



Published in final edited form as:

*Clin Cancer Res.* 2016 March 15; 22(6): 1341–1347. doi:10.1158/1078-0432.CCR-15-1441.

## Correlating Surrogate Endpoints with Overall Survival at the Individual Patient Level in BRAF<sup>V600E</sup>-Mutated Metastatic Melanoma Patients Treated with Vemurafenib

Emily C. Zabor<sup>1</sup>, Glenn Heller<sup>1</sup>, Lawrence H. Schwartz<sup>2</sup>, and Paul B. Chapman<sup>3</sup>

<sup>1</sup>Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, New York

<sup>2</sup>Department of Radiology, Columbia University College of Physicians and Surgeons and New York Presbyterian Hospital, New York, New York

<sup>3</sup>Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, New York

### Abstract

**Purpose**—Surrogate endpoints are needed that correlate with overall survival (OS). We analyzed individual patient tumor data from a phase III trial of vemurafenib versus dacarbazine (BRIM3) to identify criteria for tumor measures that correlated with OS. Correlates were validated using a separate data set from a phase II trial of vemurafenib (BRIM2).

**Experimental Design**—Deidentified tumor measurements and OS data from BRIM3 and from BRIM2 were analyzed. Target tumor measurement data and nontarget tumor data were available from pretreatment, weeks 6,12, and every 9 weeks thereafter. In the BRIM3 data set, associations of OS with both early tumor response (first 12 weeks) and time to progression (TTP) were assessed. Different definitions of response and progression were explored. Findings were validated using the BRIM2 data set.

**Results**—Thresholds of early response were explored ranging from any degree of tumor shrinkage to 100% tumor shrinkage. Correlation was weak at all thresholds tested. TTP, however, was more strongly correlated with OS. The strongest correlation was seen when progression was

---

Corresponding Author: Paul B. Chapman, Memorial Sloan Kettering Cancer Center, 300 E. 66th Street, Room 1549, New York, NY 10065. Phone: 646-888-4162; Fax: 646-888-4253; chapmanp@mskcc.org.

**Note:** Supplementary data for this article are available at Clinical Cancer Research Online (<http://clincancerres.aacrjournals.org/>).

#### Disclosure of Potential Conflicts of Interest

L.H. Schwartz is a consultant/advisory board member for Bioclinica, Celgene, Icon Medical, and Novartis. No potential conflicts of interest were disclosed by the other authors.

#### Authors' Contributions

**Conception and design:** E.C. Zabor, G. Heller, L.H. Schwartz, P.B. Chapman

**Development of methodology:** E.C. Zabor, G. Heller, L.H. Schwartz

**Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.):** P.B. Chapman

**Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis):** E.C. Zabor, G. Heller, L.H. Schwartz, P.B. Chapman

**Writing, review, and/or revision of the manuscript:** E.C. Zabor, G. Heller, L.H. Schwartz, P.B. Chapman

**Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases):** E.C. Zabor, L.H. Schwartz

**Study supervision:** P.B. Chapman

defined as 50% increase in the sum of tumor diameters or appearance of new tumors. This was confirmed by similar analyses in the BRIM2 cohort.

**Conclusions**—TTP defined as 50% increase in the sum of tumor diameters or appearance of new tumors was more strongly associated with OS than early tumor shrinkage in melanoma patients treated with RAF inhibitor. In future trials, consideration should be given to replacing response rate with TTP or PFS as preferable clinical endpoints in early-phase studies.

## Introduction

Primary clinical endpoints to establish benefit of an anticancer drug include improvement in overall survival (OS), quality of life, tumor-related symptoms, or physical functioning (1). All other endpoints, such as progression-free survival (PFS), tumor shrinkage measured as response proportion, or biomarkers, are considered potential surrogate endpoints and may or may not correlate with primary endpoints (2). Response proportion, as defined by RECIST criteria (3), is commonly used as a surrogate endpoint for OS in early-phase studies. However, tumor response has not always been a reliable endpoint (4–6). This is not surprising as our current definitions of response are essentially unchanged from their development decades ago when investigators struggled to identify minimal tumor responses using only plain X-rays and physical exams (7, 8). The development of response criteria have never been based on correlation with metrics of clinical benefit such as OS.

The RAF inhibitor vemurafenib was the first treatment for melanoma that showed both a high objective response proportion (9) and a significant improvement in OS in a large randomized phase III trial (10). As a result, this trial provided the first opportunity (and perhaps the last, because of multiple effective therapies now available in the event of relapse) to assess surrogate endpoints such as antitumor response and time to progression (TTP) for correlation with OS in melanoma. We report here our findings analyzing measures of response and progression, and their strength of association with OS at the individual patient level. This unique data analysis provides the first evidence in melanoma at the individual patient level for TTP as a surrogate endpoint that correlates with OS. These results could have important implications for the choice of surrogate endpoints in future clinical trials in melanoma.

