

# Health-Related Quality of Life in Patients Treated With Temozolomide Versus Procarbazine for Recurrent Glioblastoma Multiforme

By D. Osoba, M. Brada, W.K.A. Yung, and M. Prados

**Purpose:** To determine whether chemotherapy with temozolomide (TMZ) versus procarbazine (PCB) for recurrent glioblastoma multiforme (GBM) was associated with improvement in health-related quality of life (HRQOL).

**Patients and Methods:** HRQOL was assessed at baseline and during treatment using the European Organization for Research and Treatment of Cancer Quality of Life Questionnaire C30 and a Brain Cancer Module (BCM20) in two clinical trials that enrolled a total of 366 patients. Two hundred eighty-eight patients provided HRQOL data that could be used for analysis; 109 patients received TMZ in a phase II study, whereas 89 patients received TMZ and 90 received PCB in a randomized phase III study. Changes from baseline in the scores of seven preselected HRQOL domains (role and social functioning, global quality of life [QOL], visual disorders, motor dysfunction, communication deficit, and drowsiness) were calculated for all groups. Statistical significance, effect sizes, and proportions of patients with improved HRQOL scores (changes of  $\geq 10$  points) were calculated.

**Results:** Before disease progression, patients treated with TMZ were found to have an improvement in most of the preselected HRQOL domain scores compared with their baseline (pretreatment) scores. Those who were progression-free on TMZ at 6 months had improvement in all the preselected HRQOL domains. Conversely, patients treated with PCB reported deterioration in HRQOL that was independent of whether or not the disease had progressed by 6 months. Patients with disease progression, regardless of treatment, experienced a sharp decline in all domains at the time of progression.

**Conclusion:** Treatment with TMZ was associated with improvement in HRQOL scores compared with treatment with PCB. The deterioration reported by PCB-treated patients was likely because of toxicity. Delaying disease progression by treatment with TMZ is beneficial to the HRQOL status of patients with recurrent GBM.

*J Clin Oncol 18:1481-1491. © 2000 by American Society of Clinical Oncology.*

THE BURDEN IMPOSED on patients' lives by glioblastoma multiforme (GBM) may be described in terms of the impact on their health-related quality of life (HRQOL).<sup>1-3</sup> Although no HRQOL data has been published on the burden immediately after diagnosis, there is information on HRQOL status at first recurrence of the disease after surgery and radiation therapy.<sup>3</sup> The HRQOL status of patients with GBM at recurrence is similar to that of patients with a variety of other cancers who have advanced or metastatic disease.<sup>4</sup>

The treatment of GBM after recurrence is problematic. There is little improvement in length of survival, and the toxicity of the treatment on the patients' well-being has not been systematically described in detail. In this setting, where the treatment is palliative, it seems evident that the measurement of HRQOL should be a primary objective. However, few, if any, formal studies have been performed to determine whether treatment of recurrence by chemotherapy is associated with HRQOL benefit.

We report on the HRQOL status, as measured by the European Organization for Research and Treatment of Cancer (EORTC) Quality of Life Questionnaire (QLQ-C30)<sup>5-9</sup> and a Brain Cancer Module (BCM 20),<sup>10</sup> of patients with GBM who were treated with temozolomide (TMZ) in a phase II study<sup>11</sup> or TMZ or procarbazine (PCB) in a phase III study.<sup>12</sup> The main goals of these studies were to assess

the magnitude of the HRQOL changes reported during the course of treatment and to determine whether there was an association between improvement in HRQOL and improvement in disease status and the treatment given. It was expected that improved disease status would be associated with improvement in most or all of seven preselected HRQOL domains, regardless of the treatment given.

## PATIENTS AND METHODS

The HRQOL measurements were carried out as a component of two studies in patients with recurrent GBM, a phase II study evaluating the effect of TMZ after recurrence and a phase III study comparing the efficacy of TMZ versus PCB.

---

*From Quality of Life Consulting, Vancouver, British Columbia, Canada; Institute of Cancer Research and Royal Marsden National Health Services Trust, Sutton, Surrey, United Kingdom; Department of Neuro-Oncology, M.D. Anderson Cancer Center, Houston, TX; and University of California, San Francisco, CA.*

*Submitted August 2, 1999; accepted December 2, 1999.*

*Supported by Schering-Plough Corporation, Kenilworth, NJ.*

*Address reprint requests to David Osoba, MD, QOL Consulting, 4939 Edendale Ct, West Vancouver, British Columbia, Canada V7W 3H7; email dosoba@bc.sympatico.ca.*

*© 2000 by American Society of Clinical Oncology.*

*0732-183X/00/1807-1481*

### Eligibility Criteria

The eligibility requirements for both studies were similar. Patients were at least 18 years old with a Karnofsky performance score (KPS) of  $\geq 70$ , had a histologically proven supratentorial high-grade glioma, and, at first relapse after external radiation therapy, showed unequivocal evidence of tumor recurrence or progression by gadolinium magnetic resonance imaging (MRI) or contrast-enhanced computed tomography (CT) scanning. The patients were on a stable dose of corticosteroids for at least 3 days before chemotherapy and had an estimated life expectancy of at least 12 weeks. Patients were excluded if they had significant renal, hepatic, or bone marrow impairment, previous chemotherapy with a nitrosourea-containing regimen, PCB, or dacarbazine, or were still experiencing toxicity from any previous therapy, had previous tumors at other sites (excluding basal cell carcinoma), or were human immunodeficiency virus–positive. Other eligibility details unlikely to have an effect on HRQOL status will be presented elsewhere (Yung et al, unpublished data).

### Chemotherapy and Toxicity

Full details of the chemotherapy given are reported elsewhere (Yung et al, unpublished data). In both studies, TMZ was given orally to fasting patients. The dose was 200 mg/m<sup>2</sup>/d (total dose per cycle, 1000 mg/m<sup>2</sup>) for those who had not been given any prior chemotherapy and 150 mg/m<sup>2</sup>/d (total dose per cycle, 750 mg/m<sup>2</sup>) for those who had been given prior chemotherapy. Treatment cycles were repeated every 28 days. In the phase III study, PCB was given in an oral dose of 125 mg/m<sup>2</sup>/d (for those with prior eligible chemotherapy) or 150 mg/m<sup>2</sup>/d (for those with no prior chemotherapy) for 28 consecutive days (days 1 to 28) of each 56-day treatment cycle. Chemotherapy was scheduled for 2 years unless there was disease progression or unacceptable toxicity.

