Progression-free survival as a surrogate endpoint for overall survival in glioblastoma: a literature-based meta-analysis from 91 trials

Kelong Han, Melanie Ren, Wolfgang Wick, Lauren Abrey, Asha Das, Jin Jin, and David A. Reardon

Genentech, South San Francisco, California (K.H., M.R., A.D., J.J.); University Medical Center & DKFZ, Heidelberg, Germany (W.W.); F. Hoffmann–La Roche, Basel, Switzerland (L.A.); Dana-Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts (D.A.R.)

Corresponding Author: Kelong Han, PhD, Genentech Inc, 1 DNA Way, South San Francisco, CA 94080 (han.kelong@gene.com).

Background. The aim of this study was to determine correlations between progression-free survival (PFS) and the objective response rate (ORR) with overall survival (OS) in glioblastoma and to evaluate their potential use as surrogates for OS.

Method. Published glioblastoma trials reporting OS and ORR and/or PFS with sufficient detail were included in correlative analyses using weighted linear regression.

Results. Of 274 published unique glioblastoma trials, 91 were included. PFS and OS hazard ratios were strongly correlated; $R^2 = 0.92$ (95% confidence interval [CI], 0.71–0.99). Linear regression determined that a 10% PFS risk reduction would yield an 8.1% ± 0.8% OS risk reduction. $R^2$ between median PFS and median OS was 0.70 (95% CI, 0.59–0.79), with a higher value in trials using Response Assessment in Neuro-Oncology (RANO; $R^2 = 0.96$, $n = 8$) versus Macdonald criteria ($R^2 = 0.70$; $n = 83$). No significant differences were demonstrated between temozolomide- and bevacizumab-containing regimens ($P = .10$) or between trials using RANO and Macdonald criteria ($P = .49$). The regression line slope between median PFS and OS was significantly higher in newly diagnosed versus recurrent disease (0.58 vs 0.35, $P = .04$). $R^2$ for 6-month PFS with 1-year OS and median OS were 0.60 (95% CI, 0.37–0.77) and 0.64 (95% CI, 0.42–0.77), respectively. Objective response rate and OS were poorly correlated ($R^2 = 0.22$).

Conclusion. In glioblastoma, PFS and OS are strongly correlated, indicating that PFS may be an appropriate surrogate for OS. Compared with OS, PFS offers earlier assessment and higher statistical power at the time of analysis.

Keywords: glioblastoma, meta-analysis, overall survival, progression-free survival, regression, response rate, surrogate endpoint.
analyzed in glioblastoma. A recent pooled analysis focused on phase I and single-arm phase II trials, with specific treatments, and evaluated the PFS/OS correlation at the individual level in 5 glioblastoma trials. While good individual-level correlation was demonstrated, the small sample sizes precluded any conclusions regarding correlation at trial level. Our analysis evaluates the validity of PFS and ORR as surrogate endpoints for OS using a meta-analysis of completed published phase II and III glioblastoma trials and consideration of a greater range of variables than previously evaluated.

Materials and Methods

Literature Search and Data Extraction

Completed phase II, III, and IV trials in glioblastoma published between January 1, 1991 and June 4, 2012 were identified through a systematic search on MEDLINE/PubMed and Trialtrove (Citeline) using the following keywords: “oncology and CNS” OR “glioblastoma” OR “GBM” OR “glioblastoma multiforme” AND “survival” OR “PFS” OR “progression free survival” OR “progression-free survival” OR “overall survival” OR “OS” OR “progression”. Relevant sources identified in the bibliographies of reviewed papers were also included. Publications reporting PFS and OS data from unique glioblastoma trials utilizing standard tumor response criteria were included. Abstracts and other nonjournal publications were included if sufficient detail was provided. Duplicate publications of the same trial, pediatric studies, non–English language papers, sources lacking methodology detail, and review/summary papers were excluded. The analysis was performed using the original authors’ and per protocol endpoint definitions. When available, hazard ratios (HRs) for PFS and OS were recorded. Treatment, patient, and clinical endpoint data from each study were included in the database. Endpoints of interest were OS, PFS, and ORR. Due to the small number of glioblastoma trials available, the analysis was not limited to randomized trials, and studies with high-grade glioma (World Health Organization [WHO] grades III and IV) patient populations were included; these patients were denoted as “All” in the database, since they contained both glioblastoma and anaplastic glioma histologic subpopulations.

