Comparison of Power Between Randomized Discontinuation Design and Upfront Randomization Design on Progression-Free Survival

Pingfu Fu, Afshin Dowlati, and Mark Schluchter

A B S T R A C T

Purpose

Enrichment based on molecular characteristics has emerged as an important inclusion criterion in phase II trials of targeted anticancer agents. In this study, we evaluate a well-described method of population enrichment by tumor growth characteristics in the early development stage of targeted cytostatic agents.

Methods

For some solid tumors, such as pancreatic carcinoma, using a time-to-event end point (eg, time to disease progression) to evaluate the efficacy of a cytostatic agent in a phase II trial is more relevant than clinical response by Response Evaluation Criteria in Solid Tumors. In this setting, we compared the power of the randomized discontinuation and upfront randomization designs under two previously proposed tumor growth models for treatment effect when the end point is time-to-event.

Results

By selecting patients with more homogeneous tumor growth characteristics, the randomized discontinuation design is more efficient than the upfront randomization design when treatment benefit is restricted to slow-growing tumors. Under a model where only a subset of patients expressing the molecular target are sensitive to the agent, the randomized discontinuation design is more powerful than the upfront randomization design when the treatment effect is small; and vice versa when the treatment effect is moderate to large.

Conclusion

For selected targeted agents where a bioassay to select patients expressing the specific molecular target is not available, the randomized discontinuation design is a feasible alternative patient enrichment strategy in certain disease settings and provides a reasonable platform to evaluate drugs before phase III testing.

J Clin Oncol 27:4135-4141. © 2009 by American Society of Clinical Oncology

INTRODUCTION

Over the past decade, we have witnessed a transition in drug discovery from classic cytotoxics to agents targeting defined molecular pathways, of which many are thought to have potential cytostatic effects. These so-called novel targeted agents usually affect specific proteins that play a direct role in malignant transformation and progression, including tumor growth, angiogenesis, and metastases. In preclinical models many of these drugs exhibit growth inhibitory patterns with a resultant cytostatic picture and lack the cytotoxicity observed with traditional chemotherapy.¹ The clinical counterpart of such an observation is the lack of significant tumor shrinkage.² More careful observation of the tumor growth pattern of pa-

tients enrolled on clinical trials utilizing targeted agents has led to the conclusion that although responses are not seen by the classic Response Evaluation Criteria in Solid Tumors end point, a significant number of patients have stable disease as depicted by “waterfall graphs.”³ Because of this finding, traditional phase II evaluation based on tumor regression as demonstrated by physical, radiologic, or imaging examination has been challenged.⁴ As a growing number of targeted agents are tested in clinic trials, there has been pressure to develop new study designs that could both better predict outcome and also be more efficient.⁵

As evidenced by many genomic and proteomic analyses of cancer, the cancer patient population is quite heterogeneous.^{6,7} In certain situations, patients’ clinical and biologic phenotype

From the Departments of Epidemiology and Biostatistics and Hematology/Oncology, Case Comprehensive Cancer Center, Case Western Reserve University, Cleveland, OH.

Submitted August 15, 2008; accepted March 12, 2009; published online ahead of print at www.jco.org on July 27, 2009.

Supported by Grant No. U01 CA62502 from the National Institutes of Health.

Authors’ disclosures of potential conflicts of interest and author contributions are found at the end of this article.

Corresponding author: Pingfu Fu, PhD, Department of Epidemiology and Biostatistics, Case Western Reserve University, 10900 Euclid Ave, Cleveland, OH 44106; e-mail: pxf@bfox.cwru.edu.