The criteria for reducing dosage or stopping treatment were grade 3 hematologic toxicity or other unacceptable toxicity as determined by the National Cancer Institute common toxicity criteria, progression of disease, or the completion of the prespecified lengths of treatment.

### HRQOL Measurement

HRQOL status was assessed by the EORTC QLQ-C30 (+3)<sup>5</sup> (a later version of the EORTC QLQ-C30<sup>6</sup> that was used to develop version 2.0<sup>5</sup>) and BCM20.<sup>10</sup> Both have been shown previously to be reliable and valid instruments in the setting of recurrent high-grade gliomas.<sup>5,10</sup> The QLQ-C30 (+3) consists of 33 items that form five functioning domains (physical, role, emotional, social, and cognitive), a global quality-of-life (QOL)/overall health domain, three symptom domains (fatigue, pain, and nausea and vomiting), and six single items (dyspnea, diarrhea, constipation, anorexia, insomnia, and financial impact). The version 2.0 scoring system<sup>13</sup> was used in these studies, and the instrument will be referred to as version 2.0 in the remainder of this report. The BCM20 contains 20 items grouped into four domains (future uncertainty, visual disorder, communication deficit, and motor dysfunction) and seven single items (headache, seizure, drowsiness, hair loss, itching, weakness of both legs, and difficulty controlling bladder function). Some of these symptoms (ie, hair loss and itching) were related to toxicity of previous treatment by radiation therapy or side effects of TMZ, whereas the remainder are associated with the recurrence of GBM.<sup>10</sup>

The HRQOL questionnaires (in the appropriate languages) were given to patients for completion before the first cycle of chemotherapy (baseline assessment) and, subsequently, just before each cycle of

chemotherapy for all patients and also before all the analogous nonchemotherapy visits for patients taking PCB.

### Clinical Assessment, Scanning Procedures, and Tumor Status

Using standard criteria, complete response (CR), partial response (PR), stable disease, and progressive disease were determined by CT or MRI scans scheduled every 2 months. A clinical response or progression could be declared by neurologic examination in the intervals between scheduled scanning. In actual practice, fewer than 5% of clinical responses were assessed only by neurologic status.

### Statistical Procedures

The QLQ-C30 was scored according to methods described in the second edition of the EORTC QLQ-C30 Scoring Manual.<sup>13</sup> All raw scores were converted to lie in a range between 0 and 100. For the functioning scales and the global QOL scale, a higher score indicates better functioning, whereas for the symptom scales/items, a higher score indicates more of the symptom or difficulty. The BCM20 was scored in a manner analogous to the QLQ-C30 (version 2.0)<sup>10</sup>; higher scores indicate more of the symptom or more difficulty. The changes in scores between baseline and any subsequent scores were calculated by subtracting each patient's baseline score from his/her subsequent scores. As patients dropped out of the study, the change in scores for each patient was calculated for only those remaining on study. In the TMZ-treated patients, all dropouts were a result of disease progression. In the PCB-treated patients, dropouts were a result of disease progression and toxicity from treatment. If a score was not available at the time that progression was declared, then the previous score (4 weeks earlier) was used as the score nearest progression. (For the exact numbers of patients to whom this applied, see Results.) Changes in HRQOL status were correlated with changes in tumor and disease status to determine whether there was an association between them.

To decrease the possibility of finding statistically or clinically significant associations by chance alone, the clinical investigators decided a priori to limit the number of HRQOL domains that would be examined by significance testing. The seven chosen domains were role functioning, social functioning, global QOL, visual disorder, motor dysfunction, communication deficit, and drowsiness. It was expected that these were likely to be most affected in GBM and that, if significant changes were not detectable during treatment in these domains, then changes in others would be unlikely to be of more clinical importance. Subsequent to examining the changes in the seven prespecified domains, other domains were also examined. A Bonferroni correction for multiple testing was not carried out because this method assumes independence of the variables involved, and HRQOL variables show small to modest correlations.<sup>6-9</sup> Another method is to set the *P* value limit higher than the conventional limit of .05 (eg, .01). We did not do this, however, because we wished to compare the results in terms of conventional *P* values and effect sizes (see below). A direct comparison of the TMZ and PCB groups in the phase III study was hampered by the high attrition rates and, thus, small numbers of patients near the end of the treatment period. To overcome this difficulty, in part, we used changes in scores for each individual from baseline to specified times and determined whether there was improvement over the baseline score in each of the groups.

To examine the change from baseline in HRQOL scores at the time of and before disease progression, the mean change scores at specified times were plotted backward from progression. First, the changes in scores from baseline to the time of progression (as determined by

**Table 1. Patient Characteristics**

Characteristic	Patients in Phase II Study (n = 109)		Patients in Phase III Study (n = 179)			
	TMZ		TMZ (n = 89)		PCB (n = 90)	
	No.	%	No.	%	No.	%
Age, years						
Mean	53.2		51.2		49.3	
Range	24-77		21-72		23-73	
Sex						
Male	64	59	62	70	56	62
Female	45	41	27	30	34	38
KPS						
100	5	5	2	2	10	11
90	36	33	27	30	30	33
80	27	25	32	36	21	23
70	41	38	28	32	28	31
Missing	0	0	0	0	1	1

radiologic imaging) were determined. Then, the change from baseline to the treatment cycle 4 weeks before progression and 8 weeks, 12 weeks, and so on, before progression were also determined. The change scores at progression were plotted at the right side of a bar graph, and the times before progression (-4, -8, -12, and so on) were plotted to the left of the progression time.

A clinically significant change in scores was defined as a change of  $\geq 10$  (on a scale of 0 to 100) lasting for at least two HRQOL assessments 4 weeks apart. This criterion was based on previous studies involving a formal assessment of how much change in QLQ-C30 scores is perceptible to patients (subjectively significant)<sup>14</sup> and how much change is associated with changes in disease status.<sup>15</sup> Other approaches using other questionnaires in other illnesses also indicate that a change of  $\geq 10$  (on a 0 to 100 scale) may be interpreted as being clinically significant.<sup>16-18</sup> Accordingly, the numbers of patients in our studies who were deemed to have experienced a subjectively significant improvement for at least 8 weeks were enumerated to obtain the proportions with improvement in HRQOL status, and the mean durations of change were calculated.

