

Tumor Status at 12 Weeks Predicts Survival in Advanced Colorectal Cancer: Findings from NCCTG N9741

JAMES M. HEUN,^a AXEL GROTHEY,^a MEGAN E. BRANDA,^a RICHARD M. GOLDBERG,^b DANIEL J. SARGENT^a

^aMayo Clinic, Rochester, Minnesota, USA; ^bUniversity of North Carolina, Chapel Hill, North Carolina, USA

Key Words. Colorectal cancer • Tumor status • Endpoints • Clinical trial

Disclosures

James M. Heun: None; **Axel Grothey:** *Consultant/advisory role:* Amgen, Pfizer, Genentech, Bayer; *Research funding/contracted research:* Genentech, Daiichi, Bayer; **Megan E. Branda:** None; **Richard M. Goldberg:** *Consultant/advisory role:* sanofi-aventis; **Daniel J. Sargent:** None.

Section Editor **Patrick Johnston** discloses employment with Almac Diagnostics; a consulting relationship with Almac, Roche, Chugai Pharmaceuticals, and sanofi-aventis; honoraria received from AstraZeneca, Chugai Pharmaceuticals, Pfizer, sanofi-aventis, and Roche; research funding from AstraZeneca and Amgen; and ownership interests in Almac Diagnostics and Fusion Antibodies.

Section Editor **Peter O'Dwyer** discloses a consulting relationship with Tetralogic Pharmaceuticals, PrECOG, and AstraZeneca; an advisory relationship with Nereus Pharmaceutical, Tetralogic Pharmaceuticals, and PrECOG; research support from Pfizer, Bristol-Myers Squibb, Methylgene, Novartis, Genentech, Bayer, Merck, Kosan, Ardea, and Exelixis; honoraria received from Genentech, Bayer, Methylgene, and Bristol-Myers Squibb; and ownership interest with Tetralogic Pharmaceuticals.

Reviewers "A" and "B" disclose no financial relationships.

The content of this article has been reviewed by independent peer reviewers to ensure that it is balanced, objective, and free from commercial bias. On the basis of disclosed information, all conflicts of interest have been resolved.

LEARNING OBJECTIVES

After completing this course, the reader will be able to:

1. Explain the difference in survival prediction between response criteria (WHO) when used as a two-level variable (CR/PR vs. other) and as a three-level variable (CR/PR vs. SD vs. PD).
2. Describe the limited benefit of using actual tumor measurements over traditional criteria (as a three-level variable) in predicting survival in colorectal cancer.

CME

This article is available for continuing medical education credit at CME.TheOncologist.com.

ABSTRACT

Purpose. We explored the prognostic value of actual tumor measurements (TM) versus World Health Organization (WHO) criteria as three-level (responder, stable, and progression) and two-level (responder and non-responder) variables at 12 and 24 weeks as predictors of

survival in Intergroup Trial N9741, a phase III trial in metastatic colorectal cancer (CRC).

Methods. All patients with measurable disease ($N = 1,188$) were included. The percentage changes in TM from baseline to 12 and 24 weeks were calculated. The

Correspondence: Dan Sargent, Ph.D., Mayo Clinic, 200 First Street SW, Rochester, Minnesota 55905, USA. Telephone: 507-284-8803; Fax: 507-266-2477; e-mail: sargent.daniel@mayo.edu Received March 3, 2011; accepted for publication March 30, 2011; first published online in *The Oncologist Express* on May 31, 2011. ©AlphaMed Press 1083-7159/2011/\$30.00/0 doi: 10.1634/theoncologist.2011-0064

prognostic values of TM versus WHO criteria (as three- and two-level variables) at 12 and 24 weeks were compared, using Cox models for overall survival (OS) in a landmark analysis, adjusting for baseline tumor size, performance status, and treatment arm.

Results. Tumor status at 12 weeks by WHO criteria (three or two levels) or actual TM were all strongly associated with OS. Actual TM provided no meaningful additional benefit compared with the three-level WHO criteria. Tumor status at 24 weeks was also strongly associated with survival, but added no additional prog-

nostic value compared with the 12-week assessment. At 12 weeks, actual TM improved prognostic characterization of patients with WHO status of response, but provided no additional value in patients with stable disease or progression.

Conclusions. In N9741, the use of actual TM, or following tumor status beyond 12 weeks, did not improve survival prediction compared with a single three-level response assessment at 12 weeks, suggesting that 12-week tumor status could be an appropriate phase II trial endpoint in metastatic CRC. *The Oncologist* 2011;16:859–867

INTRODUCTION

Endpoints used in colorectal cancer (CRC) trials have come under intense scrutiny as we continue to determine clinically meaningful outcomes [1–8]. The controversy pertains not only to phase III trials, but also to phase II trials, which are critical in providing insight into which drugs should advance to phase III testing [3, 9]. Phase II trials are employed to identify promising therapeutics, and to eliminate as efficiently as possible those agents doomed to ultimately fail. There are numerous phase II studies which showed promising activity for single agents and combinations of agents that when tested in larger and costly phase III studies led to negative results [10, 11].

