Evaluation of Randomized Discontinuation Design

Boris Freidlin and Richard Simon

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.

INTRODUCTION

Recent advances in biotechnology have resulted in a shift toward molecularly targeted anticancer agents. In the early stages of development, reliable assays to identify the sensitive patients that express the target are often not available. 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. Furthermore, some of these agents are thought to be cytostatic and are only expected to inhibit tumor growth without shrinking existing tumors. 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. Thus there is a need for development of new statistical approaches for the early development of targeted cytostatic agents.

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. An alternative approach for testing cytostatic agents in heterogeneous populations is the randomized discontinuation design (RDD; Rosner et al): 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\textsuperscript{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.\textsuperscript{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,\textsuperscript{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), \( \lambda_i \) denotes \( i \)th 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 \( \lambda \) 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 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 are assumed to have a log normal distribution across the population.

Table 1 shows that in most cases under 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 et al\textsuperscript{15}; 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
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\textsuperscript{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
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 (i.e., 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

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