VOLUME 23 · NUMBER 22 · AUGUST 1 2005

JOURNAL OF CLINICAL ONCOLOGY ORIGINAL REPORT

Evaluation of Randomized Discontinuation Design

Boris Freidlin and Richard Simon

From the Biometric Research Branch, National Cancer Institute, Bethesda, MD.

Submitted February 18, 2004; accepted August 16, 2004.

Authors' disclosures of potential conflicts of interest are found at the end of this article.

Address reprint requests to Boris Freidlin, PhD, Biometric Research Branch, Division of Cancer Treatment and Diagnosis, National Cancer Institute, 6130 Executive Blvd, EPN 8122, MSC 7434, Bethesda, MD 20892-7434; e-mail: freidlinb@ctep.nci.nih.gov.

0732-183X/05/2322-5094/$20.00

DOI: 10.1200/JCO.2005.02.520

ABSTRACT

Purpose
Single-arm phase II trials may not be appropriate for testing cytostatic agents. We evaluate two kinds of randomized designs for the early development of target-based cytostatic agents.

Methods
We compared power of the randomized discontinuation and upfront randomization designs under two models for the treatment effect of targeted cytostatic agents.

Results
The randomized discontinuation design is not as efficient as upfront randomization if treatment has a fixed effect on tumor growth rate or if treatment benefit is restricted to slower-growing tumors. On the other hand, the randomized discontinuation design can be advantageous under a model where only a subset of patients, those expressing the molecular target, is sensitive to the agent. To achieve efficiency, the design parameters must be carefully structured to provide adequate enrichment of the randomly assigned patients.

Conclusion
With careful planning, the randomized discontinuation designs can be useful in some settings in the early development of targeted agents where a reliable assay to select patients expressing the target is not available.

J Clin Oncol 23:5094-5098. © 2005 by American Society of Clinical Oncology

INTRODUCTION

Recent advances in biotechnology have resulted in a shift toward molecularly targeted anticancer agents.[1-3] In the early stages of development, reliable assays to identify the sensitive patients that express the target are often not available.[4,5] This complicates evaluation of targeted agents as a result of the dilution of the treatment effect by the presence of the patients who do not benefit from the agent.[6] Furthermore, some of these agents are thought to be cytostatic and are only expected to inhibit tumor growth without shrinking existing tumors.[7,8] Traditionally, clinical development of cytotoxic agents involved a single arm phase II evaluation of the response rate. In most cases, this approach is no longer adequate for the development of cytostatic agents.[9] Thus there is a need for development of new statistical approaches for the early development of targeted cytostatic agents.[10,11]

In the development of cytostatic agents, where reliable historical control is not available, one approach would be to randomly assign patients between the new agent or placebo.[10,12,13] An alternative approach for testing cytostatic agents in heterogeneous populations is the randomized discontinuation design (RDD; Rosner et al[14]): initially all patients are given the drug. After some fixed period (eg, 4 months), patients are evaluated: patients who respond to treatment continue on the drug, those whose disease progresses are taken off study, and patients with stable disease are randomly assigned between continued administration of the drug or observation/placebo typically for another fixed period (eg, 4 months). This

design uses an enrichment strategy[15] to focus on the patients who are more likely to benefit from the drug.

The goal of phase II development is to screen out inactive agents and identify the most promising ones for a definitive testing in randomized phase III trials.[16] In this article, we compare RDD and upfront randomization approaches with phase II testing of targeted cytostatic agents. Because cytostatic agents are expected to slow progression and are not expected to result in a significant number of responses, the evaluation will be with respect to progression rates (in phase II setting, progression rates at some fixed time point are generally more reliable than time to progression). Possible applications of the enrichment approach to the response end point are addressed in the discussion. We assume that both upfront randomization design and RDD are acceptable and review the relative merits of the two approaches. The practical implementation issues are addressed in the discussion.