The traditional binary endpoint in CRC of response rate (RR) as determined by established criteria (Response Evaluation Criteria in Solid Tumors [RECIST] or World Health Organization [WHO]) [12, 13] has helped to identify promising cytoreductive agents to promote to future studies. However, RR alone has been a poor predictor of eventual success in phase III trials [11, 14]. Using the two-level approach of RR (responder, non-responder) does not include stable disease (SD) in response. We know, nonetheless, that patients treated with traditional cytoreductive chemotherapy can still experience a clinical benefit without a significant change in tumor size [8]. With many new biologic agents now being tested that may offer significant clinical benefit with a cytostatic rather than cytoreductive mechanism of action [3, 8], the use of RR as the primary endpoint in early phase II trials may overlook the potential benefit of such agents.

Though other endpoints have been proposed, progression-free survival (PFS) currently remains the most widely accepted alternative endpoint to overall survival in phase III trials for CRC. The PFS endpoint takes into account not only those patients with tumor shrinkage but also those with stable disease who may still garner a meaningful clinical benefit from cytostatic drugs [1, 15, 16]. A similar endpoint for phase II trials in CRC that includes the benefit of stable

disease is desirable. Simply moving from the binary endpoint of RR to the time-to-event variables of PFS or time to progression (TTP) in phase II trials includes in the estimate of benefit those agents capable of producing stable disease. However, this also adds significant time and expense to an early stage trial. Because median PFS times with current combination therapies in first-line CRC are now approximately 8–10 months, the identification of other, earlier endpoints for phase II trials is desirable. The use of PFS or TTP at a specific time point is one appealing strategy for phase II trials and has the potential to provide clinically useful information in a cost-efficient and timely manner. Such a model has been studied in glioblastoma multiforme (GBM) [17], resulting in validation of PFS at 6 months as the primary endpoint in many phase II trials in that setting [18–21]. An 8-week assessment of tumor status has been studied in lung cancer and found to be predictive of survival [22]. Mathematical models incorporating change in tumor size at 7 weeks have been developed in CRC to help predict survival in future studies [23]. To our knowledge, no study has yet evaluated the survival prediction of tumor status at specific time points in CRC as a possible endpoint for future studies.

It has been suggested that the use of actual tumor size (often expressed via waterfall plots) may provide an additional valuable measure of antitumor activity [24]. The use of a continuous endpoint of change in actual tumor size has also been suggested as an alternative to the traditionally established criteria, such as RECIST or WHO, in predicting the true benefit of a drug [25]. The thought is that a patient with tumor *shrinkage* of 15% will likely do better than a patient with tumor *growth* of 15%, yet they are both listed as “stable disease” under the established RECIST criteria [12]. In many studies, however, continuous endpoints of actual change in tumor size report the “best” response for each patient throughout the study rather than at an equal and specific time point, which can make interpretation of the re-

Table 1. Baseline patient characteristics (*N* = 1,188)

	IFL (<i>N</i> = 338)	FOLFOX (<i>N</i> = 554)	IROX (<i>N</i> = 296)	Overall (<i>N</i> = 1,188)
Assessment time 12 wks: mean (SD)	11.9 (2.3)	11.6 (2.6)	11.7 (2.3)	11.7 (2.5)
Assessment time 24 wks: mean (SD)	23.8 (1.8)	23.6 (2.2)	23.9 (2.1)	23.8 (2.1)
Follow-up months in patients alive at the end of study: median (95% CI)	45.1 (38.3, 53.7) <i>N</i> = 15	46.0 (43.7, 49.5) <i>N</i> = 83	54.5 (37.2, 55.7) <i>N</i> = 20	43.8 (43.9, 49.2) <i>N</i> = 118
Tumor status at 12 wks				
Responder (CR/PR)	114 (34%)	237 (43%)	98 (33%)	449 (38%)
Stable	146 (43%)	251 (45%)	132 (45%)	529 (45%)
Progression	78 (23%)	66 (12%)	66 (22%)	210 (18%)
Tumor status at 24 wks				
Responder	103 (30%)	232 (42%)	90 (30%)	425 (30%)
Stable	87 (26%)	165 (30%)	80 (27%)	332 (27%)
Progression	148 (44%)	157 (28%)	126 (43%)	431 (43%)
Overall confirmed status				
Responder	121 (36%)	270 (49%)	114 (39%)	505 (43%)
Age: mean (SD)	59.5 (11.1)	60.5 (11.5)	59.8 (11.8)	60 (11.5)
Male	225 (67%)	332 (60%)	180 (61%)	737 (62%)
Prior adjuvant chemotherapy	51 (15%)	87 (16%)	39 (13%)	177 (15%)
PS				
0–1	327 (97%)	530 (96%)	280 (95%)	1137 (96%)
2	11 (3%)	24 (4%)	16 (5%)	51 (4%)
Race				
White	290 (86%)	469 (85%)	255 (86%)	1014 (85%)
Black	35 (10%)	38 (7%)	24 (8%)	97 (8%)
Hispanic	9 (3%)	24 (4%)	10 (4%)	43 (4%)
Other	4 (1%)	23 (4%)	7 (2%)	34 (3%)

Abbreviations: CI, confidence interval; CR, complete response; FOLFOX, oxaliplatin and infused fluorouracil plus leucovorin; IFL, irinotecan and bolus fluorouracil plus leucovorin; IROX, irinotecan and oxaliplatin; PR, partial response; PS, performance status; SD, standard deviation.

sults to an individual patient quite difficult, and ultimately may lessen the value of the information.