Standard statistical procedures were performed in this study using an SAS statistical package (SAS Institute, Cary, NC). *P* values were calculated for paired comparisons using the Student's *t* test. Effect sizes were calculated by dividing the SD of the mean of the baseline completion score by the mean of the second, third, and so on, completion.<sup>19,20</sup> Effect sizes of 0.2 to 0.5 are considered small, sizes between 0.5 and 0.8 are moderate, and sizes over 0.8 are large.

**RESULTS**

*Patient Characteristics*

One hundred thirty-eight and 225 patients with GBM at the time of their first recurrence were enrolled onto the phase II study and phase III study, respectively. The characteristics of the patients who completed the baseline assessment and at least one additional HRQOL assessment in both studies were similar, except that there was a higher proportion of males receiving TMZ in the phase III study than males taking TMZ in the phase II study or PCB in the

phase III study. Baseline KPS status was similar in all groups, except, in patients taking TMZ in the phase III study, there were more patients with a KPS of 80 than in the other cohorts (Table 1).

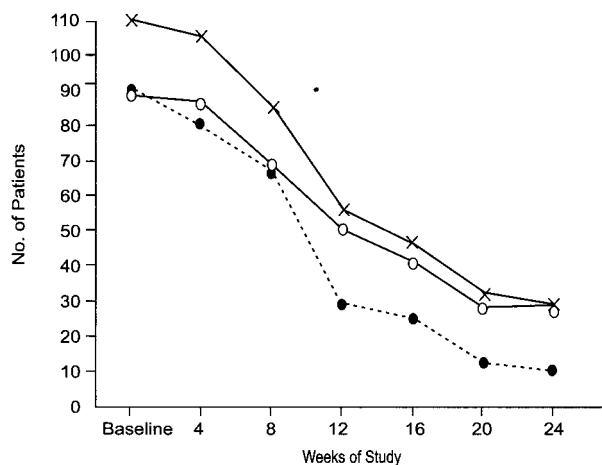
*Baseline HRQOL Scores*

The entire baseline HRQOL characteristics (including all the domains and single items) for both studies are to be presented elsewhere (unpublished data). Here, we focus on the HRQOL scores at baseline for the seven preselected domains and some additional domains of interest (Table 2).

Baseline scores were similar for all groups in both studies, with the exception that, in the phase III study,

**Table 2. Baseline Scores: GBM at Time of First Recurrence**

Domains	Baseline Scores for Phase II Study		Baseline Scores for Phase III Study	
	(n = 105-109) (mean $\pm$ SD)	(n = 86-89) (mean $\pm$ SD)	TMZ (n = 86-89) (mean $\pm$ SD)	PCB (n = 86-90) (mean $\pm$ SD)
Preselected domains				
Role functioning	54.8 $\pm$ 32.9	57.7 $\pm$ 31.6	61.0 $\pm$ 28.2	
Social functioning	63.5 $\pm$ 33.5	59.9 $\pm$ 31.9	60.8 $\pm$ 28.5	
Global QOL	55.5 $\pm$ 23.2	63.0 $\pm$ 20.6	58.6 $\pm$ 22.9	
Visual disorder	19.8 $\pm$ 23.7	22.1 $\pm$ 27.7	17.8 $\pm$ 23.1	
Motor dysfunction	25.8 $\pm$ 25.5	33.4 $\pm$ 26.0	27.8 $\pm$ 24.2	
Communication deficit	27.1 $\pm$ 30.7	35.3 $\pm$ 33.4	26.3 $\pm$ 24.4	
Drowsiness	32.7 $\pm$ 28.4	41.0 $\pm$ 27.7	39.6 $\pm$ 25.9	
Other domains				
Physical functioning	67.2 $\pm$ 28.6	71.6 $\pm$ 29.1	67.9 $\pm$ 29.0	
Emotional functioning	62.9 $\pm$ 26.4	70.5 $\pm$ 19.2	69.2 $\pm$ 20.9	
Cognitive functioning	61.8 $\pm$ 30.5	62.4 $\pm$ 28.5	63.7 $\pm$ 25.7	
Fatigue	39.0 $\pm$ 22.2	40.3 $\pm$ 23.1	42.5 $\pm$ 24.3	
Pain	15.1 $\pm$ 22.6	16.3 $\pm$ 21.3	21.5 $\pm$ 22.2	
Insomnia	21.5 $\pm$ 31.1	22.6 $\pm$ 29.4	23.4 $\pm$ 26.5	
Future uncertainty	39.8 $\pm$ 28.6	43.6 $\pm$ 28.2	40.7 $\pm$ 26.4	



**Fig 1.** HRQOL questionnaire completion rates. Phase II study, TMZ (x); phase III study, TMZ (O), PCB (●). Completion rates are synonymous with numbers of patients on study at the indicated time points.

patients in the TMZ-treated group had a greater communication deficit than did patients in the PCB-treated group and patients in the phase II study. Both groups in the phase III study experienced more drowsiness than did the group in the phase II study.

#### Questionnaire Completion Rates

In the phase II study, of the 138 patients enrolled, 29 provided only baseline scores and 109 had a baseline score plus one or more HRQOL scores while on treatment. In the phase III study, of the 225 patients enrolled, 20 did not provide any HRQOL information, 26 provided only baseline data, and 179 (89 in the TMZ group and 90 in the PCB group) completed the questionnaires at baseline and while on treatment. The reason for failing to collect baseline data was administrative error. During both studies, attrition rates were high, and the numbers of patients remaining on study at 6 months were 29 in the phase II study, 28 in the phase III TMZ group, and 10 in the phase III PCB group (Fig 1). HRQOL data were missing because patients were withdrawn from the study as a result of progressive disease or death. In addition, some patients were withdrawn for reasons of toxicity, and these were mainly in the PCB arm of the study (phase II study: TMZ, three patients [2%]; phase III study: TMZ, three patients [3%] and PCB, 11 patients [10%]). The attrition rates for patients treated with TMZ in both studies were similar, whereas attrition was more rapid for patients treated with PCB (Fig 1). There were no completions missing completely at random.

In 31 (28%) of 109 patients in the phase II study, HRQOL scores were not available at the time of progression. In the phase III study, 22 (25%) of 89 of the patients in the TMZ

arm and 23 (26%) of 90 of patients in the PCB arm did not provide scores at progression. For these patients, the last score before progression was used to determine the association of scores with change in tumor status (see below). However, only scores at progression were used to determine the effects of progression on HRQOL (see below).

