Alternate Endpoints for Screening Phase II Studies

Neesha Dhani,[1] Dongsheng Tu,[2] Daniel J. Sargent,[3] Lesley Seymour,[2] and Malcolm J. Moore[1]

Abstract Phase II trials are screening trials that seek to identify agents with sufficient activity to continue development and those for which further evaluation should be halted. Although definitive phase III trials use progression-free or overall survival to confirm clinical benefit, earlier endpoints are preferable for phase II trials. Traditionally, tumor shrinkage of a predetermined degree (response) has been used as a surrogate of eventual survival benefit based on the observation that high response rates (RR), and particularly complete responses, in the phase II setting resulted in survival benefit in subsequent phase III trials. Recently, some molecularly targeted agents have shown survival and clinical benefit despite very modest RRs in early clinical trials. These observations provide a major conundrum, with concerns of inappropriate termination of development for active agents with low RRs being balanced by concerns of inactive agents being taken to late-phase development with resultant increases in the failure rate of phase III trials. Numerous alternate or complementary endpoints have been explored, incorporating multinomial endpoints (including progression and response), progression-free survival, biomarkers, and, more recently, evaluation of tumor size as a continuous variable. In this review, we discuss the current status of phase II endpoints and present retrospective analyses of two international gastrointestinal cancer studies showing the potential utility of one novel approach. Alternate endpoints, although promising, require additional evaluation and prospective validation before their use as a primary endpoint for phase II trials.

As our knowledge of cancer biology expands, more therapeutic targets are identified, resulting in increases in the number of targeted agents entering development. Limited patient and financial resources and several high-profile failures in phase III trials (1–3) have resulted in investigators examining more efficient and effective early-phase trial designs.

Although it is important to pick the "winners" in phase II, it is equally important to identify ineffective agents not warranting further development to direct resources to more promising agents. Although response rate (RR) is commonly the primary endpoint in phase II trials, its limitations are well described, including but not limited to a lack of concordance between RR in single-center phase II trials and subsequent multicenter phase III studies (4). Further, there have been descriptions of standard chemotherapeutics (such as the combination of oxaliplatin with 5-fluorouracil and leucovorin in colorectal cancer) showing a survival advantage in nonresponding patients (5).

Recently, examples have been reported of novel agents with very modest RR resulting in prolongation of progression-free survival (PFS) or overall survival (OS) in phase III studies (6), suggesting that tumor stabilization rather than shrinkage may still result in clinical benefit. Using RR as a "go/no go" criterion in such circumstances may lead to inappropriate termination of development. A striking example is the putative Raf kinase and antivascular agent sorafenib in renal cell carcinoma and hepatocellular cancer, where RRs using Response Evaluation Criteria in Solid Tumors (RECIST) were <10%, conventionally a signal to abandon further development (7, 8). Subsequent randomized trials showed significant prolongation of both PFS and OS (6, 9). Tumor measurements represented as waterfall plots showed that many patients had either reduction in tumor size insufficient to classify as a response or lack of tumor growth [stable disease (SD)]. Other examples include previously treated non–small cell lung cancer treatment with single-agent erlotinib in which significant OS benefit was shown despite low RR (10). Thus, the simple categorization of patients into responders and nonresponders based on tumor shrinkage may fail to use all available information that could be gleaned from response evaluations.

In this review, we will consider current phase II endpoints and some more recently proposed novel endpoints and designs (11–14). We will then briefly describe retrospective analyses of previously analyzed and published phase III studies in gastrointestinal cancers to assess other approaches using tumor measurements.

Phase II Endpoints

Phase II trials are screening studies with a primary goal of identifying a signal of antitumor activity in a well-defined and relatively homogenous population of patients with a single

Authors' Affiliations: [1]Princess Margaret Hospital, Toronto, Ontario, Canada; [2]National Cancer Institute of Canada Clinical Trials Group, Kingston, Ontario, Canada; and [3]Mayo Clinic, Rochester, Minnesota

Received 10/7/08; revised 1/16/09; accepted 1/16/09; published OnlineFirst 3/10/09.

Grant support: Canadian Cancer Society.

Requests for reprints: Lesley Seymour, National Cancer Institute of Canada Clinical Trials Group, Queens University, 10 Stuart Street, Kingston, Ontario, Canada. Phone: 613-533-6430; Fax: 613-533-2411; E-mail: Lseymour@ctg.queensu.ca.

doi:10.1158/1078-0432.CCR-08-2034

tumor type. Rapid screening phase II trials are essential for efficient and cost-effective development of a new therapeutic (15, 16). Adjei et al. provide an overview of the current status of phase II studies in oncology (17).

The gold standard for evaluation of any new cancer therapy is improvement in OS. OS has limited utility in phase II trials due to the length of observation required and the confounding effects of crossover and effective second-line therapies, and other endpoints are preferable. Table 1 outlines some of the advantages and disadvantages of the efficacy endpoints described below.

Response/tumor shrinkage endpoints. Tumor shrinkage has been the most commonly used efficacy endpoint in phase II trials. Traditionally, changes in tumor size have been categorized into three defined criteria response: complete response (CR) or partial response (PR), SD, or progressive disease (PD) based on standard criteria of tumor measurement (Table 2) such as those defined by the WHO and later RECIST (18–21). The basic assumption for the use of RR has been that a higher rate of response, compared with historical or concurrent controls, is predictive for improvements in survival and that an agent would not benefit patients without resulting in significant tumor shrinkage (22–25). Although these assumptions generally appear to hold true for cytotoxic agents, there are several potential and apparent limitations with the use of RR as an endpoint for trials of molecularly targeted agents, where tumor shrinkage may be modest and may not meet the empiric criteria defined in studies of cytotoxic agents (26). Even in the case of conventional cytotoxic agents that do result in moderate RRs, it is not uncommon for higher RRs to fail to translate into a survival benefit (27–29), whereas agents such as gemcitabine have shown survival benefit without increases in RR (30). In addition, tumor shrinkage may occur late and the requirement for confirmation of any observed response may result in loss of the RR in the phase II trial setting. Also, serial response evaluations are costly, inconvenient, and subject to variability in assessment.

