-free survival after three years of median follow-up is an appropriate end point.

The relationship between a potential surrogate end point and the true end point should be analyzed on several levels. Merely assessing the relationship between progression-free survival and survival using study-level summary statistics could potentially be misleading (Buyse and Pascal, 1996). A serious problem is that it is extremely likely that progression-free survival and survival are affected by the same factors (known or unknown), so any relationship observed between progression-free survival and overall survival in different studies may be mediated by these confounding factors. As a consequence, it is also important to ascertain the strength of the relationship at the patient level. Furthermore, it has been argued that the surrogate end point must be strongly associated or correlated with the true end point (Begg and Leung, 2000), although a strong correlation between the end points is not by itself sufficient for determining the adequacy of a surrogate (Baker and Kramer, 2003).

In this study, we evaluated the relationship between PFS6 and OS12 at all three levels: patient level, study level, and correlation/concordance of study end results. It appeared as though the agreement at both the study level and patient level is at most moderate for both the recurrent GBM and the newly diagnosed GBM cohorts. However, the study decision outcomes from our simulation analyses had relatively high agreement. Furthermore, there was a substantial difference in overall survival between patients who had disease progression by six months and those who had not, as well as a large hazard ratio between progression-free survival and over-

all survival in the time-dependent Cox regression analysis. Overall, the relationship between PFS6 and OS12 appeared slightly stronger in the patients with recurrent GBM than in patients with newly diagnosed disease. A progression-free survival evaluation at six months also seemed to capture more of the OS12 event information for the recurrent cases than for the newly diagnosed cases (Fig. 2), which is supported by the larger difference in median survival times between six-month progressors and nonprogressors in the recurrent group.

It should be understood that our analysis is not a formal analysis of the adequacy of PFS6 as a surrogate for OS12. Such an analysis would require a large amount of patient-level data from multiple phase III studies. In addition, our analysis is somewhat limited by the fact that results for all the trials were negative. As a consequence, the relationship between these two end points could differ for trials with positive results. In particular, for a trial with a positive result, OS12 might be improved while PFS6 is unaffected, or PFS6 might be improved and OS12 unaffected.

An increasingly important role of a phase II trial is to provide information on the effects of therapy on a biologic or molecular level. This is especially true for targeted therapeutics. If a phase II trial measures only potential survival benefit, an agent that has activity on the biologic or molecular level might be wrongfully discarded. Obviously an active agent that does not demonstrate potential benefit would not be ready for a phase III trial. However, if the phase II trial could provide information that aids in understanding the nature of the molecular and biologic activity, this would be valuable in determining future combination therapies using this agent or in identifying subgroups of patients for which this agent is active. To achieve this goal, it is essential to use phase II designs that provide the necessary second-

ary end points for assessing the molecular and biologic activity of a therapy, such as the end points proposed by Lang et al. (2002).

Given that GBM survival experience is similar to that with late-stage disease, it is feasible to obtain a preliminary estimate of treatment efficacy from survival-based end points in a reasonable amount of time, providing investigators with a sound understanding of how the treatment will affect survival prior to the initiation of a large phase III trial. In light of our assessment of the relationship between PFS6 and OS12, it appears that PFS6 provides only a moderately reliable estimate of survival. For patients with newly diagnosed GBM, it seems judicious to use an OS12 end point. Using a PFS6 end point saves only about six months out of a complete process that requires a total of two to three years (from study concept to mature data) at the cost of a less reliable estimate of the true survival end point, which is what will be used in a subsequent phase III trial. This assumes that statistical power would be the same using the same number of patients for both end points, as was essentially true for our data. The time savings could be less or more in the future, depending on changes in the historical control values for PFS6 and OS12. The trial with the historical control value closer to 0.50 would require more patients. On the other hand, agents that are found to have activity in phase II trials for recurrent GBM are often brought forward to a phase II trial for patients with newly diagnosed GBM, either as an agent used concurrently with radiotherapy or as adjuvant treatment. Furthermore, a six-month time point for the progression-free survival evaluation in recurrent gliomas captured more OS12 information than in newly diagnosed cases. Thus, it seems reasonable to use PFS6 as the primary end point in trials of recurrent GBM.

