Neuro-Oncology

Neuro-Oncology 17(8), 1106–1113, 2015
doi:10.1093/neuonc/nou345
Advance Access date 7 January 2015

Combining progression-free survival and overall survival as a novel composite endpoint for glioblastoma trials

Lorenzo Trippa, Patrick Y. Wen, Giovanni Parmigiani, Donald A. Berry, and Brian M. Alexander

Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, Massachusetts (L.T., G.P.); Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts (L.T., G.P.); Center for Neuro-Oncology Dana-Farber/Brigham and Women's Cancer Center, Harvard Medical School, Boston, Massachusetts, (P.Y.W., B.M.A.); Department of Radiation Oncology, Dana-Farber/Brigham and Women's Cancer Center, Harvard Medical School, Boston, Massachusetts (B.M.A.); Department of Biostatistics, The University of Texas M.D. Anderson Cancer Center, Houston, Texas (D.A.B.)

Corresponding Author: Brian M. Alexander, MD, MPH, Dana Farber Cancer Institute, 450 Brookline Avenue, Boston, MA 02215 (bmalexander@lroc.harvard.edu).

Background. The use of auxiliary endpoints may provide efficiencies for clinical trial design, but such endpoints may not have intrinsic clinical relevance or clear linkage to more meaningful endpoints. The purpose of this study was to generate a novel endpoint that considers both overall survival (OS) and earlier events such as progression-free survival (PFS) and determine whether such an endpoint could increase efficiency in the design of glioblastoma clinical trials.

Methods. Recognizing that the association between PFS and OS varies depending on therapy and tumor type, we developed a statistical model to predict OS based on PFS as the trial progresses. We then evaluated the efficiency of our model using simulations of adaptively randomized trials incorporating PFS and OS distributions from prior published trials in neuro-oncology.

Results. When treatment effects on PFS and OS are concordant, our proposed approach results in efficiency gains compared with randomization based on OS alone while sacrificing minimal efficiency compared with using PFS as the primary endpoint. When treatment effects are limited to PFS, our approach provides randomization probabilities that are close to those based on OS alone.

Conclusion. Use of OS as the primary endpoint, combined with statistical modeling of the relationship between OS and PFS during the course of the trial, results in more robust and efficient trial designs than using either endpoint alone.

Keywords: adaptive clinical trials, auxiliary endpoints, clinical trial, endpoints, overall survival, progression-free survival.

Background

Glioblastoma (GBM) has a poor prognosis despite multimodality therapy, and therapeutic advances have been few.[1,2] Recently, there has been increasing frustration with inefficiencies created by the cancer research bureaucracy[3] and interest in finding ways to speed up the conduct and analysis of clinical trials. One potential way to shorten the time from trial initiation to results is to use primary endpoints that incorporate imaging-based assessments of progression, such as progression-free survival (PFS), with earlier times to event than overall survival (OS).[4] Furthermore, since experimental therapies most directly influence the time until progression, it is generally easier to detect effects on PFS, especially if there is long and heterogeneous survival post progression.[5]

There has been some concern, however, regarding the use of progression-based endpoints for clinical trials in neuro-oncology.[6] While outcomes, such as survival time, may have clear clinical relevance, endpoints based on imaging assessments, such as response or progression status, are not as clearly linked to patient benefit. Furthermore, while impacts on these endpoints may be associated with impacts on OS, this relationship is not uniformly consistent across different tumor types and different therapies,[7,8] which adds further complexity. In other words, it can be difficult to anticipate how positive effects on overall response rate or PFS will translate to effects on OS. For example, two recent phase III trials in GBM demonstrated improvements in PFS with no effect on OS,[7,8] while prior trial data seemed to suggest a stronger link.[4]

Given the controversy regarding the use of early endpoints to guide clinical trial decision-making but also recognizing the potential value of these endpoints,[4,6] we created a composite endpoint model to use in clinical trials for newly diagnosed GBM patients. This model remains anchored to identifying effects on OS but allows substantial efficiency gains when PFS

Received 14 October 2014; accepted 23 November 2014

data both support positive treatment effects on progression and anticipate OS treatment effects. We tested the composite model's performance through simulations in the context of a multiarm Bayesian, adaptively randomized clinical trial.[9,10]

Methods

We defined an adaptive randomization procedure for multiarm trials based on a model for PFS and OS outcomes. The model