Initial efforts to overcome some of these limitations included the design of studies using RR with coprimary endpoints, such as PFS or absence of progression (multinomial endpoints; ref. 12), or in novel designs such as randomized discontinuation designs (7). Alternate trial designs have been extensively reviewed elsewhere, including by Rubinstein et al. (31), and will not be considered further in this review.

A major limitation of characterizing tumor response categorically by RECIST is that change in tumor dimensions in a cohort of patients on a clinical trial is actually a continuous variable. Important information is potentially ignored when this continuum is categorized into only two or three groups (32). The use of waterfall or spider plots showing individual changes in tumor size for all patients in a study are becoming more common (Fig. 1) and have been instrumental in graphically showing the benefit of some treatments such as sorafenib in renal cell carcinoma and the

Table 1. Advantages and disadvantages of commonly used endpoints in efficacy evaluation of novel molecular anticancer agents

Endpoint	Advantages	Disadvantages
Tumor RR	Standardized, easily applicable to multicenter trials Early outcome	Measurement imprecision Difficult in some tumor types (mesothelioma and peritoneal disease) Correlation with patient benefit variable
TTP, PFS	Unlike OS, not confounded by salvage therapy	Subject to assessment and investigator bias Requires control cohort Only partially validated as a surrogate of survival benefit
OS	Clinically relevant outcome	Requires control cohort Affected by crossover designs and subsequent therapies Longer follow-up time required
Quality of life	Indicative of direct patient benefit	Multiple comparisons may lead to positive results by chance Time-intensive evaluation Analyses complex
Molecular biomarkers	May prove to be predictive and allow patient enrichment May provide additional insight into resistance mechanisms	Usually not validated as a surrogate of efficacy during early clinical development of an agent
Imaging	May allow early assessment of antitumor effect	May add little to response assessment Costly and time-consuming Difficult to combine results with multi-institutional trials

Table 2. Comparison of the RECIST and the WHO criteria

	RECIST	WHO
Target lesions	Measurable lesions to a maximum of 5 (2 per organ)	All measurable lesions
Type of measurement	Unidimensional	Bidimensional
Tumor burden assessment	Sum of greatest diameter of target lesions	Sum of products of maximum perpendicular diameters
Response		
CR	Resolution of all disease	Resolution of all disease
PR	≥30% decrease in sum	>50% decrease
SD	Neither PR nor PD met	Neither PR nor PD met
PD	20% increase AND > 5 mm absolute increase OR new lesion	25% increase OR new lesion

K-ras mutation/epidermal growth factor receptor interaction in colorectal cancer (33, 34). The information can be analyzed quantitatively by summarizing the mean (SD) and compared between groups using a *t* or Wilcoxon test. Other methods include transforming the tumor size variable to yield a log-normal distribution (32, 35). First suggested >20 years ago, this idea was recently represented again by Karrison et al. using data from four individual studies to illustrate feasibility for more contemporary trial designs (35). Although analyses using tumor size as a continuous variable are attractive as they minimize information loss, there are some issues, such as accounting for patients who progress due to new lesions and how to deal with missing data (patients who were not reassessed on time).

PFS endpoints. Time to progression (TTP) and PFS are often used interchangeably, but in the formal definition of TTP, death is excluded as an event. As PFS includes death as an event of interest, it may correlate better with OS outcomes by capturing a possible negative effect of therapy on survival due to adverse events and is therefore preferable as a regulatory endpoint (36). PFS and TTP endpoints avoid the confounding crossover effect of subsequent therapy that influences OS and have the additional benefit of requiring a shorter duration of follow-up. Both TTP and PFS are influenced by the frequency of evaluation, potential assessment bias, and investigator bias if the trial is not blinded, particularly when a promising new compound is being evaluated in a crossover design. The main limitation of PFS is in terms of defining its clinical relevance (15) and the requirement for a randomized design (unless a stable historical control is available) with relatively larger sample size (26, 37) as well as the need for frequent disease assessments, although the latter can be overcome by considering a PFS rate at a prespecified time point (e.g., 6 months) as the primary endpoint.

The growth modulation index is a variant of TTP endpoints and is a ratio of the TTP on current (study) treatment to the TTP on the most recent prior therapy. In this way, patients act as their own controls, and a study treatment is deemed to be potentially active if the growth modulation index ratio exceeds 1.33, signifying that TTP on the therapy under evaluation is at least 33% longer than that on prior treatment regimen (38).

Other proposed endpoints. Patient reported outcomes, such as quality of life, evaluate the risks and benefits of therapy from the patient's perspective. Although initially considered subjective, efforts to develop valid and reliable instruments have resulted in recommendations on standard tools and methods for patient reported outcomes in clinical trials. Patient reported outcomes have played an important role in the approval of mitoxantrone for hormone-resistant prostate cancer and gemcitabine in pancreatic cancer (30, 39). They remain, however, somewhat challenging endpoints to evaluate and interpret and require time-intensive tools administered by experienced individuals (40–43). Patient reported outcomes are rarely, if ever, used as primary endpoints in phase II trials but may have a role as a secondary endpoint.

Normal or tumor tissue can be used as a predictive or prognostic biomarker. Biomarkers may include expression of a cell surface marker such as epidermal growth factor receptor, presence of mutations, activity of a signaling pathway, or measurements of target inhibition in tumor or in more accessible surrogate tissues. More global assessments of hallmark processes of cancer such as apoptosis, angiogenesis, or proliferation index can also be evaluated. Predictive biomarkers may permit patient enrichment and a greater therapeutic and economic efficiency, such as with the use of HER-2/*neu* expression predicting for benefit from trastuzumab and K-ras mutational status predicting for benefit from epidermal growth factor receptor inhibitors. Unfortunately, biomarkers have rarely been sufficiently robust or validated to be useful as surrogates of efficacy in early clinical development. Currently, the role of biomarkers in early-phase trials is mainly exploratory, with definitive data being generated from the phase III evaluation (44–48), although they are seldom used as endpoints. Measures based on circulating tumor cells may hold particular promise in this area (49). This topic is explored in further detail in the article by McShane et al. (50).