In this study, we explored the prognostic value of tumor measurements at 12 weeks and 24 weeks as predictors of survival in Intergroup Trial N9741, which compared irinotecan and bolus fluorouracil plus leucovorin (IFL), oxaliplatin and infused fluorouracil plus leucovorin (FOLFOX), and irinotecan and oxaliplatin (IROX) as first-line therapy in advanced CRC [26]. Our aims are (a) to determine whether an early categorical endpoint of tumor assessment is capable of predicting survival in colorectal cancer and (b) to compare the prognostic value of actual tumor measurements (TM) versus the traditional WHO criteria (both as a three-level variable that incorporates SD and as a two-level variable that does not include SD in response) at these time points.

PATIENTS AND METHODS

Patients (*N* = 1,188) from the North Central Cancer Treatment Group (NCCTG) N9741 trial comparing frontline chemotherapy in metastatic colon cancer with measurable disease were included in the analysis. Baseline patient characteristics are given in Table 1. Actual tumor measurements were obtained and the percentage changes in TM for each patient from baseline to 12 weeks and 24 weeks were calculated. If a 12-week measurement was not available, tumor measurements obtained during the closest assessment made during weeks 6–18 were used. For the 12-week landmark analysis, 90 of 1,188 patients did not have a radiographic assessment within the 6- to 18-week time frame. For these 90 patients, we assigned a 50% increase in TM if they had a clinical status of “progression” during the time

frame ($N = 28$) and 0% change if they did not have progression ($N = 62$). Similarly, for the 24-week assessment, if no 24-week tumor measurement was available, we used measurements from the closest assessment made during weeks 18–30 to calculate the percentage change in TM. For the 24-week landmark analysis, 378 of 1,116 patients did not have a radiographic assessment within the 18- to 30-week time frame. For these 378 patients, we again assigned a 50% increase in TM if they had a clinical status of “progression” during the time frame ($N = 226$) and 0% change if they did not have progression ($N = 152$). For modeling, the transformation $\log(1.01 + \% \text{ change})$ and $\log(\text{baseline measurement/absolute change})$ was used to provide normality.

WHO criteria assess the sum of the products of index lesions, requiring a 50% reduction for partial response (PR), complete disappearance for complete response (CR), and a 25% increase for progressive disease (PD), with all other measurements representing stable disease (SD) [13]. WHO criteria require confirmation of response after 4 weeks for patients with PR or CR. In this study, WHO disease status at 12 and 24 weeks was considered as both a two-level (response [CR/PR] versus no response [SD, PD]) and a three-level (response, stable, and progression) variable. The WHO status reported at the time of the tumor measurement was used with the time frames described above.

Tumor measurements in N9741 were obtained by the treating institution and measures were taken to review data quality beyond the normal study data validation. In the original study, 10% of cases were selected for quality assurance with review of clinical reports and data collection by the primary investigator with no concerns recognized. The FDA also performed audits on selected sites (as N9741 was the registration study for the first-line indication for FOLFOX) with no concerns in data collection identified.

The prognostic value of actual change in tumor size, versus WHO disease status (both as a two-level and a three-level variable), were compared at 12 and 24 weeks using Cox models for OS after 12 weeks and after 24 weeks, respectively, in a landmark analysis, adjusting for baseline tumor size, performance status, and treatment arm. The prognostic value for each model was assessed using the concordance index C . The concordance index is similar to a correlation coefficient from a simple two-way association, or R^2 from a linear regression. The concordance index is a useful supplement to a p value in a Cox regression. p values may be statistically significant but the result not clinically meaningful. The concordance index provides information regarding the strength of the relationship between the variables and outcomes, where 0.5 means there is no association and 1.0 means that there is a perfect association or

Table 2. Multivariate analysis for overall survival (12-week landmark, $N = 1,188$)

Variable	HR (95% CI)	<i>p</i>
12-week tumor status (three-level variable); $C = 0.66$		
Stable ^a	1.09 (0.93, 1.27)	.28
Progression ^a	2.89 (2.48, 3.37)	<.01
PS ^c	1.48 (1.07, 2.07)	.02
Arm F (FOLFOX) ^b	0.69 (0.60, 0.80)	<.01
Arm G (IROX) ^b	0.83 (0.70, 0.98)	.03
Log(baseline measure)	1.27 (1.21, 1.34)	<.01
12-week tumor status (two-level variable); $C = 0.65$		
Non-responder ^a	2.81 (2.40, 3.29)	<.01
PS ^c	1.59 (1.19, 2.13)	<.01
Arm F (FOLFOX) ^b	0.68 (0.59, 0.79)	<.01
Arm G (IROX) ^b	0.85 (0.73, 1.00)	.05
Log(baseline measure)	1.27 (1.20, 1.34)	<.01
12-week actual TM percentage change; $C = 0.64$		
Log(% change at 12 wks)	1.49(1.37, 1.63)	<.01
PS ^c	1.68 (1.26, 2.25)	<.01
Arm F (FOLFOX) ^b	0.69 (0.59, 0.79)	<.01
Arm G (IROX) ^b	0.88 (0.75, 1.03)	.11
Log(baseline measure)	1.24 (1.18, 1.31)	<.01

^aResponse of CR/PR is the reference group for confirmed response comparisons.