#### Association of Changes in HRQOL Scores With Changes in Tumor Status

The relationships between changes in HRQOL scores and changes in disease status were assessed in patients during 6 months of therapy. In the interval from baseline up to and including 6 months, 22 patients who remained progression-free and 87 who experienced progression in the phase II study provided HRQOL data. In the phase III study, 19 progression-free patients and 70 patients with progression and seven progression-free patients and 83 patients with progression provided HRQOL data in the TMZ arm and PCB arm, respectively.

In the TMZ-treated patients, freedom from progression was consistently associated with improvement or stability in the seven preselected HRQOL domains (Fig 2). The exception was in visual disorder in the phase III study, where progression-free patients had deterioration that was not statistically significant but had an effect size of 0.36. Improvement reached statistical significance ( $P \leq .05$ ) for global QOL, communication deficit, and drowsiness in the phase II study; but, in the phase III study, improvement reached statistical significance only for drowsiness. Effect sizes were small (0.2 to 0.4) for most changes between baseline and 6 months, except in global QOL and drowsiness, where effect sizes were moderate (0.48 and 0.56, respectively).

Patients treated with TMZ who had progression of disease (as shown by MRI/CT scan) almost always showed deterioration in the seven HRQOL domains between baseline and the time of progression (Fig 2). The exception was visual disorder in the phase III TMZ-treated patients. The effect sizes were small for changes in role functioning, social functioning, global QOL, and visual disorder and moderate for motor dysfunction and drowsiness. Thus, in patients who were treated with TMZ and whose disease did not progress by 6 months, there was improvement in HRQOL, whereas in patients whose disease progressed by 6 months, there was deterioration in HRQOL.

In domains other than the preselected ones, the most striking changes seemed to be improvement in emotional functioning and reduction of future uncertainty in those patients who were progression-free and a worsening of physical functioning in both groups of patients. However,

### Phase II Study (TMZ)

### Phase III Study (TMZ)

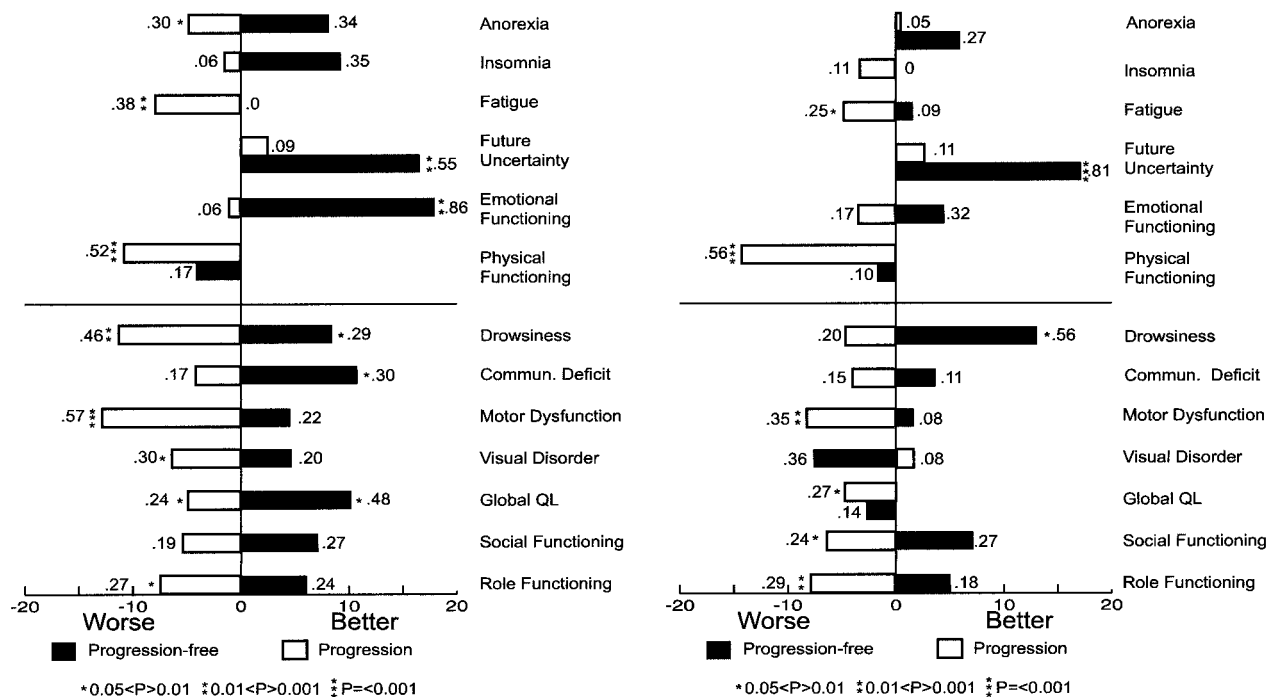


Fig 2. Magnitude of HRQOL score changes from baseline to 6 months for patients treated with TMZ. The seven preselected domains are shown below the horizontal broken line. The P values are denoted by asterisks, and the effect sizes are shown numerically adjacent to each bar. Effect sizes of ≥ 0.2 are considered to be clinically significant.

because these domains had not been chosen a priori, no firm conclusions are possible.

A similar analysis of patients treated with PCB in the phase III study gave results that were in sharp contrast to those reported by patients treated with TMZ. PCB-treated patients who were progression-free at 6 months reported a deterioration of functioning or symptoms in all of the seven preselected domains, and the deterioration was of the same magnitude as reported by patients with progression of disease (Fig 3). Although the number of patients providing HRQOL data who were progression-free is small (seven), which precludes meaningful statistical testing of the null hypotheses, the effect sizes for changes from baseline ranged from small to large. Because effect sizes are only modestly dependent on the number of patients in the sample, they suggest that the deterioration in these patients was significant. Furthermore, the worsening was independent of changes in disease status.

#### Effect of Progression on HRQOL Scores

Each patient's scores at the time of disease progression, regardless of treatment, were determined. Similarly, the

scores of the previous HRQOL completions before the time of progression were also determined, and the changes from baseline were calculated for each patient. At the time of progression, the mean change in all domain scores deteriorated below their baseline levels except the score for visual disorder in patients treated with TMZ in the phase III study (Fig 4). During the weeks preceding progression, when the disease was either responding or stable in a proportion of the patients, the changes in scores from baseline for patients treated with TMZ showed an improvement (communication deficit and drowsiness in both studies and global QOL in the phase II study). There was a tendency toward improvement (motor dysfunction in both studies and role functioning in the phase III study), or no change, or no significant worsening from baseline (social functioning and visual disorder).