METHODS

Similar to Rosner et al,[14] we use an exponential tumor growth model: tumor diameter for patient i at time t is

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

where k denotes the treatment effect ($k = 0$ for placebo patients), λ_i denotes ith patient tumor growth rate, and t denotes time (in weeks) measured from the baseline. As will be discussed, the treatment effect will differ among the patients. The tumor growth rates λ are assumed to have a log normal distribution across the population. The same tumor growth population parameters as in Rosner et al were used: the mean and variance of the growth rates were set to have approximately 70% of patients progress by 16 weeks, with an average of 32% increase in tumor diameter by 16 weeks. Progression was defined as 20% increase in tumor diameter from the baseline. We considered two models for the effect of treatment. The first model, the growth rate cutoff model (GRC), presumes that only patients with tumor growth rate, λ_i, below a certain cutoff value benefit from the treatment. The cutoff value, c_0, was defined in terms of the percentage increase in diameter of an untreated tumor over 16 weeks; in other words, tumors that would have grown (without treatment) by less than c_0% at 16 weeks benefit from the treatment, whereas tumors that would have grown (without treatment) by more than c_0% at 16 weeks are unaffected by the treatment. The second model, the sensitive fraction model (SF), presumes that only a fraction, p_r, of the patient population is sensitive to the treatment and that treatment sensitivity is independent of growth rate. The GRC model corresponds to a scenario, assumed in Rosner et al, where because of the tumor growth rate heterogeneity, a portion of the study population will have disease that is too aggressive to benefit from the treatment. The SF model corresponds to the targeted therapy setting where there is no reliable assay to select patients expressing the target. For the SF model, we assumed that both sensitive and nonsensitive populations have the same distribution of the tumor growth rates

We considered a generalization of RDD where all patients receive the drug of interest during the first stage (we used 16 weeks for the calculation). Patients with tumors that increased by less than the given percentage, c_1%, at the end of the first stage are randomly assigned to continuing or discontinuing the therapy (second stage). At the end of the second stage (we used 32 weeks from the baseline for the calculations), progression rates in the two randomized arms are compared (progression was defined as 20% increase in tumor diameter). The accrual to this trial continues until the required number of patients are randomly assigned. RDD was compared with the upfront randomization design, where all patients are randomly assigned at baseline and progression rates at 32 weeks are compared.

To compare the designs, we first derived analytic expressions for the progression rates on the control and the experimental arms under the GRC and SF models. The power of the one-sided .05 level test of equality of progression rates was then calculated. For each setting, the overall sample size was selected to provide 85% power for the test with the highest power. We considered a range of possible treatment effects k corresponding to the cytostatic mechanism of action; the results are presented for $k = 0.3$, corresponding to a 30% reduction in tumor growth rate; $k = 0.5$ corresponding to a 50% reduction in tumor growth rate and $k = 0.7$ corresponding to a 70% reduction in tumor growth. For the SF model, powers were tabulated for a range of fraction of drug-sensitive patients: $p_r = 100\%, 70\%, 50\%$, and 30%.

RESULTS

Table 1 presents results for the GRC model, with $c_1 = 20\%$; in other words, all patients with no more than 20% increase in tumor after 16 weeks go to the second stage. Column 2 shows the growth rate cutoff for drug activity. No cutoff means that all tumors are sensitive to treatment as in Rosner at al[14]; 13% cutoff means that only tumors that would grow over 16 weeks by less than 13% without treatment are sensitive. "Proportion going to second stage" (column 4) is the expected proportion of the overall sample size that continues to the second stage. It reflects the distribution of growth rate for untreated tumors, the size of the treatment effect, and the growth rate cutoff for classifying which tumors are sensitive. The proportion of patients going to the second stage generally decreases as the growth cutoff for drug activity decreases. Eventually, however, the only patients whose tumors are sensitive to the treatment are patients with such a low growth rate that they would be included in the second stage even in the absence of treatment. Reducing the growth rate cutoff beyond that point has no further effect on the proportion going to the second stage.