Statistical Methods and Analysis

Weighted linear regression analysis through the origin of the plot was used to evaluate correlation between the following pairs of endpoints: (i) HR in PFS and OS, (ii) median PFS (mPFS) and median OS (mOS), (iii) ORR and mOS, (iv) 6-month PFS and mOS, and (v) 6-month PFS and 1-year OS. Pearson’s $R^2$ coefficient was used as a correlation measure between these endpoints. An $R^2$ value of greater than 0.9 is an indicator for a strong correlation, 0.89 to 0.6 for a good to moderate correlation, and below 0.6 for a weak correlation. Points were weighted by the number of patients in the intent-to-treat population. Linear scale or log scale was selected based on the data distribution. Confidence intervals (CIs) for $R^2$ and the weighted fit were calculated using a bootstrapping method (resampling 1000 times) assuming a sample size of 60 patients in additional trials (mean sample size per arm in the included glioblastoma trials).

Differences in the PFS/OS correlations by response criteria that include tumor response and clinical symptomatology (Macdonald, Response Assessment in Neuro-Oncology [RANO]), treatments (TMZ- or bevacizumab [BEV]-containing regimens), line settings (newly diagnosed, recurrent), and histology (glioblastoma only, mixed high-grade glioma populations) were evaluated using analysis of covariance for weighted linear regression. Glioblastoma-only trials were classified as those with ≥95% glioblastoma histology in actual accrual. Statistical analyses were performed using R 2.14.2. Lead-time that could be gained by using PFS instead of OS as the endpoint was calculated as mOS minus mPFS for each arm and compared with other cancer types. Digitization was performed using Plot Digitizer 2.6.2 (plotdigitizer.sourceforge.net).

Results

Database

A total of 274 references were identified in the initial search. After review, 183 were excluded: 139 due to insufficient information or publication type, 11 due to lack of survival data, 9 review/summary papers, 8 pediatric studies, 8 similar papers for the same studies, and 8 due to nonstandard criteria, unclear methodology, or non–English language. Ninety-one unique trials were eligible for the meta-analysis, including 7125 patients and 115 study arms (Table 1). Of these, 2 were published in nonjournal sources: 1 abstract and 1 trial summary. Eleven HR pairs for PFS and OS were available from 10 trials; historical data were used for the control arm to calculate HR in 3 single-arm trials, but in each case, the historical patient population had similar baseline and treatment characteristics compared with the study population. None of these 10 trials contained any agents targeting vascular endothelial growth factor (VEGF) or VEGF receptor (VEGFR).

Correlation Between HR for PFS and OS

$R^2$ was 0.92 (95% CI, 0.71–0.99) for the weighted linear regression of HR in OS as a function of HR in PFS (Fig. 1), indicating strong correlation. Linear regression demonstrated that a 10% risk reduction for PFS would yield an 8.1% ± 0.8% risk reduction for OS.

Correlation Between Median PFS and OS

There was a good correlation between mPFS and mOS, with an $R^2$ of 0.70 (95% CI, 0.59–0.79; Fig. 2A). There was no significant difference in the slope of the regression line between trials using Macdonald and RANO response criteria ($P = .49$), with good correlation between mPFS and mOS observed with both (Fig. 2B). Some trials (eg, Friedman et al. used criteria that were similar to RANO and thus were classified as RANO criteria. When the correlation between mPFS and mOS was evaluated by treatments, $R^2$ was 0.70 (95% CI, 0.50–0.85) and 0.75 (95% CI, 0.60–0.86) for TMZ-containing and non-TMZ-containing regimens, respectively (Fig. 2C). No significant difference in the slope of the regression line was demonstrated between these 2 treatment types ($P = .10$). There was no significant difference in the slope of the regression line between mPFS and mOS and between line settings (newly diagnosed vs recurrent, $P = .04$; Fig. 2E) and between histology types (glioblastoma only vs mixed histology, $P = .02$; Fig. 2F). When the correlation between mPFS and mOS was evaluated in trials conducted at different time periods (1991 to present), no significant difference in the slope of the regression line was demonstrated.
Correlation Between Other Endpoints and OS