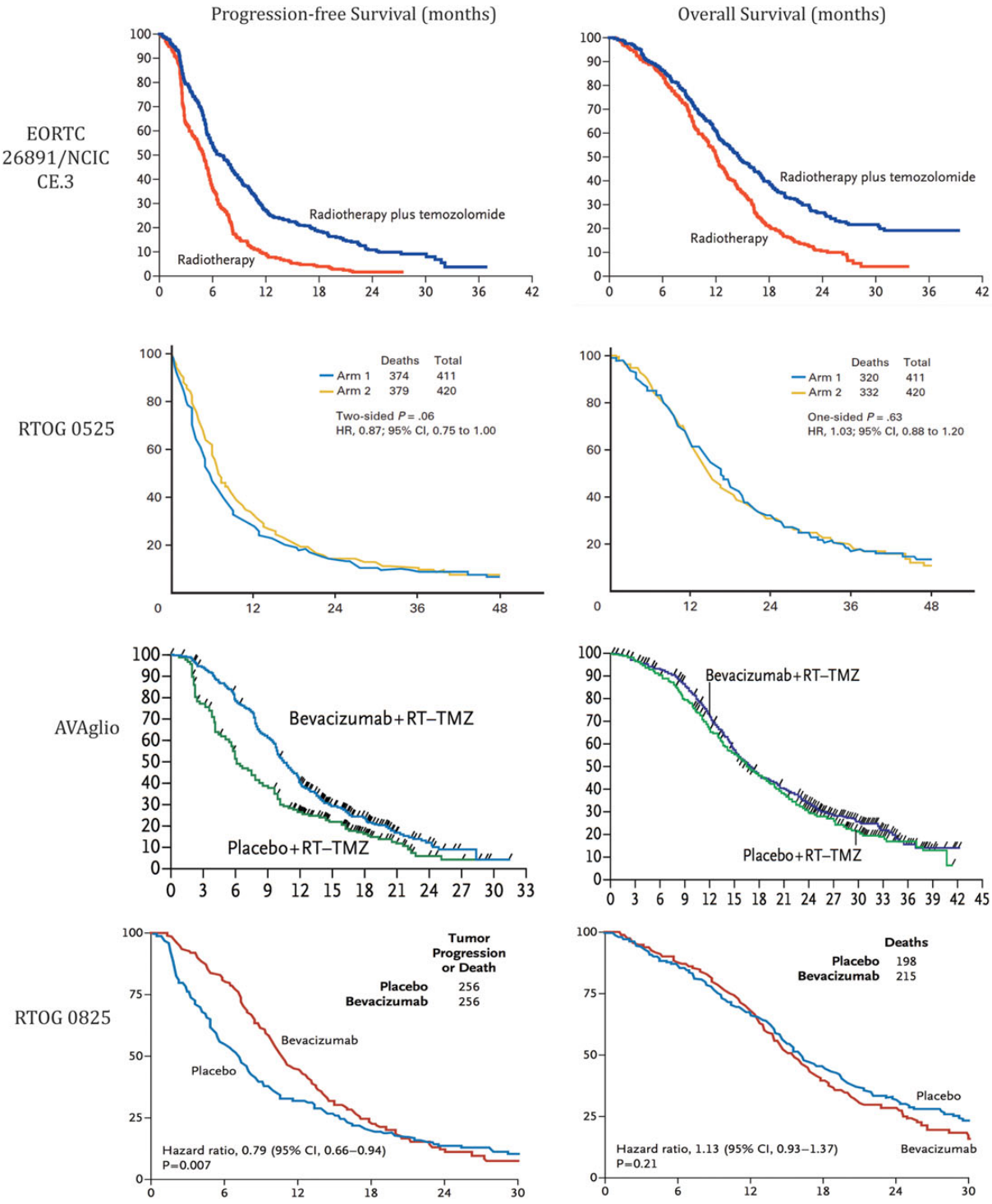

Fig. 1. Kaplan-Meier survival curves from EORTC 26981/NCIC CE.3 (reprinted with permission from Massachusetts Medical Society),[14] RTOG 0525 (reprinted with permission. © 2013 American Society of Clinical Oncology. All rights reserved.),[20] AVAglio (reprinted with permission from Massachusetts Medical Society),[7] and RTOG 0825 (reprinted with permission from Massachusetts Medical Society)[8] trials.

includes (K + 1) PFS distributions ($G_0, \ldots, G_K$) and (K + 1) OS distributions ($F_0, \ldots, F_K$), one for each of the K experimental arms, and $k = 0$ corresponding to the control arm. We evaluated two models for auxiliary and primary endpoints: an additive model[5] and a joint proportional hazards (PHs) model. Under the additive model, G_k can differ from G_0, but survival post progression (SPP) (defined at the individual level as SPP = OS-PFS) is assumed to be independent of treatment assignment. This leads to an adaptive randomization procedure based entirely on PFS data.

The PH model is defined by PHs for PFS and OS distributions, via positive parameters $\lambda = (\lambda_1, \ldots, \lambda_K)$ and $\theta = (\theta_1, \ldots, \theta_K)$ such that $G_k(\cdot \geq t) = G_0(\cdot \geq t)^{\lambda_k}$ and $F_k(\cdot \geq t) = F_0(\cdot \geq t)^{\theta_k}$. Our posterior computations use the factorization:

$$p(\theta, \lambda|\text{data}) \propto p(\theta, \lambda) \times PL(PFS.\text{data}|\lambda) \times PL(OS.\text{data}|\theta) \quad (1)$$

where $p(\theta, \lambda)$ is the prior distribution on the unknown parameters, and PL stands for partial likelihood. Two choices that simplify computations are the use of partial likelihoods and the separation of PFS and OS data into two distinct terms. The distribution $p(\theta, \lambda|\text{data})$ generated by (1) is a practical approximation that we used for adapting randomization probabilities. The prior, $p(\theta, \lambda)$, is the model component through which PFS data from the early stage of the trial can affect prediction of the treatment effects on OS. Typically, θ_k and λ_k are positively correlated a priori, so that promising PFS data from the k-th arm translates into optimistic prediction for the OS effect in the same arm.

