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

Theodore G. Karrison, Michael L. Maitland, Walter M. Stadler, Mark J. Ratain

Background The primary objective of phase II cancer clinical trials is to determine whether a new regimen has sufficient activity to warrant further study, with activity generally defined as tumor shrinkage. However, oncology drug development has been limited by high rates of failure (lack of efficacy) in subsequent phase III testing. This high failure rate may reflect the process by which antineoplastic agents are usually evaluated in phase II trials, i.e., via single-arm studies in which the primary efficacy measure is the proportion of patients who achieve a complete or partial response to the treatment. This design may efficiently eliminate truly ineffective therapy but may not reliably indicate whether subsequent phase III testing is warranted.

Methods We describe the design of a randomized phase II clinical trial of sorafenib in combination with erlotinib for the treatment of patients with non–small-cell lung cancer using change in tumor size, measured on a continuous scale, as the primary outcome variable. For the purpose of determining the sample size of the trial, we made assumptions as to the likely magnitude of treatment effect and the variability in tumor size changes based on data from four previous trials using these agents.

Results The study design includes two different dosage arms and a placebo group with a total sample size of 150 patients and is powered to detect a modest reduction in the mean tumor size burden in the high-dose sorafenib arm compared with a slight increase in the placebo group.

Conclusions Clinical trial designs that treat change in tumor size as a continuous variable rather than categorizing the changes are feasible, and by inclusion of a prospective control group they offer advantages over conventional single-arm trials.

J Natl Cancer Inst 2007;99:1455–61

The primary objective of phase II cancer clinical trials is to determine whether a new regimen has sufficient activity to warrant further study. The most widely adopted method for assessing antitumor activity in solid tumors is a measure of tumor shrinkage that is based on a set of standardized criteria, the Response Evaluation Criteria in Solid Tumors (RECIST) (1). A patient who achieves a complete or partial response by these criteria is defined as an objective responder, and the proportion of objective responders, i.e., the response rate, is the primary endpoint in the design and analysis of phase II cancer trials.

In its simplest application, a phase II trial with response rate as the primary endpoint would seek to determine whether the drug has a nonzero rate in a specified population. This approach to assessing a drug's therapeutic activity derives from the historic experience in oncology that only a limited number of tested drugs had any activity and many disease conditions had no effective therapy. In this setting, a historical control was sufficient because one could reliably assume that a response rate of, say, more than 5% would not occur using previously available therapy. In general, phase II trials in oncology have been conducted in two stages, with a first stage of minimum trial size set to stop early if the true response rate is equal to some uninteresting level p_0 (2,3). For the second stage, if it occurs, a total sample size is set such that if the true response rate equals some desirable target level p_A, the probability that the regimen would be declared inactive is sufficiently small.

In addition to trial size, one must also consider the inherent error in assessing the primary endpoint. Moertel and Hanley (4) demonstrated that there is considerable measurement error when assessing tumor size by clinical examination and that this error

Affiliations of authors: Department of Health Studies (TGK), Department of Medicine (MLM, WMS, MJR), Cancer Research Center (TGK, MLM, WMS, MJR), and Committee on Clinical Pharmacology and Pharmacogenomics (MLM, MJR), University of Chicago, Chicago, IL.

Correspondence to: Mark J. Ratain, MD, Department of Medicine, University of Chicago, 5841 South Maryland Ave, MC 2115, Chicago, IL 60637 (e-mail: mratain@medicine.bsd.uchicago.edu).

See "Funding" and "Notes" following "References."

DOI: 10.1093/jnci/djm158

CONTEXT AND CAVEATS

Prior knowledge

Phase II trials of promising cancer therapies are generally single-arm studies in which the primary measure of efficacy is the proportion of patients who achieve a complete or partial response to treatment.

Study design

Using data from previous trials to estimate the effects of erlotinib and sorafenib on tumor size, the authors designed a phase II trial of these drugs that contained two different dosage arms and a placebo group. The primary outcome variable was tumor size change measured on a continuous scale.

Contribution

The author's design is an alternative to the common single-arm phase II study and has the advantages of preserving information on tumor size changes and incorporating a control arm.

Implications

Phase II clinical trial that treat change in tumor size as a continuous variable and do not rely solely on historical controls may not require prohibitively large sample sizes and may offer advantages over current designs.