^bArm A (IFL) is reference group for treatment comparisons.

^cReference is 0–1.

Abbreviations: CI, confidence interval; CR, complete response; FOLFOX, oxaliplatin and infused fluorouracil plus leucovorin; HR, hazard ratio; IFL, irinotecan and bolus fluorouracil plus leucovorin; IROX, irinotecan and oxaliplatin; PR, partial response; PS, performance status; TM, tumor measurement.

prediction ability. The added value of confirmation of response was assessed using a landmark analysis starting at 24 weeks after randomization, comparing 12-week disease status with disease status at 24 weeks, where to be declared a responder at 24 weeks, the response must have been maintained for at least 4 weeks.

RESULTS

Multivariate analyses for overall survival with the 12-week landmark time point are reported in Table 2. WHO status at 12 weeks was strongly associated with survival when used as either a three-level ($p < .01$) or two-level variable ($p < .01$). Actual TM were also strongly associated with OS at 12 weeks ($p < .01$), but provided no meaningful additional

Table 3. Multivariate analysis for overall survival (24-week landmark, $N = 1,116$)

Variable	HR (95% CI)	<i>p</i>
Tumor status (three-level variable); $C = 0.68$		
Stable ^a	1.09 (0.93, 1.27)	.28
Progression ^a	2.89 (2.48, 3.37)	<.01
PS ^c	1.48 (1.07, 2.07)	.02
Arm F (FOLFOX) ^b	0.69 (0.60, 0.80)	<.01
Arm G (IROX) ^b	0.83 (0.70, 0.98)	.03
Log(baseline measure)	1.25 (1.19, 1.33)	<.01
Tumor status (two-level variable); $C = 0.69$		
Non-responder ^a	2.78 (2.42, 3.18)	<.01
PS ^c	1.51 (1.08, 2.10)	.02
Arm F (FOLFOX) ^b	0.70 (0.61, 0.82)	<.01
Arm G (IROX) ^b	0.83 (0.70, 0.99)	.04
Log(baseline measure)	1.23 (1.17, 1.30)	<.01
Actual TM percentage change; $C = 0.65$		
Log(% change at 24 wks)	1.36 (1.27, 1.45)	<.01
PS ^c	1.42 (1.02, 1.97)	.04
Arm F (FOLFOX) ^b	0.71 (0.61, 0.82)	<.01
Arm G (IROX) ^b	0.88 (0.75, 1.05)	.15
Log(baseline measure)	1.23 (1.17, 1.30)	<.01

^aResponse of CR/PR is the reference group for 24-week response comparisons.
^bArm A (IFL) is reference group for treatment comparisons.
^cReference is 0–1.
Abbreviations: CI, confidence interval; CR, complete response; FOLFOX, oxaliplatin and infused fluorouracil plus leucovorin; HR, hazard ratio; IFL, irinotecan and bolus fluorouracil plus leucovorin; IROX, irinotecan and oxaliplatin; PR, partial response; PS, performance status; TM, tumor measurement.

predictive value ($C = 0.64$) compared to the WHO three-level assessment ($C = 0.66$).

A confirmatory analysis with the 24-week landmark was performed with similar results (Table 3). WHO status as a two-level variable and actual TM remained strongly associated with survival when analyzed at 24 weeks; however, the 24-week status provided no meaningful additional predictive value ($C = 0.68$) compared to 12 weeks ($C = 0.66$). The three-level WHO status at 24 weeks provided no added value compared to tumor status as a two-level variable at 24 weeks, as the difference in survival for stable patients versus responding patients disappeared ($p = .28$). Restricting the analysis to consider only confirmed responses in the 24-week landmark analysis did not seem to add value ($p < .01$, $C = 0.69$).

Kaplan-Meier estimates of OS based on 12- and 24-week tumor status are found in Figure 1. There is clear separation ($p < .01$) in the survival curves when a three-level variable is used for the WHO criteria at 12 weeks, but there is no difference in survival between responding and stable patients at 24 weeks.

To further explore any potential benefit of actual TM over WHO status as a three-level variable, a subset analysis was performed within each of the 3 categorizations of response, where overall survival was modeled by percent change of tumor measurements alone (Table 4). Of the three categories, tumor measurements were associated with survival within CR/PR patients ($p < .01$), but not within patients with SD ($p = .87$) or progression ($p = .81$) (Figure 2). Of note, within CR/PR patients the survival curves separated only after 15 months.

DISCUSSION

The search for promising endpoints for phase II trials in oncology is highly relevant and is the subject of active investigation. In this report, we have provided data from a large randomized phase III trial in CRC that provides hypothesis generating results for the development of novel endpoints for randomized phase II studies. First, in our data set, tumor status, as a three-level variable at 12 weeks, that incorporates SD, was an important predictor of survival and provided identification of additional clinical benefit compared to RR alone. There are distinct separations in the survival curves when comparing patients whose 12-week status was CR/PR versus SD versus PD. This indicates that using RR alone (CR/PR) as an endpoint in phase II trials will miss potentially clinically beneficial agents that may offer disease control. We know from previous studies that patients in N9741 without a tumor response still received a clinical benefit from these traditionally cytoreductive therapies [8], but this analysis now shows further separation among “non-responders” by separating survival curves for those patients with stable disease and those with progressive disease. This finding, demonstrated in a large study using response inducing agents, is likely to be even more relevant in the era of cytostatic biologic agents used alone or in combination with traditional chemotherapy.