Our Bayesian adaptive randomization (BAR) procedure has a similar interpretation to that previously described.[11] Specifically, the probability, $\pi_{i,k}$, that the i-th enrolled patient will be randomized to the k-th arm, is a function of the posterior distribution:

$$\pi_{i,k} \propto \begin{cases} \dfrac{p(OS_0 \ll OS_K \mid \text{available data})^{\gamma(i)}}{\sum_{j=1}^{K} p(OS_0 \ll OS_K \mid \text{available data})^{\gamma(i)}} & k = 1, \ldots, K, \\ \dfrac{1}{K}\exp(\max(n_{i,1}, \ldots, n_{i,K}) - n_{i,0})^{\eta(i)} & k = 0, \end{cases}$$

where $n_{i,k}$ is the count of randomizations to arm k before the enrollment of the i-th patient, $OS_0 \ll OS_K$ indicating a positive OS treatment effect of the k-th arm, and both $\gamma(i)$ and $\eta(i)$ are increasing functions.

We analyzed BAR with auxiliary and primary endpoints. We considered both scenarios in which a positive OS effect is and is not anticipated by a PFS improvement, as seen in prior glioma data.[6–8,12,13] Because our goal was to evaluate BAR robustness, the ratios of the accrual rate to PFS and OS times are critical. For instance, adaptation is impractical if PFS times exceed the accrual period. Therefore, we used an extensive set of scenarios to investigate BAR.

We considered several relationships between PFS and OS based on actual clinical trial data in GBM. Examples of these scenarios from large phase III trials are summarized in Fig. 1 and Table 1. The left column of Fig. 2 shows scenarios with 3 experimental arms and 1 control arm. The sample size was 240, and we assumed an accrual rate of 15 patients per month. In scenario 1, hypothetical treatment 1 is detrimental to PFS but has a positive effect on OS, as would be hypothetically seen in pseudoprogression related to standard treatment[13] or possibly immunotherapy. In scenario 2, treatment 1 has positive effects on both PFS and OS, while in scenario 3 an improvement in PFS is accompanied by no effect on OS. Scenario 4 shows no impact on either PFS or OS. To evaluate adaptive randomization with the joint PH model, we also considered scenarios where the additive model held. Scenarios similar to the first panel of Fig. 2 were defined with PFS distributions identical to the EORTC/NCIC CE.3 trial[14] in Fig. 1 and gamma-distributed SPP times, with mean equal to 1, 3, or 6 months. We compared BAR based on the joint PH model to BAR based on PFS (in accordance with the additive model) or OS.

Results

We first evaluated the performance of the BAR design with the composite endpoint model considering different accrual rates. The x-axis in column 2 of Fig. 2 shows the ratio of the actual accrual rates to the estimate of 15 patients per month. In the most extreme case shown, accrual is more than 4-fold faster. The solid lines in the right column show the mean number of patients randomized to each arm under various scenarios, with adaptive randomization based on the joint PH model. These panels also provide the frequentist power of the adaptive design (dashed lines) for rejecting the null hypothesis of zero or the detrimental treatment effect on OS; for these computations, we assumed complete

Table 1. Relationship between progression-free survival and overall survival from 3 clinical trials in newly diagnosed glioblastoma

Trial	Arm	Median PFS	PFS, HR	Median OS	OS, HR
EORTC 26981/NCIC CE.3[14]	RT	5.0		12.1	
	RT/TMZ	6.9	0.56	14.6	0.63
RTOG 0525[20]	RT/TMZ	5.5		14.9	
	RT/ddTMZ	6.7	0.87	16.6	1.03
RTOG 0825[8]	RT/TMZ	7.3		16.1	
	RT/TMZ/bevacizumab	10.7	0.79	15.7	1.13
AVAglio[7]	RT/TMZ	6.2		16.7	
	RT/TMZ/bevacizumab	10.6	0.64	16.8	0.88

Abbreviations: dd, dose-dense; HR, hazard ratio; OS, overall survival; PFS, progression-free survival; RT, radiotherapy; TMZ, temozolomide.