Imaging modalities may allow for noninvasive assessments of response or target effects. Changes in these endpoints may be detected earlier than RECIST and some tumors seem particularly relevant for the use of primary imaging endpoints (51, 52). Multiple imaging modalities, such as fluorodeoxyglucose-positron emission tomography, dynamic contrast-enhanced magnetic resonance imaging, and volumetric imaging, have shown promise in single institution series and multicenter evaluation is now under way for the fluorodeoxyglucose-positron emission tomography in lymphoma and lung cancer. The current status of imaging tools in phase II studies is reviewed in the article by Shankar et al. (53).

Which endpoints are commonly used? El-Maraghi et al. reviewed 89 single-agent phase II trials of 19 targeted agents in 6 solid tumor sites (54). The majority used a nonrandomized single-arm design with RR as the primary or coprimary endpoint <20% used PFS as the primary endpoint and very few included a multinomial endpoint. The mean sample size for studies using a RR, PFS, or a multinomial endpoint was 56, 115, and 41 patients, respectively. Twenty percent of the studies attempted enrichment using a biomarker thought to predict for benefit. Successful drugs (Food and Drug Administration approval) had higher RR in phase II. Some agents with RR of <10%, however, still went on to regulatory approval after phase III. This analysis concluded that standard response thresholds used for cytotoxic drugs (e.g., RR > 20%)

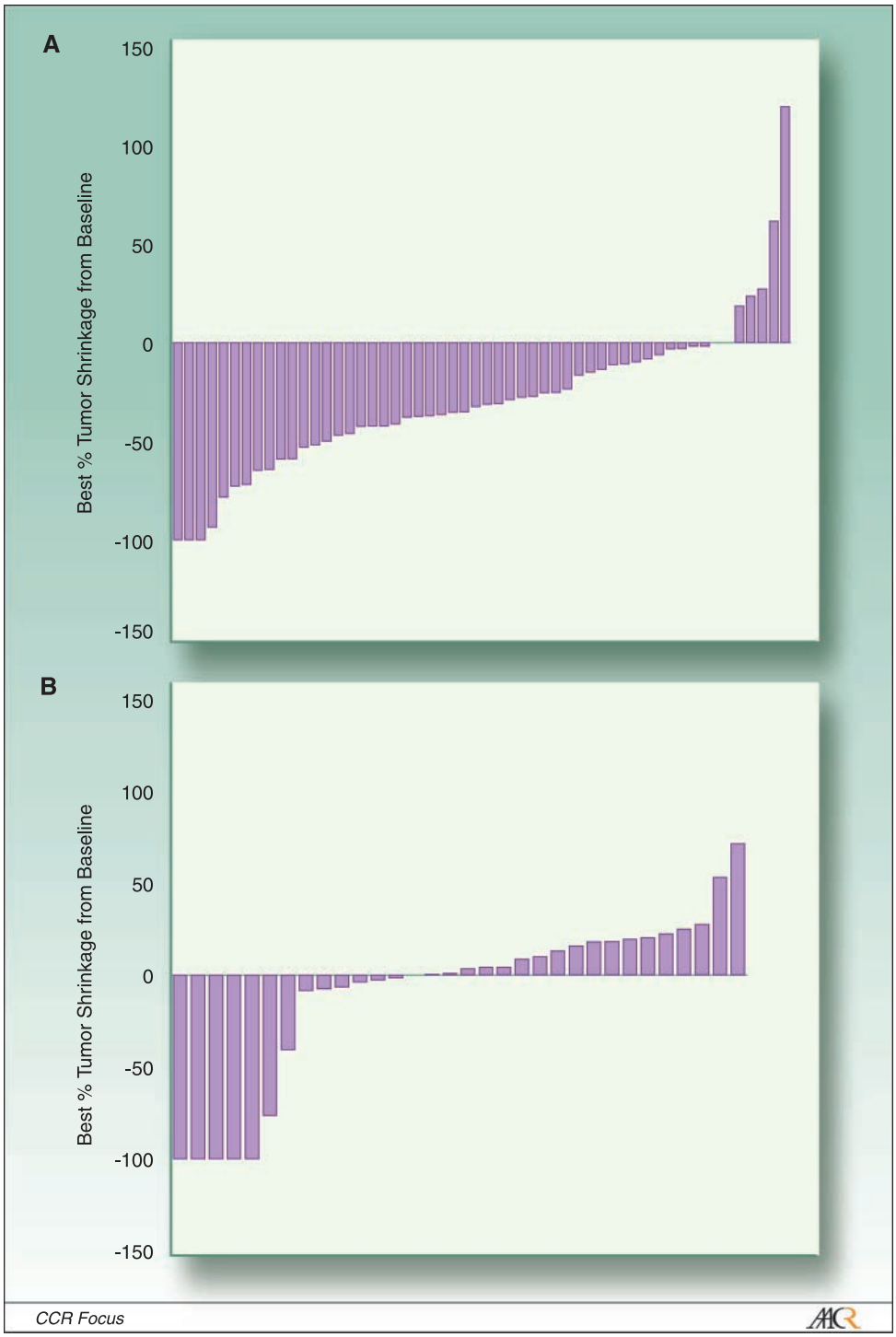

Fig. 1. Simulated examples of waterfall and spider plots. *A* and *B,* waterfall plots depicting the best percentage reduction in tumor size as evidenced by the sum of diameters of target lesions. The agent in *A* appears to have more antitumor activity than the agent in *B*. Reprinted and adapted from Eur J Cancer 2008;44:25–9. Booth CM, et al. Design and conduct of phase II studies of targeted anticancer therapy: recommendations from the task force on methodology for the development of innovative cancer therapies (MDICT). ©2008 with permission from Elsevier.

may result in inappropriate rejection of agents that actually have meaningful clinical benefit. If all agents with low RR were to proceed to phase III testing, however, the number of negative phase III trials would likely further increase.

Another review analyzed phase III studies of targeted therapies (1985-2005) to identify potential predictive factors in the preceding phase II trials (55). Less than half of positive phase II trials resulted in positive findings in phase II.