Limitations

The relationship between mean change in tumor size and patient benefit is unknown. Therefore, the potential of widespread use of this design to reduce the number of unsuccessful phase III trials is not clear.

increases with the size of the tumor. For imaging techniques, Erasmus et al. (5) evaluated interobserver variability in the reading of computerized tomography scans of lung lesions and found an average relative difference in unidimensional measurements of a single tumor by two different readers of 12%, with a range of 0%–194%. This variability led Moertel and Hanley (4) to recommend that for a patient to be considered a responder, the percentage decrease in bidimensional tumor size should be at least 50%. Miller et al. (6) subsequently recommended that for a patient to be considered to have progressive disease, the percentage increase should be at least 25%. The univariate RECIST criteria for partial response, a 30% or greater decrease in the sum of the longest diameter of all target lesions, and for progressive disease, a 20% or greater increase in the sum of the longest diameter of target lesions, are approximate mathematical mappings of these bidimensional criteria under the assumption of a perfect sphere (1). Nevertheless, the choice of any cut point is arbitrary.

Consideration of measurement error when assessing tumor size is clearly important, both in clinical decision-making for an individual patient and when making inferences among groups of patients. It should be noted, however, that although a 15% reduction in tumor size in an individual patient may be inconsequential or within the range of measurement error, an average 15% reduction in a group of patients may well be statistically significant and indicative of a treatment effect. Of course, whether an effect of this magnitude would lead to an improvement in overall survival is a separate question and generally the subject of phase III trials.

More recently, oncology has had the advantage of a far greater number of agents and validated targets as well as an increasing number of populations and disease scenarios for which effective therapy exists. Therefore, phase II trials will often seek to determine whether the response rate for a new treatment exceeds that for standard therapy or whether the addition of a new therapy to a standard one is beneficial. In such a situation, one would generally perform a randomized clinical trial of the standard versus the experimental therapy. There has, however, been a reluctance to conduct randomized comparisons in oncology phase II trials, in large part because the sample size required is deemed greater than desired at this stage of drug development. Thus, most phase II trials are single-arm studies designed to determine whether the response rate for the new agent exceeds some prespecified rate set equal to or slightly below the response rate of the standard therapy.

There are several problems with this approach. First, the historical response rate for a given treatment can be quite variable and highly dependent on the enrolled patient population, even when standardized criteria for evaluating response such as RECIST are used. Second, although single-arm phase II designs are generally efficient in determining whether a regimen is not worthy of further study, due to the lack of internal controls they may not provide sufficiently reliable information as to whether the new regimen has a better and clinically meaningful response rate in comparison with the standard therapy. Partly as a result, oncology drug development has been limited by high rates of failure in subsequent phase III testing (7,8).

Finally, the response rate endpoint itself may be problematic. Citing bevacizumab and cetuximab as examples, Ratain and Eckhardt (9) noted that drugs may be active even if they do not lead to high-level tumor regression. Also, categorizing a continuous variable discards information (10,11), and in the case of a cytostatic agent this may be all of the information pertinent to the drug's effectiveness if the categorization is response or no response. This problem, i.e., detecting the effects of agents that are more cytostatic in nature, can be partly overcome by redefining the endpoint as stable disease or better, but this categorization still discards valuable information. Furthermore, when smaller degrees of tumor shrinkage or even lack of growth are defined as a drug benefit, the variability in patient and cancer natural history makes it even more incumbent on investigators to use a concurrent rather than a historical control group.

These concerns prompted us to propose, not a new design approach, but the revival of an old one. We illustrate the advantages of such an approach by describing the design of a randomized phase II clinical trial of sorafenib in combination with erlotinib for the treatment of patients with non–small-cell lung cancer (NSCLC). Our proposal is based on the work of Lavin (12), who treated tumor size as a continuous variable for assessing antitumor activity.

Methods and Results

Preserving Information

Lavin's (12) proposal to use tumor size changes as a continuous rather than dichotomous variable should not be surprising because categorization of a continuous variable results in a loss of information and consequently a reduction in statistical efficiency, particularly if the distribution is split into just two classes. Categorization

is often performed to simplify data analysis and interpretation of results, but such simplicity can come at a high cost and may well create new problems (13). For example, MacCallum et al. (14) illustrated how dichotomization of a continuous measure can both obscure important differences between individuals and inappropriately accentuate small differences. Issues of measurement error aside, a patient whose tumor shrinks by 25% has a different outcome than one whose tumor increases by 10%, yet both are labeled by RECIST as having stable disease. Conversely, there is not much difference between 35% shrinkage and 25% shrinkage, yet the former patient is classified as a responder but the latter is not.