© 2009 by American Society of Clinical Oncology

0732-183X/09/2725-4135/\$20.00

DOI: 10.1200/JCO.2008.19.6709

may predict benefit from novel molecularly targeted agents. One major challenge is to determine the predictive markers in the early phase of drug development. However, in this early stage of drug development, either the predictive target cannot be defined or the assays have not been validated.⁸ Although enrichment of the study population based on molecular markers is an ideal strategy for drug development, it is uncommonly used in phase I designs given the problems with defining and measuring the marker. Furthermore, given the low response rate to many new agents, traditional study designs such as the Simon's two-stage design of phase II trials,⁹ may not be appropriate and/or efficient. Indeed, in recent years, given the failure of many new agents in phase III clinical trials, phase II trials which screen drugs for decisive testing of the drugs in phase III trial have come under heavy scrutiny with the hope of better predicting success in subsequent phase III trials. For this purpose, alternative designs for phase II studies, such as randomized phase II studies,¹⁰ randomized discontinuation designs,^{11,12,13} multinomial designs,¹⁴ and alternative short-term end points other than clinical response, such as time-to-disease progression, have been put forward.

Although the need for a control arm in the phase II development of cytostatic agents has not been agreed on by majority of investigators,¹⁵ it has been strongly advocated by some,¹⁶ especially when a reliable historical control is not available. In this study, we evaluate the randomized discontinuation design (RDD) in the testing of cytostatic agents (agents which are thought to have a low response rate by Response Evaluation Criteria in Solid Tumors) in a heterogeneous cancer population where a molecular enrichment strategy is not available at the time of phase II development.^{11,13,17} The RDD is an example of an enrichment strategy that is not based on measuring a specific biologic marker but rather a subpopulation with common tumor growth characteristics from the original patient population is selected. In the RDD, the enrichment is accomplished by treating all patients with the given drug for a short period of time (run-in period). After the run-in period, patients who have stable disease or better response (complete or partial response) are randomly assigned to either continue or discontinue with the agent for a fixed period of time. Those with disease progression are taken off study. The run-in period of RDD seeks to select a more homogeneous group of patients, consisting of those who are more likely to benefit from the investigational drug. The goal of such an enrichment strategy is to increase statistical power over the original patient population with heterogeneity of tumor growth characteristics, biomarkers, or genetic signatures since patients who are not sensitive to the agent will dilute the treatment effect over the whole population.

In the following sections, we compare the power between the RDD and the upfront randomization design (RD) when the clinical end point is time-to-disease progression rather than clinical response¹² or progression rate at fixed time point.¹³ Instead of simulating survival data directly from a specific parametric distribution,¹⁸ such as the Weibull, our time-to-disease progression data were generated based on simple exponential tumor growth models,¹² one of many classes of mathematical models of tumor growth¹⁹ describing the tumor microenvironment and offering therapeutic guidance. The models we choose yield valuable insights

into the relationship among tumor size, growth characteristics, and therapeutic response.

METHODS

Without loss of generality, we assume the baseline tumor diameter is 1; and using the same notations and tumor growth models studied by Freidlin and Simon¹³ the exponential tumor growth model can be specified as follows:

$$D_i(t) = \exp[(1-k)\lambda_i t], \quad (1)$$

where k denotes the treatment effect ($k = 0$ for placebo, 1 for active drug), λ_i denotes tumor growth rate of patient i and t denotes time (in weeks) measured from the onset of treatment. In the simulation, the tumor growth rates λ_i are assumed to have a log normal distribution in the patient population. The level of heterogeneity of cancer patients depends on the variation of the growth rates λ_i and this will also reflect the treatment effect on individual patients. Following Freidlin and Simon,¹³ the mean and variance of the growth rates were set to have approximately 70% of patients progress by 16 weeks, with a median of 32% increase in tumor diameter by 16 weeks. Progression was defined as a 20% increase in tumor diameter from the baseline.

Using the two models for the effect of treatment proposed by Freidlin and Simon (ie, the growth rate cutoff model [GRC] and the sensitive fraction model [SF]), we derive the formula for calculating the time to disease progression. In the GRC model, only patients with tumor growth rate λ_i below a certain cutoff value benefit from treatment. The cutoff value c_0 was defined in terms of the percentage increase in diameter of an untreated tumor over 16 weeks. The GRC model is meant to correspond to some diseases, such as non-Hodgkin's lymphoma where the tumor for some of the patients is slow to develop (indolent) but grows rapidly for other patients (aggressive). In the SF model, only a fraction, p_r , of the patient population is sensitive to the treatment and treatment sensitivity is independent of growth rate. The SF model corresponding to the targeted therapy setting, as described by Freidlin and Simon,¹³ where there is no reliable assay to select patients expressing the targeted biomarkers. In the simulation, we assumed that both sensitive and nonsensitive populations have the same distribution of the tumor growth rates.