Table 1 shows that in most cases under the GRC model, the upfront randomization design is superior to RDD. The powers for the upfront randomization design and RDD are determined by the number of patients randomly assigned and the proportion of the randomly assigned patients who are sensitive to treatment. The RDD generally loses a large

Table 1. Randomized Discontinuation Design Versus Upfront Randomization: Growth Rate Cutoff Model

Treatment Effect	Growth Cutoff for Drug Activity (%)	Overall Sample Size (No. of patients)	Randomized Discontinuation Design		Upfront Randomization (power)
			Proportion Going to Second Stage	Power	
.3	No cutoff	390	.47	.53	.85
.3	25	390	.39	.56	.85
.3	20	390	.30	.63	.85
.3	15	390	.30	.63	.85
.3	13	600	.30	.48	.85
.5	No cutoff	92	.63	.59	.85
.5	30	92	.47	.67	.85
.5	25	92	.39	.76	.85
.5	20	63	.30	.72	.85
.5	15	262	.30	.38	.85
.7	No cutoff	26	.84	.63	.85
.7	40	26	.59	.82	.85
.7	30	31	.47	.85	.81
.7	25	48	.39	.85	.83
.7	20	92	.30	.80	.85

number of randomly assigned patients because their tumors progress too rapidly during the first stage. This is not sufficiently counterbalanced by enrichment of patients with sensitive tumors that have large enough untreated growth rates to progress during the second stage. For a narrow range of cutoff values in the presence of a strong treatment effect, RDD has a marginally higher power. Because the exact cutoff value is unlikely to be known in practice and the power advantage is marginal, RDD use under GRC model is not optimal.

Tables 2 and 3 present results for the SF model. The SF model represents a situation often occurring in early development of targeted agents where only a subset of the patients is expected to be sensitive to the drug, but a reliable assay to identify the sensitive patients is not available. Under this model, sensitivity to the treatment is based on molecular characteristic of the disease and is independent of the tumor growth rate. In contrast, under the GRC model, sensitivity to the treatment is based on the tumor growth rate. The population proportion of sensitive patients is shown in column 2 and ranges from 30% to 100%. We considered two versions of the RDD design. In the first (Table 2), all patients with no more than 20% increase in tumor diameter by the end of the first stage are randomly assigned (ie, $c_1 = 20\%$; as in Rosner et al[14]). In the second version (Table 3), patients with no more than 10% increase in tumor diameter (ie, $c_1 = 10\%$) are randomly assigned. The second version uses a more stringent eligibility for the second stage, thus allowing a more efficient enrichment of the randomized sample. Column 5 shows the expected

Table 2. Randomized Discontinuation Design Versus Upfront Randomization: Sensitive Fraction Model—Version 1 (less than 20% growth required for stage 2)

Treatment Effect	Population: Fraction of Sensitive Patients (%)	Overall Sample Size (No. of patients)	Randomized Discontinuation Design			Upfront Randomization (power)
			Proportion Going to Second Stage	Second Stage: Fraction of Sensitive Patients (%)	Power	
.3	100	390	.47	100	.53	.85
.3	70	720	.42	78	.53	.85
.3	50	1,340	.38	61	.55	.85
.3	30	3,400	.35	40	.56	.85
.5	100	92	.63	100	.59	.85
.5	70	162	.53	83	.61	.85
.5	50	280	.46	68	.62	.85
.5	30	680	.40	48	.63	.85
.7	100	26	.83	100	.63	.85
.7	70	48	.67	87	.68	.85
.7	50	82	.57	74	.70	.85
.7	30	188	.46	55	.71	.85

Table 3. Randomized Discontinuation Design Versus Upfront Randomization: Sensitive Fraction Model—Version 2 (less than 10% growth required for stage 2)