Transforming Size Data to Achieve a Normal Distribution

In addition to treating tumor size as a continuous variable, Lavin (12) considered transformations of this variable, a standard data-analytic tool (15). He presented data from a study of 46 patients with advanced gastric cancer showing that the distribution of tumor size ratios (ratio of tumor size at 1 month after treatment to that at baseline) was approximately log-normal; in other words, the log of the ratio was approximately normally distributed. Using statistical notation, if y_t denotes tumor size at a fixed time t and y_0 denotes baseline tumor size, then

$$\log\left(\frac{y_t}{y_0}\right) = \log(y_t) - \log(y_0) \sim N(\mu, \sigma^2),$$

where "$\sim N(\mu, \sigma^2)$" means that the variable on the left-hand side has a normal distribution with mean μ and variance σ^2. It is not uncommon to find that ratios have a skewed distribution which, after transformation to the log scale, conforms more closely to the symmetric normal distribution. It is also important to note that exact conformity to the normal distribution is not required; statistical theory (the central limit theorem) informs us that as the sample size increases, the distribution of the sample mean will tend toward normality. Thus, a comparison of means from two treatment groups can be accomplished by applying a simple t test, and the approximate normality often afforded by the log transformation means that the sample size need not be very large for this test to be valid (a sample size of n = 15 or 30 per group generally suffices). Other alternatives for comparing tumor size data in different treatment groups would be a covariance analysis, in which the baseline tumor size is treated as a covariate, or a repeated measures analysis of variance (ANOVA). Brogan and Kutner (16) discussed the assumptions associated with each type of analysis, as well as their similarities and differences, and concluded that the investigator should choose the method based on his/her research objectives. Fleiss (17) argued that analysis of covariance (ANCOVA) is preferred over analysis of change scores because the former will almost always be associated with greater variance reduction. Although ANCOVA is worth considering when analyzing tumor size data, evaluation of the change in tumor size is so familiar to clinical investigators that we have elected to base the primary analysis of our proposed study on this measure.

Sample Size Considerations

If tumor size is to be treated as a continuous variable, a concurrent control group is almost surely necessary, for two reasons. First, unlike the situation in which response rates are analyzed, previous studies in the literature usually do not provide the data needed for historical comparison. Second, even if historical controls were available, the inherent intraindividual and interindividual variability in tumor size together with potential patient selection effects would limit such an approach. As Estey and Thall (18) pointed out, historical comparisons are confounded because "variables that have a substantive impact on response to treatment usually vary a great deal between trials …. Consequently, when the results of separate single-arm trials of different treatments are compared, an apparent treatment difference may be due to a trial effect. Conversely, the apparent absence of a treatment effect may be due to an actual treatment effect being canceled out by a trial effect."

The major downside to performing a randomized trial is that a larger sample size is required. Gehan and Freireich (19) noted that a two-arm study requires four times the number of patients as a single-arm study in which the control response rate is treated as a known parameter. However, by using a continuous endpoint the size of the trial can be kept to a feasible number. As Lavin (12) showed, for a two-arm study, the sample size required for a continuous endpoint can be 44%–64% less than that needed for a dichotomous variable. For example, suppose one wishes to detect an improvement in the response rate from 20% for standard therapy to 40% for an investigational agent with 80% statistical power, using a one-sided test at the $\alpha = .05$ significance level. This would require 73 patients per treatment arm, or a total of 146. If instead, we analyze the change in tumor size and assume that on a log scale these changes have an approximately normal distribution with a standard deviation (SD), σ, of 0.64 [the value obtained in Lavin's (12) gastric cancer example], then response rates of 20% and 40% correspond to mean changes of –0.155 and –0.529, respectively, or a difference of 0.374. The sample size required to detect a difference of this magnitude with 80% power is 36 per group, or a total of 72, less than half of that needed for a comparison of response rates. Interestingly, as far back as 1960, Zubrod et al. (20) pointed out that a measured or graded response would require smaller sample sizes than the yes–no (quantal) approach.