In the RD, where all patients are randomly assigned to placebo or treatment at baseline, the time to disease progression (20% increase in tumor) of each patient under either treatment effect model (ie, GRC or SF) is observed if the patient has disease progression by 16 weeks after initiation of treatment and is censored at 16 weeks if the patient is progression free. We also consider the scenario where the follow-up is 32 weeks long (in the event that 16-week follow-up is too short). In the RDD where all patients initially receive the agent of interest during the run-in period (we used 16 weeks regardless of the length of follow-up in the second stage), patients with stable disease or response (ie, those who have not progressed) in the run-in period are randomly assigned to either continuing or discontinuing the therapy (second stage). As in the RD, we consider two different follow-ups: 16 weeks and 32 weeks from the time of random assignment at the end of the run-in period, time-to-disease progression was calculated: each patient either has disease progression or administrative censoring by the end of follow-up. For both the RD and RDD, the event is defined by disease progression which was defined as 20% increase in tumor diameter. The sample size was fixed in advance.

Under each design and treatment effect model, we first present expressions that give the rate of disease progression at time t for the model, where t is 16 or 32.

With the expression for the tumor diameter $D_i(t)$ as in (1), λ was simulated from a lognormal distribution with mean $\mu = -4.196$ and standard variation $\sigma = 0.5326$ (the μ and σ are chosen so that approximately 70% of patients progress by 16 weeks, with a median 32% increase in tumor diameter by 16 weeks).

GRC Model

In the RD, the probability of disease progression ($\geq 20\%$ increase in tumor) at 16 weeks under the GRC model for the placebo arm is

$$\Pr\{D(16) > 1.2\} = \Pr\left\{\lambda > \frac{\log(1.2)}{16}\right\} = 0.7. \quad (2)$$

For the treatment arm, the tumor diameter under the GRC model can be expressed as

$$D(16) = \begin{cases} \exp(16\lambda), & \text{if } \lambda > \frac{\log(1 + c_0/100)}{16} \\ \exp[(1-k)16\lambda], & \text{if } \lambda \leq \frac{\log(1 + c_0/100)}{16} \end{cases} \quad (3)$$

where c_0 is the tumor growth rate cutoff value above which patients will not benefit from the treatment. Thus, from (2) and (3) and the lognormal distribution of λ , the probabilities of disease progression at 16 and 32 weeks can be calculated.

Under the RDD, every patient is treated by the agent of interest in the run-in period. Thus, at the end of the run-in period, the tumor size can be expressed by equation (3). At the second stage, only patients without disease progression are randomly assigned to placebo or the agent of interest. In order to evaluate the power of treatment effect based on an analysis of time to disease progression, we start the clock at the time of random assignment. The probability of disease progression at 16 weeks or any fixed time points including 32 weeks after random assignment for the placebo and treatment arms can also be calculated. Because of the features of the RDD and the memoryless property of exponential tumor growth model used, patients who did not progress in the run-in period and who were random assigned to the agent of interest in the second stage will not progress in the first 16 weeks after random assignment. This means that no events will be observed in the second stage of the RDD with 16 weeks follow-up. Thus, instead of 16 weeks follow-up for the second stage used by others on different setting,^{12,13} we focus on simulations with 32 weeks follow-up after random assignment for both the RD and RDD.

For each simulation, the power was calculated based on a one-sided log-rank test with type I error $\alpha = .05$. In our simulations, the range of the possible treatment effect parameter k is from 0.1 to 0.9, the growth cutoff rate c_0 changes from 100% to 15% under the GRC model, and the sensitive proportion of patients to treatment ranges from 1 to 0.3 under the SF model. The simulation results are based on $N = 1,000$ replications. In the same simulations, we also investigated the power of the one-sided .05 level χ^2 test to examine the equality of progression rates (proportions) between the two arms for both the GRC model and the SF model.