References

- Baker, S.G., and Kramer, B.S. (2003) A perfect correlate does not a surrogate make. *BMC Med. Res. Methodol.* **3**, 16.
- Batchelor, T., Stanley, K., and Andersen, J. (2001) Clinical trials in neuro-oncology. *Curr. Opin. Neurol.* **14**, 689–694.
- Begg, C.B., and Leung, D.H.Y. (2000) On the use of surrogate end points in randomized trials. *J. R. Stat. Soc. A* **163**(pt. 1), 15–28.
- Brada, M., and Yung, W.K.A. (2000) Clinical trial end points in malignant glioma: Need for effective trial design strategy. *Semin. Oncol.* **27**(suppl. 6), 11–19.
- Buckner, J.C., Brown, L.D., Cascino, T.L., Gerstner, J.B., Krook, J.E., Westberg, M.W., Wiesenfeld, M., O'Fallon, J.R., and Scheithauer, B. (1990) Phase II evaluation of infusional etoposide and cisplatin in patients with recurrent astrocytoma. *J. Neurooncol.* **9**, 249–254.
- Buckner, J.C., Brown, L.D., Kugler, J.W., Cascino, T.L., Krook, J.E., Mailliard, J.A., Kardinal, C.G., Tschetter, L.K., O'Fallon, J.R., and Scheithauer, B.W. (1995) Phase II evaluation of recombinant interferon alpha and BCNU in recurrent glioma. *J. Neurosurg.* **82**, 430–435.
- Buckner, J.C., Burch, P.A., Cascino, T.L., O'Fallon, J.R., and Scheithauer, B.W. (1998) Phase II trial of recombinant interferon-alpha-2a and eflo-nithine in patients with recurrent glioma. *J. Neurooncol.* **36**, 65–70.
- Buckner, J.C., Michalak, J.C., Schomberg, P.J., Burton, G.V., Sandler, H.M., Cascino, T.L., Hawkins, R.B., Scheithauer, B.W., and O'Fallon, J.R. (2001a) Phase III trial of BCNU plus cisplatin (CDDP) versus BCNU alone, and standard radiation therapy (SRT) versus accelerated radiation therapy (ART) in glioblastoma (GBM) patients (Pts): NCCTG/SWOG results. *Proc. Am. Soc. Clin. Oncol.* **20**, 51a (abstract 202).
- Buckner, J.C., Schomberg, P.J., McGinnis, W.L., Cascino, T.L., Scheithauer, B.W., O'Fallon, J.R., Morton, R.F., Kuross, S.A., Mailliard, J.A., Hatfield, A.K., Cole, J.T., Steen, P.D., and Bernath, A.M. (2001b) A phase III study of radiation therapy plus carmustine with or without recombinant interferon-alpha in the treatment of patients with newly diagnosed high-grade glioma. *Cancer* **92**, 420–433.
- Burch, P.A., Bernath, A.M., Cascino, T.L., Scheithauer, B.W., Novotny, P., Nair, S., Buckner, J.C., Pfeifle, D.M., Kugler, J.W., and Tschetter, L.K. (2000) A North Central Cancer Treatment Group phase II trial of topotecan in relapsed gliomas. *Invest. New Drugs* **18**, 275–280.