Lavin (12) used a one-sided test, which is common in the phase II setting. Furthermore, we believe it would be acceptable to relax the α level from .05 to .10 because a positive result will be followed by a confirmatory phase III trial (21). These measures also reduce the sample size required relative to that needed for a conventional, two-sided test at the .05 significance level. However, if one desires stronger evidence before proceeding to phase III, a lower α level can be used.

Other Considerations in Treating Tumor Size Data

Lavin (12) made two other important points regarding the implementation of a trial in which the tumor size ratio is the endpoint considered. First, patients may die or drop out of the study due to toxicity or other reasons before the chosen time point for measuring tumor size change. Second, if a patient has a complete response, the log ratio is undefined. For analysis on a continuous scale, a simple solution in both of these situations is to rank these two classes of patients at the extreme ends of the distribution (worst possible outcome for deaths and dropouts and best possible outcome for complete responders) and to replace the t test with a nonparametric test. When the data are normally distributed, use of a nonparametric (e.g., Wilcoxon) test entails a very small loss of

Table 1. Four trials used for sample size calculation*

First author, year (reference)	Cancer type	Treatment	No. of patients	% Change		Log ratio		Median PFS (mo)	No. of CR	No. of early deaths†
				Mean	SD	Mean	SD			
Rudin, 2006 (24)	NSCLC	Erlotinib	33	2.1%	18.8%	0.004	0.190	3.4	0	2
Ratain, 2006 (25)	RCC	Sorafenib	193	–18%	33%	–0.198	0.402	6.7	0	1
Shepherd, 2005 (23)	NSCLC	Erlotinib	405	10.1%	34.2%	0.048	0.340	2.2	3	20‡
Gatzemeier, 2006 (26)	NSCLC	Sorafenib	48	1.2%	19.7%	–0.009	0.215	2.7	0	3

* SD = standard deviation; PFS = progression-free survival; CR = complete response; NSCLC = non–small-cell lung cancer; RCC = renal cell cancer.

† Within first two cycles of therapy (2 months).

‡ Approximate number determined from overall survival curve at 2 months.

efficiency, generally no more than 5%, relative to a *t* test (22). For nonnormally distributed data, the nonparametric test will usually be associated with an increase in efficiency. Therefore, switching from a parametric to a nonparametric test should have minimal impact on the statistical power of the comparison. We recommend calculating the sample size based on a *t* test and increasing the number of patients by 5% to allow for the use of a nonparametric procedure should that prove necessary. Finally, for patients with multiple lesions, Lavin (12) suggested choosing the most clearly measurable tumor mass. Consistent with RECIST criteria, we propose instead that one calculate the sum of the longest diameters of all target lesions and then take the log of the ratio of the sums obtained before and after treatment. In the event that a new lesion emerges, we suggest simply adding its longest diameter to the sum. Whereas RECIST would label such a case "progressive disease," we see no reason for making any further modification to the index of total tumor burden.

A Specific Example: The Design of a Phase II Trial of Sorafenib in Combination With Erlotinib for Non–Small-Cell Lung Cancer

Three agents—docetaxel, pemetrexed, and erlotinib—have been approved as monotherapy in the second-line setting of NSCLC. As an orally available, noncytotoxic agent, and the only one for which a large, randomized, placebo-controlled phase III trial demonstrated a statistically significant improvement in overall survival (23), the small-molecule tyrosine kinase inhibitor erlotinib may be considered the best agent for further combination therapy development strategies. The objective of our proposed study is to determine whether a combination regimen consisting of the oral agents erlotinib and sorafenib, an inhibitor of the vascular endothelial growth factor signaling pathway, has sufficiently greater activity than erlotinib alone to merit a subsequent phase III trial. This study will also seek to determine whether a 200 mg twice daily and/or a 400 mg twice daily dose of sorafenib in the combination regimen should be brought to the phase III setting.

Patients will be randomly assigned to one of three treatment arms: erlotinib, 150 mg daily plus placebo (E150/S0); erlotinib, 150 mg daily plus 200 mg sorafenib twice daily (E150/S200); or erlotinib, 150 mg daily plus 400 mg sorafenib twice daily (E150/S400). All treatments will be administered in a double-blind fashion. Clinical and laboratory/toxicity evaluations will be conducted every 4 weeks, and computerized tomography every 8 weeks (once every two cycles). The primary endpoint will be the change in tumor size burden from baseline to 8 weeks; specifically, if a patient has *m* target lesions identified at baseline, the primary outcome variable will be

$$\log(y_{81} + y_{82} + \ldots + y_{8m}) - \log(y_{01} + y_{02} + \ldots + y_{0m}),$$

where y_{tj} denotes the tumor size at time *t* for lesion *j*.