## Materials and Methods

Data on target lesion size at each time point measured and OS were provided by Genentech on patients from both the BRIM3 (10, 11) and BRIM2 (9) trials. BRIM3 ( $N=675$ ) was a multi-institutional phase III trial in which previously untreated patients with BRAF<sup>V600E</sup>-mutated metastatic melanoma were randomized between vemurafenib and dacarbazine (10). Appropriate cross-sectional imaging of the chest, abdomen, and pelvis was performed at baseline, week 6, week 12, and every 9 weeks thereafter. Treatment cross-over was not permitted initially but after the planned OS interim analysis (once 50% of the projected deaths had occurred), the protocol was amended upon the Data Safety Monitoring Board's recommendation in January 2011 to allow cross-over. Ultimately, 25% of the dacarbazine patients crossed over to vemurafenib at progression (11). All outcome data were used for dacarbazine patients who crossed over to vemurafenib, including outcomes data after

crossover. After participation on this trial, 22% and 18% of the dacarbazine and vemurafenib patients, respectively, received ipilimumab as melanoma therapy (11). BRIM2 was a multi-institutional single-arm phase II trial of vemurafenib in 132 patients with BRAF<sup>V600E</sup>-mutated metastatic melanoma (9). For this trial, previous treatment was required and in fact, 49% of patients had received more than one prior therapy. Cross-sectional imaging was performed at baseline and every 6 weeks thereafter.

In both BRIM3 and BRIM2, up to five target lesions were assessed per patient, according to RECIST 1.1. The primary objective of our analysis was to determine the strength of association between overall survival time and quantitative radiographic measures of early (within 12 weeks) tumor response and progression at any time. To assess the association between early response and OS, we first summed the target lesions at each scan time during the first 12 weeks. Next, we calculated the relative percent reduction in tumor burden at each time as  $100 \times (\text{sum baseline} - \text{sum follow-up}) / (\text{sum baseline})$ . Finally, we defined early best response for each patient as the maximum relative percent reduction in tumor burden within the first 12 study weeks. The 12-week time point was used to define early best response as both BRIM3 and BRIM2 had scans scheduled at this time point and significant tumor responses beginning after week 12 are uncommon on either vemurafenib or DTIC.

Thresholds of early response were explored ranging from any degree of tumor shrinkage to 100% tumor shrinkage, in 10% increments. Because sample size was limited, only thresholds containing at least 10% of the data in each group were considered. To assess the association between early response and OS, we used a nonparametric weighted C-index (12) that is appropriate for use with censored survival data. The weighted C-index was normalized to range between 0 and 1 (with 0 indicating no association) to create comparability with the TTP association metric discussed below. The SD of this normalized version was estimated using bootstrap methodology. A landmark analysis was used estimating survival from the landmark time point of 12 weeks. Patients who died or were censored prior to 12 weeks could not be evaluated for correlation of early response with OS (13).

For the assessment of TTP, cutoffs for progressive disease were defined as any increase, 25%, 50%, or 100% increase in the sum of the target tumor diameters. All patients with at least one follow-up scan were evaluable for the association between TTP and OS. Progression was calculated both from baseline and from a response nadir, if one existed. In addition, we considered the development of new lesions and evaluated both scenarios, including and excluding the development of new lesions, as a progression. At each scan time, the nadir was defined as the minimum of the sum lesion size at all preceding scans. To avoid large percentage changes when the nadir was a very small value, nadir values below 2 mm were analyzed as 2 mm for the purposes of this analysis. TTP was calculated from date of study enrollment to date of progression, by each definition, or last follow-up. OS was calculated from the date of study enrollment to the date of death or last follow-up. As TTP and OS are both right censored endpoints, the Kendall tau ( $\tau$ ), derived from the Clayton copula and scaled between 0 and 1, was used as the measure of association (14).

Both BRIM3 study arms were pooled for primary analyses. Subanalyses examined patients treated with vemurafenib or dacarbazine separately. The BRIM2 data were used for independent validation of the findings from BRIM3. Data cutoff for both studies was February 1, 2012.

Statistical analyses were conducted using SAS software version 9.4 (SAS Institute), R software version 3.1.1 (R Core Development Team, Vienna, Austria) including the “survival” and “CPE” packages, and the FORTRAN programming language.