The results reported by patients treated with PCB in the phase III study are in marked contrast to those from patients treated with TMZ. The PCB-treated patients, as a group, did not show improvement from baseline in most of the seven domains. In fact, these patients actually showed deterioration that persisted through most of the treatment period,

## Phase III Study (PCB)

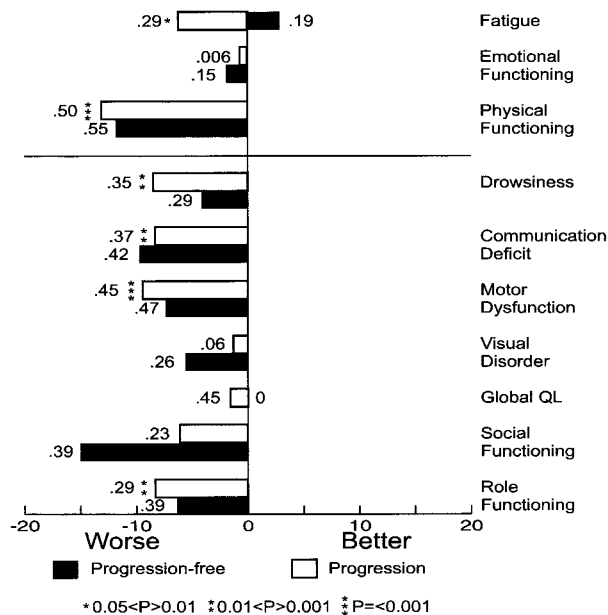


Fig 3. Magnitude of HRQOL score changes from baseline to 6 months for patients treated with PCB. See Fig 2 for footnotes.

with a tendency toward improvement only in the last two visits before disease progression was declared (Fig 4). This improvement is in keeping with the view that the HRQOL scores improved when PCB was stopped.

### Proportions of Patients With Clinically Significant Changes in HRQOL

Previous work in patients with cancer<sup>14,15</sup> indicates that a change in QLQ-C30 scores of  $\geq 10$  (on a 0 to 100 scale) is probably of clinical significance because it is perceived subjectively as a moderate change by patients and also reflects differences between stages of disease. In our studies, the proportions of patients experiencing a change of  $\geq 10$  in the scores of the seven preselected domains were calculated. Because one of the aims of this study was to determine whether there was improvement in HRQOL during treatment, only patients whose baseline scores were  $\leq 90$  for the functioning scales and  $\geq 10$  for the symptom scales were included in the analysis. Patients at the maximum and minimum scores (100 for functioning scales and 0 for symptom scales) would not have had an opportunity for improvement. Furthermore, to minimize chance variation, only those in whom improvement lasted  $\geq 8$  weeks were enumerated. Of the TMZ-treated patients who were eligible for improvement, the proportions

reporting improvement varied from a low of 15% (for global QOL) to a high of 40% (for communication deficit). For PCB-treated patients, the proportions were lower, varying from 14% (for drowsiness) to 24% (for motor dysfunction) (Table 3).

### Duration of HRQOL Improvement

The duration of an HRQOL response was calculated for each patient if the criterion of an improvement of  $\geq 10$  points lasting  $\geq 8$  weeks in eligible patients was met. The mean duration of response ( $\pm$  SD) was then calculated in weeks for the subgroups of patients in each treatment arm. We chose to show the mean rather than the median duration because of the small numbers of patients involved and because the median would not reflect the long duration of response in some patients. For example, when we calculated the median duration, it was at the minimum threshold level (8 weeks) for all seven domains in the patients treated with PCB, but the range of the duration was 8 to 36 weeks.

The mean duration of response was greater in patients treated with TMZ than in patients treated with PCB, except for visual disorder, which was similar for both subgroups in the phase III study (Table 4). As might be expected, the duration of HRQOL response (mean  $\pm$  SD) was longest in patients achieving a CR/PR ( $25$  to  $33 \pm 13$  to  $15$  weeks), was a little shorter in patients with stable disease ( $14$  to  $23 \pm 6$  to  $13$  weeks), and shortest in patients with progressive disease ( $8$  to  $10 \pm 0$  to  $4$  weeks).

## DISCUSSION

The measurement of HRQOL outcomes in recurrent GBM is still new, but it is highly appropriate. Because chemotherapy for recurrence has modest effects on overall survival, achieving an improvement in HRQOL could be an important benefit to patients. It is particularly necessary to be certain that the benefits of improved HRQOL in some domains are not offset by deterioration in others as a result of side effects of treatment.

We wished to avoid making a large number of multiple comparisons during the analysis of our data. Therefore, at the outset, we chose the seven domains that we thought would be most likely to reflect the clinical changes that experienced neuro-oncologists have observed in clinical practice. Thus, we did not make comparisons of scores in all possible domains and apply a Bonferroni correction. In retrospect, the numerical differences in some of the domain scores suggest that a few domains other than those selected may also have been of interest. Therefore, the statistical analysis of these domains have been shown in two of the figures (Figs 2 and 3), but we caution against firm conclusions from these data because further studies will need to be

