Randomized Discontinuation Design: Application to Cytostatic Antineoplastic Agents

By Gary L. Rosner, Walter Stadler, and Mark J. Ratain

Purpose: Propose a phase II study design to evaluate the activity of a putative cytostatic agent, acknowledging heterogeneity of tumor growth rates in the population of patients.

Methods: In the setting of renal cell carcinoma, some patients’ tumors will grow slowly naturally. An appropriate design has to distinguish antiproliferative activity attributable to the novel agent from indolent disease. We propose a randomized discontinuation design that initially treats all patients with the study agent (stage 1) and then randomizes in a double-blind fashion to continuing therapy or placebo only those patients whose disease is stable (stage 2). This design allows the investigators to determine if apparent slow tumor growth is attributable to the drug or to selection of patients with naturally slow-growing tumors.

Results: By selecting a more homogeneous population, the randomized portion of the study requires fewer patients than would a study randomizing all patients at entry. The design also avoids potential confounding because of heterogeneous tumor growth. Because the two randomly assigned treatment groups each comprise patients with apparently slow growing tumors, any difference between the groups in disease progression after randomization is more likely a result of the study drug and less likely a result of imbalance with respect to tumor growth rates. Stopping rules during the initial open-label stage and the subsequent randomized trial stage allow one to reduce the overall sample size. Expected average tumor growth rate is an important consideration when deciding the duration of follow-up for the two stages.

Conclusion: The randomized discontinuation design is a feasible alternative phase II study design for determining activity of possibly cytostatic anticancer agents.


In recent years, new avenues have opened up for developing molecularly targeted anticancer agents. Some of these newer agents, such as putative angiogenesis inhibitors, may not act by killing tumor cells directly and, thereby, shrinking the size of the tumor. Instead, they may slow or even arrest tumor growth, thus inducing a state of stable disease. Inducing a stable tumor is a clinically useful outcome, especially if it is prolonged and associated with minimal drug-related toxicity, as has been predicted for antiangiogenic agents.

Historically, anticancer drug development has rested on the premise that agents must be cytotoxic to be effective. By equating efficacy with cytotoxicity, traditional drug development, progressing from phase I through phase III clinical studies, has sought the highest effective dose that does not induce intolerable levels of toxicity. Thus, phase I studies seek the maximally tolerated dose of a new agent or combination, and phase II studies screen new therapies for antitumor activity. Publications discussing phase II design in oncology assume the agent under study to be cytotoxic and the study end point to be tumor shrinkage as demonstrated by physical or radiologic measurements. In this paradigm, drugs felt likely to result in tumor shrinkage in an equal or greater percentage of patients than expected with the currently available agents are labeled active.

Newer agents, especially those with specific molecular targets, however, may not be cytotoxic and may, therefore, require new study designs to address the questions of interest. For example, some agents may slow tumor growth or induce stasis, and study design paradigms must shift. Recently, Korn et al discussed several different proposed trial designs for evaluating newer forms of anticancer therapy. In this article, we discuss a phase II design for screening possibly cytostatic agents via randomized discontinuation as an example of such a shift away from traditional designs.

The randomized discontinuation design is an example of an enrichment design. The intent is to select a subset of enrolled patients who are relatively homogeneous with respect to important prognostic factors and randomize only these patients. Examples of such enrichment strategies abound, as in prevention trials in which all patients receive either the study drug or placebo for a short run-in period, but only those patients who adhere to treatment or who do not experience excess toxicity proceed to randomization. By randomizing only those patients who appear most likely to follow their prescribed treatment plan, the clinical trial attempts to avoid dilution of statistical power because of lack of adherence to prescribed treatment plans.

Randomized discontinuation seeks to produce a more homogeneous group of patients, consisting of those who are more likely to show a treatment benefit if it exists. When evaluating a drug for cytostatic activity, such enrichment is helpful because there is great heterogeneity of tumor growth rates across a group of cancer patients. This heterogeneity exists among primary tumors and metastatic lesions. For example, median survival in metastatic renal cell carcinoma is approximately 10 months, but...
5% of patients survive for 5 or more years. A patient’s disease may appear stable with or without the drug if the patient has a particularly slow-growing tumor. In addition, if there is great heterogeneity in growth rates, the evaluation of an agent as active or inactive may be made by chance, if patients with stable disease are considered as responders. A study population with predominantly slow-growing tumors might suggest a treatment benefit, whereas a population with faster-growing tumors would suggest little or no cytostatic activity. Observation of apparently delayed times to progression in this uncontrolled study might reflect the effect of the therapy, the singular characteristics of the enrolled patients, or a bit of each factor. One would need some way to estimate the natural history of the disease for these patients in the absence of the drug under study and adjust the analysis accordingly. Prognostic indicators can only account for a portion of this variability, making measurement of a stable disease rate difficult to interpret.