## Results

### Correlation of tumor response with OS in BRIM3

Our initial hypothesis was that some degree of early (within the first 12 weeks of treatment) tumor shrinkage would correlate with OS. A total of 489 patients alive and on study at the 12 week landmark time point were used to determine if early response correlated with OS. We tested a series of response thresholds for association with OS ranging from 10% to 60% maximum reduction of the sum of the target lesions (Fig. 1). We found that early tumor shrinkage had almost no correlation with OS no matter what threshold we looked at (normalized weighted C-index values 0.10–0.15). Similar results were observed in a sensitivity analysis using a 6-week landmark time point (Supplementary Table S1). In case the low response proportion and high cross-over rate of the DTIC cohort was confounding this analysis, we analyzed the vemurafenib cohort separately. We found a similar range of normalized weighted C-index values (0.09–0.18), confirming the weak correlation of tumor response with OS even among the vemurafenib cohort (Supplementary Fig. S1).

### Correlation of TTP with OS in BRIM3

We next examined the association between TTP and OS. A total of 582 BRIM3 patients had at least one follow-up scan at any time and were evaluable for the TTP endpoint. Progression was measured either from baseline or nadir tumor size, with and without new lesions considered as a criteria for progression. Within each of the four settings, we tested a series of tumor progression thresholds for association with OS (Table 1). We found the strongest association between TTP and OS when progression was defined as a 50% increase in the target lesions whether or not new lesions were considered as progression, either from baseline (without new lesions  $\tau = 0.675$ ; with new lesions  $\tau = 0.588$ ) or from best response (without new lesions  $\tau = 0.638$ ; with new lesions  $\tau = 0.568$ ). An increase of 25% also correlated with survival, but not quite as strongly.

When we examined the association between TTP and OS in the vemurafenib cohort separately, ( $n = 325$ ; Table 2) many patients were classified as having progressed due to the appearance of new tumors rather than due solely to the growth of target lesions. For example, when defining progression as a 25% increase in the sum of the target lesion size from nadir, 107 of 252 patients (42%) progressed solely due to appearance of new lesions. The strongest correlation with OS was seen when progression was defined as 50% increase in the sum of the target lesions or appearance of new lesions ( $\tau = 0.655$ ). Lower thresholds of progression, such as any increase or 25% also correlated with OS, but less strongly ( $\tau =$

0.518 and 0.608, respectively). A similar pattern of results was seen when TTP from best response was evaluated. If progression was defined without consideration of the appearance of new lesions, the correlation with OS was considerably weaker.

For the DTIC cohort (Table 3), target lesion progression from baseline had a greater impact on the association between TTP and OS than in the vemurafenib cohort. This magnitude of this correlation may have been slightly diminished by the 25% of DTIC patients who ultimately crossed over to vemurafenib. The strongest correlation with survival was seen when progression was defined as a 50% increase in the target lesions without considering the appearance of new lesions, whether evaluated from baseline ( $\tau = 0.727$ ) or from best response ( $\tau = 0.697$ ). In this cohort, allowing the appearance of new lesions to define progression reduced the correlation with OS. This indicates that for patients treated with DTIC, meaningful progression is mostly due to the enlargement of target lesions and less commonly due to the appearance of new tumors.

### Validation of findings using the BRIM2 data set

For validation, the 132 patients treated with vemurafenib on BRIM2 (9) were evaluable for both early best response and progression endpoints. We analyzed the BRIM2 data set for the association of the TTP with OS (Table 4). Similar to the vemurafenib group in BRIM3, the association with survival was highest when progression was defined as a 50% increase in the sum of the target diameters or appearance of new tumors ( $\tau = 0.723$ ). Although, a strong association with OS was identified with any increase from baseline in the BRIM2 patients. A similar pattern of results was seen if progression was defined as an increase from best response of 25% ( $\tau = 0.635$ ), 50% ( $\tau = 0.708$ ), or 100% ( $\tau = 0.710$ ).

As with the BRIM3 data, we tested a series of early best response thresholds for their association with OS in the BRIM2 patients (Fig. 2). In the BRIM2 patient cohort, there was a much stronger association of response with survival. This discordance between BRIM2 and BRIM3 with regard to the association of response with survival may be due to the fact that BRIM3 patients were treatment-naïve, whereas the BRIM2 patients had received prior therapy. In fact, 49% of BRIM2 patients had received two or more prior therapies. Tumor shrinkage may have some correlation with OS in patients who have already progressed on multiple prior therapies and yet are still healthy enough to participate on a clinical trial. This suggests that the correlation of response with OS is not robust. The strongest association between response and OS in BRIM2 was seen when tumor shrinkage greater than 10% was considered a response (normalized weighted C-index = 0.49). With more stringent response criteria (i.e., shrinkage of target lesions by more than 20%), the association became weaker (normalized weighted C-index values 0.16–0.30), although these values were still greater than those seen in BRIM3.