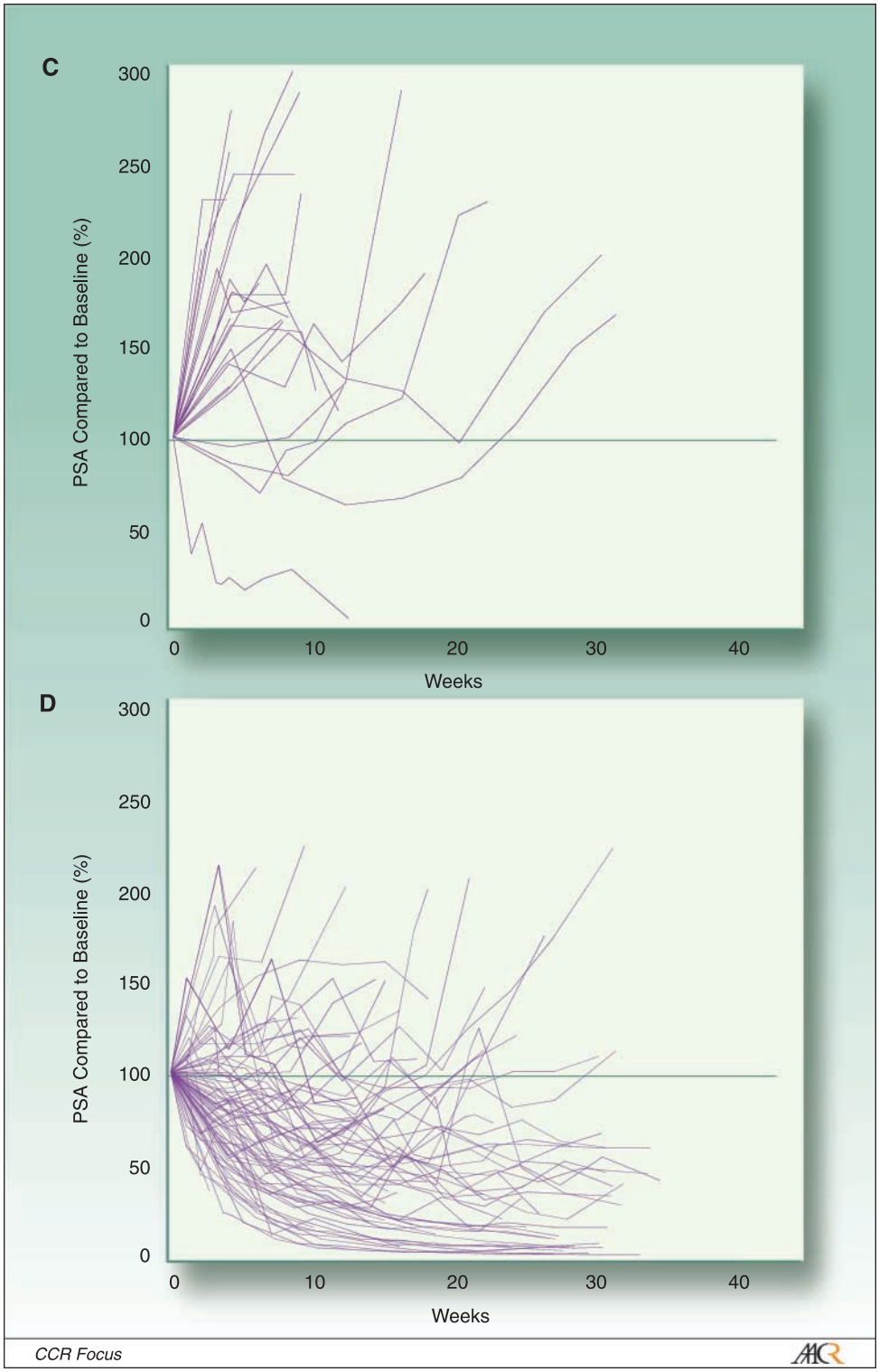

Fig. 1 *Continued. C* and *D,* examples of spider plots, showing changes in prostate-specific antigen levels as a percentage compared with baseline (*horizontal line*). The agent in *C* appears to be less active compared with the agent in *D*.