Second, actual tumor measurements provided little additional benefit over WHO status (considered as a three-level variable; CR/PR, SD, PD) at 12 weeks to predict subsequent patient survival. This is another critical finding that may provide guidance in designing future phase II trials. In evaluating the survival curves separately for patients with 12-week status of CR/PR, SD, and PD, actual tumor measurements provided additional predictive ability for only one group—those patients achieving a response. This

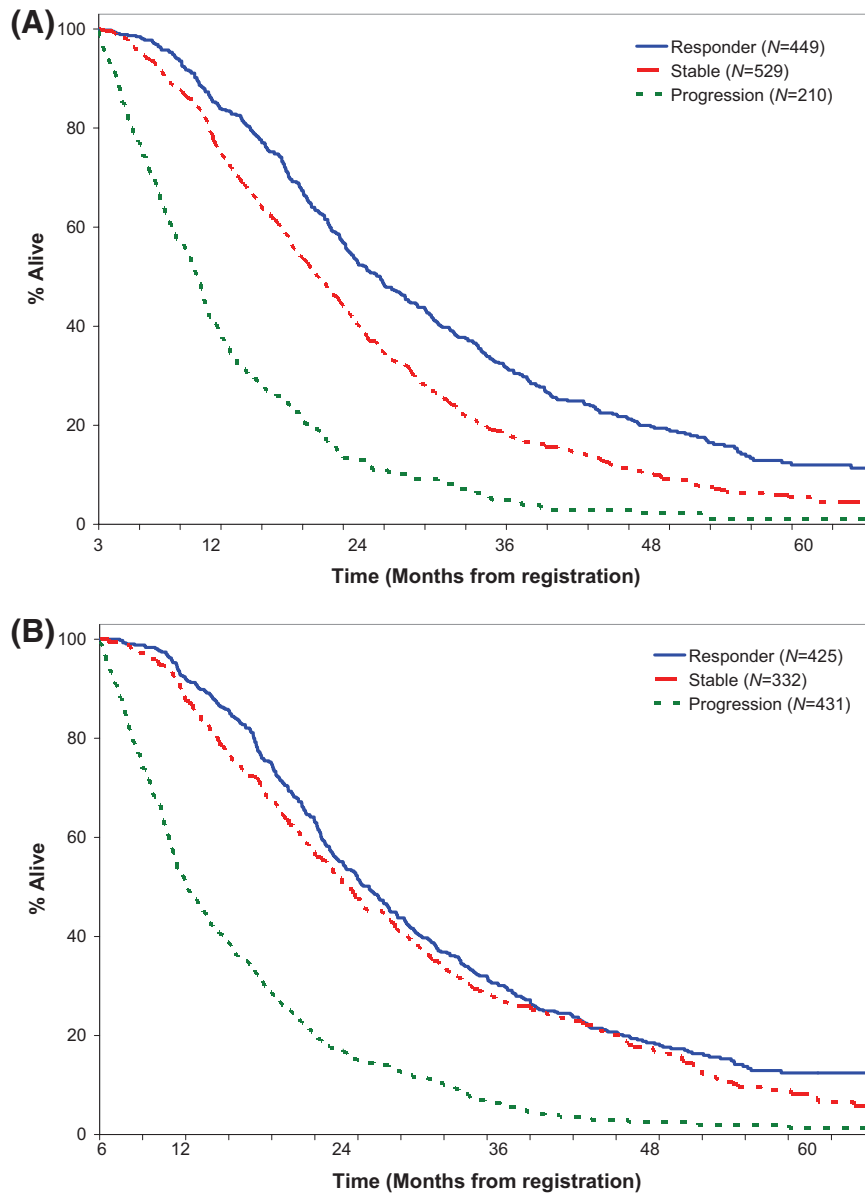


Figure 1. Kaplan-Meier estimates of overall survival based on 12- and 24-week tumor status. **(A):** Overall survival based on 12-wk tumor status (World Health Organization criteria, three-variable model). Median survival: responder = 20.7 mos; stable = 16.2 mos; progression = 7.2 mos. **(B):** Overall survival based on 24-week tumor status (World Health Organization criteria, three-variable model). Median survival: response = 18.8 mos; stable = 17.0 mos; progression = 6.0 mos.

indicates a potential difference in survival benefit based on the extent of the response, but the usefulness of this finding may be of only a modest benefit in early trials. Especially in phase II trials, the key determination is whether there is truly a clinical benefit of a therapy to promote further confirmatory studies, not to further clarify the range of benefit among those patients responding. Further, the survival advantage for patients with greater response emerged only after extended follow-up. Thus, the use of traditional established disease status criteria, when expanded to a three-level factor as opposed to the traditional responder/

non-responder classification in phase II trials, appears to be adequate. Whether there is a role for actual tumor measurements in larger phase III trials as well as whether the prognostic value added to those patients responding will be of benefit for clinicians in practice remains to be determined.