### Discussion

Tumor shrinkage was among the first biomarkers used in oncology clinical trials and among the first surrogate endpoints for OS. The first widely adopted standard tumor response criteria, published by the World Health Organization in 1981 (7), used a 50% shrinkage of the sum of the products of bidimensional tumor diameters as the requirement for a partial

response. This threshold was based on guidelines originally designed to limit interobserver variability when measuring tumor shrinkage by physical exam (8). Subsequent iterations of response criteria generally kept this magnitude of tumor shrinkage as the threshold for a partial response (15) to allow historical trial comparison even though it was never based on correlation with OS. Several meta-analyses have shown that response can correlate with survival at the clinical trial level (6, 16, 17) but few analyses have shown correlation between response and survival at the individual patient level (18, 19) and no such analysis has been done in melanoma. Recently, Shi and colleagues using similar methodology, reported a comparable magnitude of correlation between PFS and OS in colorectal carcinoma patients treated in 12 different studies (19). Although they did not explore different criteria for PFS as we have done, our results validate these conclusions in a different tumor type. There has been effort to correlate early tumor shrinkage with OS in other cancer types at the individual patient level (20–22). However, these studies are limited by relatively low rates of tumor response and no measure of the magnitude of correlation.

The BRIM2 and BRIM3 trials were the first opportunities in melanoma to look at correlations between response and OS at the individual patient level because vemurafenib was the first drug for melanoma associated with a high objective response rate and improved OS. At the time of these trials, there was limited availability of other treatments (eg, ipilimumab) after progression that might affect OS. We found that the correlation of tumor response with OS in BRIM3 was weak no matter what threshold of response we used.

There was, however, a correlation between TTP and OS. In BRIM3, we found the strongest correlation with OS when we defined progression as >50% increase in the sum of the target lesion size. This suggests that patients with <50% increase in tumor size need not be considered to have progressed. In patients who progressed on DTIC, most progressed on the basis of target lesion growth. This is consistent with the observation that at treatment start, there is global resistance to DTIC in most patients as implied by the low response proportions. In contrast, progression on vemurafenib was often due to appearance of new tumors. This could be due to the fact that in most patients treated with vemurafenib, there is shrinkage of target lesions. When the melanoma becomes resistant, slight growth of previously undetected tumors could result in the appearance of new tumors before the target tumors had yet grown sufficiently to qualify as progression. An alternative hypothesis is that smaller, undetectable tumor deposits more readily develop resistance to RAF inhibitors than larger tumors. The correlation between TTP and OS in patients treated with vemurafenib on BRIM3 was validated by data from BRIM2 suggesting that this correlation is robust.

Recently, using standard RECIST criteria, Chan and colleagues reported that only 18% of patients treated on a variety of clinical trials either with dabrafenib or vemurafenib progressed solely on the basis of new lesion formation (23). However, their patient cohort, which was smaller than ours, was also substantially different. Their cohort had a high incidence of CNS metastases (28%) compared with <1% in our two trials combined. A substantial proportion had received prior therapy compared with no prior therapy in BRIM3. Some of their patients had BRAF V600E-mutated malignancies other than melanoma.

Previously, Flaherty and colleagues examined the correlation between PFS and OS in 12 melanoma trials at the meta-analytic level (17). The challenge with interpretation of these data is the variability in treatment and mechanism of action of the therapeutic agents in the 12 trials, and the limited follow up in some of the trials.

There are some limitations of our study. The imaging data were originally collected and curated for analysis of the study trial endpoints; we did not remeasure any of the lesions on the scans. Other lesion selection might have produced other sums of lesions and categorical responses (24). RECIST criteria for progressive disease may be met by one or more of three criteria: tumor size increase, unequivocal progression of non target disease, or new lesions. It is possible that data on new lesions may not have been consistently collected in cases where there was obvious progression as a result of target tumor measurements, or vice versa. This could affect the ratio of progressions attributed to target lesion growth versus progressions attributed to new lesions alone. Finally, we did not consider the category of non-target lesion progression alone. The category is qualitative indeed and subjective and would be difficult to categorize and conceptualize in the context of the quantitative correlative nature of this article.

There has been a revolution in the treatment of metastatic melanoma. We now have six FDA-approved drugs for metastatic melanoma, four of which have been demonstrated to improve OS compared with DTIC. It is likely that we will have additional highly effective drugs soon. This progress provides patients with multiple treatment options that can prolong OS and as a result, it may be difficult to assess the effect of any new therapy for melanoma on OS in future trials because of postprogression treatment options. In melanoma, BRIM3 may be the only opportunity to examine the correlation between the intermediate endpoints of response and TTP with OS at the individual patient level. The data indicate that TTP defined as growth of the sum of target lesions by 50% or appearance of new lesions, best correlated with OS. This correlation of TTP with OS was validated in the BRIM2 data set. Early tumor response correlated only weakly with OS.