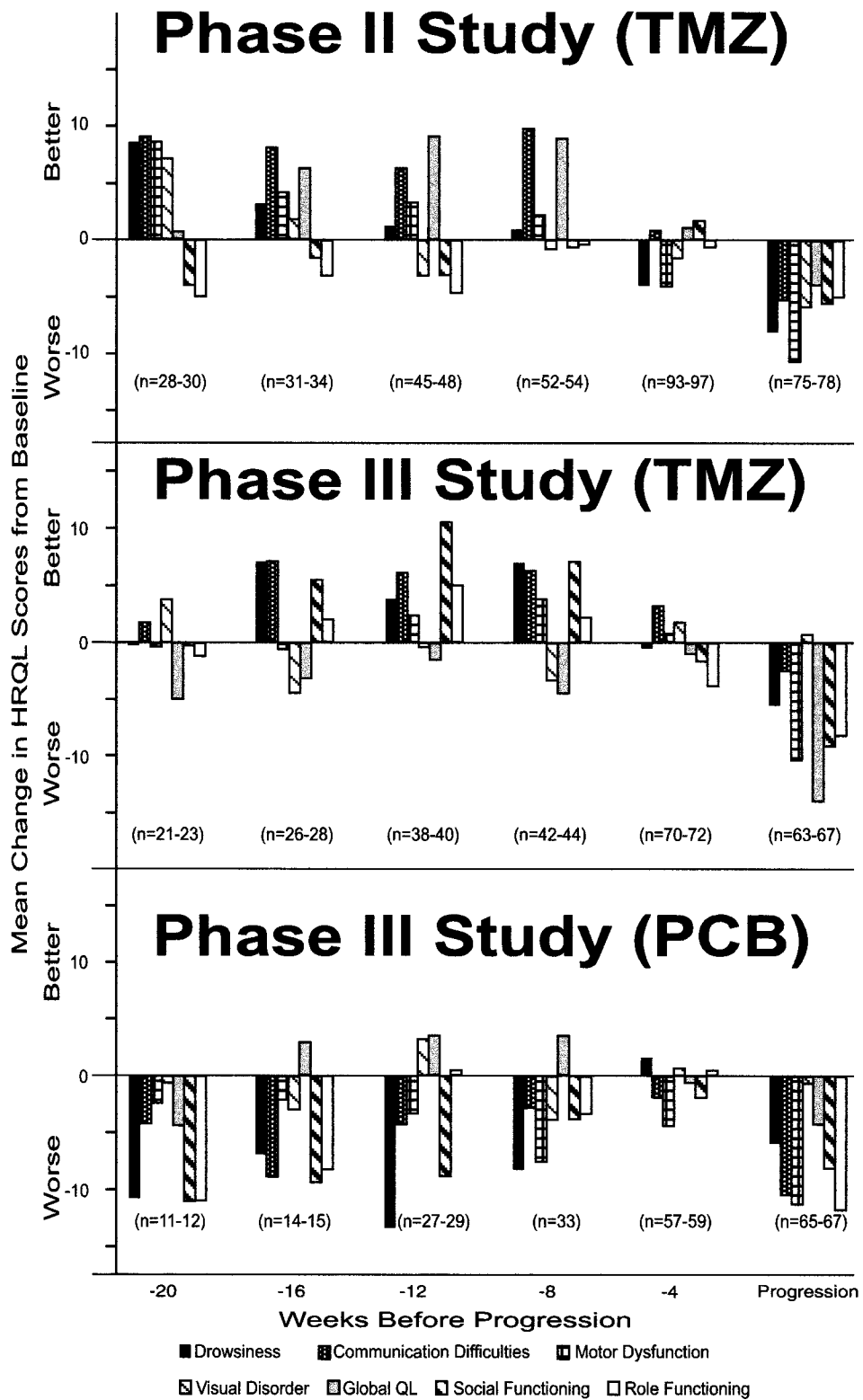


Fig 4. Changes in HRQOL scores from baseline to times before disease progression in all patients treated with TMZ and PCB. Changes in scores from baseline to time of progression are plotted at the right-hand side of each panel; scores for the times before progression are then plotted from right to left.

**Table 3. Proportions of Patients Who Reported Improvement of  $\geq 10$  in Preselected HRQOL Scores**

Domain	Patients in Phase II Study			Patients in Phase III Study					
	TMZ			TMZ			PCB		
	No. Eligible*	No.	%	No. Eligible*	No.	%	No. Eligible*	No.	%
Role functioning	84	23	27†	65	17	27	67	12	18
Social functioning	74	29	39	57	18	32	66	12	18
Global QOL	97	25	26	66	10	15	74	13	18
Visual disorder	61	23	38	44	15	34	46	11	24
Motor dysfunction	75	20	27	63	20	32	66	12	18
Communication deficit	67	25	37	57	23	40	57	11	19
Drowsiness	71	16	23	64	12	19	65	9	14

\*Patients with maximum (100) functioning scores and minimum (0) baseline symptom scores are excluded because improvement for these groups was not possible.

†The proportions of patients with no change or deterioration in HRQOL scores is not 100 minus the proportion shown above with improvement because all patients were eligible for deterioration.

performed in which these additional domains are selected at the outset.

A difficulty in clinical trials with rapid attrition rates caused by high treatment failure rates and short survival is that attrition may be unequal in the arms of the study. When there is a high proportion of missing HRQOL data at preset analysis time points (such as with each cycle of chemotherapy) because of patient drop-out as a result of progressive disease and the attrition rates differ in each arm of a study, it is difficult to make comparisons between the arms. It can be argued that the patients who possessed the most important determinants for survival will be on the study longest and will be the most likely to display the best HRQOL results. Thus, a comparison of patients in a two-arm study will be a comparison of best with best and not truly reflect the HRQOL of all the patients in each arm. However, if long-surviving patients are deriving any net benefit in HRQOL, then they should show, at the very least, an improvement in their on-treatment or posttreatment HRQOL scores compared with their pretreatment scores.

Therefore, we used each patient as his/her own control to determine whether those patients who continued on treatment had improved HRQOL scores compared with their own baseline scores.

Because virtually all patients with recurrent GBM will eventually experience progression of disease at some point, analysis of HRQOL data before disease progression in all patients seems reasonable. Each time point before progression, as displayed in Fig 4, represents a mixture of patients with a clinical response or stable disease and patients about to have disease progression. These proportions will change with time during the trial as more and more patients develop progression. Our analysis tried to account for these changes by analyzing the scores of patients at each treatment cycle preceding progression. The data for TMZ-treated patients strongly support the notion that HRQOL is either improved or preserved if the disease does not progress and that maintenance of patients in a state of freedom from progression is of benefit to them. However, this notion is not supported by the data from the PCB-treated patients. They

**Table 4. Mean Duration of HRQOL Responses\* for Seven Preselected Domains**

Domain	Phase II Study		Phase III Study			
	TMZ		TMZ		PCB	
	No. of Patients	Duration of Response (weeks, mean $\pm$ SD)	No. of Patients	Duration of Response (weeks, mean $\pm$ SD)	No. of Patients	Duration of Response (weeks, mean $\pm$ SD)
Role functioning	23	16.7 $\pm$ 10.0	21	15.0 $\pm$ 9.7	13	11.1 $\pm$ 4.4
Social functioning	29	16.8 $\pm$ 11.0	24	17.5 $\pm$ 10.6	13	11.1 $\pm$ 4.1
Global QOL	25	21.6 $\pm$ 12.7	14	16.0 $\pm$ 14.7	13	10.2 $\pm$ 3.1
Visual disorder	23	16.3 $\pm$ 10.8	17	11.3 $\pm$ 4.1	11	12.7 $\pm$ 8.4
Motor dysfunction	19	19.4 $\pm$ 12.5	24	16.7 $\pm$ 14.1	12	10.7 $\pm$ 3.6
Communication deficit	24	19.3 $\pm$ 10.4	29	18.1 $\pm$ 13.1	13	9.8 $\pm$ 3.1
Drowsiness	16	16.0 $\pm$ 9.5	16	12.5 $\pm$ 5.6	11	10.9 $\pm$ 4.4

\*Improvement in scores of  $\geq 10$  (10 to 100 scale) for patients whose baseline functioning scale scores were  $\leq 90$  and symptom scale scores were  $\geq 10$ . Duration of improvement had to be a minimum of 8 weeks in all groups before it was deemed to constitute an HRQOL response.

experienced deterioration in most domain scores despite nonprogression of disease. Thus, there seemed to be some other factor affecting their HRQOL, and this likely was the toxicity associated with taking PCB. Thus, the minimal toxicity associated with TMZ treatment was also important, in addition to regression or nonprogression of disease, for achieving the improvements in HRQOL seen in this population.