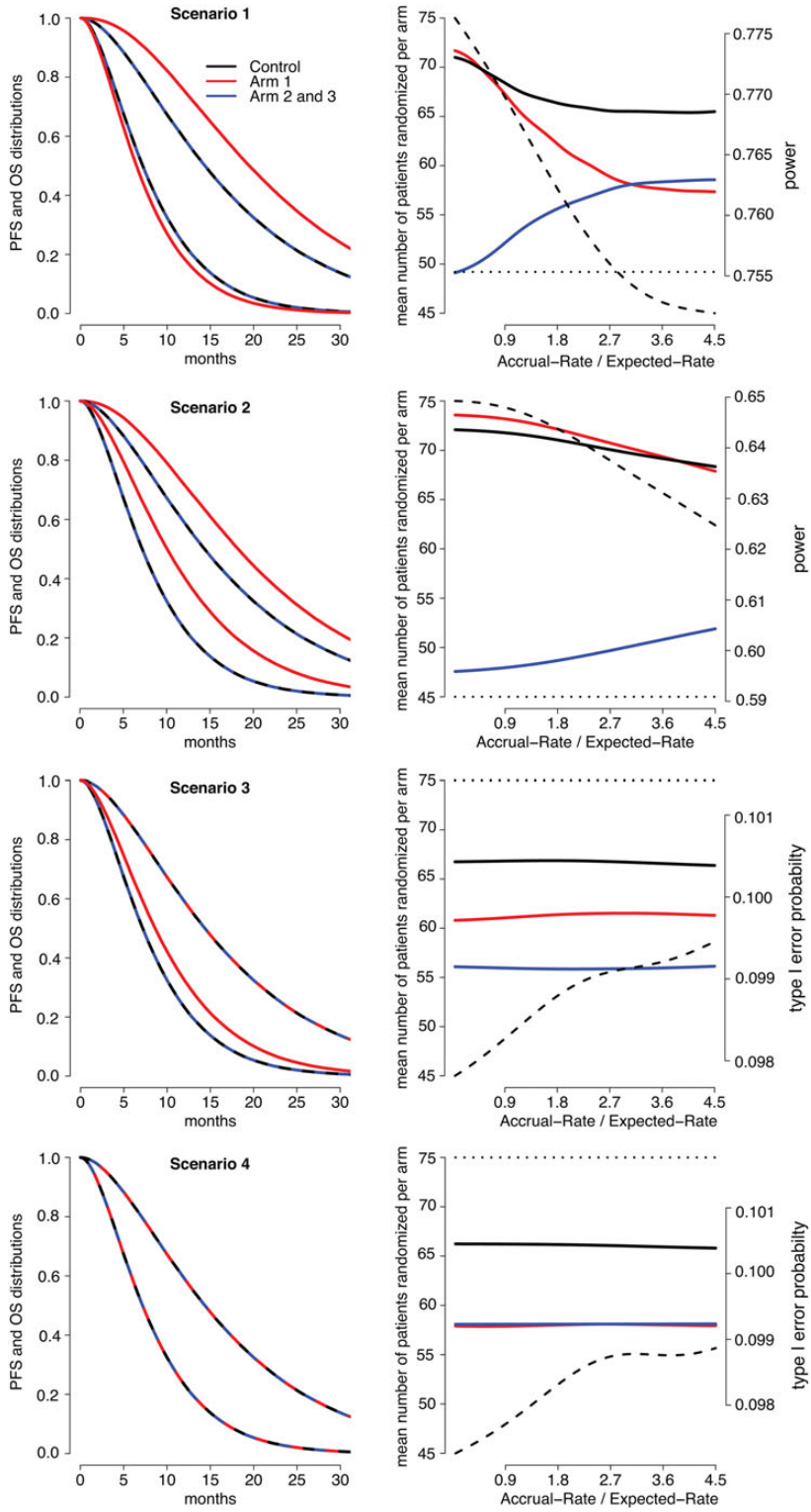

Fig. 2. Comparison of average arm-specific sample sizes for various scenarios. The left column represents scenarios with different outcome distributions. Each panel contains multiple curves. The lower curves are progression-free survival (PFS), while the upper curves are overall survival (OS). Each arm is denoted by a different color. The right column shows both the arm-specific expected sample sizes and the frequentist power of detecting a positive treatment effect with a significance level of 0.1 for our adaptive design (dashed lines) and balanced randomization (dotted lines) for varying accrual rates (x-axis). In the absence of a positive OS treatment effect (Scenarios 3 and 4), the type I error probabilities are displayed instead of power. The estimates in each panel were obtained using 10^6 simulated trials.

follow-up after completion of the accrual period and an α-level of 0.1. For comparison, the dotted lines show the power of a standard balanced randomized design. We used a previously described procedure for hypothesis testing in this setting.[15,16] Our results have an intuitive explanation: the higher the accrual rate, the more difficult it becomes to augment randomization toward the best treatment option. In some of our scenarios, we further increased this difficulty by assuming treatment effects on PFS disagreed with the OS distributions. Despite these challenges, BAR is remarkably robust across all scenarios we considered. When we compared BAR with balanced randomization, in only a few implausible cases did BAR fail to improve the power or to increase the number of patients assigned to the best available treatment.

We then summarized the sensitivity of the operating characteristics to several variations of the BAR trial design. We first evaluated the consequences of adding experimental arms. Additional arms with no treatment effect result in a further increase of the average sample size for the effective arm and the control, which in turn results in an increase in power. For example, in scenarios 1 and 2 of Fig. 2, the probability of rejecting the null hypothesis of no treatment effects with one additional experimental arm increases by 1.3% and 0.8%, respectively. The choice of scenarios with at most a single effective treatment is consistent with the limited number of drugs approved for GBM in the last two decades.

Next, we added early stopping rules, dropping an experimental arm when the posterior probability of a treatment effect on OS with hazard ratio (HR) < 0.9 became < 0.1. We observed negligible variation of the operating characteristics after this change. We then evaluated robustness of the design, assuming a data acquisition delay of up to 4 weeks due to management inefficiencies. The randomization probabilities were updated monthly when we observed small changes in the average number of patients assigned to each arm (<2 patients across all scenarios).

We then evaluated the testing procedure: the *P* value for the *k*-th arm was obtained through a bootstrap estimate of the log-rank statistics, accounting for the adaptive design as described previously.[15,16] We observed differences $< 0.35\%$ between the type I error rate and the nominal significance level at $\alpha = 0.05$, 0.1, and 0.15 across all scenarios as well as variations of scenarios 1 and 2 without any effective arm.