Using actual tumor measurements as a continuous endpoint in phase II trials has been suggested to offer the benefit of smaller sample sizes, but the value of these measurements in predicting survival had not been validated [25]. The assumption that a patient with tumor growth of 15% will do worse than one with tumor shrinkage of 15%

Table 4. Univariate analysis of percentage change within 12-week tumor status categories

Tumor status	Variable	HR (95% CI)	<i>p</i>
CR/PR	Log(% change)	1.29 (1.12, 1.48)	<.01
Stable	Log(% change)	1.03 (0.72, 1.49)	.87
Progression	Log(% change)	1.03 (0.82, 1.30)	.81

^a12-week landmark analysis.
Abbreviations: CI, confidence interval; CR, complete response; HR, hazard ratio; PR, partial response.

was not demonstrated in our analyses and calls into question the added value of actual tumor measurements over traditional categorical disease assessment criteria at least in early phase studies. The suggestion to treat actual tumor measurements as longitudinal data over time [25] may ultimately require a longer study duration and likely negate the benefit of the reduced sample size in regards to trial feasibility.

Third, neither continuous tumor measurements at 24 weeks, nor WHO status at 24 weeks, nor confirmed response within the first 24 weeks added prognostic value compared to tumor status at 12 weeks in patients with advanced CRC. The traditional WHO and RECIST criteria require confirmation of response at four weeks, but the data here suggest that may not be required. The updated RECIST 1.1 criteria still require confirmation of response if the trial's primary endpoint is tumor response rate, but no longer require confirmation in randomized studies [27]. The findings from this study suggest that as a primary endpoint, a 12-week assessment in patients with CRC alone may provide an adequate measure of the benefit of therapy, thus improving feasibility and controlling costs otherwise incurred if using other time-dependent variables.

Using a fixed-time disease status variable has also been studied in other tumors with favorable results [22]. Ballman et al. compared the relationship between 6-month PFS and 12-month overall survival in phase II trials in patients with glioblastoma multiforme and determined that PFS at 6 months was a reasonable endpoint for phase II trials in patients with GBM [17]. This assessment has led to a shift in phase II trial design in neurooncology and many GBM phase II trials now use 6-month PFS rate as the primary endpoint. Lara et al. evaluated disease control rate (CR/PR/SD) at 8 weeks in patients with advanced non-small cell lung cancer and found that disease control rate at 8 weeks not only provided an early assessment of subsequent outcomes, but was also a more powerful predictor of survival than the RR alone [22].

Our study also provides potential benefit to clinicians in

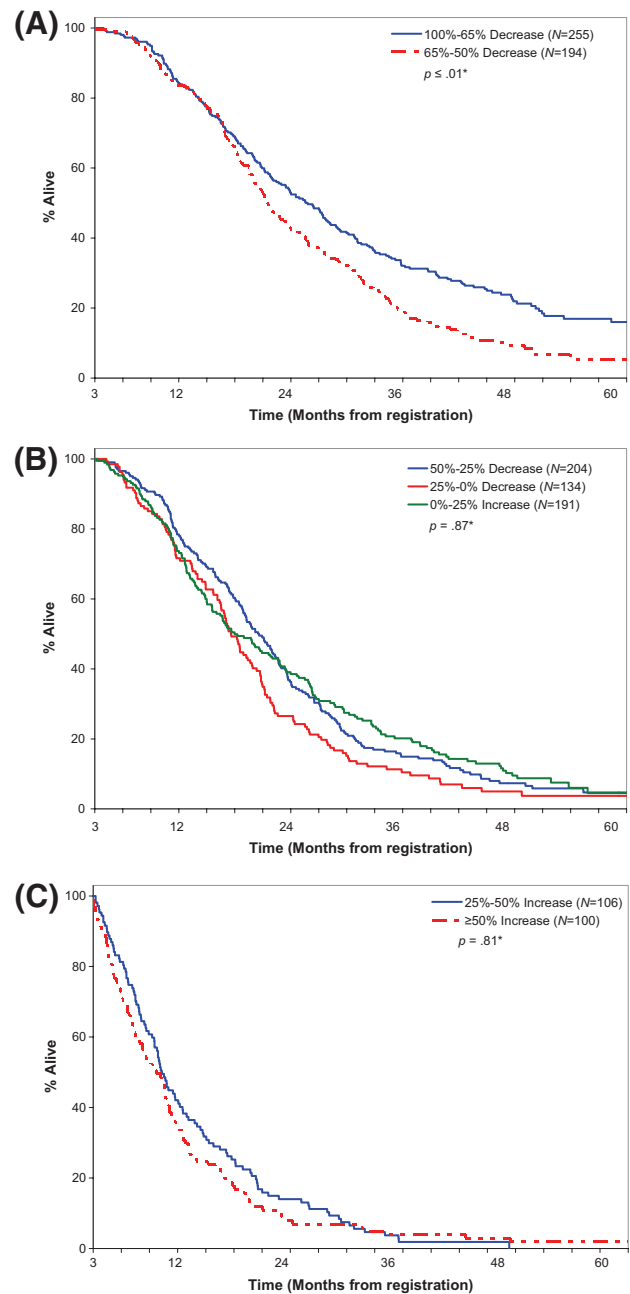


Figure 2. Added benefit of percentage of change in tumor measurement at 12 wks by tumor status. **(A)**: Responsive disease (complete response/partial response). Median time to survival: 100%–65% decrease = 22.8 mos; 65%–50% decrease = 18.5 mos. **(B)**: Stable disease. Median time to survival: 50%–25% decrease = 17.6 mos; 0–25% decrease = 14.6 mos; 0–25% increase = 15.0 mos. **(C)**: Progressive disease. Median time to survival: ≥50% increase = 6.8 mos; 25%–50% increase = 7.3 mos.