Objective response rate was poorly correlated with mOS ($R^2 = 0.22$; Fig. 3). For 6-month PFS versus 1-year OS (by study arm), $R^2$ was 0.60 (95% CI, 0.37–0.77), indicating a moderate correlation between the 2 survival rates (Fig. 4A). The correlation between 6-month PFS and mOS yielded an $R^2$ of 0.64 (95% CI, 0.42–0.77; Fig. 4B).

Lead-time Analysis

The lead-time that could be gained by using PFS instead of OS as the endpoint averaged 7.4 months (max 17.6 mo) and 4.2 months (max 8.1 mo) in newly diagnosed and recurrent cases, respectively (Fig. 5). The lead-time increased with increasing mOS: for newly diagnosed cases, it increased from 6–7 months for a mOS of 1 year to ~9–10 months for a mOS of 1.5 years; for recurrent patients, it increased from 3–4 months for a mOS of half a year to ~5–6 months for a mOS of 9 months.

Discussion

This is a systematic evaluation of whether PFS is an appropriate surrogate endpoint for OS in glioblastoma clinical trials. We assembled the largest literature glioblastoma trial database to date, which included almost all published glioblastoma trials (phase II and beyond) since 1991, as well as the latest advances in treatment...
Neuro-Oncology 699

The lead-time in newly diagnosed glioblastoma is comparable to using PFS instead of OS as an endpoint in glioblastoma. Notably, we demonstrated that a significant lead-time benefit is achieved for OS. Firstly, PFS offers the opportunity for early assessment. The sizes of the symbols are proportional to the number of patients included in the trial for this and subsequent figures. Treatments (year of publication) are: 1. procarbazine vs TMZ (2000)\(^8\); 2. carmustine plus radiotherapy (RT) with or without cisplatin (2003)\(^9\); 3. cis-retinoic acid vs thalidomide, both in combination with TMZ and RT (2005)\(^10\); 4. cilengitide 500 mg vs 2000 mg (2008)\(^11\); 5. TMZ plus RT with or without pegylated liposomal doxorubicin (2009)\(^12\); 6. erlotinib vs thalidomide/cis-retinoic acid, all in combination with TMZ and RT (2009)\(^13\); 7. RT with or without TMZ (2009)\(^14\); 8. procarbazine, lomustine, and vincristine vs TMZ (2010)\(^15\); 9. TMZ 200 mg/m\(^2\) for 5 days versus 100 mg/m\(^2\) for 21 days (2010)\(^16\); 10. hydroxyurea with or without imatinib (2010)\(^17\); 11. enzastaurin vs lomustine, modified Levin criterion (2010).\(^18\) All trials used Macdonald or RANO criteria except for number 11.

There are several advantages of using PFS as a surrogate endpoint for OS. Firstly, PFS offers the opportunity for early assessment. We demonstrated that a significant lead-time benefit is achieved using PFS instead of OS as an endpoint in glioblastoma. Notably, the lead-time in newly diagnosed glioblastoma is comparable to that in metastatic colorectal carcinoma (mean \(\approx 8.5\) mo, max \(\approx 13.8\) mo) based on digitized data.\(^1\) Secondly, delaying progression may represent a clinically significant benefit for glioblastoma patients. Glioblastoma is a highly infiltrative and destructive tumor that generates significant peritumoral edema and mass effect. Progressive underlying tumor is frequently associated with new or worsening neurologic deficits, which may in turn impact overall function and quality of life. Thirdly, PFS offers higher statistical power, since more PFS events usually have occurred by the time of analysis than OS events, especially given the significant PFS lead-time in glioblastoma. Notably, in our analysis, the observed percent risk reduction for PFS was higher than the percent risk reduction for OS, indicating that PFS is a more sensitive endpoint for treatment effect. Finally, PFS is independent of subsequent postprogression treatment.