Our simulations under different concordance levels of PFS and OS times, evaluated by Harrell's C index, suggested little influence on the arm-specific sample sizes distributions. Fig. 3 contrasts arm-specific sample sizes when assuming a high concordance index of 1, versus a low concordance of 0.6, for scenario 1. In Fig. 3, dark and light gray are associated to high and low concordance, respectively, with an intermediate shade for the overlapping part of the histograms showing a very modest difference.

Fig. 4 demonstrates how our adaptive algorithm, based on the joint model, performs in comparison with models that adapt only using PFS or OS. The left column displays two different scenarios. In the top row, a positive OS effect for one arm is predicted by the PFS effect, while in the bottom row a positive OS effect is combined with a contraction on PFS times. The middle and right columns show how these two scenarios differ midway through the trial with respect to the estimation of HR and posterior probability of a treatment effect, respectively. In the top scenario, because there is agreement between PFS and OS, the joint model approximates the OS HR, leveraging on PFS data. The estimates based only on PFS data (dotted lines) match the inferred effects under the additive model. This provides a comparison of the two modeling approaches. We observed only small differences in the

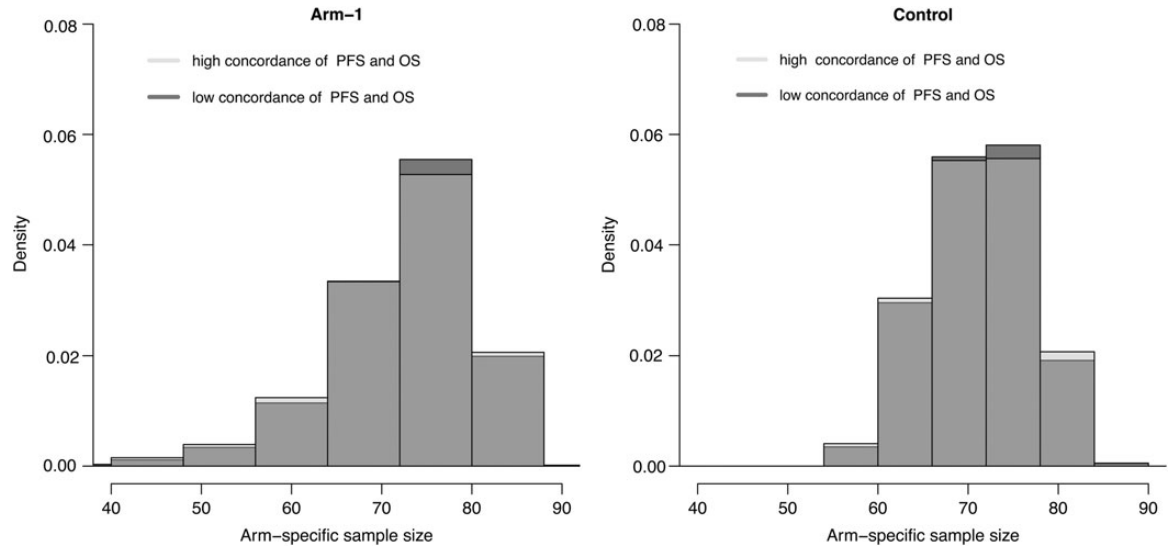

Fig. 3. Negligible effects on the trial operating characteristics of different concordance levels between progression-free survival (PFS) and overall survival (OS). We considered PFS and OS to be marginal distributions, as displayed in Scenario 1 of Fig. 2, and constructed joint distributions with high concordance (Harrell's C index equal to 1) and low concordance (Harrell's C index equal to 0.6) between PFS and OS. The panels contrast these joint distributions and display the number of patients randomized to Arm 1 and to the control arm across 10^5 simulations.