Treatment Effect	Population: Fraction of Sensitive Patients (%)	Overall Sample Size (No. of patients)	Randomized Discontinuation Design			Upfront Randomization (power)
			Proportion Going to Second Stage	Second Stage: Fraction of Sensitive Patients (%)	Power	
.3	100	300	.19	100	.85	.76
.3	70	480	.16	83	.85	.70
.3	50	770	.14	67	.85	.65
.3	30	1,650	.12	47	.85	.59
.5	100	78	.32	100	.85	.80
.5	70	120	.25	89	.85	.74
.5	50	186	.20	78	.85	.70
.5	30	370	.16	60	.85	.63
.7	100	25	.56	100	.85	.84
.7	70	38	.42	94	.85	.77
.7	50	58	.33	86	.85	.72
.7	30	110	.23	73	.85	.65

proportion of the second-stage randomly assigned patients who are sensitive. Comparison of column 5 and column 2 shows the degree of enrichment achieved. Tables 2 and 3 demonstrate that with proper enrichment, RDD has higher power than the upfront randomization design. The degree of enrichment is the key to efficiency of the RDD design. The first version of RDD ($c_1 = 20\%$) does not provide sufficient enrichment for the second stage and has lower power than the corresponding upfront randomization design. On the other hand, the use of a more stringent $c_1 = 10\%$ in the second version of RDD results in better enrichment and a consistently higher power relative to the corresponding upfront randomization design.

DISCUSSION

Enrichment can provide an efficient way for early screening of targeted cytostatic agents in the setting where only a fraction of the patient population is expected to be sensitive to the agent but no reliable assay to identify the sensitive patients is available. Successful application of the RDD requires careful planning to ensure that the second-stage population is sufficiently enriched. If all patients are sensitive or if sensitivity is determined by the growth rate, then our Table 1 shows that the RDD can be considerably less efficient than the upfront randomization design. A similar conclusion was reached by Capra.[17] This power disadvantage cannot be adequately corrected by a more aggressive enrichment strategy: using a more stringent $c_1 = 10\%$ for the GRC model did not result in a consistent improvement over upfront randomization (results not shown). The above conclusions still hold for the versions of RDD that include early stopping rules for promising/disappointing progression rates after the first stage, because similar rules can be incorporated in the upfront randomization design.

In some settings, upfront randomization to placebo is not feasible, and alternative designs such as RDD may be necessary.[7] However, the claims that the upfront randomization to placebo is unfeasible should be carefully examined. For example, it has been argued that upfront randomization to placebo in metastatic renal cancer is not practical.[18] However, Yang et al[19] successfully conducted a placebo-controlled, randomized phase II trial of bevacizumab in metastatic renal cancer. Upfront randomization may be made more attractive to patients by including a cross-over provision where control arm patients are crossed over to the new agent either at the time of progression or after a fixed number cycles. Another approach to make the upfront randomization acceptable to patients is to use an active control, for example, to randomly assign between a standard agent and standard agent plus the new agent. In any case, different investigators may come to different conclusions about the feasibility of the upfront randomization to placebo in a given setting.

Although carefully designed RDD can increase efficiency in some settings, a number of concerns with its interpretation need to be addressed.[20,7,10,21] If the study is positive, it may be difficult to generalize the study results and design a follow-up phase III trial (ie, what should be its target population). A randomized phase III trial might not even be practical at that point; for example, if an RDD comparing a new agent to placebo in metastatic renal cancer is positive, a follow-up placebo-controlled phase III trial would probably be unfeasible. On the other hand, a negative study might be a result of a carry-over effect or development of drug resistance during the first stage. Moreover, the difference between 4 versus 8 months of treatment may

not be representative of the full beneficial effect of the agent. Thus the application of the RDD should be limited to the situations where these concerns are thoroughly addressed (eg, if there is no biologic basis for the carry-over and drug resistance).

The potential value of the enrichment approach may not be limited to cytostatic agents. An enrichment strategy may be adapted to targeted agents that are expected to shrink tumors. If a less stringent nonstandard response criterion is applied, a control arm may be needed to control for spurious responses resulting from the measurement error. A modified RDD may be useful in this setting.