- Buysse, M., and Pascal, P. (1996) On the relationship between response to treatment and survival time. *Stat. Med.* **15**, 2797–2812.
- Cascino, T.L., Brown, L.D., Morton, R.F., Everson, L.K., Marschke, R.F., Dinapoli, R.P., and O'Fallon, J.R. (1988) Evaluation of fludarabine phosphate in patients with recurrent glioma. *Am. J. Clin. Oncol.* **11**, 586–588.
- Cascino, T.L., Veeder, M.H., Buckner, J.C., O'Fallon, J.R., Wiesenfeld, M., Levitt, R., Goldberg, R.M., Kuross, S.A., Morton, R.F., and Scheithauer, B.W. (1996) Phase II study of 5-fluorouracil and leucovorin in recurrent primary brain tumor. *J. Neurooncol.* **30**, 243–246.
- Cohen, J. (1968) Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol. Bull.* **70**, 213–220.
- Dinapoli, R.P., Brown, L.D., Arusell, R.M., Earle, J.D., O'Fallon, J.R., Buckner, J.C., Scheithauer, B.W., Krook, J.E., Tschetter, L.K., and Maier, J.A. (1993) Phase III comparative evaluation of PCNU and carmustine combined with radiation therapy for high-grade glioma. *J. Clin. Oncol.* **11**, 1316–1321.
- Elliott, T.E., Buckner, J.C., Cascino, T.L., Levitt, R., O'Fallon, J.R., and Scheithauer, B.W. (1991) Phase II study of ifosfamide with mesna in adult patients with recurrent diffuse astrocytoma. *J. Neurooncol.* **10**, 27–30.
- Elliott, T.E., Dinapoli, R.P., O'Fallon, J.R., Krook, J.E., Earle, J.D., Morton, R.F., Levitt, R., Tschetter, L.K., Scheithauer, B.W., Pfeifle, D.M., Twito, D.I., and Nelimark, R.A. (1997) Randomized trial of radiation therapy (RT) plus dibromodulcitol (DBD) versus RT plus BCNU in high grade astrocytoma. *J. Neurooncol.* **33**, 239–250.
- Fazzari, M., and Heller, G. (2000) The phase II/III transition: Toward the proof of efficacy in cancer clinical trials. *Control. Clin. Trials* **21**, 360–368.
- Galanis, E., Buckner, J.C., Burch, P.A., Schaefer, P.L., Dinapoli, R.P., Novotny, P.J., Scheithauer, B.W., Rowland, K.M., Vukov, A.M., Mailliard, J.A., and Morton, R.F. (1998) Phase II trial of nitrogen mustard, vincristine, and procarbazine in patients with recurrent glioma: North Central Cancer Treatment Group results. *J. Clin. Oncol.* **16**, 2953–2958.
- Galanis, E., Buckner, J.C., Maurer, M.J., Kreisberg, J.I., Ballman, K., Boni, J., Peralba, J.M., Jenkins, R.B., Dakhil, S.R., Morton, R.F., Jaeckle, K.A., Scheithauer, B.W., Dancey, J., Hidalgo, M., and Walsh, D.J., for the North Central Cancer Treatment Group (2005a) Phase II trial of temsirolimus (CCI-779) in recurrent glioblastoma multiforme: A North Central Cancer Treatment Group Study. *J. Clin. Oncol.* **23**, 5294–5304.
- Galanis, E., Buckner, J.C., Maurer, M.J., Reid, J.M., Kuffel, M.J., Ames, M.M., Scheithauer, B.W., Hammack, J.E., Pipoly, G., and Kuross, S.A. (2005b) Phase I/II trial of pyrazoloacridine and carboplatin in patients with recurrent glioma: A North Central Cancer Treatment Group trial. *Invest. New Drugs* **23**, 495–503.
- Hess, K.R., Wong, E.T., Jaeckle, K.A., Kyritsis, A.P., Levin, V.A., Prados, M.D., and Yung, W.K. (1999) Response and progression in recurrent malignant glioma. *Neuro-Oncology* **1**, 282–288.
- Kaplan, E.L., and Meier, P. (1958) Non-parametric estimation from incomplete observations. *J. Am. Stat. Assoc.* **53**, 457–481.
- Lang, F.F., Gilbert, M.R., Puduvalli, V.K., Weinberg, J., Levin, V.A., Yung, W.K., Sawaya, R., Fuller, G.N., and Conrad, C.A. (2002) Toward better early-phase brain tumor clinical trials: A reappraisal of current methods and proposals for future strategies. *Neuro-Oncology* **4**, 268–277.
- Levitt, R., Buckner, J.C., Cascino, T.L., Burch, P.A., Morton, R.F., Westberg, M.W., Goldberg, R.M., Gallagher, J.G., O'Fallon, J.R., and Scheithauer, B.W. (1995) Phase II study of amonafide in patients with recurrent glioma. *J. Neurooncol.* **23**, 87–93.