The easiest way to account for the natural history of the disease in the study design is by including some sort of control in the study design. The use of historical controls has been adequately discussed and criticized in the literature. Cross-over studies, another possibility, demand an unchanging disease state (in the absence of effective treatment) during the study period to allow assessment of treatment effects regardless of treatment sequence. This assumption would not hold for a majority of cancer patients, especially in the metastatic setting. One also needs to be certain the agent’s effects do not carry over into subsequent treatment periods when carrying out a cross-over study. Changes in disease state and carryover effects can reduce the advantage of a cross-over study over a parallel-group study.

A more standard and preferable design to control for the natural history of the disease is a randomized clinical trial comparing a novel agent to placebo. Problems will arise in a randomized comparative trial, however, if there is substantial heterogeneity in tumor growth rates across the population. In particular, heterogeneity may lead to a reduction in statistical power (i.e., the probability of finding a statistically significant difference), because some fraction of the study population benefits from the drug while the remainder does not. One would need to increase the sample size to overcome the diluted statistical power. For example, with median time to progression as the primary end point in metastatic renal cell carcinoma, one would need to randomize 50 patients to each treatment to have 95% probability to detect a doubling from 2.3 months to 4.6 months in median time to progression. (These calculations assume that 100 patients enter the study annually, the study lasts 20 months from the time the first patient enters the study, and a log-rank test, two-sided, with a 0.05 level of significance.) If, however, there is great heterogeneity in tumor growth rates around the averages implied by the median times to progression, the power will drop. Suppose that the agent in question extends the time to progression for 30% of the patients receiving it, but it does not affect the time to progression for 70% of the patients who enter the study because of their rapidly progressive disease. This level of heterogeneity will cause the power of the randomized trial to drop to 18%. One would have to increase the sample size to around 330 patients to regain a 95% probability of finding the 2.3-month increase in median survival statistically significant.

We discuss randomized discontinuation as a design in which all patients receive the drug of interest and randomization proceeds only for the patients with stable disease at a specified time point and who tolerate the drug. Random treatment assignments consist of continuing or discontinuing therapy. The design seeks to identify activity, defined as stabilizing disease or inhibiting growth, in the experimental as opposed to the control group. The study is explanatory rather than pragmatic, because we are looking for evidence of biologic activity and are not yet evaluating the clinical effectiveness of the treatment strategy. The next section presents the details of the randomized discontinuation design as it could be applied to the phase II analysis of cytostatic agents in cancer. We also illustrate the statistical properties of a randomized discontinuation design in this setting based on computer simulations of clinical trials. As a specific example, we use an ongoing study of carboxyaminoimidazole (CAI) in metastatic renal cancer sponsored by the Cancer and Leukemia Group B (CALGB).