counseling patients regarding their clinical response to therapy at 12 weeks. Early tumor evaluation in predicting clinical outcomes has been evaluated in other tumor sites [22, 28–30] as well as colorectal cancer [31–34]. It is notable, however, that this study did not include biologic agents that

are now a standard part of therapy in advanced CRC. Because the therapeutic agents used (i.e., biologic or other cytostatic agents) may also influence the overall prognostic value of tumor status in CRC [34], caution must be exercised by clinicians before using this information in counseling patients until further analysis is available.

The inclusion of stable disease at 12 weeks as an endpoint in phase II trials in CRC heightens the importance of randomization to a concurrent control arm to help distinguish natural tumor biology and supportive care benefits from potentially disease controlling agents [3, 14]. Though randomization will increase the sample size required, using the endpoint of a 12-week tumor assessment in CRC will limit the duration of the study, ensuring that these early studies will be not only feasible, but also more fruitful.

There are limitations to our study that should be acknowledged. These findings are from a single, large clinical trial, with response-inducing agents. Most patients (75%) had ≤ 3 assessed lesions; thus, the sum of tumor measurements may not reflect the entire tumor burden. Greater than 50% of patients received second-line and later-line therapies, possibly reducing the relationship between first-line tumor shrinkage and eventual survival. Only a small number of patients with CR ($N = 23$) were included. Though central radiologic review was not required, quality assurance procedures were followed, allowing this study to yield results from “real world” data. Importantly, the findings from this study are based on agents that frequently induce tumor shrinkage and therefore, analysis should be repeated from large clinical trials using cytostatic agents. Results of

this study are potentially useful in phase II trials for colon cancer, but there should be caution in extrapolating results to other tumor types as therapeutic agents and tumor biology will differ. None of the measures assessed here provided highly accurate prediction, as assessed by the concordance index, indicating the continued need to develop new early clinical trial endpoints.

CONCLUSION

In this large phase III randomized trial, tumor status as a three-level variable (CR/PR, SD, PD) at 12 weeks after randomization predicted subsequent overall survival and may be an appropriate and feasible endpoint for further consideration in future phase II clinical trials for CRC. Actual tumor measurements versus a three-level WHO criteria at 12 weeks offered only a mild improvement in survival prediction, limited only to those patients with a tumor response. Following tumor status for >12 weeks, with either actual tumor measurements or by WHO criteria, and the requirement of confirming a response did not improve survival prediction compared to a single assessment at 12 weeks.

AUTHOR CONTRIBUTIONS

Conception/Design: Axel Grothey, Daniel Sargent, Richard M. Goldberg
Provision of study material or patients: Axel Grothey, Daniel Sargent, Richard M. Goldberg
Collection and/or assembly of data: James Heun, Axel Grothey, Megan Branda, Daniel Sargent
Data analysis and interpretation: James Heun, Axel Grothey, Megan Branda, Daniel Sargent, Richard M. Goldberg
Manuscript writing: James Heun, Axel Grothey, Megan Branda, Daniel Sargent
Final approval of manuscript: James Heun, Axel Grothey, Megan Branda, Daniel Sargent, Richard M. Goldberg

REFERENCES

- Allegra C, Blanke C, Buyse M et al. End points in advanced colon cancer clinical trials: a review and proposal. *J Clin Oncol* 2007;25:3572–3575.
- Fleming TR, Rothmann MD, Lu HL. Issues in using progression-free survival when evaluating oncology products. *J Clin Oncol* 2009;27:2874–2880.
- 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;22:4442–4445.
- Sargent DJ, Wieand HS, Haller DG et al. Disease-free survival versus overall survival as a primary end point for adjuvant colon cancer studies: individual patient data from 20,898 patients on 18 randomized trials. *J Clin Oncol* 2005;23:8664–8670.
- Grothey A. Surrogate endpoints for overall survival in early colorectal cancer from the clinician’s perspective. *Stat Methods Med Res* 2008;17:529–535.
- Piedbois P, Buyse M. Endpoints and surrogate endpoints in colorectal cancer: a review of recent developments. *Curr Opin Oncol* 2008;20:466–471.
- Buyse M. Use of meta-analysis for the validation of surrogate endpoints and biomarkers in cancer trials. *Cancer J* 2009;15:421–425.
- Grothey A, Wieand HS, Haller DG et al. Response-independent survival benefit in metastatic colorectal cancer: a comparative analysis of N9741 and AVF2107. *J Clin Oncol* 2008;26:183–189.
- Rubinstein LV, Korn EL, Freidlin B et al. Design issues of randomized phase II trials and a proposal for phase II screening trials. *J Clin Oncol* 2005;23:7199–7206.
- DiMasi JA, Grabowski HG. Economics of new oncology drug development. *J Clin Oncol* 2007;25:209–216.
- Roberts TG Jr., Lynch TJ Jr., Chabner BA. The phase III trial in the era of targeted therapy: unraveling the “go or no go” decision. *J Clin Oncol* 2003;21:3683–3695.
- Therasse P, Arbuck SG, Eisenhauer EA et al. New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada. *J Natl Cancer Inst* 2000;92:205–216.
- Miller AB, Hoogstraten B, Staquet M et al. Reporting results of cancer treatment. *Cancer* 1981;47:207–214.
- Gan HK, Grothey A, Pond GR et al. Randomized phase II trials: inevitable or inadvisable? *J Clin Oncol* 2010;28:2641–2647.
- Tang PA, Bentzen SM, Chen EX et al. Surrogate end points for median