RESULTS

By the methods described earlier, we present results for two models: GRC and SF with 32 weeks follow-up and the event (disease progression) is defined as a 20% increase of tumor diameter for both time-to-event and binary outcomes. Table 1 presents results for the GRC model with 16 weeks run-in period (for RDD) and 32 weeks follow-up after random assignment (for both RDD and RD). Under the GRC model sensitivity to treatment is based on tumor growth characteristics; and we assume that the growth cutoff rate is the same for both the run-in period (RDD) and in the second stage of RDD and in the RD. Column 1 indicates the treatment effect. Column 2 shows the growth rate cutoff for drug activity. No cutoff means that all patients (tumors) are sensitive to the treatment of interest; and a 70% cutoff means that only tumors that would grow over 16 weeks by less than 70% without treatment will benefit from the treatment. Column 3 is the overall sample size (fixed in advance) for the RD and the run-in period of classifying which

tumors are sensitive to the treatment. The proportion of patients going to the second stage usually increases when the treatment effect increases; and it decreases when the growth cutoff for drug activity decreases. For most scenarios, the RDD has more power than the RD; for some cases with large treatment effects and large growth cutoff rates, the power of the RD is comparable to that of the RDD. When the growth cutoff rate is below 20%, the power for both RDD and RD drops dramatically (when the growth cutoff is decreased, power decreased, and larger sample size are needed). Although the overall sample size is reduced significantly in the RDD, patients are greatly enriched during the run-in period in terms of tumor sensitivity to the agent of interest and the advantage of enrichment overpowers the advantage of larger sample size that the RD enjoys. When comparing the log-rank test used with the time-to-event end point to the χ^2 test used with the binary end point (proportion of disease progression at 32 weeks after random assignment) we found that in most cases the log-rank test was more powerful than the χ^2 test. Exceptions to this, where the log-rank test was less powerful, occurred in some extreme cases when the treatment effect was large and the growth cutoff rate was low (eg, 25%, 15%).

Table 2 shows the results for the SF model with 32 weeks follow-up after random assignment (16 weeks run-in period for the RDD). In contrast to the GRC model under which sensitivity to the treatment is based on the tumor growth characteristics, the SF model represents a situation in the early development of targeted agents where only a subset of patients with targeted molecular biomarkers is expected to be sensitive to the drug of interest, but a reliable assay to identify the sensitive patients is not available. This design assumes that the patients who are sensitive to the agent of interest in the run-in period of the RDD will also be sensitive in the second stage if they are eligible for that stage. Column 2 in Table 2 shows the proportion of sensitive patients, ranging from 30% to 100%. When the percentage of sensitive patients is low (eg, $\leq 60\%$), the power of the RD is comparable to or better than the power of the RDD. However, when the size of the treatment effect is small and the percentage of sensitive patients is greater than 60%, the RDD is more powerful than the RD. In these latter scenarios, the benefits of enrichment outweigh the advantage of the larger sample size in the RD. In contrast, when the size of the treatment effect is intermediate to large and the proportion of sensitive patients is not 100%, the power of the RD is higher than that of the RDD, and the two designs are equivalent in power when all patients are sensitive to the agent of interest. In this scenario, the percentage of patients going to the second stage is relatively large (45% to 99%). Although the original patient population is quite heterogeneous in terms of tumor growth characteristics, the patient population is quite homogeneous in terms of sensitivity to the agent. Thus, the benefits of enrichment are outweighed by the larger sample size enjoyed by the RD. Similarly, when comparing the log-rank test based on the time-to-event end point to the χ^2 test based on the binary outcome, the log-rank test has more power in general.