**METHODS**

The CALGB has chosen to evaluate the activity of CAI (NSC 609974) in renal cell cancer using a randomized discontinuation design. CAI is a novel oral agent whose mechanism of action is postulated to be inhibition of calcium-mediated signal transduction. It is an inhibitor of receptor-gated calcium channels, resulting in the inhibition of phospholipase C-γ and phospholipase A2 phosphorylation. Blockade of this pathway prevents the release of arachidonic acid that has been correlated with inhibition of malignant proliferation and metastasis. In vitro, this has been shown to lead to inhibition of tumor cell mobility and invasion. CAI also has been shown to inhibit angiogenesis by inhibiting basic fibroblast growth factor stimulation of endothelial cell proliferation, adhesion, motility, and tube formation. In mouse models, CAI inhibited subcutaneous growth and pulmonary metastases of a variety of human cancer cell lines. We wrote a computer program to determine power and sample size relationships by simulating clinical trials designed with randomized discontinuation under different study settings. In the simulations, tumor growth was exponential and outcome consisted of response categories. One could also model tumor growth as Gompertzian or some other model one feels most appropriate for the disease under study. Tumor growth rates in the simulations were distributed across the population according to a lognormal distribution (i.e., the natural logarithm of the growth rates is normally distributed in the population). Further, the mean of the growth rates was set (initially) to produce 13% tumor growth rate over the first 16 weeks of follow-up (assuming no CAI effect). The mean and variance of the growth rate’s lognormal distribution led to an average of 70% of patients progressing by 16 weeks, whereas the remaining 30% of patients had stable disease by 16 weeks, by the response evaluation criteria in solid tumors (RECIST) criteria. (None of the simulated patient tumors responded spontaneously.) These assumed values are based on retrospective observations of stable disease rates in phase II trials with classic cytotoxic agents found to be ineffective in metastatic renal cell cancer. For each simulated patient, the computer program randomly generated a tumor growth rate and determined the relative tumor size after 16 weeks of therapy. The program then classified the patient’s tumor as having progressed if the diameter increased by more than 20% of baseline by the time of each evaluation. If the diameter was between 70% and 120% of its baseline diameter, the tumor was considered stable. Greater than 99% reduction in tumor diameter corresponded to a complete response, and a partial response corresponded to reducing tumor diameter between 30% and 99%. In the model, no effect of CAI on tumor growth was represented by a zero value for the CAI effect parameter, with positive values corresponding to growth inhibition. A value of 1 meant that the agent inhibited all further tumor growth. If CAI were associated with tumor regression, then the CAI effect would be larger than 1 in the model. As in the CALGB study, all simulated patients received CAI during the first stage (generally 16 weeks or four 4-week cycles). Patients with stable disease at the end of four cycles of CAI went on to randomization and the second stage of the study. Accrual to the entire study (i.e., the first stage) stopped once the required number of randomized patients was reached in each simulation. Figure 1 shows a schematic drawing of the study design. If a patient should progress after randomization to placebo in the actual
CALGB study, that patient may receive CAI once again. This step was added to facilitate patient compliance. Because the cross-over occurs after the placebo failure, the simulations did not include the possible return to CAI. Also, patients who experience a complete or partial response in the CALGB study while on CAI would not go on to the randomization stage but would continue receiving CAI until disease progression or toxicity.

For each configuration of growth rate, duration of study stage (ie, follow-up before and after randomization), and sample size, we ran 5,000 simulations. The simulated trials included two main analyses: the overall risk of progression during the initial nonrandomized stage of the study via interim analyses. If the risk of progression was greater than expected (ie, the estimated risk that a patient’s tumor will progress by 16 weeks was 85% or greater with 90% certainty, given the simulated trial’s data thus far), then the study stopped and concluded that the agent was not likely active. If, on the other hand, the evolving estimate of the risk of progressive disease was less than expected (ie, the risk of progressive disease was 40% or less, with 90% certainty, given the data at the time of the interim analysis), then the study stopped, as well. In this latter case, the conclusion would be that the agent is active.

The randomized comparison of the agent to placebo in the simulations consisted of a simple $\chi^2$ test of the equality of the respective proportions of patients with progressive disease. One might want instead to compare the CAI and placebo patients with respect to proportions of patients with aggressive disease, stable disease, and objective response after four cycles of therapy following randomization (or eight cycles of therapy from study entry) using a test for trends in proportions.$^{23}$ We report here only the results comparing the treatment-specific risks of progression. The simulated trials also included an interim analysis halfway through the randomized trial. The O’Brien-Fleming stopping boundary$^{24}$ for an overall two-sided 0.05-level test determined early termination boundaries for the randomized stage of the study in the simulations. One could implement other interim analysis schedules and different early stopping guidelines, of course, particularly including stopping because of a lack of a treatment effect during the randomized part of the study. In the simulations, analyses of the randomized portion of the study occurred as soon as accrual to the randomized part of the study allowed, even if patients were still entering the open-label stage.