However, the use of PFS is associated with several limitations that must also be considered. Firstly, it is important to standardize response criteria; a number of standard criteria are being used in glioblastoma trials and have been used in trials building the basis of our knowledge, such as Macdonald, Levin, Response Evaluation Criteria In Solid Tumors (RECIST), and RANO. Secondly, the discrepancy between the time of clinical event (progression or death) and radiologic assessment could be a confounding factor, and therefore the time interval between clinical and radiologic assessments should be minimized and consistent across studies. Thirdly, the association between radiologic progression, clinical benefit, and quality of life remains open for discussion.

The debate about the definition of a valid surrogate endpoint is ongoing, with many proposals under consideration.\(^19\) For example, it has been proposed that the conclusion of the statistical test based on the surrogate endpoint should be consistent with that based on the gold standard endpoint, and/or the treatment effect on the surrogate endpoint should predict the treatment effect on the gold standard endpoint.\(^20\) However, there is a general consensus that there should be good correlation between the surrogate and the gold standard endpoints and that the treatment effect on the gold standard endpoint should be captured by the surrogate endpoint.\(^21\)

Our analysis demonstrated that the percentage risk reduction calculated from the HR of PFS is highly correlated with the percentage risk reduction calculated from the HR of OS in glioblastoma trials, indicating that the treatment effect on PFS can predict the treatment effect on OS in glioblastoma. A great portion (92%) of variability in OS difference can be explained by the PFS difference (\(R^2 = 0.92\)). Notably, the 95% CI and prediction interval were relatively narrow. Median PFS and OS were also well correlated (\(R^2 = 0.70\)). Taken together, these results lend substantial support to the use of PFS as a surrogate endpoint for OS in glioblastoma trials.

While the validity of PFS as a surrogate for OS can be demonstrated in some tumor types, the effect is not consistent.\(^22\) Broglio et al.\(^23\) attributed differences to variations in survival post-progression (SPP), where SPP is the time difference between OS and PFS. Overall survival in cancers with long SPPs are more affected by the presence of confounding factors and therefore demonstrate weaker correlations with PFS. Glioblastoma patients have a median SPP of around 7 months,\(^24\) and therefore the PFS versus OS correlation should be fairly strong. Our analysis supports this premise.

Although sample sizes were small in our analysis, every attempt was made to standardize the data, with exclusion of trials including...
Fig. 2. (A) Correlation between mPFS and mOS by study arm. (B) Correlation between mPFS and mOS in trials using Macdonald or RANO criteria for response evaluation. All RANO trials contain BEV test regimens, and the 3 BEV-containing trials using Macdonald criteria are indicated in blue. There are 83 arms using Macdonald criteria (red circle or square) and 8 arms (7 unique trials) using RANO criteria (black circle or square). (C) Correlation between mPFS and mOS separated by treatment (TMZ [red] vs non-TMZ [black]). Trials included used the Macdonald/RANO criteria for tumor assessment. (D) Correlation between mPFS and mOS separated by treatment (BEV [red] vs non-BEV [black]). Trials included used the Macdonald/RANO criteria for tumor assessment. (E) Correlation between mPFS and mOS separated by line settings (newly diagnosed vs recurrent). (F) Correlation between mPFS and mOS separated by histology (glioblastoma only vs mixed histology). Abbreviation: PI, prediction interval.
ไร้ที่มีการibling methodological detail และไม่ใช้การประเมินการตอบแบบ standar-
ดิและไม่ใช้การวัดผลการตอบสนองที่เป็นมาตรฐาน.

The effects of TMZ and BEV on the correlation between mPFS and mOS were selected for study because these 2 treatments appeared most often in the literature, and too few trials report other specific treatments. Interestingly, despite differing mechanisms of action, these treatments demonstrated consistent correlations between mPFS and mOS. Although the small sample number in our analysis precludes any definitive conclusions with regard to treatment effect, the results warrant future studies.