We used TTP to obtain an unencumbered estimate of the correlation between progression on imaging and survival. In future melanoma trials, PFS may be more appropriate as a surrogate endpoint for OS as it includes death without documented progression as an event. Although changing our chief surrogate endpoint from tumor response to PFS with progression defined as we have done in this analysis would represent a major adjustment, our data indicate that it is the surrogate endpoint that more likely correlates with OS. It remains to be seen if PFS defined in this way is correlates with OS in patients treated with immunotherapy using ipilimumab and/or antibodies against PD1.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

The authors thank Betty Nelson from Genentech, Inc., who enthusiastically supported this project, facilitated data transfer from Genentech to MSKCC, and made important editing contributions to the article.

### Grant Support

P.B. Chapman was supported in part by the John K. Figge Fund.

### References

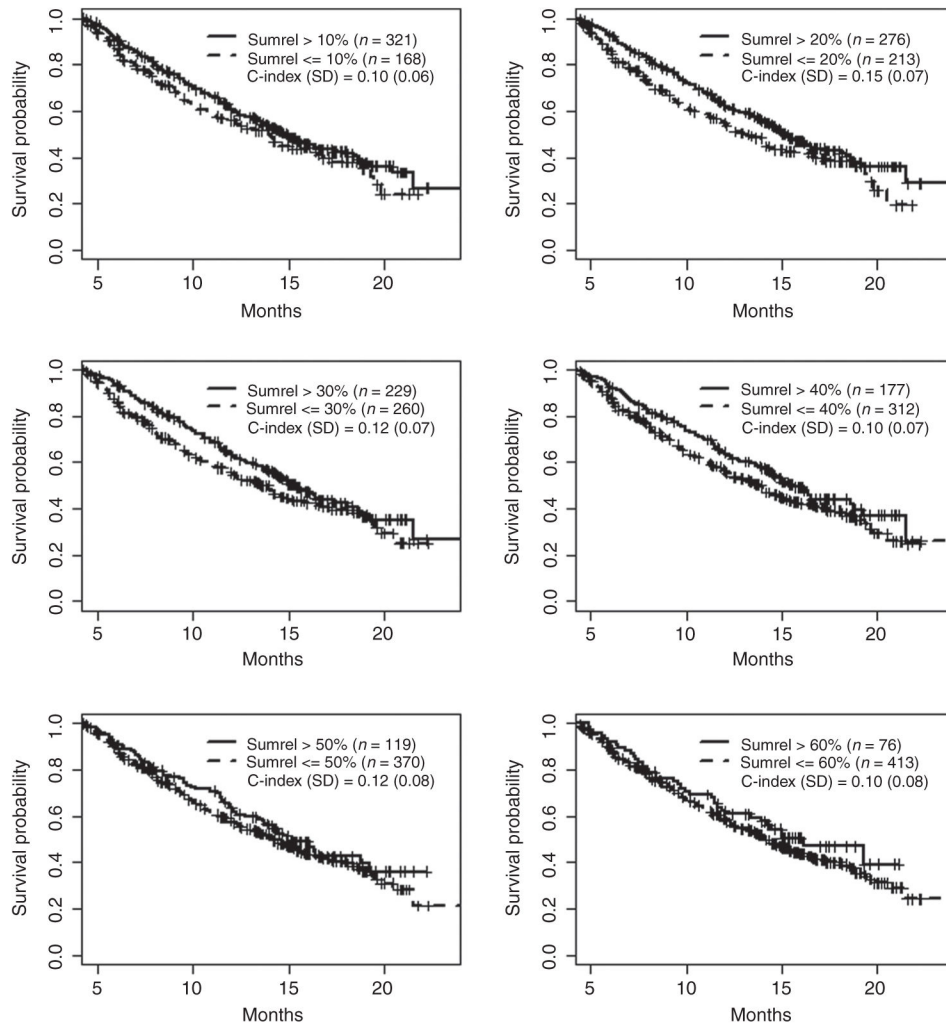
1. Pazdur R. Endpoints for assessing drug activity in clinical trials. *Oncologist*. 2008; 13:19–21. [PubMed: 18434634]
2. Fleming TR, DeMets DL. Surrogate end points in clinical trials: are we being misled? *Ann Intern Med*. 1996; 125:605–13. [PubMed: 8815760]
3. Schwartz LH, Bogaerts J, Ford R, Shankar L, Therasse P, Gwyther S, et al. Evaluation of lymph nodes with RECIST 1.1. *Eur J Cancer*. 2009; 45:261–7. [PubMed: 19091550]
4. Yanagawa M, Tatsumi M, Miyata H, Morii E, Tomiyama N, Watabe T, et al. Evaluation of response to neoadjuvant chemotherapy for esophageal cancer: PET response criteria in solid tumors versus response evaluation criteria in solid tumors. *J Nucl Med*. 2012; 53:872–80. [PubMed: 22582049]
5. Ratain MJ, Eckhardt SG. Phase II studies of modern drugs directed against new targets: if you are fazed, too, then resist RECIST. *J Clin Oncol*. 2004; 22:4442–5. [PubMed: 15483011]
6. Buyse M, Thirion P, Carlson RW, Burzykowski T, Molenberghs G, Piedbois P. Relation between tumour response to first-line chemotherapy and survival in advanced colorectal cancer: a meta-analysis. *Meta-Analysis Group in Cancer. Lancet*. 2000; 356:373–8. [PubMed: 10972369]