Determination of Sample Size

To determine sample size for this trial, we examined data from four single-agent studies of erlotinib or sorafenib to assess the likely magnitude of treatment effect and variability in tumor size changes as well as to relate the mean change with subsequent degree of clinical benefit (Table 1). These four studies are described below.

In a trial of erlotinib conducted at our own institution and Johns Hopkins University (24), data from 33 patients with NSCLC were available with tumor size measurements at cycle 2 (8 weeks after the start of treatment). The mean percentage change in tumor size from baseline was +2.1%, with an SD of 18.8%. On the log ratio scale, the mean change was +0.004 with an SD of 0.190. The median progression-free survival time among the patients in this single-arm study was 3.4 months (Table 1).

A phase II placebo-controlled trial of sorafenib in patients with metastatic renal cell carcinoma reported by Ratain et al. (25) demonstrated a statistically significant benefit of sorafenib in metastatic renal cancer using a randomized discontinuation design. All patients (n = 202) received sorafenib initially during a 12-week open-label run-in period. Data on percentage change in tumor size during the run-in period were available for 193 patients and were displayed in a waterfall plot (a plot of percent change in tumor size ordered from largest increase to largest decrease). The mean percent change was –18% with an SD of 33%. Using the delta method, a mathematical procedure for approximating the mean and SD of a transformed variable, we calculated that these values correspond to a mean change of –0.198 with an SD of 0.402 on the log ratio scale. The estimated median progression-free survival time with sorafenib measured from entry into the study was 29 weeks (6.7 months) (Table 1).

In the randomized, placebo-controlled clinical trial of erlotinib as a treatment for NSCLC patients who had been previously treated with one or two regimens of combination chemotherapy reported by Shepherd et al. (23), the rates of complete response, partial response, stable disease, and progressive disease in the erlotinib arm (n = 488) were 0.7%, 8.2%, 36.1%,

and 38%, respectively (17% of outcomes were "not confirmed"). Descriptive statistics on changes in tumor size treated as a continuous variable were not given. To approximate the mean change in tumor size, rather than assigning the midpoints of the RECIST categories, we used the waterfall plot from Ratain et al. (25) to approximate the typical change within each RECIST category. Based on these typical changes and the percentage distribution across the four RECIST categories, we calculated an overall mean and variance. (For the three complete responders in the trial, a conservative value of –0.693, corresponding to a 50% reduction in tumor size, was imputed.) This yielded +10.1% (SD = 34.2%) on the percent change scale and +0.048 (SD = 0.340) on the log ratio scale (Table 1). The median progression-free survival time in the erlotinib group was 2.2 months, compared with 1.8 months in the placebo group ($P<.001$), and the median overall survival time was 6.7 months, compared with 4.7 months ($P<.001$).

Finally, in an unpublished phase II trial of sorafenib for treatment of patients with advanced NSCLC conducted by Gatzemeier et al. (26), data on tumor size changes were graphically presented for 48 patients. From this graph we determined that the mean change in tumor size was 1.2% (SD = 19.7%) on the percent change scale and –0.009 (SD = 0.215) on the log ratio scale. Median progression-free survival was 2.7 months (Table 1).

The estimates of the effect of sorafenib in the renal cell cancer trial suggest that modest mean shrinkage (a decrease of 18% in tumor mass in the initial phase of treatment) is associated with a doubling in time to progression for this typically indolent disease. The data from the Shepherd et al. (23) trial of erlotinib as a treatment for NSCLC suggest that a slowing of tumor growth, even without mean tumor mass shrinkage (evidenced by a mean increase in tumor size of 10.1% during the initial treatment period), may be associated with a survival benefit. For our trial, we assumed a mean log ratio of 0.05 for E150/S0 [as observed in Shepherd et al. (23)], –0.07 for E150/S200, and –0.13 for E150/S400. As proposed, the study is powered to detect an effect of combination treatment with erlotinib and high-dose sorafenib in NSCLC patients that produces a change in tumor size equal to two-thirds (–0.13/–0.198) of that seen with sorafenib as a treatment for renal cell cancer. The assumed log ratio for E150/S200 group is between that assumed for the placebo and high-dose groups, but closer to the latter.