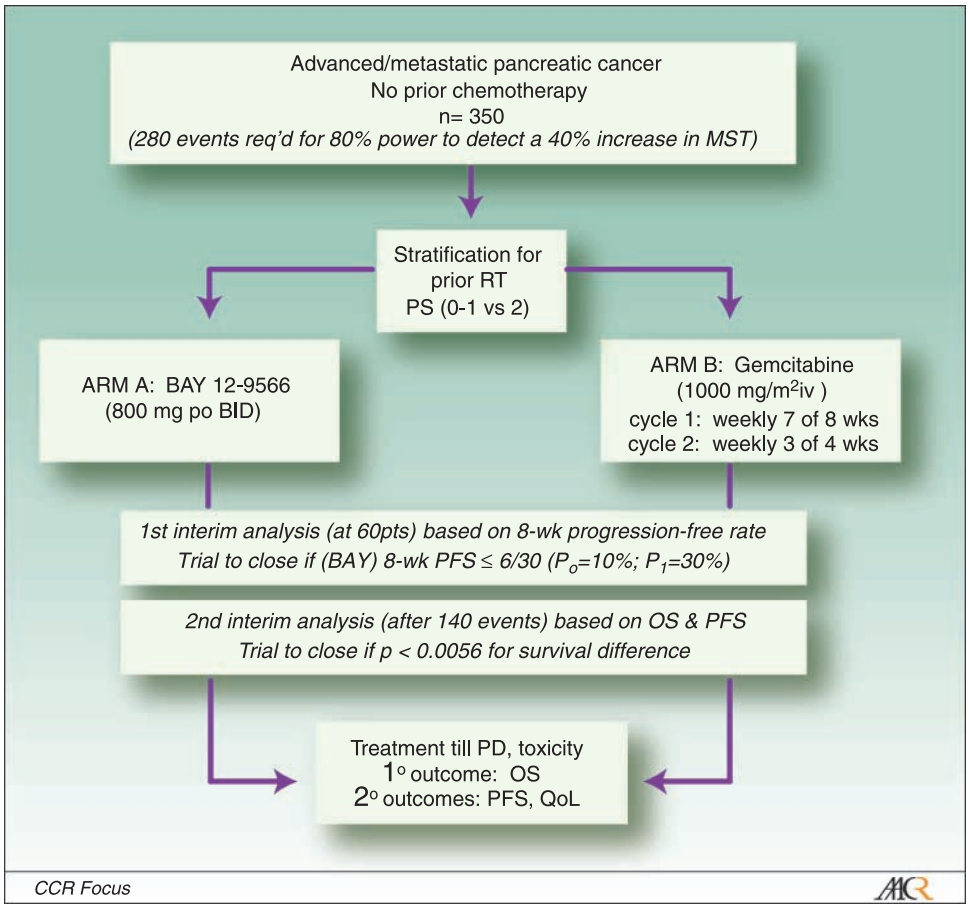

Fig. 2. Design and statistical analysis of NCIC CTG PA.1 comparing gemcitabine with the matrix metalloproteinase BAY 12-9566 for chemotherapy-naive advanced pancreatic cancer. *MST,* median survival time; *BID,* twice daily; *QoL,* quality of life; *PS,* performance status; *RT,* radiation therapy.

Predictive phase II trials were multicentered and industry sponsored and had a shorter time interval between phase II and III publications. There was a nonsignificant trend for phase II trials with PFS or TTP as endpoints to predict for positive phase III trials compared with phase II RR trials. Higher phase II RR did not predict for a positive phase III. More recent phase II trials were more likely to predict for a positive phase III study, suggesting improvement in design or decision-making.

Exploratory Efficacy Analyses

Overall, the data described previously support the validation and implementation of efficacy endpoints other than RR for the evaluation of targeted therapies. We retrospectively tested an early assessment of tumor size as a continuous variable as a surrogate for eventual survival improvement in two previously analyzed and published randomized clinical trials in gastrointestinal malignancies conducted by the National Cancer Institute of Canada Clinical Trials Group (NCIC CTG) and the North Central Clinical Treatment Group. In the North Central Clinical Treatment Group trial, response as a two-category versus three-category variable was also evaluated.

NCIC CTG PA.1. NCIC CTG PA.1 (Fig. 2) compared BAY 12-9566 to gemcitabine (at the time not approved nor standard of care) in chemotherapy-naive patients with advanced

Table 3. Results of NCIC CTG PA.1

	BAY 12-9566	Gemcitabine	*P*
PFS (mo)	1.68	3.5	< 0.001
Hazard ratio (95% confidence interval)	0.530 (0.412-0.681)		
OS (mo)	3.74	6.59	< 0.001
Hazard ratio (95% confidence interval)	0.574 (0.445-0.740)		
Response	***n* = 108, *n* (%)**	***n* = 115, *n* (%)**	
PR	1 (1)	6 (5)	
SD	31 (29)	62 (54)	

Table 4. Tumor measurement as a continuous variable: differences in the logarithm of tumor size at baseline and 8-wk assessment in all patients of PA.1

	BAY 12-9566		Gemcitabine		*P* value
	n	Mean (SD)	*n*	Mean (SD)	
Logarithm of tumor size at baseline	75	1.71 (0.61)	98	1.87 (0.60)	
Logarithm of tumor size at 8 wk	75	1.80 (0.82)	98	1.81 (0.77)	
Difference in logarithm of tumor size	75	0.087 (0.44)	98	-0.066 (0.47)	<0.0001

pancreatic cancer. Disease was assessed at baseline and at 8-week intervals using standard WHO-based criteria. Crossover to gemcitabine was permitted at PD. A first interim analysis was conducted after 60 patients were enrolled. The Data Safety Monitoring Committee recommended continuation of the trial, as the protocol defined requirement for ≤6 patients (of 30 patients on the experimental arm) not to have had PD at 8 weeks was met. The second interim analysis took place after 277 patients were accrued and 140 deaths had occurred. The study was terminated at that point based on the inferior survival in BAY 12-9566 patients compared with those on gemcitabine (3.74 versus 6.59 months; $P < 0.001$; Table 3; ref. 2).

In this retrospective, exploratory analysis of patients with measurable disease enrolled in PA.1, the logarithm of the sum of the longest diameters of target lesions was calculated at baseline and at 8 weeks as well as the difference in logarithms to give an indication of change in tumor size over the 8 weeks. The Wilcoxon rank-sum test was then used to compare the difference between the two treatment groups (Table 4). Our analysis showed a statistically significant difference in tumor size in patients treated with gemcitabine (log difference of -0.066) in comparison with those treated with BAY 12-9566 (log difference of 0.087) with $P < 0.0001$, consistent with the final OS results. Table 5 illustrates the results of differences in logarithms of tumor size at baseline and at 8 weeks post-treatment of the patients included in the first interim analysis. Again, there was a significant difference in tumor response in patients on gemcitabine (log difference of -0.031) compared with those on BAY 12-9566 (log difference of 0.077) with $P = 0.007$. Thus, considering actual tumor measurements as a continuous variable was a more sensitive early predictor of the lack of clinical efficacy of BAY 12-9566 than PD at 8 weeks.

Intergroup trial N9741. Intergroup N9741 (Fig. 3) was a three-arm randomized trial comparing the then standard regimen of irinotecan with 5-fluorouracil and leucovorin to the experimental regimens of oxaliplatin with 5-fluorouracil and leucovorin and oxaliplatin with irinotecan as first-line treatment for metastatic colorectal cancer (56). Tumor response was assessed by WHO criteria and categorized as CR, PR, SD, or PD.