DISCUSSION

In the development of targeted therapeutic agents, attempts including novel designs (eg, adaptive random assignment based on clinical outcomes), and other strategies to better characterize tumors with the

Table 1. RDD Versus RD: Growth Rate Cutoff Model

Treatment Effect	Growth Cutoff for Drug Activity (%)	Overall Sample Size	RDD			RD	
			Proportion Going to Second Stage	Power (based on TTP)*	Power (based on proportion)*	Power (based on TTP)*	Power (based on proportion)*
0.1	No cutoff	120	0.37	0.16	0.02	0.15	0.02
0.1	70	120	0.37	0.2	0.017	0.17	0.015
0.1	50	120	0.37	0.18	0.017	0.16	0.013
0.1	40	120	0.37	0.18	0.022	0.15	0.019
0.1	30	160	0.37	0.21	0.025	0.14	0.023
0.1	25	200	0.37	0.28	0.039	0.15	0.038
0.1	15	400	0.3	0.15	0.081	0.09	0.082
0.2	No cutoff	60	0.46	0.38	0.015	0.32	0.013
0.2	70	60	0.46	0.36	0.017	0.31	0.008
0.2	50	60	0.46	0.39	0.018	0.31	0.015
0.2	40	60	0.46	0.37	0.036	0.28	0.01
0.2	30	80	0.46	0.46	0.049	0.26	0.037
0.2	25	100	0.45	0.52	0.051	0.23	0.04
0.2	15	200	0.3	0.31	0.203	0.1	0.169
0.3	No cutoff	40	0.56	0.53	0.045	0.48	0.017
0.3	70	40	0.56	0.57	0.046	0.5	0.024
0.3	50	40	0.56	0.56	0.043	0.47	0.012
0.3	40	40	0.56	0.57	0.049	0.41	0.025
0.3	30	60	0.56	0.68	0.107	0.38	0.069
0.3	25	80	0.44	0.79	0.191	0.37	0.14
0.3	15	160	0.3	0.44	0.524	0.13	0.413
0.5	No cutoff	20	0.78	0.81	0.092	0.76	0.109
0.5	70	20	0.78	0.79	0.108	0.72	0.115
0.5	50	20	0.78	0.79	0.095	0.67	0.098
0.5	40	40	0.73	0.96	0.497	0.83	0.452
0.5	30	40	0.56	0.97	0.565	0.66	0.453
0.5	25	60	0.44	0.99	0.864	0.64	0.718
0.5	15	120	0.3	0.44	0.633	0.12	0.495
0.6	No cutoff	20	0.88	0.93	0.388	0.91	0.4
0.6	70	20	0.89	0.94	0.396	0.89	0.399
0.6	50	20	0.83	0.93	0.426	0.84	0.427
0.6	40	20	0.73	0.94	0.435	0.72	0.397
0.6	30	40	0.57	0.99	0.962	0.74	0.881
0.6	25	60	0.44	0.99	0.99	0.71	0.981
0.6	15	120	0.3	0.44	0.65	0.13	0.503
0.7	No cutoff	20	0.96	0.99	0.823	0.99	0.804
0.7	70	20	0.93	0.99	0.829	0.98	0.813
0.7	50	20	0.84	0.99	0.871	0.92	0.812
0.7	40	20	0.74	0.99	0.913	0.82	0.797
0.7	30	40	0.56	1	1	0.79	0.976
0.7	25	60	0.44	1	1	0.69	0.969
0.7	15	120	0.3	0.42	0.607	0.13	0.509
0.8	No cutoff	20	0.99	0.99	0.99	0.99	0.99
0.8	70	20	0.93	1	0.994	0.99	0.987
0.8	50	20	0.83	1	0.999	0.94	0.966
0.8	40	20	0.73	0.99	0.99	0.81	0.894
0.8	30	40	0.56	1	1	0.78	0.971
0.8	25	60	0.45	1	1	0.72	0.975
0.8	15	120	0.3	0.42	0.636	0.1	0.479
0.9	No cutoff	18	0.99	1	1	1	1
0.9	70	20	0.93	1	1	0.99	0.995
0.9	50	20	0.83	1	1	0.95	0.968
0.9	40	30	0.74	1	1	0.94	0.993
0.9	30	40	0.56	1	1	0.79	0.973
0.9	25	60	0.44	1	1	0.68	0.97
0.9	15	120	0.3	0.47	0.657	0.13	0.527

NOTE. Patients with stable disease or clinical response enter second stage (16 weeks run-in and 32 week follow-up after random assignment).