We pooled the variance estimates from the four trials and obtained an SD of 0.346 for the log ratio. We then chose a 1 df trend test for assessing dose response. To have 85% power, based on a one-sided test at the $\alpha = .10$ significance level, 40 patients per arm would be required, for a total of 120 patients. However, because the dose–response relationship may not be linear, we increased the sample size to 48 per group (144 total), thus providing 80% power to detect a true difference of 0.18 between the mean log ratio of any two groups (equal to the assumed difference between the high-dose and placebo groups). This calculation incorporates a Tukey allowance for multiple comparisons and again uses a one-sided test at the $\alpha = .10$ significance level. Finally, to maintain power if a nonparametric test is needed, we further increased the sample size to 150 (50 patients per treatment arm).

Data Analysis

The tumor size data will be analyzed by fitting a regression model of log ratio of tumor size against sorafenib dose. If a linear model does not fit the data, we will conduct a 2 df ANOVA F test, followed by pairwise group comparisons. In the event that nonparametric tests become necessary, Cuzick's (27) trend test or a Kruskal–Wallis test will be employed. Across the four cited trials (Table 1), there were 3 complete responders and 26 early deaths. If there are similar occurrences in our trial, we will use nonparametric tests.

Discussion

In this article, we have described the design of a randomized phase II trial using the change in tumor size treated as a continuous measure as the primary outcome variable. The study includes two different dosage arms and a placebo group with a total sample size of 150 patients. Thus, it is not an unduly large trial, and it will provide a stronger basis for determining whether to proceed to phase III than would a single-arm study based on RECIST-derived response rates. Our endpoint will also be sensitive to a sorafenib effect that is more cytostatic than cytotoxic in nature, should that turn out to be the case. In addition, this design will allow us to test an intermediate dose level, which may prove equally (or more) efficacious and less toxic than the currently presumed optimal dose. Another advantage of the proposed design is that it will allow us to examine changes in drug target biomarkers in a randomized fashion. Although not discussed here, an early stopping rule for futility could be incorporated into this design if desired (28).

Efforts to change from traditional single-arm phase II trial designs to obtain more conclusive results raise conflicting goals. Whereas investigators and sponsors seek results that are more predictive of phase III outcomes, and conducting randomized phase II trials is one way to achieve this goal, these trials require more patients. However, as illustrated above, with analysis of a continuous endpoint the increase in sample size is not prohibitive. Moreover, the use of classical, categorical endpoints sets a high and not necessarily useful threshold for advancing a drug and increases the likelihood for both false-positive and false-negative outcomes (9). In addition to improving the efficiency of antineoplastic drug development in the phase II setting, assessment of combination regimens is increasingly important (29) but even more problematic than for single agents. With growing numbers of agents and agent classes available for evaluation, it is crucial to develop phase II clinical trial strategies that use patient resources efficiently and provide data more predictive of phase III results than current phase II approaches.

Our targeted effect size, i.e., an average tumor shrinkage with erlotinib plus sorafenib that is equal to two-thirds of that seen with sorafenib in patients with renal cell cancer (compared with a small average increase under erlotinib alone), is one that we believe is likely to be associated with a clinically important difference. This is because an average 18% reduction during the run-in period of the renal cancer trial of sorafenib (25) ultimately led to a substantial difference in the median progression-free survival time during the randomized discontinuation phase of the study. The confirmatory phase III "up-front" randomized trial of sorafenib also

detected a 12-week improvement in median progression-free survival (30). Even if a new agent is purely cytostatic in nature, with a mean tumor size change of zero, a treatment effect could still be detected by the proposed design if standard therapy results in an average increase, although the statistical power will depend on the magnitude of the difference. We do not assume that a statistically significant difference in mean tumor size change will necessarily translate into a benefit in terms of the more clinically relevant endpoints of progression-free or overall survival. Rather, we assert that controlled studies using this design are more likely to predict clinically meaningful results in phase III trials than are single-arm phase II studies that rely on historical control response rates (31). In any case, a positive result from a randomized phase II trial should not be taken as definitive evidence supporting the adoption of a new therapy in absence of a confirmatory phase III trial, as some have cautioned (32,33).