RESULTS

Table 1 shows the average proportion of responses in the two stages of the trial for a few hypothesized values of the CAI effect on tumor growth. We also show the number of patients one would need to enroll, on average, in the first stage to randomize 100 or 200 patients in the second stage. Adding interim analyses based on the fraction progressing during the open-label part of the study leads to some savings in resources as the underlying effect of CAI increases. If, in fact, the agent is 100% cytostatic (ie, CAI effect = 1.0 in the Table), the trial stops after enrolling just 25 patients. Table 1 also shows that a trial with CAI effect equal to 0.55, which corresponds to increasing the percentage of patients with stable disease at 16 weeks from 30% to 70%, will be found significant 92% of the time with 100 patients in the randomized portion of the study. This treatment effect also leads to study termination during the open-label first stage 46% of the time, concluding a treatment effect with an average total enrollment of 105 patients. Smaller treatment effects would require larger samples to have high probability of leading to conclusions of treatment activity in the clinical trial.

In Table 1 for the second stage, one notices that the proportion of placebo patients with stable disease in the 16-week period after randomization drifts slightly lower as the drug’s activity increases—a regression to the mean effect. As the drug becomes more active, more patients with aggressive disease enter the randomized stage. Once active treatment stops and placebo begins, this higher percentage of patients with fast-growing tumors becomes evident in the placebo group as an increasing proportion with progressive disease as one goes down the rows of Table 1.

Changing the stopping criteria applied to data from the open-label part of the study will change the study characteristics (not shown). For example, one may feel fairly confident that one will see approximately 70% of patients progressing in the

---

Table 1. Some Characteristics of the Design

<table>
<thead>
<tr>
<th>CAI Effect</th>
<th>CR (%)</th>
<th>PR (%)</th>
<th>SD (%)</th>
<th>PD (%)</th>
<th>CR (%)</th>
<th>PR (%)</th>
<th>SD (%)</th>
<th>PD (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0/0</td>
<td>27/27</td>
<td>73/73</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td>30/70</td>
<td>30/70</td>
<td>70/73</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.2</td>
<td>41/59</td>
<td>33/26</td>
<td>67/74</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.25</td>
<td>44/56</td>
<td>33/26</td>
<td>67/74</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.3</td>
<td>47/53</td>
<td>36/24</td>
<td>64/76</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.35</td>
<td>51/49</td>
<td>39/24</td>
<td>61/76</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.4</td>
<td>55/45</td>
<td>42/24</td>
<td>58/76</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.45</td>
<td>59/41</td>
<td>44/24</td>
<td>56/76</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>64/36</td>
<td>47/24</td>
<td>53/76</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.55</td>
<td>70/30</td>
<td>51/24</td>
<td>49/76</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.6</td>
<td>75/25</td>
<td>55/23</td>
<td>45/77</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.65</td>
<td>80/20</td>
<td>59/24</td>
<td>41/76</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.7</td>
<td>85/15</td>
<td>65/24</td>
<td>35/76</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td>0/0</td>
<td>0/0</td>
<td>0/0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

NOTE. 1st stage = 16 weeks; 2nd stage = 16 weeks.

Abbreviations: CR, complete response; PR, partial response; SD, stable disease; PD, progressive disease.
absence of effective therapy. This confidence could translate into a stopping rule based on the posterior certainty that the risk of progression remains approximately 70% during the open-label part of the study. If one is willing to stop the study when one is 99% certain that the risk of progression is between 0.6 and 0.8, then the expected sample size in the absence of a CAI effect drops to around 200, whether one wishes to randomize 100 or 200 patients. We assumed initially that roughly 70% of the patients will progress by 16 weeks in the absence of effective therapy, but we acknowledged uncertainty of this estimate. Our prior uncertainty about the underlying risk of progression had a risk of progression of 85% or higher at the extreme of what we might expect to see, and we applied this value as a cutoff corresponding to no activity for the tables. If one feels more certain about the underlying risk of progression by 16 weeks, one would incorporate that knowledge into the design, for instance, by altering the cutoff for deciding if one should stop the study. For example, if one will consider the drug inactive if there is 90% certainty that the risk of progression is at least two thirds by 16 weeks, then one will only enroll 164 patients, on average, to attain 100 randomized patients if the drug is not active (not shown); a cutoff of 60% risk of progression with 90% certainty leads to an overall average sample size of 74 patients when the agent has no activity.