Abbreviations: RDD, randomized discontinuation design; RD, upfront randomization design; TTP, time to disease progression.

*The first number is the power based on TTP as the end point and one-sided .05 level log-rank test; the second number is the power based on progression rate as the outcome end point (binary) and one-sided .05 level χ^2 test of equality of progression rates between two arms.

Table 2. RDD Versus RD: Sensitive Fraction Model

Treatment Effect	Population Fraction of Sensitive Patients for Both RD and Two Stages of RDD (%)	Overall Sample Size	RDD			RD	
			Proportion Going to Second Stage	Power (based on TTP)*	Power (based on proportion)*	Power (based on TTP)*	Power (based on proportion)*
0.1	100	300	0.37	0.39	0.078	0.36	0.063
0.1	80	400	0.36	0.33	0.064	0.3	0.064
0.1	60	500	0.34	0.26	0.06	0.26	0.064
0.1	40	600	0.33	0.15	0.055	0.17	0.049
0.1	30	700	0.32	0.1	0.039	0.09	0.033
0.2	100	100	0.46	0.52	0.056	0.47	0.052
0.2	80	120	0.43	0.43	0.047	0.41	0.058
0.2	60	140	0.4	0.29	0.044	0.28	0.049
0.2	40	160	0.36	0.18	0.023	0.18	0.031
0.2	30	180	0.35	0.12	0.029	0.14	0.025
0.3	100	80	0.56	0.82	0.125	0.76	0.134
0.3	80	100	0.51	0.72	0.161	0.71	0.147
0.3	60	120	0.46	0.53	0.127	0.56	0.112
0.3	40	140	0.4	0.33	0.076	0.35	0.064
0.3	30	160	0.38	0.25	0.07	0.26	0.063
0.5	100	40	0.78	0.97	0.445	0.96	0.443
0.5	80	60	0.68	0.94	0.572	0.95	0.515
0.5	60	80	0.59	0.85	0.508	0.91	0.456
0.5	40	100	0.49	0.57	0.342	0.72	0.305
0.5	30	120	0.45	0.43	0.279	0.58	0.235
0.6	100	30	0.89	0.99	0.71	0.98	0.659
0.6	80	30	0.77	0.89	0.436	0.91	0.421
0.6	60	60	0.65	0.87	0.565	0.95	0.677
0.6	40	60	0.53	0.49	0.223	0.71	0.332
0.6	30	80	0.48	0.38	0.183	0.62	0.296
0.7	100	20	0.96	0.99	0.81	0.99	0.825
0.7	80	20	0.83	0.86	0.47	0.90	0.583
0.7	60	40	0.7	0.82	0.594	0.94	0.808
0.7	40	120	0.57	0.87	0.795	0.99	0.98
0.7	30	160	0.5	0.73	0.671	0.96	0.945
0.8	100	18	0.99	0.99	0.984	0.99	0.981
0.8	80	20	0.86	0.93	0.775	0.96	0.851
0.8	60	20	0.72	0.58	0.338	0.79	0.573
0.8	40	40	0.58	0.47	0.296	0.77	0.688
0.8	30	60	0.51	0.41	0.256	0.76	0.697
0.9	100	18	0.99	0.99	0.99	0.99	0.99
0.9	80	18	0.86	0.93	0.848	0.97	0.905
0.9	60	20	0.72	0.63	0.422	0.82	0.709
0.9	40	30	0.58	0.38	0.202	0.66	0.538
0.9	30	40	0.51	0.32	0.115	0.59	0.509

NOTE. Patients with stable disease or clinical response enter second stage (16 weeks run-in and 32 week follow-up after random assignment).

Abbreviations: RDD, randomized discontinuation design; RD, upfront randomization design.