- Mick, R., Crowley, J.J., and Carroll, R.J. (2000) Phase II clinical trial design for noncytotoxic anticancer agents for which time to disease progression is the primary endpoint. *Control. Clin. Trials* **21**, 343–359.
- Moynihan, T.J., O'Fallon, J.R., Krook, J.E., Schomberg, P., Dinapoli, R.P., Kazem, I., Scheithauer, B., and Buckner, J.C. (2002) A phase II trial of pre-irradiation chemotherapy with BCNU, cisplatin and oral etoposide combined with radiation therapy in the treatment of glioblastoma (GBM). *Proc. Am. Soc. Clin. Oncol.* **21**, 77a (abstract 306).
- Peto, R., and Peto, J. (1972) Asymptotically efficient rank invariant procedures (with discussion). *J. R. Stat. Soc. [A]* **135**, 185–207.
- Rajkumar, S.V., Buckner, J.C., Schomberg, P.J., Cascino, T.L., Burch, P.A., and Dinapoli, R.P. (1998a) Phase I evaluation of radiation combined with recombinant interferon alpha-2a and BCNU for patients with high-grade glioma. *Int. J. Radiat. Oncol. Biol. Phys.* **40**, 297–302.
- Rajkumar, S.V., Buckner, J.C., Schomberg, P.J., Pitot IV, H.C., Ingle, J.N., and Cascino, T.L. (1999a) Phase I evaluation of preirradiation chemotherapy with carmustine and cisplatin and accelerated radiation therapy in patients with high-grade gliomas. *Neurosurgery* **44**, 67–73.
- Rajkumar, S.V., Buckner, J.C., Schomberg, P.J., Reid, J.M., Bagniewski, P.J., Ames, M.M., Cascino, T.L., and Marks, R.S. (1998b) Phase I and pharmacokinetic study of preirradiation chemotherapy with BCNU, cisplatin, etoposide, and accelerated radiation therapy in patients with high-grade glioma. *Int. J. Radiat. Oncol. Biol. Phys.* **42**, 969–975.
- Rajkumar, S.V., Burch, P.A., Nair, S., Dinapoli, R.P., Scheithauer, B., O'Fallon, J.R., Ezzell, P.S., Leitch, J.M., Morton, R.F., and Marks, R.S. (1999b) Phase II North Central Cancer Treatment Group study of 2-chlorodeoxyadenosine in patients with recurrent glioma. *Am. J. Clin. Oncol.* **22**, 168–171.
- Rajkumar, S.V., Reid, J.M., Novotny, P.J., Safgren, S.L., Scheithauer, B.W., Johnson, P.S., Nair, S., Morton, R.F., Hatfield, A.K., Krook, J.E., Ames, M.M., and Buckner, J.C., for the North Central Cancer Treatment Group (2000) A randomized phase II and pharmacokinetic study of dacarbazine in patients with recurrent glioma. *J. Neurooncol.* **49**, 255–261.
- Rao, R.D., Thome, S.D., O'Fallon, J., Earle, J.D., Dinapoli, R.P., and Buckner, J.C. (2002) Safety of thrice-daily hyperfractionated radiation and BCNU for high-grade gliomas. *Int. J. Radiat. Oncol. Biol. Phys.* **53**, 376–384.
- Reid, J.M., Buckner, J.C., Schaaf, L., Cha, S., Wright, K., Marks, R., Wiesenfeld, M., Pfeifle, D., Hatfield, A., Krook, J., Duncan, B., and Miller, L. (2000) Anticonvulsants alter the pharmacokinetics of irinotecan (CPT-11) in patients with recurrent glioma. *Proc. Am. Soc. Clin. Oncol.* **19**, 160a (abstract 620).
- Sargent, D.J., Wieand, H.S., Haller, D.G., Gray, R., Benedetti, J.K., Buysse, M., Labianca, R., Seitz, J.F., O'Callaghan, C.J., Francini, G., Grothey, A., O'Connell, M., Catalano, P.J., Blanke, C.D., Kerr, D., Green, E., Wolmark, N., Andre, T., Goldberg, R.M., and De Gramont, A. (2005) Disease-free survival versus overall survival as a primary end point for adjuvant colon cancer studies: Individual patient data from 20,898 patients on 18 randomized trials. *J. Clin. Oncol.* **23**, 8664–8670.
- Sekine, I., Tamura, T., Kunitoh, H., Kubota, K., Shinkai, T., Kamiya, Y., and Saijo, N. (1999) Progressive disease rate as a surrogate endpoint of phase II trials for non-small-cell lung cancer. *Ann. Oncol.* **10**, 731–733.
- Simon, R. (1989) Optimal two-stage designs for phase II clinical trials. *Control. Clin. Trials* **10**, 1–10.
- Uhm, J.H., Ballman, K.V., Giannini, C., Krauss, J.C., Buckner, J.C., James, D., Scheithauer, B.W., O'Fallon, J.R., and Jaeckle, K.A. (2004) Phase II study of ZD1839 in patients with newly diagnosed grade 4 astrocytoma. *J. Clin. Oncol.* **22**(suppl.), 108s (abstract 1505).