7. Miller AB, Hoogstraten B, Staquet M, Winkler A. Reporting results of cancer treatment. *Cancer*. 1981; 47:207–14. [PubMed: 7459811]
8. Hayward JL, Carbone PP, Heuson JC, Kumaoka S, Segaloff A, Rubens RD. Assessment of response to therapy in advanced breast cancer: a project of the Programme on Clinical Oncology of the International Union Against Cancer, Geneva, Switzerland. *Cancer*. 1977; 39:1289–94. [PubMed: 912660]
9. Sosman JA, Kim KB, Schuchter L, Gonzalez R, Pavlick AC, Weber JS, et al. Survival in BRAF V600-mutant advanced melanoma treated with vemurafenib. *N Engl J Med*. 2012; 366:707–14. [PubMed: 22356324]
10. Chapman PB, Hauschild A, Robert C, Haanen JB, Ascierto P, Larkin J, et al. Improved survival with vemurafenib in melanoma with BRAF V600E mutation. *N Engl J Med*. 2011; 364:2507–16. [PubMed: 21639808]
11. McArthur GA, Chapman PB, Robert C, Larkin J, Haanen JB, Rinventors Dummer, et al. Safety and efficacy of vemurafenib in BRAF(V600E) and BRAF(V600K) mutation-positive melanoma (BRIM-3): extended follow-up of a phase 3, randomised, open-label study patent 1474-5488 (Electronic) 1470-2045 (Linking). 2014
12. Uno H, Cai T, Pencina MJ, D'Agostino RB, Wei LJ. On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Stat Med*. 2011; 30:1105–17. [PubMed: 21484848]
13. Anderson JR, Cain KC, Gelber RD. Analysis of survival by tumor response. *J Clin Oncol*. 1983; 1:710–9. [PubMed: 6668489]
14. Oakes D. On consistency of Kendall's tau under censoring. *Biometrika*. 2008; 95:997–1001.
15. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer*. 2009; 45:228–47. [PubMed: 19097774]
16. El-Maraghi RH, Eisenhauer EA. Review of phase II trial designs used in studies of molecular targeted agents: outcomes and predictors of success in phase III. *J Clin Oncol*. 2008; 26:1346–54. [PubMed: 18285606]
17. Flaherty KT, Hennis M, Lee SJ, Ascierto PA, Dummer R, Eggermont AM, et al. Surrogate endpoints for overall survival in metastatic melanoma: a meta-analysis of randomised controlled trials. *Lancet Oncol*. 2014; 15:297–304. [PubMed: 24485879]
18. Paesmans M, Sculier JP, Libert P, Bureau G, Dabouis G, Thiriaux J, et al. Response to chemotherapy has predictive value for further survival of patients with advanced non-small cell