If, on the other hand, one demands greater evidence based on data from the open-label part of the study before stopping to declare activity, one will need to enroll more patients than shown in Table 1. By way of example, changing the stopping rule to declare activity, one will need to enroll more patients than shown; a cutoff of 60% risk of progression with 90% certainty that the risk of progression is at least two thirds by 16 weeks; 2nd stage

The term “tumor” refers to neoplastic disease, and we used the term “nonrandomized” to mean patients who were not randomized before the study began, regardless of whether the patient was a control (or placebo) or active treatment. We also investigated the effect on power of changing the duration of the two stages of the randomized discontinuation design. Table 3 summarizes simulation results when the treatment lacks activity and we changed the lengths of the initial nonrandomized stage and the follow-up after randomization. Shortening the initial stage of the study yields a more heterogeneous population going on to randomization than with the longer initial follow-up during the open-label stage. If the initial stage of the study is too short, then very few patients will progress before randomization. In the simulations, this situation led to an increase in the proportion of studies falsely declaring activity because too many trials stopped before randomization with a lower than expected proportion of patients progressing during the initial stage. If the duration of the initial stage is too long, on the other hand, more patients will progress before randomization, and many more patients would be enrolled to have the desired number of randomized patients. Altering the length of follow-up after randomization has no effect on the number of patients initially enrolled. Lengthening the postrandomization follow-up too much, however, may lead to many more patients progressing on both treatments. In this situation, the study will be less likely to show a treatment effect and will be too conservative statistically (ie, the randomized part of the study will incorrectly declare a treatment difference statistically sig-

<table>
<thead>
<tr>
<th>Assumed Growth</th>
<th>CR/PR (%)</th>
<th>SD (%)</th>
<th>PD (%)</th>
<th>CR/PR (%)</th>
<th>SD (%)</th>
<th>PD (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Faster growing tumors</td>
<td>0/10</td>
<td>90/0</td>
<td>0/16</td>
<td>0/16</td>
<td>84/84</td>
<td>94/99</td>
</tr>
<tr>
<td>Slower growing tumors</td>
<td>0/50</td>
<td>50/0</td>
<td>0/38</td>
<td>0/38</td>
<td>62/62</td>
<td>5.6/5.4</td>
</tr>
</tbody>
</table>

Note: 1st stage = 16 weeks; 2nd stage = 16 weeks.
nificant with probability less than 5% when using a nominal 5% significance cutoff). Shortening the follow-up time after randomization does not seem to have as great an effect. In the situations we simulated and show in Table 3, extending the duration of follow-up after randomization to 20 weeks did not affect the operating characteristics of the design (data not shown).

**DISCUSSION**

The randomized discontinuation design addresses limitations of uncontrolled single-arm studies and is an appealing design strategy when screening for activity among possibly cytostatic agents in cancer. The study’s design calls for random cessation of treatment with the study drug in patients whose disease appears stable while receiving the drug. Major benefits of this design are that all patients are given the drug initially, and the study drug appears stable while receiving the drug. Major benefits of this design are that all patients are given the drug initially, and the study drug appears stable while receiving the drug.

Other options that have been proposed for phase II trials are summarized in Table 4. Each of these proposed options is imperfect. Historical controls are often unreliable.9,10 The validity of cross-over designs requires an assumption of constant tumor growth in the absence of any drug effect. The same concern applies to the use of a patient’s prior clinical course as a basis for comparison.25,26

Because of the difficulties encountered when assessing activity of a cytostatic compound in phase II trials, many pharmaceutical companies have moved directly from phase I to phase III testing. This is not prudent, because phase III trials have significant costs associated with them, both for the patients and the sponsoring organization. It seems better to establish the agent’s activity and potential for clinical benefit using novel phase II approaches before launching a definitive phase III trial. The goal of alternate designs is to help select drugs likely to lead to growth inhibition and disease stabilization in the most efficient manner possible, taking the heterogeneity of natural tumor growth rates into account. These designs must also be sensitive enough to differentiate a drug’s disease stabilizing effect from naturally indolent disease. We feel that this latter determination is possible only in a randomized setting.