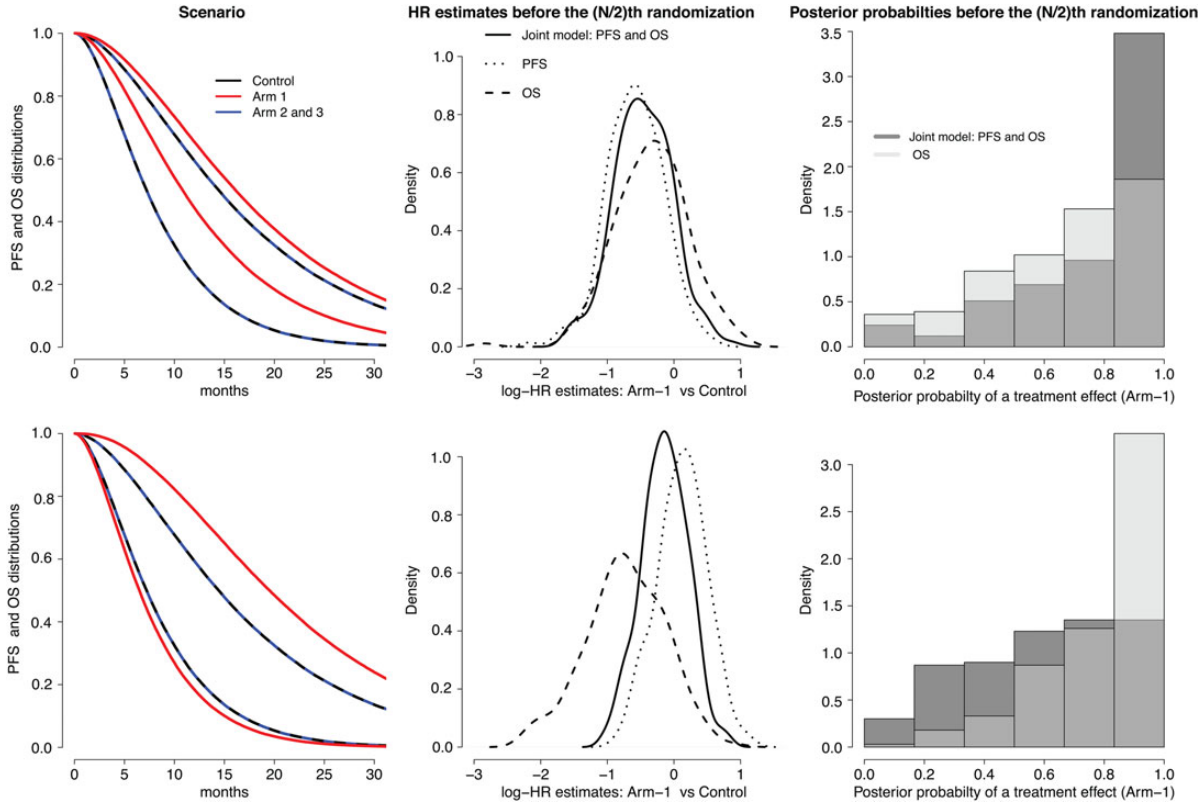

Fig. 4. Randomization probabilities using adaptation on overall survival (OS) only versus the proposed joint use of progression-free survival (PFS) and OS outcomes. The panels on the left show 2 hypothetical scenarios. The panels in the middle show early hazard ratio (HR) estimates contrasting Arm 1 with the control arm across 10^5 simulations. The estimates are given at the enrollment of the 120th patient out of 240 patients. Dotted and dashed lines refer to estimates based only on PFS and OS data, respectively. The solid lines refer to OS HR estimates based on a joint prior distribution that incorporates correlation of PFS and OS treatment effects. The right panels contrast early posterior probabilities of a positive treatment effect for Arm 1 across simulations when the investigator only uses OS data versus the OS and PFS joint model.

number of patients assigned to each arm across various scenarios consistent with the additive model. In the most extreme case, we observed an average reduction of 4.1 out of 180 patients assigned to the effective arm for the PH model compared with the additive model. This increased efficiency from the additive model comes at a price when the treatment effects on PFS and OS do not agree. The lower row displays one example.

Figure 4 illustrates the estimates across stimulations at a chosen time point during the trial. Fig. 5 shows how randomization probabilities changed during the course of the trial. The scenarios underlying the left and right panels are the empirical distributions of the EORTC 26981/NCIC CE.3 and RTOG 0825 trials (displayed in Fig. 1). In the first case, PFS and OS treatment effects agree, while there is a PFS benefit without an OS benefit in the second scenario. In the left panel, the joint model (solid lines) results in efficiency gains compared with an OS-only model (dashed lines) as the effective (red) arm has a higher probability of randomization (y-axis) at earlier points in the trial (x-axis). When the early PFS signal is misleading, as in the right panel scenario, the initial increase in randomization probability (red solid line) generated by the PFS signal wanes over the course of the study.

Discussion

Improving the length and quality of life are clear goals of clinical research. For trials in GBM, the variability and length of post-progression survival is unfortunately limited, arguing against the substitution with alternative endpoints such as PFS for more clinically meaningful ones such as OS.[5,6] Even so, effective use of surrogate endpoints offers the promise of more efficient clinical trials by providing earlier answers to questions of therapeutic efficacy. In particular, auxiliary endpoints are valuable when considering an adaptively randomized trial in which timely information is brought to bear on decision-making during the course of the study. However, adaptively randomized studies are potentially vulnerable from incorporating misleading results into the adaptive procedure. While treatment effects on PFS may have correlated well with OS effects in past GBM studies,[4,17,18] there is no guarantee that experimental therapies will maintain these previously described associations. There is also some evidence against a generalizable correlation between auxiliary endpoint and OS effects that remains valid across treatments.[7,8]

One solution to addressing auxiliary endpoint uncertainty is to employ a composite model that 'learns' the relationship with