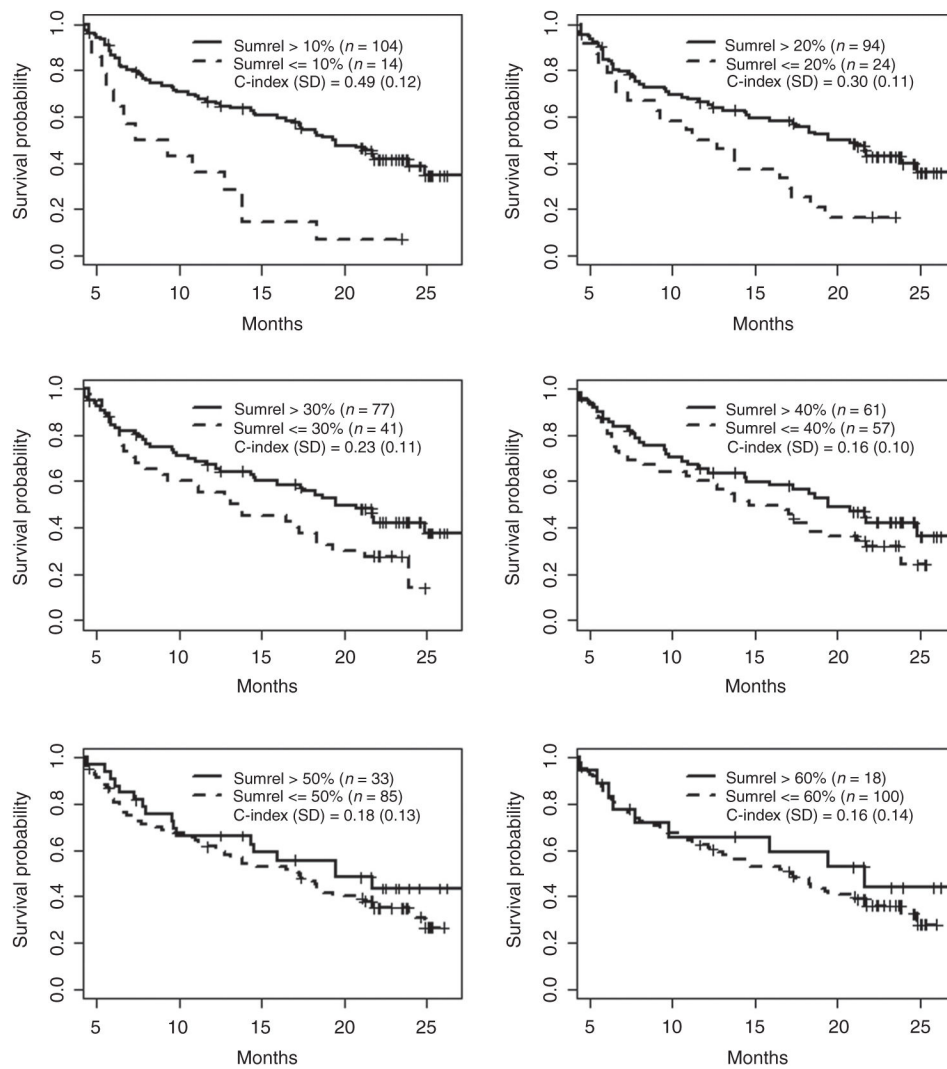
- lung cancer: 10 years experience of the European Lung Cancer Working Party. *Eur J Cancer*. 1997; 33:2326–32. [PubMed: 9616276]
19. Shi Q, de Gramont A, Grothey A, Zalcborg J, Chibaudel B, Schmoll H-J, et al. Individual patient data analysis of progression-free survival versus overall survival as a first-line end point for metastatic colorectal cancer in modern randomized trials: findings from the analysis and research in cancers of the digestive system database. *J Clin Oncol*. 2014; 33:22–8. [PubMed: 25385741]
  20. Piessevaux H, Buyse M, Schlichting M, Van Cutsem E, Bokemeyer C, Heeger S, et al. Use of early tumor shrinkage to predict long-term outcome in metastatic colorectal cancer treated with cetuximab. *J Clin Oncol*. 2013; 31:3764–75. [PubMed: 24043732]
  21. Seidel C, Busch J, Weikert S, Steffens S, Bokemeyer C, Grunwald V. Tumour shrinkage measured with first treatment evaluation under VEGF-targeted therapy as prognostic marker in metastatic renal cell carcinoma (mRCC). *Br J Cancer*. 2013; 109:2998–3004. [PubMed: 24169357]
  22. Jain RK, Lee JJ, Ng C, Hong D, Gong J, Naing A, et al. Change in tumor size by RECIST correlates linearly with overall survival in phase I oncology studies. *J Clin Oncol*. 2012; 30:2684–90. [PubMed: 22689801]
  23. Chan MM, Haydu LE, Menzies AM, Azer MW, Klein O, Lyle M, et al. The nature and management of metastatic melanoma after progression on BRAF inhibitors: Effects of extended BRAF inhibition. *Cancer*. 2014; 120:3142–53. [PubMed: 24985732]
  24. Schwartz LH, Mazumdar M, Brown W, Smith A, Panicek DM. Variability in response assessment in solid tumors: effect of number of lesions chosen for measurement. *Clin Cancer Res*. 2003; 9:4318–23. [PubMed: 14555501]