It should be noted that in cases where the fraction of sensitive patients is low, the required sample size to detect a moderate treatment effect may become prohibitively large for a phase II investigation. For example, overall sample size of 1,650 patients would be required to detect a 30% decrease in tumor growth rate if only 30% of patients are sensitive to the treatment. This, however, applies equally to the RDD and the upfront randomization design.[5,22]

In conclusion, the RDD can be useful in the early development of targeted agents where a reliable assay to identify the sensitive patients is not available. Its application should be carefully structured to provide sufficient enrichment to the randomly assigned patients, limited to the settings where tumor's natural history and biology allow for a meaningful interpretation of study results, and properly incorporated in the overall agent development process, including the plan for subsequent definitive phase III testing.

Acknowledgment

We thank the referee for helpful suggestions.

Authors' Disclosures of Potential Conflicts of Interest

The authors indicated no potential conflicts of interest.

REFERENCES

1. Balis FM: Evolution of anticancer drug discovery and the role of cell based screening. J Natl Cancer Inst 94:78-79, 2002

2. Schilsky RL: End points in cancer clinical trials and the drug approval process. Clin Cancer Res 8:935-938, 2002

3. Rothenberg ML, Carbone DP, Johnson DH: Improving the evaluation of new cancer treatments: Challenges and opportunities. Nat Rev Cancer 3:303-309, 2003

4. Johnson DH: Targeted therapy in non-small cell lung cancer: Myth or reality. Lung Cancer 41:s3–s8, 2003

5. Dancey JE, Freidlin B: Targeting epidermal growth factor receptor: Are we missing the mark? Lancet 362:62-64, 2003

6. Betensky RA, Louis DN, Cairncross JG: Influence of unrecognized molecular heterogeneity on randomized clinical trials. J Clin Oncol 20:2495-2499, 2002

7. Stadler WM, Ratain MJ: Development of target-based antineoplastic agents. Invest New Drugs 18:7-16, 2000

8. Fox E, Curt GA, Balis FM: Clinical trial design for target-based therapy. Oncologist 7:401-409, 2002

9. Eisenhauer EA: Phase I and II trials of novel anti-cancer agents: Endpoints, efficacy and existentialism. Ann Oncol 9:1047-1052, 1998

10. Korn EL, Arbuck SG, Pluda JM, et al: Clinical trial designs for cytostatic agents: Are new approaches needed? J Clin Oncol 19:265-272, 2001

11. Millar AW, Lynch KP: Rethinking clinical trials for cytostatic drugs. Nat Rev Cancer 3:540-545, 2003

12. Simon R, Wittes RE, Ellenberg SS: Randomized phase II clinical trials. Cancer Treat Rep 69:1375-1381, 1985

13. Buyse M: Randomized designs for early trials of new cancer treatments: An overview. Drug Inf J 34:387-396, 2000

14. Rosner GL, Stadler W, Ratain MJ: Randomized discontinuation design: Application to cytostatic antineoplastic agents. J Clin Oncol 20:4478-4484, 2002

15. Temple RJ: Special study designs: Early escape, enrichment, studies in non-responders. Commun Statist-Theory Meth 23:499-531, 1994

16. Nottage M, Siu LL: Principles of clinical trial design. J Clin Oncol 20:42s-46s, 2002 (suppl)

17. Capra WB: Comparing the power of the randomized discontinuation design to that of the classic randomized design on time-to-event endpoints. Control Clin Trials 25:168-177, 2004

18. Ratain MJ, Stadler WM: Clinical trials designs for cytostatic agents. J Clin Oncol 19:3154, 2001

19. Yang JC, Haworth L, Sherry RM: A randomized trial of bevacizumab, an anti–vascular endothelial growth factor antibody, for metastatic renal cancer. N Engl J Med 349:427-434, 2003

20. Leber PD, Davis CS: Threats to validity of clinical trials employing enrichment strategies for sample selection. Control Clin Trials 19:178-187, 1998

21. Korn EL, Arbuck SG, Pluda JM, et al: Clinical trials designs for cytostatic agents. J Clin Oncol 19:3154-3155, 2001

22. Freidlin B, Korn EL: A testing procedure for survival data with few responders. Stat Med 21:65-78, 2002