In this retrospective analysis, we explored the additional prognostic value of actual tumor measurements versus WHO criteria when assessed at 12 weeks and over the course of therapy. The sum of tumor measurements was calculated at each evaluation and log-transformed for analysis. Thus, 1,164 patients (irinotecan with 5-fluorouracil and leucovorin 325, oxaliplatin with 5-fluorouracil and leucovorin 546, and oxaliplatin with irinotecan 293) had measurable disease and are included in this analysis. The percentage change in tumor measurements from baseline to 12 weeks was calculated. Patients with PD before 12 weeks were assigned a 100% increase in tumor measurement; results were not sensitive to the amount of increase such patients were assigned (20%, 50%, or 100%). The prognostic value of absolute and percentage change in tumor size versus WHO tumor status (CR/PR, SD, and PD) both at 12 weeks and over the entire course were compared, using Cox models for OS in a landmark analysis, adjusting for baseline tumor size and treatment arm. We also compared the ability of the traditional two-category WHO disease response status (responder versus nonresponder) at 12 weeks to predict OS with a three-category WHO response status (CR/PR, SD, and PD) at 12 weeks.

The three-category variable, which considers SD separately, provided significantly higher association with survival in the landmark analysis compared with the two-category WHO status ($P < 0.0001$). In addition, in a univariate analysis, actual 12-week tumor measurements were strongly associated with OS in the landmark analysis ($P < 0.0001$). In a joint model however, actual tumor measurements provided very modest and nonclinically meaningful additional prognostic value after accounting for three-category WHO disease status at 12 weeks

Table 5. Tumor measurement as a continuous variable: differences in the logarithm of tumor size at baseline and 8-wk assessment in patients included in first interim analysis of NCIC CTG PA.1

	BAY 12-9566		Gemcitabine		*P* value
	n	Mean (SD)	*n*	Mean (SD)	
Logarithm of tumor size at baseline	39	1.70 (0.71)	48	1.93 (0.64)	
Logarithm of tumor size at 8 wk	39	1.77 (1.02)	48	1.90 (0.67)	
Difference in logarithm of tumor size	39	0.077 (0.59)	48	-0.031 (0.20)	0.007

(P = 0.007), with an increase in the concordance index (a measure of model fit) of only 1%, from 0.58 to 0.59, after adding the continuous measurement to the model with a three-category disease status included. This additional benefit in prognostic value due to actual tumor measurements was limited to patients with WHO disease status of CR/PR (P = 0.005); actual tumor measurements provided no improved prognostic value in patients who were WHO SD status (P = 0.35) or PD (P = 0.57) at 12 weeks. Overall WHO best confirmed response status (over the entire course of treatment) had no improved prognostic ability compared with WHO tumor status at 12 weeks in predicting OS. Similar exploration was done using 6-week tumor size and status, with poor results. We concluded that there was clear additional value in considering a three-category variable for disease status (CR/PR, SD, and PD) at a fixed, early time point (12 weeks) versus a two-category consideration (responder versus nonresponder) but that the use of actual tumor measurements provided little additional value beyond that provided by the three-level tumor status indicator.

Conclusions and Future Directions

Traditional RR-based endpoints have done reasonably well historically, but simpler and more robust and efficient phase II endpoints are needed, particularly when evaluating novel therapeutics. PFS, although a validated surrogate of OS in many tumors, generally results in larger phase II trials. Some investigators have attempted to improve the efficiency of RR by including coprimary endpoints (multinomial endpoints) or by exploring novel uses of tumor measurement data.

In our exploratory analyses in pancreatic cancer, the use of early tumor size as a continuous variable appeared superior to 8-week PFS variable and would have led to the appropriate cessation of this trial at the first interim analysis.

In our colon cancer example, although the continuous change in tumor size was also found to be predictive of a survival endpoint, interestingly, the use of a three-category assessment of disease status at 12 weeks provided similar predictive ability. The use of a three-category tumor status variable adds only minimal additional complexity to the standard of two categories (responder versus nonresponder). Of note, this trial compared chemotherapy regimens and did not include novel agents. It is plausible that the continuous variable analyses are better suited to evaluate agents for which a significant amount of tumor shrinkage is not expected. Nonetheless, the data for colon cancer suggest that the continuous tumor variable in this chemotherapy trial added predictive ability compared with traditional two-category response evaluation. Recording true continuous tumor measurements results in substantial additional workload. Multiple assumptions are required to deal with noisy data, missing data, and presence of new lesions. Usually, patients who have new lesions or definitive progression in one or more lesions