### Translational Relevance

For patients with metastatic melanoma, there are now several treatment options that improve overall survival. In the future, clinical trials developing new treatments may not be able to rely on overall survival (OS) as a primary endpoint. The field will need surrogate endpoints that correlate with OS *at the individual patient level*. Our data show that early tumor shrinkage (within the first 12 weeks) was only weakly correlated with OS no matter what threshold of tumor response we tested (ranging from any response to 100% shrinkage). On the other hand, time to tumor progression (TTP) was much more strongly correlated at the individual patient level, especially if progression was defined as 50% tumor growth. These results suggest that TTP (or progression-free survival) defined in this way might be a better surrogate for OS than tumor response.



**Figure 1.** Correlation of tumor response over the first 12 weeks with overall survival in all patients on BRIM3. Responses were defined as 10%, 20%, 30%, 40%, 50%, or 60% reduction in the sum of the target lesions. The weighted C-index was normalized to range between 0 and 1, and SDs were calculated using bootstrapping. Tick marks, censored patients.



**Figure 2.** Correlation of tumor response over the first 12 weeks with overall survival in all patients on BRIM2. Responses were defined as 10%, 20%, 30%, 40%, 50%, or 60% reduction in the sum of the target lesions. The weighted C-index was normalized to range between 0 and 1, and SDs were calculated using bootstrapping. Tick marks, censored patients.

**Table 1**

Summary of progression events defined by various cut points of relative increase in sum of lesion size, and correlation with OS in the full BRIM3 study population ( $n = 582$ )

	From baseline				From nadir			
	Without new lesions		With new lesions		Without new lesions		With new lesions	
	Number of events	Correlation with OS (SE)	Number of events	Correlation with OS (SE)	Number of events	Correlation with OS (SE)	Number of events	Correlation with OS (SE)
Relative increase								
Any	287	0.330 (0.045)	451	0.411 (0.032)	NA	NA	NA	NA
25%	180	0.543 (0.046)	404	0.541 (0.032)	306	0.547 (0.038)	468	0.524 (0.029)
50%	104	0.675 (0.065)	376	0.588 (0.030)	207	0.638 (0.044)	423	0.568 (0.029)
100%	57	0.072 (0.142)	361	0.526 (0.033)	132	0.522 (0.075)	396	0.576 (0.029)

Abbreviation: NA, not assessable.

**Table 2**

Summary of progression events defined by various cut points of relative increase in sum of lesion size, and correlation with OS among patients randomized to vemurafenib in the BRIM3 study ( $n = 325$ )

	From baseline				From nadir			
	Without new lesions		With new lesions		Without new lesions		With new lesions	
	Number of events	Correlation with OS	Number of events	Correlation with OS	Number of events	Correlation with OS	Number of events	Correlation with OS
Relative increase								
Any	81	0.341 (0.077)	221	0.518 (0.042)	NA	NA	NA	NA
25%	43	0.443 (0.099)	201	0.608 (0.039)	145	0.521 (0.049)	252	0.564 (0.036)
50%	18	NA	192	0.655 (0.034)	96	0.605 (0.053)	227	0.614 (0.034)
100%	8	NA	188	0.664 (0.035)	57	0.513 (0.087)	212	0.630 (0.035)

Abbreviation: NA, not assessable.

**Table 3**

Summary of progression events defined by various cut points of relative increase in sum of lesion size, and correlation with OS among patients randomized to dacarbazine in the BRIM3 study ( $n = 257$ )

	From baseline				From nadir			
	Without new lesions		With new lesions		Without new lesions		With new lesions	
	Number of events	Correlation with OS	Number of events	Correlation with OS	Number of events	Correlation with OS	Number of events	Correlation with OS
Relative increase								
Any	206	0.353 (0.054)	230	0.355 (0.051)	NA	NA	NA	NA
25%	137	0.600 (0.054)	203	0.508 (0.045)	161	0.587 (0.053)	216	0.504 (0.045)
50%	86	0.727 (0.073)	184	0.546 (0.048)	111	0.697 (0.069)	196	0.541 (0.047)
100%	52	0.617 (0.138)	173	0.552 (0.046)	75	0.577 (0.138)	184	0.545 (0.046)

Abbreviation: NA, not assessable.

**Table 4**

Summary of progression events defined by various cut points of relative increase in sum of lesion size, and correlation with OS among BRIM2 study participants ( $n = 132$ )

Relative increase	From baseline				From nadir			
	Without new lesions		With new lesions		Without new lesions		With new lesions	
	Number of events	Correlation with OS	Number of events	Correlation with OS	Number of events	Correlation with OS	Number of events	Correlation with OS
Any	28	0.634 (0.091)	89	0.668 (0.053)	NA	NA	NA	NA
25%	14	NA	84	0.693 (0.046)	69	0.613 (0.063)	111	0.635 (0.047)
50%	8	NA	84	0.723 (0.037)	36	0.720 (0.069)	95	0.708 (0.039)
100%	3	NA	83	0.717 (0.038)	16	NA	89	0.710 (0.041)

Abbreviation: NA, not assessable.