Although there was some toxicity reported by patients taking TMZ,<sup>12</sup> there were sufficient numbers of patients with either improvement or stability of scores in the seven preselected HRQOL domains to counteract any worsening of HRQOL scores as a result of toxicity. Thus, the net result was a benefit, as shown in the group scores. This interpretation is also reinforced by the longer durations of HRQOL responses seen in the patients treated with TMZ compared with patients treated with PCB.

As expected, patients who were treated with TMZ and achieved CR/PR or stable disease had longer durations of HRQOL responses. On the one hand, this strongly supports the medical paradigm that achieving tumor regression is an important end point, even though it may not be always associated with improvement in survival. An alternative explanation, however, is that patients experiencing nonprogression of disease may have the opportunity of achieving longer durations of HRQOL benefit simply because they have gained time in which to experience benefit.

Progression of disease was associated with a sharp deterioration in HRQOL scores. However, during the visit 4 weeks before disease progression was declared, there was evidence of a worsening of group scores (ie, to approximately pretreatment levels) compared with the scores on previous visits in the TMZ-treated group. This suggests that deterioration in HRQOL scores may have signaled disease progression before it was declared on the basis of evidence from CT and MRI scans. One explanation for this finding is that regrowth of the tumors may have started but had not yet reached the increase of 25% required to be called tumor progression. Another possibility is that this result may be explained by the study design. HRQOL assessments were completed every 4 weeks, but scans were performed every 8 weeks. Thus, it is possible that recurrence on scans would have been detected earlier had they been performed more frequently. More work will be required to resolve this issue.

As a group, patients treated with PCB did not experience the beneficial HRQOL changes during treatment seen in their counterparts treated with TMZ. This was also true for those patients who had improvement or nonprogression of disease on MRI/CT scans. An apparent paradoxical result was that the HRQOL scores were better at 4 weeks before progression of disease than at earlier times. However, this

was likely the result of some of the patients having been off treatment at 4 weeks before progression because of the 8-week cycle of PCB. Thus, it is thought to represent recovery from PCB toxicity. If this observation is correct, then this study is an example of a situation where a decrease in tumor size is not necessarily accompanied by improvement in HRQOL. It serves as a reminder that the concept that tumor regression is a surrogate for improvement in HRQOL will not hold if there is a counter-balancing increase in the toxicity of therapy that outweighs any benefit.

Another analysis problem may arise when one of the two groups of data being compared for statistically significant differences is small and the other one is large. Tests of the null hypothesis depend primarily on sample size, and, if one of the samples is small, no statistically significant difference will be found even if the numerical difference between mean scores seems to be large. Also, because statistically significant *P* values are highly dependent on sample size, very large samples can give rise to very small *P* values despite rather small numerical differences. Alternatively, effect sizes are dependent primarily on the degree of variation (SD) in the scores and are not as sample size-dependent as *P* values. Thus, they are useful for comparing differences between baseline and on-treatment or posttreatment scores in small samples resulting from attrition during the study; they are not a substitute for enrolling appropriate sample sizes at the onset of the study.<sup>16,17</sup> We used effect sizes as well as *P* values to compare the scores at 6 months with pretreatment (baseline scores) in each group of patients. We found, in both studies, that patients treated with TMZ whose disease had not progressed at 6 months showed improvement in the majority of preselected HRQOL scores over their baseline HRQOL scores. However, statistically significant improvements, as determined by *P* values, were seen in only three of the seven domains in the phase II study and in one domain in the phase III study. Effect sizes were significant ( $\geq 0.2$ ) in all of these domains in the phase II study and two of the domains in the phase III study.

Although none of the mean changes in HRQOL scores deteriorated in patients treated with TMZ who remained progression-free at 6 months in the phase II study, the patients with progression on MRI/CT scans all showed deterioration of scores in the seven preselected domains. Deterioration was statistically significant in five of the seven domains and effect sizes were significant in all five domains. In the phase III study, there was significant deterioration in the visual disorder domain as judged by effect size in the patients with nonprogression of disease. In patients with disease progression, HRQOL scores deteriorated statistically significantly in four of the seven domains, whereas

effect sizes were significant in five of the domains. In patients treated with PCB whose disease did not progress, there was no statistically significant deterioration in HRQOL scores using *P* values, but effect sizes were significant in all seven domains. For those with disease progression, the deterioration in four of the seven domains was statistically significant by *P* values and significant in six of the domains by effect sizes.

Although effect sizes were usually small or moderate for the preselected domains, the differences between those patients who were progression-free and those who were not is probably large, indicating that HRQOL status was distinctly different in those groups of patients. Also, as discussed above, the group with progression of disease was much larger than the group with nonprogression. Thus, tests of the null hypothesis are significant only in the larger group, whereas they were not significant in the much smaller group of patients with nonprogression. In contrast, effect sizes indicated significant changes in the smaller group of patients as well. Using both criteria, patients treated with TMZ who remained progression-free derived significant HRQOL benefits, particularly when compared with the deterioration reported by patients with disease progression.

The proportions of patients with an HRQOL response were higher and the durations of responses were longer in the patients treated with TMZ compared with patients treated with PCB. These findings are complementary to the results discussed earlier. Together, they provide strong evidence that TMZ is preferable to PCB in the treatment of first recurrence of GBM. Indeed, PCB is not only of no benefit but may actually be deleterious to patients' HRQOL.