*The first number is the power based on time-to-disease progression as the end point and one-sided .05 level log-rank test; the second number is the power based on progression rate as the outcome end point (binary) and one-sided .05 level χ^2 test of equality of progression rates between two arms.

goal of finding more specific therapies that are tailored to a particular tumor type as determined by molecular and genetic signatures will continue to play a big role. With the development of targeted therapies, personalized medicine using information and data from a patient, such as genotype, or level of gene expression to stratify disease, select a medication, provide a therapy, will become reality. However, molecular targeted agents pose substantial challenges in contemporary clinical trials and in early drug development. Since most cancers are known to be heterogeneous as demonstrated by recent genomics and proteomics studies,^{6,7} enrichments including molecular enrichment and identification of putative predictive markers are important in screening targeted agents in phase II studies. In the absence of a

reliable assay to identify sensitive patients, enrichment through trial designs, such as the RDD, is an alternative strategy. Our simulation study shows when the majority of patients are sensitive to the target agents, the RDD is less efficient than the RD when time-to-disease progression is the end point. The same conclusion was reached by Freidlin and Simon¹³ when the end point is proportion of patients with disease progression at a fixed time point. However, when the percentage of sensitive patients is small or the sensitivity is determined by tumor growth characteristics, our results show the degree of enrichment by the RDD is sufficient to counter the sample size advantage of the RD. Using time-to-event as the end point and generating data from a Weibull distribution, the results of simulations by Capra¹⁸

show the RD is more efficient than the RDD. Using the same type of end point, but generating data by a tumor growth model and two models of treatment effect (ie, GRC and SF models) proposed by Freidlin and Simon,¹³ our simulation study indicates there are scenarios where the RDD is more efficient than the RD. The different conclusions reached by Capra,¹⁸ Freidlin and Simon¹³ and us, among others show that the results presented in these articles are model dependent. Our simulations also show that using time-to-event end point is more efficient than using binary outcome end point in general for both GRC and SF models—as would be expected since survival analysis utilizes more of the information in the data.

A limitation of choosing time-to-event as the end point, for example, the event defined by 20% tumor increase, is that it is usually not known exactly when a tumor increases by 20%. With rigorous follow-up and state-of-art techniques of monitoring patients (eg, imaging), the estimation of time-to-event will be more accurate. In addition, there are two ways in the RDD to start the clock for determining the end points in general and time to disease progression in particular: (a) at the beginning of the run-in period, and (b) at the time of random assignment (the end of run-in period). Either choice has its limitations. For choice (a), patients in the placebo arm of stage 2 of the RDD benefit from the new agent in the run-in period and dilute the difference between placebo and new agent; for choice (b), patients in both arms on the second stage of the RDD benefit from the treatment of new agent in the run-in period. So the patient populations for RD and RDD (at the second stage) are different unless the carry-over effect or drug resistance generated during the run-in period is very limited. Our simulation results are based on choice (b); and for some slow-growing tumors it may be justifiable. Although our simulation results under the GRC and SF models are based on the assumption that patients with slow growing tumors benefit from the active drug, the simulation procedure can be easily modified if the treatment actually benefits the patients with aggressive tumors.

There are some successful applications of the RDD²⁰ and it also is currently in use in ongoing clinical trials. A number of concerns regarding the RDD, however, have been raised in the literature.^{12,13,21} These include the generalizability of the study results if positive, the carry-over effect or drug resistance generated during the run-in period, and ethical concerns of forcing half of those with stable disease or clinical response to switch to placebo. The last concern may be accept-

able, as pointed out by Rosner et al¹² in that it can be thought of as patients being randomly assigned to a drug holiday versus continuing treatment. Clearly, the implementation of the RDD requires the careful planning.

In conclusion, the RDD in the early drug development is more efficient in some settings when a particular molecular tumor subtype is not easily identifiable. Its application is appropriate only for certain disease settings and for certain targeted anticancer agents. In the design stage of a trial, the limitations outlined above should be addressed before its application.

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

Although all authors completed the disclosure declaration, the following author(s) indicated a financial or other interest that is relevant to the subject matter under consideration in this article. Certain relationships marked with a "U" are those for which no compensation was received; those relationships marked with a "C" were compensated. For a detailed description of the disclosure categories, or for more information about ASCO's conflict of interest policy, please refer to the Author Disclosure Declaration and the Disclosures of Potential Conflicts of Interest section in Information for Contributors.