There are further advantages to treating tumor size as a continuous variable. As pointed out by Lavin (12), it is easy to incorporate covariates into the analysis, and if these covariates are predictive of outcome this will serve to reduce unexplained variability and increase the power to detect a treatment effect. Whereas Lavin (12) suggested choosing a fixed posttreatment time at which to measure tumor size change, one can go a step further and treat all of the tumor size measurements as longitudinal data (34). Indeed, a drawback to our design is that the optimal time point for posttreatment evaluation is unknown and will likely vary by disease. A longitudinal analysis would make maximal use of the data, enabling a comparison of the pattern of tumor size changes between treated and control groups over time. Yet another extension of the longitudinal approach would be to fit a bivariate model for time to death and the longitudinal tumor size measurements during the period that the patient is alive. For example, Schlucter (35) has proposed a bivariate random-effects model that takes into account the correlation between a subject's rate of change in the longitudinal variable and survival time that could potentially be employed here. Hogan and Laird (36) and Henderson et al. (37) describe additional methods for the joint modeling of longitudinal and survival data. Finally, the continuous tumor size endpoint could also be incorporated into enrichment designs such as the randomized discontinuation design.

In summary, we have described the main features of a randomized, phase II clinical trial for patients with NSCLC that uses an alternative model for the evaluation of antitumor activity first proposed by Lavin (12) in 1981. The approach treats tumor size as a continuous variable (on a transformed scale) rather than categorizing the changes, thereby maintaining efficiency and reducing the number of subjects required for a comparative study. This approach may offer advantages over conventional phase II cancer trial designs.

References

(1) Therasse P, Arbuck SG, Eisenhauer EA, Wanders J, Kaplan RS, Rubinstein L, et al. New guidelines to evaluate the response to treatment in solid tumors. J Natl Cancer Inst 2000;92:205–16.

(2) Fleming TR. One sample multiple testing procedure for phase II clinical trials. Biometrics 1982;38:143–51.

(3) Simon R. Optimal two-stage designs for phase II clinical trials. Control Clin Trials 1989;10:1–10.

(4) Moertel CC, Hanley JA. The effect of measuring error in the results of therapeutic trials in advanced cancer. Cancer 1976;38:388–94.

(5) Erasmus JJ, Gladish GW, Broemeling L, Sabloff BS, Truong MT, Herbst RS, et al. Interobserver and intraobserver variability in measurement of non-small-cell carcinoma lung lesions: implications for assessment of tumor response. J Clin Oncol 2003;21:2574–82.

(6) Miller AB, Hoogstraten B, Staquet M, Winkler A. Reporting results of cancer treatment. Cancer 1981;47:207–14.

(7) Kola I, Landis J. Can the pharmaceutical industry reduce attrition rates? Nat Rev Drug Discov 2004;3:711–5.

(8) Ratain MJ. Phase II oncology trials: let's be positive [editorial]. Clin Cancer Res 2005;11:5661–2.

(9) 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.

(10) Altman DG. Categorizing continuous variables. In: Armitage P, Colton T, editors. Encyclopedia of biostatistics. Chichester (U.K.): Wiley; 1998. p. 563–7.

(11) Harrell FE. Regression modeling strategies. New York: Springer; 2001. p. 6, 379–80.

(12) Lavin PT. An alternative model for the evaluation of antitumor activity. Cancer Clin Trials 1981;4:451–7.

(13) Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. Stat Med 2006;25:127–41.

(14) MacCallum RC, Zhang S, Preacher KJ, Rucker D. On the practice of dichotomization of quantitative variables. Psychol Methods 2002;7:19–40.

(15) Tukey JW. Exploratory data analysis. Reading (MA): Addison-Wesley; 1977.

(16) Brogan DR, Kutner MH. Comparative analyses of pretest-posttest research designs. Am Stat 1980;34:229–32.

(17) Fleiss JL. The design and analysis of clinical experiments. New York: John Wiley & Sons; 1986. p. 186–92.

(18) Estey EH, Thall P. New designs for phase 2 clinical trials. Blood 2003;102:442–8.

(19) Gehan EA, Freireich EJ. Non-randomized controls in cancer clinical trials. New Engl J Med 1974;290:198–203.