- overall survival in metastatic colorectal cancer: literature-based analysis from 39 randomized controlled trials of first-line chemotherapy. *J Clin Oncol* 2007;25:4562–4568.
- 16 Buyse M, Burzykowski T, Carroll K et al. Progression-free survival is a surrogate for survival in advanced colorectal cancer. *J Clin Oncol* 2007;25:5218–5224.
 - 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;9:29–38.
 - 18 Reardon DA, Desjardins A, Vredenburgh JJ et al. Metronomic chemotherapy with daily, oral etoposide plus bevacizumab for recurrent malignant glioma: a phase II study. *Br J Cancer* 2009;101:1986–1994.
 - 19 Groves MD, Puduvalli VK, Gilbert MR et al. Two phase II trials of temozolomide with interferon-alpha2b (pegylated and non-pegylated) in patients with recurrent glioblastoma multiforme. *Br J Cancer* 2009;101:615–620.
 - 20 Aoki T, Mizutani T, Nojima K et al. Phase II study of ifosfamide, carboplatin, and etoposide in patients with a first recurrence of glioblastoma multiforme. *J Neurosurg* 2010;112:50–56.
 - 21 Galanis E, Jaeckle KA, Maurer MJ et al. Phase II trial of vorinostat in recurrent glioblastoma multiforme: a north central cancer treatment group study. *J Clin Oncol* 2009;27:2052–2058.
 - 22 Lara PN Jr., Redman MW, Kelly K et al. Disease control rate at 8 weeks predicts clinical benefit in advanced non-small-cell lung cancer: results from Southwest Oncology Group randomized trials. *J Clin Oncol* 2008;26:463–467.
 - 23 Claret L, Girard P, Hoff PM et al. Model-based prediction of phase III overall survival in colorectal cancer on the basis of phase II tumor dynamics. *J Clin Oncol* 2009;27:4103–4108.
 - 24 Ratain MJ, Eisen T, Stadler WM et al. Phase II placebo-controlled randomized discontinuation trial of sorafenib in patients with metastatic renal cell carcinoma. *J Clin Oncol* 2006;24:2505–2512.
 - 25 Karrison TG, Maitland ML, Stadler WM et al. Design of phase II cancer trials using a continuous endpoint of change in tumor size: application to a study of sorafenib and erlotinib in non small-cell lung cancer. *J Natl Cancer Inst* 2007;99:1455–1461.
 - 26 Goldberg RM, Sargent DJ, Morton RF et al. A randomized controlled trial of fluorouracil plus leucovorin, irinotecan, and oxaliplatin combinations in patients with previously untreated metastatic colorectal cancer. *J Clin Oncol* 2004;22:23–30.
 - 27 Eisenhauer EA, Therasse P, Bogaerts J et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer* 2009;45:228–247.
 - 28 Spaepen K, Stroobants S, Dupont P et al. Early restaging positron emission tomography with (18)F-fluorodeoxyglucose predicts outcome in patients with aggressive non-Hodgkin's lymphoma. *Ann Oncol* 2002;13:1356–1363.
 - 29 Haioun C, Itti E, Rahmouni A et al. [18F]fluoro-2-deoxy-D-glucose positron emission tomography (FDG-PET) in aggressive lymphoma: an early prognostic tool for predicting patient outcome. *Blood* 2005;106:1376–1381.
 - 30 Mikhaeel NG, Hutchings M, Fields PA et al. FDG-PET after two to three cycles of chemotherapy predicts progression-free and overall survival in high-grade non-Hodgkin lymphoma. *Ann Oncol* 2005;16:1514–1523.
 - 31 Piessevaux H, Buyse M, De Roock W et al. Radiological tumor size decrease at week 6 is a potent predictor of outcome in chemorefractory metastatic colorectal cancer treated with cetuximab (BOND trial). *Ann Oncol* 2009;20:1375–1382.
 - 32 Bystrom P, Berglund A, Garske U et al. Early prediction of response to first-line chemotherapy by sequential [18F]-2-fluoro-2-deoxy-D-glucose positron emission tomography in patients with advanced colorectal cancer. *Ann Oncol* 2009;20:1057–1061.
 - 33 Janssen MH, Ollers MC, Riedl RG et al. Accurate prediction of pathological rectal tumor response after two weeks of preoperative radiochemotherapy using (18)F-fluorodeoxyglucose-positron emission tomography-computed tomography imaging. *Int J Radiat Oncol Biol Phys* 2010;77:392–399.
 - 34 Piessevaux H, Bokemeyer C, Schlichting M et al. Impact of early tumor shrinkage on long-term outcome in metastatic colorectal cancer (mCRC) treated with FOLFOX4 with or without cetuximab: lessons from the OPUS trial. *J Clin Oncol* 2011;29(suppl 4; abstr 398).