Employment or Leadership Position: None **Consultant or Advisory Role:** None **Stock Ownership:** None **Honoraria:** Afshin Dowlati, Genentech, Eli Lilly **Research Funding:** Afshin Dowlati, Celgene, Glaxo-Smith-Kline, Genentech **Expert Testimony:** None **Other Remuneration:** None

AUTHOR CONTRIBUTIONS

Conception and design: Pingfu Fu, Afshin Dowlati, Mark Schluchter **Financial support:** Pingfu Fu, Afshin Dowlati, Mark Schluchter **Administrative support:** Pingfu Fu **Provision of study materials or patients:** Pingfu Fu, Afshin Dowlati **Collection and assembly of data:** Pingfu Fu, Afshin Dowlati, Mark Schluchter **Data analysis and interpretation:** Pingfu Fu, Afshin Dowlati, Mark Schluchter **Manuscript writing:** Pingfu Fu, Afshin Dowlati, Mark Schluchter **Final approval of manuscript:** Pingfu Fu, Afshin Dowlati, Mark Schluchter

REFERENCES

- Goyeneche AA, Carón RW, Telleria CM: Mifepristone inhibits ovarian cancer cell growth in vitro and in vivo. *Clinical Cancer Res* 13:3370-3379, 2007
- Fox E, Curt GA, Balis FM: Clinical trial design for targeted therapy. *Oncologist* 7:401-409, 2002
- 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* 99:1455-1461, 2007
- 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* 22:4442-4445, 2004
- Booth CM, Calvert AH, Giaccone G, 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). *Eur J Cancer* 44:25-29, 2008
- Shi H, Guo J, Duff DJ, et al: Discovery of novel epigenetic markers in non-Hodgkin's lymphoma. *Carcinogenesis* 28:60-70, 2007
- Wooster R, Weber BL: Breast and ovarian cancer. *N Engl J Med* 348:2339-2347, 2003
- Geyer CE, Forster J, Lindquist D, et al: Lapatinib plus capecitabine for HER2-positive advanced breast cancer. *N Engl J Med* 355:2733-2743, 2006
- Simon R: Optimal two-stage designs for phase II clinical trials. *Control Clin Trials* 10:1-10, 1989
- 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* 23:7199-7206, 2005
- Temple RJ: Special study designs: Early escape, enrichment, studies in non-responders. *Commun Statist-Theory Meth* 23:499-531, 1994
- Rosner GL, Stadler W, Ratain MJ: Randomized discontinuation design: Application to cytostatic antineoplastic agents. *J Clin Oncol* 20:4478-4484, 2002
- Freidlin B, Simon R: Evaluation of randomized discontinuation design. *J Clin Oncol* 23:5094-5098, 2005
- Dent S, Zee B, Dancey J, et al: Application of a new multinomial phase II stopping rule using response and early progression. *J Clin Oncol* 19:785-791, 2001
- RH El-Maraghi, EA Eisenhauer: Review of phase II: Trial designs used in studies of molecular targeted agents: Outcomes and predictors of success in phase III. *J Clin Oncol* 26:1346-1354, 2008
- Ratain MJ, Humphrey RW, Gordon GB, et al: Recommended changes to oncology clinical trial design: Revolution or evolution? *Eur J Cancer* 44:8-11, 2008

17. Kopec JA, Abrahamowicz M, Esdaile JM: Randomized discontinuation trials: Utility and efficiency. *J Clin Epidemiol* 46:959-971, 1993

18. Capra WB: Comparing the power of the discontinuation design to that of the classic randomized design on time-to-event endpoints. *Control Clin*

Trials 25:168-177, 2004

19. Jiang Y, Pjesivac-Grbovic J, Cantrell C, et al: A multiscale model for avascular tumor growth. *Biophys J* 89:3884-3894, 2005

20. 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* 24:2505-2512, 2006

21. Leber PD, Davis CS: Threats to the validity of clinical trials employing enrichment strategies for sample selection. *Control Clin Trials* 19:178-187, 1997