It may seem that randomized discontinuation will lead to a phase II study that enrolls an unusually large number of patients, and one might not want to allow the sample size to be that large in the event the agent is not very active. It is true that the study design may call for enrolling a large number of patients. The expected sample size will change by altering the criteria for deciding when to terminate the study. One may choose a more pessimistic expected probability of stable disease associated with the disease’s natural history a priori and thereby change the study’s operating characteristics. One can also reduce the expected sample size by stopping the study if one becomes fairly certain that the proportion of patients falling in the various response categories is very close to the original estimates or that the risk of progression (nonrandomized phase) is not better than initially assumed, as discussed in the Results section. As with all clinical trial designs, simulations assessing the characteristics of the design under various possible scenarios are illuminating, and we recommend their use.

With uncertainty about the proportion of patients with stable disease in the absence of active therapy, it seems fair to rely on the randomized part of the study to help determine if the drug is active or if, in fact, approximately 30% of patients will appear to have stable disease at 16 weeks. Another possible course of action is to develop dynamic eligibility criteria. For example, by analyzing the relationship between baseline patient characteristics and the risk of progression by 16 weeks, one can potentially develop and apply a predictive model to each patient on registration to the study. If the model predicts that the patient seems to be at higher risk of progressing during the first 16 weeks of open-label therapy, one may not enroll the patient, thereby reducing the number of patients receiving a drug that may not be active given their disease characteristics. We simply show one possible set of design criteria and the large-sample properties of implementing these criteria as estimated by simulation. Other possibilities exist, but we chose not to show them in the interest of space.

One may also be concerned about generalizing the results to a larger population of patients, such as those patients who, for one reason or another, did not have stable disease after 16 weeks of the study drug. The point of the proposed study design is to demonstrate activity in an explanatory, phase II sense, rather than determine the best way to utilize the anticancer agent in clinical practice. We do not feel that the ability to generalize a positive result is any more of a concern for this study design than it is for any other phase II design. There are two ways the study can indicate activity via a positive result. Either a higher proportion of patients during the first stage of the study (before randomization) will have stable disease than expected or fewer patients randomized to continue the study therapy will suffer disease progression after randomization, when compared with placebo. If the study shows a benefit for the use of the cytostatic agent in either case, then the likely subsequent evaluation will be a large, randomized phase III clinical trial—a pragmatic trial with a more long-term clinical end point such as survival. One would be more concerned about generalizing the study results in the case of the large pragmatic phase III trial rather than with a phase II study intended to see if the agent has any anticancer activity.

The study design should include blinding of patients and of persons evaluating end points. If either group of participants is aware of the individual’s treatment assignment, bias may arise. It...
is not common practice in phase II studies of anticancer agents to keep enrolled patients, their caregivers, or the people evaluating their clinical course uninformed of the patient’s therapy. In a phase II study designed with randomized discontinuation, however, one will often want, if not need, to institute double or triple blinding to avoid biases entering into the study’s outcome.

Patients who progress after randomization to placebo may want the opportunity to receive the study agent. In the CALGB study, we allow patients who progress on placebo to switch back to CAI. With masked treatment assignments, such a cross-over will require breaking the blind for this patient. Careful thought must go into deciding how to let the patient know if he or she had been receiving placebo or the study drug, so that blinding can continue for the remaining patients on the study, as well as any patients who subsequently enroll.

There are also statistical issues relating to how best to analyze the data. The original article on the randomized discontinuation design proposed analyzing the data accumulated after randomization, ignoring information gathered during the prerandomization stage. In the setting we consider here, namely, possibly cytostatic agents, we believe that there is information in the fraction of patients with stable disease at the end of the prerandomization period. For a particular disease setting, such as metastatic renal-cell cancer, one may have some idea what proportion of progressive disease to expect after several months of supposedly inactive therapy. In our example CALGB study, we expect approximately 70% of patients will progress during the first stage, leaving 30% with stable disease after 16 weeks. If a greater percentage of patients have stable disease after 16 weeks of CAI, then either CAI slows tumor growth or we underestimated the probability of progressing in 16 weeks of follow-up. The randomized phase of the design provides information about whether the drug is working. Additionally, one can increase the length of the prerandomization follow-up period and see the effect on the risk of disease progression. Similarly, if more patients progress than anticipated, then either the study’s underlying assumptions are mistaken or, perhaps, there is a patient safety issue to consider and the trial should stop. Thus, the analysis should not focus solely on data from the randomized stage. One should strive to include data collected while all patients are receiving the study drug as we described. With a full tumor growth model, one might even be able to incorporate information about growth during the nonrandomized part of the study into the analysis of the randomized portion, increasing efficiency. We chose not to assume any particular tumor growth model in our analysis, thereby making the study more robust.