Historically, it has been shown that patients with anaplastic glioma (WHO grade III) have a much better prognosis and survival than patients with glioblastoma. It would be logical to assume that a mixed grade III–IV group would have better survival because the anaplastic glioma patients' survival would increase the median values for the entire group. Our results confirmed this assumption: the slope of the regression line between mPFS and mOS is significantly higher in trials with mixed grade III–IV glioma compared with glioblastoma only. However, the difference was marginal. Possible explanations for this observation may include the fact that the lower left corner data points, which represent the mixed histology group, also represent recurrent trials (poor PFS and OS) predominantly, and most recurrent glioma patients have progressed from grade III to glioblastoma. Conversely, data in the upper right corner (high PFS, high OS) represent newly diagnosed cases predominantly and support a trend toward better OS for mixed patients (better prognosis) compared with glioblastoma-only patients.

The accrual period in the trials included in this analysis ranged from 1991 to the present. During this period, advances have been made in many aspects of glioblastoma clinical management, such as diagnostics, surgical and imaging technology, treatments, recurrence monitoring, and standard supportive care. However, the correlation between median PFS and OS seems to be consistent across different time periods, which supports the applicability of these results to future trials.

There was only a moderate correlation between 6-month PFS and mOS, which is consistent with the results of Ballman et al., who investigated the relationship between 6-month PFS and 1-year OS in phase II glioblastoma trials. However, it is impossible in our analysis to identify at which time point the PFS rate would be a good predictor for OS because most trials report only 6-month PFS and mPFS and individual patient data are not available to us. In addition, ORR and OS were poorly correlated ($R^2 = 0.22$).

The applicability of our estimate of the linear relationship between the HR of OS and HR of PFS for trials evaluating anti-VEGF
agents (ie, agents targeting VEGF or VEGFR) may require further validation because none of the trials in the HR correlation analysis contained an anti-VEGF agent, such as BEV (an anti-VEGF antibody) or cediranib (a VEGFR tyrosine kinase inhibitor). VEGF blockade decreases vascular permeability and normalizes vascular perfusion and the blood–brain barrier, often causing decreased contrast. Radiochemotherapy could induce an inflammatory reaction (edema) and abnormal vessel permeability, causing new or increased contrast enhancement without affecting the underlying tumor, a phenomenon called pseudoprogression. Furthermore, radiochemotherapies preceding or following anti-VEGF therapies could further complicate MRI evaluation given the rapid onset of pseudoprogression and pseudoresponse. Therefore, models based purely on radiochemotherapies should not be extrapolated to anti-VEGF therapies without cautious validation.

There are some limitations in our analysis. Notably, we used literature instead of individual patient data, and HRs were reported in only a small number of studies due to lack of large, phase III, randomized glioblastoma trials. Furthermore, modifications to the standard response criteria were made in some trials, and details of treatments after progression were rarely reported, making it difficult to assess the potential confounding effects of subsequent treatments and crossover therapies on OS. The number of studies incorporating the RANO criteria was also small. Although these criteria have not been formally validated, they importantly address the phenomenon of pseudoprogression and pseudoresponse. Although we noted a consistent relationship between mPFS and mOS regardless of radiologic assessment methods, future analyses may further evaluate this relationship in more trials incorporating the RANO criteria.

In conclusion, our meta-analysis of 91 unique glioblastoma trials demonstrated a strong correlation between improvements in PFS and OS. There is also a good correlation between median PFS and OS in glioblastoma trials, regardless of response criteria, treatment, line settings, and histology. However, poor correlation was observed between ORR and OS, indicating that a high ORR may not translate into improved OS. Together these findings indicate that PFS may be an appropriate surrogate for OS in glioblastoma trials. Compared with OS, PFS offers the opportunity for earlier assessment of efficacy and higher statistical power, so establishment of these correlations may facilitate interpretation of interim analyses and future trial design.

### Acknowledgments
We acknowledge Bert Lum for assistance. Support for editorial assistance for this manuscript was provided by F. Hoffmann–La Roche Ltd. These data were previously presented at the American Society of Clinical Pharmacology Therapeutics 2013 annual meeting.

### References