(20) Zubrod GG, Schneiderman SM, Frei E III, Brindley C, Gold GL, Schnider B, et al. Appraisal of methods for the study of chemotherapy of cancer in man: comparative therapeutic trial of nitrogen mustard and thio phosphoamide. J Chronic Dis 1960;11:7–33.

(21) Simon R. New methods for the design and analysis of clinical trials. In: Devita VT, Hellman S, Rosenberg SA, editors. Updates: principles & practice of oncology. 5th ed. Philadelphia (PA): Lippincott-Raven; 1999. p. 4.

(22) Bickel PJ, Doksum KA. Mathematical statistics: basic ideas and selected topics. San Francisco (CA): Holden-Day, Inc; 1977. p. 351–3.

(23) Shepherd FA, Rodrigues Pereira J, Ciuleanu T, Tan EH, Hirsh V, Thongprasert S, et al. Erlotinib in previously treated non-small-cell lung cancer. N Engl J Med 2005;353:123–32.

(24) Rudin CM, Desai AA, Janisch L, Carducci M, Karrison T, Liu W, et al. A prospective pharmacogenomic (PG), pharmacodynamic (PD), and pharmacokinetic (PK) study of determinants of erlotinib toxicity. J Clin Oncol 2006;24:3080.

(25) Ratain MJ, Eisen T, Stadler WM, Flaherty KT, Kaye SB, Rosner GL, 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.

(26) Gatzemeier U, Blumenschein G, Fosella F, Simantov R, Elting J, Bigwood D, et al. Phase II trial of single-agent sorafenib in patients with advanced non-small cell lung carcinoma. J Clin Oncol 2006;24:7002.

(27) Cuzick J. A Wilcoxon-type test for trend. Stat Med 1985;4:87–90.

(28) Dignam JJ, Bryant J, Wieand S. Early stopping of cancer clinical trials. In: Crowley J, editor. Handbook of statistics in clinical oncology. New York: Dekker; 2001. p. 189–209.

(29) Dancey JE, Chen HX. Strategies for optimizing combinations of molecularly targeted anticancer agents. Nat Rev Drug Discov 2006;5:649–59.

(30) Escudier B, Eisen T, Stadler WM, Szczylik C, Oudard S, Siebels M, et al. Sorafenib in advanced clear-cell renal-cell carcinoma. New Engl J Med 2007;356:125–34.

(31) Fazzari M, Heller G, Scher HI. The phase II/III transition: toward the proof of efficacy in cancer clinical trials. Control Clin Trials 2000; 21:360–8.
(32) Liu PY. Selection designs. In: Crowley J, editors. Handbook of statistics in clinical oncology. New York: Dekker; 2001. p. 119–26.
(33) Dignam J, Karrison TG, Bryant J. Design and analysis of oncology trials. In: Chang AE, Ganz PA, Hayes DF, Kinsella TJ, Pass HI, Schiller JH, et al., editors. Oncology—an evidence based approach. New York: Springer; 2006. p. 112–26.
(34) Diggle PJ, Heagerty P, Liang KY, Zeger SL. The analysis of longitudinal data. 2nd ed. Oxford: Oxford University Press; 2002.
(35) Schlucter MD. Methods for the analysis of informatively censored longitudinal data. Stat Med 1991;11:1861–70.
(36) Hogan JW, Laird NM. Model-based approaches to analyzing incomplete longitudinal and failure time data. Stat Med 1997;16:259–72.
(37) Henderson R, Diggle P, Dobson A. Joint modeling of longitudinal measurements and event time data. Biostatistics 2000;1:465–80.

Funding

National Cancer Institute (P30 CA14599, U01 CA69852, U01 CA07003).

Notes

Dr Ratain is a consultant to Onyx Pharmaceuticals and Dr Stadler is a consultant to and has received research funding from Onyx and Bayer Pharmaceuticals. Dr Maitland has received an honorarium (less than $10 000) and research funding from Bayer, Inc. Bayer Pharmaceuticals and Onyx Pharmaceuticals have jointly developed the drug sorafenib as a cancer therapeutic.

The authors thank James Dignam for helpful comments during the preparation of the manuscript. We also thank the Associate Editor, reviewers, and Journal Senior Editor for many helpful suggestions and comments.

The authors take full responsibility for the decision to submit the manuscript for publication and for the writing of the manuscript.

Manuscript received January 31, 2007; revised July 5, 2007; accepted August 22, 2007.