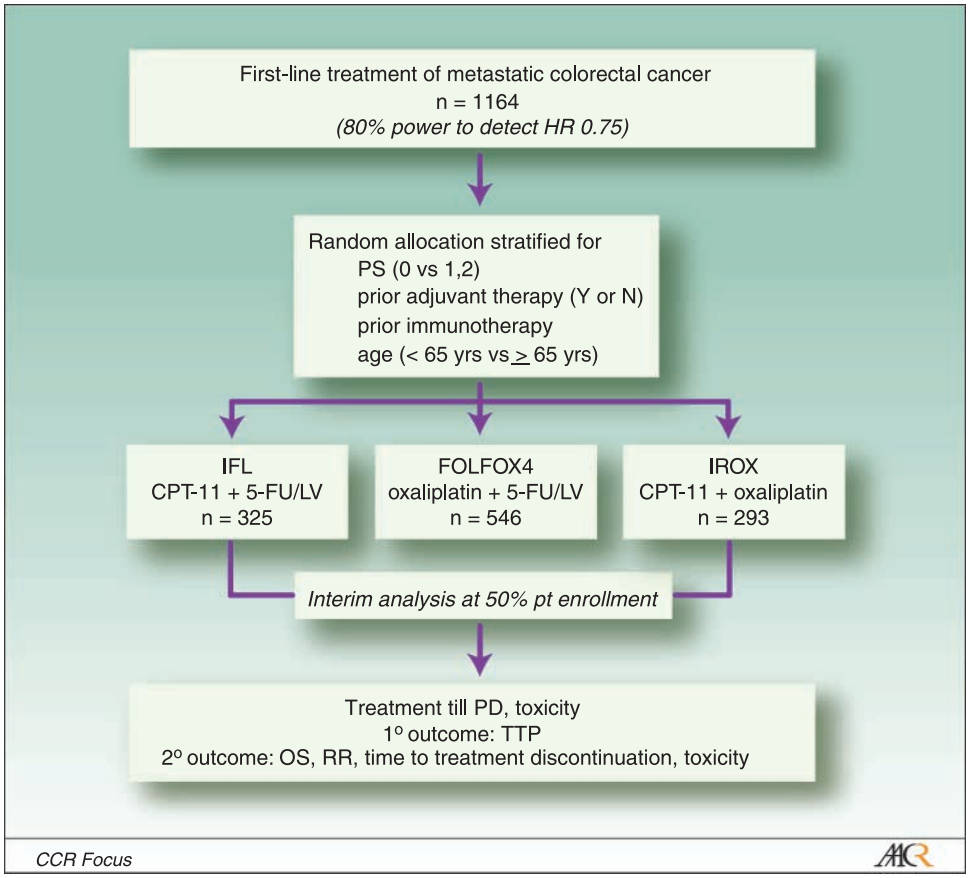

Fig. 3. Design and analysis of North Central Clinical Treatment Group N9741 comparing oxaliplatin with 5-fluorouracil and leucovorin (*FOLFOX*) 4 and oxaliplatin with irinotecan (*IROX*) to irinotecan with 5-fluorouracil and leucovorin (*IFL* ; then standard of care) for treatment-naive colorectal cancer.

are accepted as having PD, and measurements for all identified lesions may not be provided. Similarly, account must be taken of new lesions, which may define PD even when shrinkage in measurable disease has occurred (mixed response). Assessment bias may occur and patients do still need to undergo response evaluation for clinical management purposes. Clearly, such an approach is only justified if substantial benefit compared with standard methodology can be shown.

Our results, although intriguing, are retrospective and suggest that, at least in the case of chemotherapy trials, three-category response evaluation at a specific point in time may perform as well as continuous tumor measurements. Prospective testing, or individual patient data-based meta-analyses of existing randomized trials, including trials of novel therapeutics, is necessary before either endpoint could be considered validated. Continued research retrospectively validating and then testing phase II endpoints is critical to improve both robustness of the endpoint and trial efficiency. Neither of the two analyses reviewed in this article have to date been validated in other retrospective studies and as such should not yet be used as primary endpoints for phase II studies. Further testing of these methodologies on existing databases is planned.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

References

1. Leighl NB, Paz-Ares L, Douillard JY, et al. Randomized phase III study of matrix metalloproteinase inhibitor BMS-275291 in combination with paclitaxel and carboplatin in advanced non-small-cell lung cancer: National Cancer Institute of Canada-Clinical Trials Group Study BR.18. J Clin Oncol 2005;23:2831–9.
2. Moore MJ, Hamm J, Dancey J, Eisenberg PD, et al. Comparison of gemcitabine versus the matrix metalloproteinase inhibitor BAY 12-9566 in patients with advanced or metastatic adenocarcinoma of the pancreas: a phase III trial of the National Cancer Institute of Canada Clinical Trials Group. J Clin Oncol 2003;21:3296–302.
3. Shepherd FA, Giaccone G, Seymour L, et al. Prospective, randomized, double-blind, placebo-controlled trial of marimastat after response to first-line chemotherapy in patients with small-cell lung cancer: a trial of the National Cancer Institute of Canada-Clinical Trials Group and the European Organization for Research and Treatment of Cancer. J Clin Oncol 2002;20:4434–9.
4. Zia MI, Siu LL, Pond GR, Chen EX. Comparison of outcomes of phase II studies and subsequent randomized control studies using identical chemotherapeutic regimens. J Clin Oncol 2005;23:6982–91.
5. Grothey A, Hedrick EE, Mass RD, et al. Response-independent survival benefit in metastatic colorectal cancer: a comparative analysis of N9741 and AVF2107. J Clin Oncol 2008;26:183–9.
6. Escudier B, Eisen T, Stadler WM, et al. Sorafenib in advanced clear-cell renal-cell carcinoma. N Engl J Med 2007;356:125–34.
7. 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;24:2505–12.
8. Abou-Alfa GK, Schwartz L, Ricci S, et al. Phase II study of sorafenib in patients with advanced hepatocellular carcinoma. J Clin Oncol 2006;24:4293–300.
9. Llovet JM, Ricci S, Mazzaferro V, et al. Sorafenib in advanced hepatocellular carcinoma. N Engl J Med 2008;359:378–90.
10. Shepherd FA, Rodrigues Pereira J, Ciuleanu T, et al. Erlotinib in previously treated non-small-cell lung cancer. N Engl J Med 2005;353:123–32.
11. Jin H. Alternative designs of phase II trials considering response and toxicity. Contemp Clin Trials 2007; 28:525–31.
12. Dent S, Zee B, Dancey J, Hanauske A, Wanders J, Eisenhauer E. Application of a new multinomial phase II stopping rule using response and early progression. J Clin Oncol 2001;19:785–91.
13. Rosner GL, Stadler W, Ratain MJ. Randomized discontinuation design: application to cytostatic antineoplastic agents. J Clin Oncol 2002;20:4478–84.
14. Stone A, Wheeler C, Barge A. Improving the design of phase II trials of cytostatic anticancer agents. Contemp Clin Trials 2007;28:138–45.
15. Schilsky RL. End points in cancer clinical trials and the drug approval process. Clin Cancer Res 2002;8: 935–8.
16. Parulekar WR, Eisenhauer EA. Novel endpoints and design of early clinical trials. Ann Oncol 2002;13 Suppl 4:139–43.
17. Adjei AA, Christian M, Ivy P. Novel designs and endpoints for phase II clinical trials. Clin Cancer Res 2009;15:1866–72.
18. 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;92:205–16.
19. Miller AB, Hoogstraten B, Staquet M, Winkler A. Reporting results of cancer treatment. Cancer 1981; 47:207–14.
20. Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumors: revised RECIST guideline (version 1.1). Eur J Cancer 2008;45: 228–47.
21. WHO. WHO handbook for reporting results of cancer treatment. Offset publication no. 48. Geneva: WHO Press; 1979.
22. Paesmans M, Sculier JP, Libert P, 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.
23. Torri V, Simon R, Russek-Cohen E, Midthune D, Friedman M. Statistical model to determine the relationship of response and survival in patients with advanced ovarian cancer treated with chemotherapy. J Natl Cancer Inst 1992;84:407–14.
24. Torri V, Simon R, Russek-Cohen E, Midthune D, Friedman M. Relationship between response and survival in patients with advanced ovarian cancer. J Natl Cancer Inst 1992;84:899–900.
25. Graf W, Pahlman L, Bergstrom R, Glimelius B. The relationship between an objective response to chemotherapy and survival in advanced colorectal cancer. Br J Cancer 1994;70:559–63.
26. Michaelis LC, Ratain MJ. Measuring response in a post-RECIST world: from black and white to shades of grey. Nat Rev Cancer 2006;6:409–14.
27. Burzykowski T, Buyse M, Piccart-Gebhart MJ, et al. Evaluation of tumor response, disease control, progression-free survival, and time to progression as potential surrogate end points in metastatic breast cancer. J Clin Oncol 2008;26:1987–92.
28. 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.
29. Goffin J, Baral S, Tu D, Nomikos D, Seymour L. Objective responses in patients with malignant melanoma or renal cell cancer in early clinical studies do not predict regulatory approval. Clin Cancer Res 2005;11: 5928–34.
30. Burris HA III, Moore MJ, Andersen J, et al. Improvements in survival and clinical benefit with gemcitabine as first-line therapy for patients with advanced pancreas cancer: a randomized trial. J Clin Oncol 1997; 15:2403–13.
31. Rubinstein L, Crowley J, Ivy P, LeBlanc M, Sargent D. Randomized phase II designs. Clin Cancer Res 2009;15:1883–90.
32. Lavin PT. An alternative model for the evaluation of antitumor activity. Cancer Clin Trials 1981;4:451–7.
33. Riechelmann RP, Chin S, Wang L, et al. Sorafenib for metastatic renal cancer: the Princess Margaret experience. Am J Clin Oncol 2008;31:182–7.
34. Amado RG, Wolf M, Peeters M, et al. Wild-type KRAS is required for panitumumab efficacy in patients with metastatic colorectal cancer. J Clin Oncol 2008; 26:1626–34.
35. Karrison TG, Maitland ML, Stadler WM, Ratain MJ. 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;99:1455–61.
36. Pazdur R. Endpoints for assessing drug activity in clinical trials. Oncologist 2008;13 Suppl 2:19–21.
37. Rubinstein LV, Korn EL, Freidlin B, Hunsberger S, Ivy SP, Smith MA. Design issues of randomized phase II trials and a proposal for phase II screening trials. J Clin Oncol 2005;23:7199–206.
38. Mick R, Crowley JJ, Carroll RJ. Phase II clinical trial design for noncytotoxic anticancer agents for which time to disease progression is the primary endpoint. Control Clin Trials 2000;21:343–59.
39. Tannock IF, Osoba D, Stockler MR, et al. Chemotherapy with mitoxantrone plus prednisone or prednisone alone for symptomatic hormone-resistant prostate cancer: a Canadian randomized trial with palliative end points. J Clin Oncol 1996;14: 1756–64.