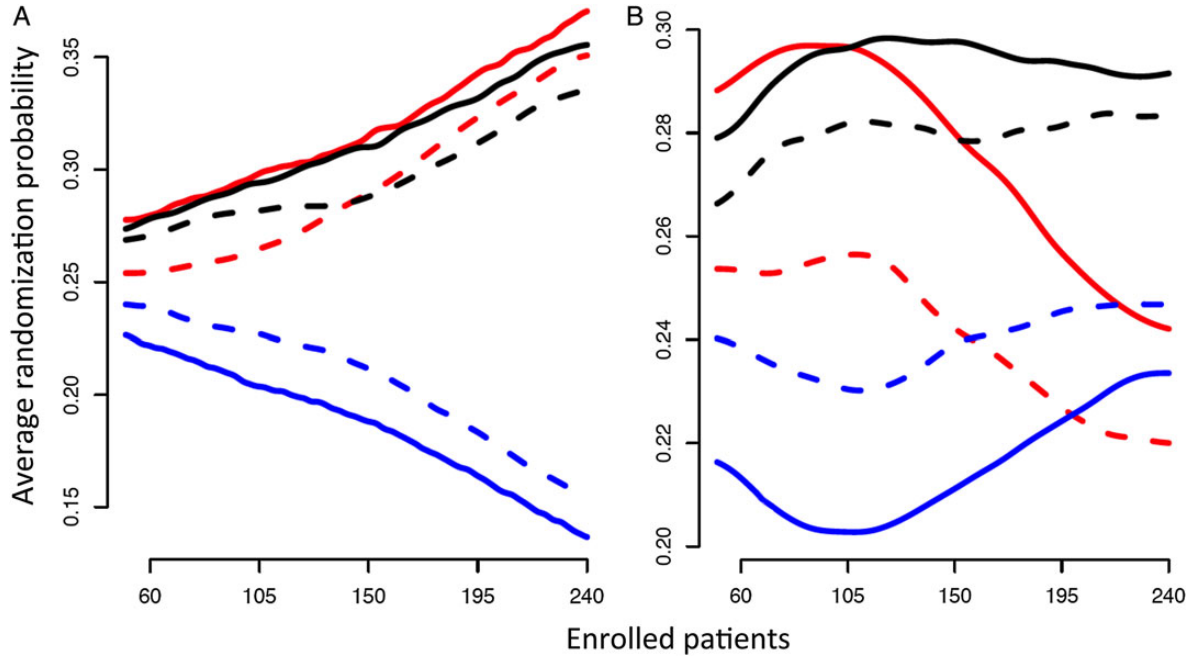

Fig. 5. OS data vs. PFS and OS model for adaptive randomization. The panels display the average probability across simulations that the i-th enrolled patient is randomized to each arm under the OS/PFS model (solid lines) or OS data alone (dashed lines). We constructed simulation scenarios that reflects the PFS and OS data generated by the EORTC 26981/NCIC CE.3 and RTOG 0825 trials by automated scanning (WebPlotDigitizer 2.6) the published Kaplan-Meier curves[8,14] and then using these curves as sampling models. Control arms (black lines) and two experimental arms (blue lines) for each panel match the empirical data control arm PFS and OS distributions. A third experimental arm (red lines) matches the empirical distributions under the superior treatment options of the EORTC 26981/NCIC CE.3 trial (left panel) and the RTOG 0825 trial (right panel).

a primary endpoint during the course of the study. A notable example of this is the I-SPY 2 trial, in which MRI assessments of breast cancer response to neoadjuvant chemotherapy are used for prediction of the primary endpoint (complete pathological response) during the course of the trial.[19] For past GBM trials, the PFS median lead time has been 7.4 months for newly diagnosed tumors and 4.2 months for recurrent tumors,[4] thus providing potentially useful earlier results to inform randomization.

OS has been modeled as the sum of PFS and SPP.[5] In some cancers and for some treatments, it may be reasonable to assume that treatment effects are limited to PFS and that the various arms have identical SPP distributions. Although such a hypothesis can be tested based on outcomes data and appropriately used to report treatment effects estimates, its use for designing adaptive and/or group-sequential trials may not be fully generalizable. Particularly in GBM, the assumption that SPP is independent of treatment arm appears in some cases to be inappropriate[7,8] for designing trials. In this context, we jointly modeled PFS and OS. During the course of a study, the model, without assuming a specific relationship between OS and PFS treatment effects, leverages accumulating data and estimates OS and PFS distributions sequentially.

In our PH model, randomization is informed by the OS estimates, which are in turn informed by the PFS data. If concordance between the PFS and OS effects is low, this is learned by the model during the course of the trial. In contrast, when the assumptions of the additive model proposed by Broglio and Berry (wherein SPP is independent of treatment)[5] hold, the PH model will randomize slightly fewer patients to an effective treatment arm. This might not be desirable for settings in which the risk is low for conflicting PFS treatment effects and OS effects. Given the small efficiency gains of using the additive model based purely on PFS, however, GBM investigators should have a high bar for prior knowledge of the PFS/OS relationship. Another important consideration is that these small efficiency differences between the additive model and the PH model are contingent upon the GBM SPP times, which are much shorter compared with other cancers. In contexts with longer SPP times, however, the efficiency gain of the additive model would become more prominent.

The additive model has additional implications since PFS effects might be diluted by SPP times. We compared testing the null hypothesis of no treatment effects based on PFS data or OS data when the additive assumption holds. Across our scenarios and with significance level at 0.05 or 0.1, inclusion of the additive assumption, which implies that only PFS data are used for testing, produced a power increase between 2.2% and 5.2% compared with testing based on OS data. The analyses should therefore include evaluation of positive treatment effects on PFS when there is support for the additive assumption from previous studies and if the trial data do not invalidate such a hypothesis.