There may be some difficulty in understanding the exact meaning of what actually constitutes improvement or deterioration in each of the various HRQOL domains if the user of the results is not familiar with the composition of the questions (items) in each of the domains. In this study, the EORTC QLQ-C30 (version 2.0) domains that showed the greatest separations between patients free of progression and those with progression by 6 months were role functioning, social functioning, and global QOL. The domains from the BCM20 showing the greatest separation were motor dysfunction, communication deficit, drowsiness, and future uncertainty. A listing of the abbreviated content of the items that are included in each of these domains is given in Table 5. For discussion here, the meaning of the differences in scores for role functioning will be used as an example. The items included in the role functioning domain ask about limitations in working or other daily activities and about limitations in pursuing hobbies or other leisure time activities. Thus, the difference in the role functioning scores

**Table 5. Abbreviated Content of Items Included in Domain**

Preselected Domains	Items Dealt With in Domain
Role functioning	Limitations in doing work or other daily activities, and in pursuing hobbies or other leisure time activities
Social functioning	Interference with family life and with social activities
Overall health and global QOL	Ratings of overall health and of overall QOL
Visual disorder	Double vision, blurred vision, and difficulty reading because of vision
Motor dysfunction	Weakness on one side of the body, trouble with coordination, and feeling unsteady on feet
Communication deficit	Finding the right words, difficulty speaking, and trouble communicating thoughts
Drowsiness	Feeling drowsy during the daytime

between patients with and without disease progression indicates that those without progressive disease were able to work, perform other daily activities, and pursue hobbies or leisure time activities to a greater extent than those who had disease progression. Furthermore, patients with disease progression reported a decrease in these activities, whereas those who remained progression-free reported an improvement in these activities during treatment of their disease. Another implied conclusion is that the side effects of treatment with TMZ were not severe enough to cause sufficient deterioration to significantly decrease the observed improvement in HRQOL.

An increase in shortness of breath (dyspnea) in patients who did not progress was an unexpected finding. It may be a chance event, but its increased severity in relation to the length of time on study and, hence, possibly to treatment, along with the unlikelihood that GBM metastasized to the lungs suggest that this finding needs to be monitored in future studies.

We conclude that TMZ used for treatment of recurrent GBM after surgery/radiation of the primary disease was of benefit if patients remained free of disease progression. Their HRQOL status was improved or, at the least, was preserved until the time of disease progression. At disease progression, HRQOL clearly deteriorated. In the phase III study, where patients were randomized to receive either TMZ or PCB, improvement in both progression-free survival at 6 months<sup>12</sup> and HRQOL favors TMZ over PCB as a treatment for the recurrence of GBM.

#### ACKNOWLEDGMENT

We thank Jill Vardy for assistance with assembling the manuscript.

## REFERENCES

1. Mackworth N, Fobair P, Prados MD: Quality of life self-reports from 100 brain tumor patients: Comparisons with Karnofsky Performance Scores. *J Neurooncol* 14:243-253, 1992
2. Weitzner MA, Meyers CA, Gelke CK, et al: The Functional Assessment of Cancer Therapy (FACT) Scale: Development of a brain subscale and revalidation of the general version (FACT-G) in patients with primary brain tumors. *Cancer* 75:1151-1161, 1995
3. Osoba D, Aaronson NK, Muller M, et al: Effect of neurological dysfunction on health-related quality of life in patients with high-grade glioma. *J Neurooncol* 31:263-278, 1997
4. Brada M, Yung WKA, Osoba D: Symptom burden in patients (pts) with recurrent high-grade gliomas. *Proc Am Soc Clin Oncol* 18:147a, 1999 (abstr 564)
5. Osoba D, Aaronson NK, Zee B, et al: Modification of the EORTC QLQ-C30 (version 2.0) based on content validity and reliability testing in large samples of patients with cancer. *Qual Life Res* 6:103-108, 1997
6. Aaronson NK, Ahmedzai S, Bullinger M, et al: The European Organization for Research and Treatment of Cancer QLQ-C30: A quality-of-life instrument for use in international trials in oncology. *J Natl Cancer Inst* 85:365-376, 1993
7. Osoba D, Zee B, Pater J, et al: Psychometric properties and responsiveness of the EORTC Quality of Life Questionnaire (QLQ-C30) in patients with breast, ovarian and lung cancer. *Qual Life Res* 3:353-364, 1994
8. Ringdal GI, Ringdal K: Testing the EORTC Quality of Life Questionnaire on cancer patients with heterogenous diagnoses. *Qual Life Res* 2:129-140, 1993
9. Niezgoda HE, Pater JL: A validation study of the domains of the core EORTC Quality of Life Questionnaire. *Qual Life Res* 2:319-325, 1993
10. Osoba D, Aaronson NK, Muller K, et al: The development and psychometric validation of a brain cancer quality-of-life questionnaire for use in combination with general cancer-specific questionnaires. *Qual Life Res* 5:139-150, 1996
11. Osoba D: Approaches to assessing health-related quality of life in clinical trials for patients with high-grade recurrent gliomas. *Qual Life Res* 6:695, 1997 (abstr 274)
12. Yung A, Levin VA, Albright A, et al: Randomized trial of Temodal (TEM) vs. procarbazine (PCB) in glioblastoma multiforme (GBM) at first relapse. *Proc Am Soc Clin Oncol* 18:139a, 1999 (abstr 532)
13. Fayers P, Aaronson N, Bjordal K, et al: EORTC QLQ-C30 Scoring Manual. Brussels, Belgium, Quality of Life Unit, European Organization for Research and Treatment of Cancer Data Centre, 1995
14. Osoba D, Rodrigues G, Myles J, et al: Interpreting the significance of changes in health-related quality-of-life-scores. *J Clin Oncol* 16:139-144, 1998
15. King MT: The interpretation of scores from the EORTC quality of life questionnaire QLQ-C30. *Qual Life Res* 5:555-567, 1996
16. Ware JE, Kosinski M, Keller SD: SF-36 Physical and Mental Health Summary Scales: A User's Manual. Boston, MA, The Health Institute, New England Medical Center, 1994
17. Jaeschke R, Singer J, Guyatt GH: Measurement of health status: Ascertaining the minimal clinically important difference. *Control Clin Trials* 10:407-415, 1989
18. Juniper EF, Guyatt GH, Willan A, et al: Determining a minimal important change in a disease-specific quality of life questionnaire. *J Clin Epidemiol* 47:81-87, 1994
19. Kazis LE, Anderson JJ, Meenan RF: Effect sizes for interpreting changes in health status. *Med Care* 27:S178-S189, 1989 (suppl 3)
20. Cohen J: *Statistical Power Analysis for the Behavioural Sciences*. New York, NY, Academic Press, 1977