Leber and Davis warn against the use of enrichment designs in which patients are initially treated in an open-label stage with the study drug, followed by randomization to the study drug or placebo. In particular, they discuss a study of tacrine’s effects on dementia. In the tacrine study, all patients went through a dose-titration stage and a placebo period, before randomization to active therapy or placebo. Only those patients who appeared to benefit from the dose-titration stage were eligible for randomization. One concern about that study is whether blinding could have been maintained, given the side effects of tacrine, because all patients received placebo and the three possible doses of tacrine. Exposure to each possible treatment increased the chance that each patient would be able to figure out their randomly assigned treatment. In our proposal, however, the situation is different. All patients receive the study drug at a single dose, but only half of those with stable disease will also be exposed to placebo. In the CALGB study, we will ask patients to guess their assigned treatment to assess the success of blinding.

A more important difference between our study and the tacrine trial, however, is that our primary end point is tumor size, which is objective, unlike the tacrine study’s primary end point, cognitive function, which has a major subjective component.

Another concern about the tacrine study and enrichment designs, in general, is the potential for carryover effects to mislead or dilute the difference seen between placebo and the study drug during the randomized stage of the study. Pharmacologic or therapeutic effects achieved during the initial stage may well interfere with the study’s ability to detect an effect of the study drug. In the case of the CALGB study, we feel fairly confident that any cytostatic effect CAI may have will not have much residual effect on the risk of progression during the 16-week follow-up period after randomization to placebo. In addition, it has been argued that this design would be unable to detect modest activity of a drug whose duration of effect is short. However, an oncologic drug that loses its activity after 16 weeks, despite continuing therapy, is of little interest for future studies.

One may be concerned about ethical issues surrounding discontinuing potentially active therapy. In a standard randomized trial, randomization takes place as close as possible to but before the time when therapy with the investigational agents begins. In standard phase III settings, we have become comfortable randomizing patients to one of two or more therapeutic regimens, as long as the participants feel they are in a state of clinical equipoise. Accept randomization, even though we know we may be assigning inferior therapy to half of the patients. With the randomized discontinuation design we discuss here, however, we randomize patients to a drug holiday (versus continuing treatment) only if their disease is stable after receiving the investigational agent for some fixed period of time. We do not know if the drug lowered the risk of progression or if the patient just happens to have very slowly changing disease. Therefore, we are truly in a state of clinical equipoise in this regard, just as when we randomize patients before we initiate therapy in other randomized studies.

One of the reviewers asked our thoughts about successor trials that might follow a randomized discontinuation study design for an explanatory study as we have described. Naturally, whether the successor study is another phase II study or a large randomized phase III study (or any study at all) depends on the current study’s results, the disease context, and the other therapies that are waiting to enter clinical evaluation. We feel that the study that might follow after the CALGB study of CAI would likely be a randomized phase III trial that compared the best standard therapy to either CAI alone or CAI with another chemotherapeutic agent. The subsequent trial’s design will certainly depend on the level of activity demonstrated in the randomized discontinuation trial.

Randomized discontinuation is certainly not appropriate for just any disease setting or for assessing activity of any anticancer agent. For the reasons outlined above, however, we feel that the randomized discontinuation design is a rational phase II study design to demonstrate activity of a possibly cytostatic agent in a disease having substantial patient-to-patient variation in tumor growth rates, and for which disease stabilization is a clinically
meaningful measure of activity. This article presents the first iteration of the application of randomized discontinuation for designing phase II studies in medical oncology, studies that regard stable disease as clinically important and useful measures of activity. There will certainly be refinements, some suggestions for which we point out earlier in this article. The next step, of course, is to see how the most important critics view the design, namely, the patients.

ACKNOWLEDGMENT

We thank two anonymous referees for their suggestions that greatly improved the manuscript.

REFERENCES

26. Mick R, Crowley JJ, Carroll RJ: Phase II clinical trial design for noncytotoxic anticancer agents for which time to disease progression is the primary endpoint. Control Clin Trials 21:343-359, 2000