The PH model does incorporate an initial linkage of PFS and OS effects, however. While PFS is no longer used when there is

evidence of a lack of concordance with OS effects, the model requires additional time and enrolled patients to learn; therefore, the randomization probabilities can be affected from a misleading PFS signal early in the trial. Our model is a compromise between the point of view that the investigator can predict a PFS/OS linkage in advance and the point of view that OS is the only outcome that should be used. This compromise gains efficiency over OS and mitigates error when there is no correspondence between PFS and OS effects.

Another alternative to our model would be to use OS only. Our results show that the OS-only adaptive design still results in efficiency gains over a balanced randomization and sacrifices some efficiency over the joint model but, as expected, is not sensitive to randomizations driven by PFS effects that do not translate into OS improvements.

Conclusions

We developed a composite endpoint model to design BAR trials using PFS data to provide efficiency while still maintaining the clinical relevance of OS. The potential for inefficient randomization resulting from conflicting PFS and OS signals is mitigated in our model compared with PFS alone, at a modest cost in terms of efficiency. When PFS and OS effects are correlated, measuring the treatment effects on PFS for decision-making results in power gains because the dilution effect of SPP is reduced. BAR based on OS still provides gains over balanced randomization, however, and the trade-offs of using surrogate endpoints such as PFS in any capacity in adaptively randomized studies should be a significant discussion topic during trial development.

Funding

This work was supported by an Innovations in Regulatory Science Award from The Burroughs Wellcome Fund.

Disclosures

D.A.B. is co-owner of and statistical consultant with Berry Consultants, LLC, a company that designs Bayesian adaptive clinical trials for pharmaceutical and medical device companies and NIH cooperative groups.

References

1. Quant EC, Drappatz J, Wen PY, et al. Recurrent high-grade glioma. *Curr Treat Options Neurol*. 2010;12(4):321–333.
2. Wen PY, Kesari S. Malignant gliomas in adults. *N Engl J Med*. 2008; 359(5):492–507.
3. Steensma DP, Kantarjian HM. Impact of cancer research bureaucracy on innovation, costs, and patient care. *J Clin Oncol*. 2014;32(5):376–378.
4. Han K, Ren M, Wick W, et al. Progression-free survival as a surrogate endpoint for overall survival in glioblastoma: a literature-based meta-analysis from 91 trials. *Neuro Oncol*. 2014;16(5):696–706.
5. Broglio KR, Berry DA. Detecting an overall survival benefit that is derived from progression-free survival. *J Natl Cancer Inst*. 2009; 101(23):1642–1649.
6. Alexander BM, Trippa L. Progression-free survival: too much risk, not enough reward? *Neuro Oncol*. 2014;16(5):615–616.
7. Chinot OL, Wick W, Mason W, et al. Bevacizumab plus radiotherapy-temozolomide for newly diagnosed glioblastoma. *N Engl J Med*. 2014;370(8):709–722.
8. Gilbert MR, Dignam JJ, Armstrong TS, et al. A randomized trial of bevacizumab for newly diagnosed glioblastoma. *N Engl J Med*. 2014;370(8):699–708.
9. Thall PF, Wathen JK. Practical Bayesian adaptive randomisation in clinical trials. *Eur J Cancer*. 2007;43(5):859–866.
10. Berry DA. Adaptive clinical trials: the promise and the caution. *J Clin Oncol*. 2011;29(6):606–609.
11. Trippa L, Lee EQ, Wen PY, et al. Bayesian adaptive randomized trial design for patients with recurrent glioblastoma. *J Clin Oncol*. 2012; 30(26):3258–3263.
12. Wen PY, Macdonald DR, Reardon DA, et al. Updated response assessment criteria for high-grade gliomas: response assessment in neuro-oncology working group. *J Clin Oncol*. 2010;28(11):1963–1972.
13. Brandes AA, Franceschi E, Tosoni A, et al. MGMT promoter methylation status can predict the incidence and outcome of pseudoprogression after concomitant radiochemotherapy in newly diagnosed glioblastoma patients. *J Clin Oncol*. 2008; 26(13):2192–2197.
14. Stupp R, Mason WP, van den Bent MJ, et al. Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. *N Engl J Med*. 2005;352(10):987–996.
15. Rosenberger WF, Hu F. Bootstrap methods for adaptive designs. *Stat Med*. 1999;18(14):1757–1767.
16. Trippa L, Rosner GL, Muller P. Bayesian enrichment strategies for randomized discontinuation trials. *Biometrics*. 2012;68(1): 203–211.
17. Ballman KV, Buckner JC, Brown PD, et al. The relationship between six-month progression-free survival and 12-month overall survival end points for phase II trials in patients with glioblastoma multiforme. *Neuro Oncol*. 2007;9(1):29–38.
18. Polley MY, Lamborn KR, Chang SM, et al. Six-month progression-free survival as an alternative primary efficacy endpoint to overall survival in newly diagnosed glioblastoma patients receiving temozolomide. *Neuro Oncol*. 2010;12(3):274–282.
19. Barker AD, Sigman CC, Kelloff GJ, et al. I-SPY 2: an adaptive breast cancer trial design in the setting of neoadjuvant chemotherapy. *Clin Pharmacol Ther*. 2009;86(1):97–100.
20. Gilbert MR, Wang M, Aldape KD, et al. Dose-dense temozolomide for newly diagnosed glioblastoma: a randomized phase III clinical trial. *J Clin Oncol*. 2013;31(32):4085–4091.