40. Bruner DW, Bryan CJ, Aaronson N, et al. Issues and challenges with integrating patient-reported outcomes in clinical trials supported by the National Cancer Institute-sponsored clinical trials networks. J Clin Oncol 2007;25:5051–7.

41. Moinpour CM, Feigl P, Metch B, Hayden KA, Meyskens FL, Jr., Crowley J. Quality of life end points in cancer clinical trials: review and recommendations. J Natl Cancer Inst 1989;81:485–95.

42. Roila F, Cortesi E. Quality of life as a primary end point in oncology. Ann Oncol 2001;12 Suppl 3:S3–6.

43. Nayfield SG, Ganz PA, Moinpour CM, Cella DF, Hailey BJ. Report from a National Cancer Institute (USA) workshop on quality of life assessment in cancer clinical trials. Qual Life Res 1992;1:203–10.

44. Greenberg AK, Lee MS. Biomarkers for lung cancer: clinical uses. Curr Opin Pulm Med 2007;13:249–55.

45. Beachy SH, Repasky EA. Using extracellular biomarkers for monitoring efficacy of therapeutics in cancer patients: an update. Cancer Immunol Immunother 2008;57:759–75.

46. Krause M, Baumann M. Clinical biomarkers of kinase activity: examples from EGFR inhibition trials. Cancer Metastasis Rev 2008;27:387–402.

47. Maier S, Dahlstroem C, Haefliger C, Plum A, Piepenbrock C. Identifying DNA methylation biomarkers of cancer drug response. Am J Pharmacogenomics 2005;5:223–32.

48. Dhani N, Siu LL. Clinical trials and biomarker development with molecularly targeted agents and radiotherapy. Cancer Metastasis Rev 2008;27:339–49.

49. Cohen SJ, Punt CJ, Iannotti N, et al. Relationship of circulating tumor cells to tumor response, progression-free survival, and overall survival in patients with metastatic colorectal cancer. J Clin Oncol 2008;26:3213–21.

50. McShane LM, Hunsberger S, Adjei AA. Effective incorporation of biomarkers into phase II trials. Clin Cancer Res 2009;15:1898–905.

51. Benjamin RS, Choi H, Macapinlac HA, et al. We should desist using RECIST, at least in GIST. J Clin Oncol 2007;25:1760–4.

52. Choi H, Charnsangavej C, Faria SC, et al. Correlation of computed tomography and positron emission tomography in patients with metastatic gastrointestinal stromal tumor treated at a single institution with imatinib mesylate: proposal of new computed tomography response criteria. J Clin Oncol 2007;25:1753–9.

53. Shankar LK, Van den Abbeele A, Yap J, Benjamin R, Schuetze S, Fitzgerald TJ. Considerations for the use of imaging tools for phase II treatment trials in oncology. Clin Cancer Res 2009;15:1891–7.

54. 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.

55. Chan JK, Ueda SM, Sugiyama VE, et al. Analysis of phase II studies on targeted agents and subsequent phase III trials: what are the predictors for success? J Clin Oncol 2008;26:1511–8.

56. 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.

Clinical Cancer Research

Alternate Endpoints for Screening Phase II Studies

Neesha Dhani, Dongsheng Tu, Daniel J. Sargent, et al.

Clin Cancer Res 2009;15:1873-1882.