The randomized discontinuation trial: a phase II design to assess growth-inhibitory agents

Walter M. Stadler

Section of Hematology/Oncology, University of Chicago, Chicago, Illinois

Abstract

An increasing number of putative anticancer targets and drugs have been identified with many of these expected to be growth inhibitory. Clinical development of these agents in the phase II setting is challenging because tumor shrinkages, or at least tumor shrinkages that meet the standard definitions of objective response, are not expected. Time to progression end points are however problematic because expected times in the absence of therapy (the null hypothesis) cannot be predicted accurately, thus requiring trials to enroll a concurrent control group. Another problem is that the patient population that will benefit from a new drug remains poorly defined in early-phase development. The randomized discontinuation trial design addresses both of these issues. All patients are initially treated with the drug; patients with an objective response continue therapy; patients who do not progress or experience excess toxicity within a prespecified “run-in” period are then randomized to continuing or discontinuing therapy in a double-blind, placebo controlled manner. Despite certain limitations that need to be recognized, the ability of this design to “select” a cohort most likely to benefit and to rigorously evaluate the disease-stabilizing activity of an investigational agent provides multiple advantages. [Mol Cancer Ther 2007;6(4):1180–5]

Introduction

There has been an explosion in the number of putative agents and drugs for the treatment of human malignancies. This can be attributed to both the increasing knowledge about molecular alterations in human cancer and the increasing capabilities of academic and pharmaceutical industry scientists to target these alterations. One outcome of research on these novel therapeutic targets is the increasing realization that effective targeting and treatment may lead to tumor growth inhibition or tumor shrinkage that does not meet the standard criteria for “objective response (1).” For many cancers, especially those that strike the elderly and those for which there are few systemic symptoms, growth inhibition, especially if induced by a well-tolerated agent, would be a clinically relevant effect.

Development of growth-inhibitory agents in oncology has, however, been challenging. The biggest challenge is the phase II setting, in which one attempts to determine whether a drug has sufficient antitumor activity to justify larger phase III trials seeking to determine whether the drug has sufficient patient benefit in an overall risk-benefit analysis. Perhaps, a better definition of phase II trials is the one provided by Sheiner (2) who described them as “learning trials” and phase III trials as “confirming trials.” Other recent reviews have addressed oncology phase II trial design in general (3–7). In this article, the limitations of traditional phase II designs for growth-inhibitory agents will be briefly reviewed followed by description and analysis of the randomized discontinuation design as an alternative approach.

Traditional Phase II Designs and Their Limitations for Growth-Inhibitory Agents

If one is to consider various phase II trial designs for growth-inhibitory agents, it is instructive to first consider phase II end points. The typical phase II cancer trial end point is a measure of robust and reliable degree of tumor shrinkage, typically termed “objective radiologic response (1).” These measures, of which the Response Evaluation Criteria in Solid Tumors guidelines are most commonly used, were originally based on the desire to reliably characterize within patient differences such that any change could be attributed to the intervention as opposed to simple measurement variability. In fact, the 50% decrease in bidimensional measurement criteria for objective response, from which the 30% Response Evaluation Criteria in Solid Tumors guidelines were derived, was based largely on the fact that oncologists were able to reliably palpate differences in the size of spheres under a thin mattress only when the difference was at least this large (8).

Although these criteria are based on measurement error in an individual, one effect is that observation of a high objective response rate implies a true drug effect in the studied population. In other words, objective responses are
not expected in the absence of any therapy or if the therapy is ineffective in this population. In statistical terms, the null hypothesis for the objective response rate is <5% with very tight confidence intervals and thus historical controls are sufficient.

As noted, the essential problem with growth-inhibitory agents is that tumor shrinkage, or at least tumor shrinkage that meets the standard criteria, is unlikely to occur. Therefore, some measure of tumor progression is likely a more appropriate phase II end point. Yet, the natural history of most cancers is so variable that it cannot generally be determined whether lack of tumor growth within an arbitrary time frame, or any observed time to progression, is the result of any administered drug. In statistical terms, the null hypothesis for a study population progression end point is highly variable and it is therefore not possible to rely on historical controls as the comparator. Assessment of a progression end point in phase II trials thus requires a concurrent control group. Importantly, similar issues apply if one seeks to determine whether an ‘‘objective response rate’’ is better than observed with standard therapy.

In addition to challenges with the end point and the requirement for concurrent rather than historical controls, the basic assumptions of traditional phase II designs also need to be considered. Traditional two-stage, single-arm designs assume that rejection of a potentially useful drug for future studies is a more significant error than acceptance of an inactive drug (9, 10). This was based on the historical lack of anticancer agents and the presumption that the value of a dubious drug will become apparent with further study but that a drug once rejected for use in a particular population is unlikely to be restudied. In the current era of multiple putative growth-inhibitory agents, however, rejection of a modestly active drug may not be as critical as recommending for definitive phase III testing a drug that has a high probability of failing such an evaluation (3). The highly expensive and much discussed failures of the metalloproteinase inhibitors and gefitinib are quite instructive in this regard (11–16). More robust data from phase II trials of growth-inhibitory agents are thus necessary before embarking on definitive phase III trials.

Finally, a growing issue in anticancer drug development is that the patient population most likely to benefit or respond is generally not known. The classic manner in which to differentiate cancer patients is by the organ of origin and phase II studies are typically limited in this same manner. This practice not only ignores the known clinical and molecular heterogeneity of various cancers arising in a single organ but also ignores similarities between cancers from different organ sites. Breast cancer, for example, not only consists of estrogen receptor–positive and estrogen receptor–negative tumors but has also been divided into ‘‘luminal A,’’ ‘‘luminal B,’’ ‘‘basal-like,’’ and ‘‘HER2 positive’’ based on molecular profiling studies (17). If trastuzumab, an anti-HER2 antibody, had initially been investigated in all breast cancers as opposed to only the HER2-positive cohort, it is unlikely that its beneficial effects would have been recognized (18). On the other hand, AKT pathway alterations are common in tumors of diverse histologic origins and are not present in all tumors of a specific histologic, or perhaps even molecular, subtype (19). Therefore, limiting evaluation of an AKT pathway targeting agent to a specific organ of origin may not be logical.

Nevertheless, it must be recognized that our knowledge of cancer molecular phenotypes remains incomplete and that current hypotheses about a particular mechanism of action of a drug are often incorrect. Thus, sole reliance on available molecular data to guide selection of patients for clinical trials can be dangerous. Perhaps, most illustrative is the development of epidermal growth factor receptor inhibitors. The initial preclinical data suggested that any degree of epidermal growth factor receptor expression was sufficient for their antitumor activity (20). The clinical data, however, suggested that the activity, even in tumors in which expression of the target is essentially 100%, is modest at best (14, 21). Ongoing research is still trying to define whether the best clinical predictive markers are epidermal growth factor receptor mutation, amplification, or molecular markers of the ‘‘epithelial-mesenchymal transition (22).’’ It thus becomes critical to begin to define the population most likely to respond and possibly benefit from treatment during phase II testing. If reasonable hypotheses for such a patient population can be generated during phase II, confirmation of the observation in phase III studies in a more precise patient population can be conducted in a very rapid and efficient manner.

The Randomized Discontinuation Trial: An Alternative Phase II Design

The overall schema for the randomized discontinuation trial (RDT) design is depicted in Fig. 1. All patients receive the drug initially. Patients who experience objective radiologic response and are apparently benefiting continue therapy. Patients who experience progressive disease or toxicity within a specified time frame discontinue therapy. Those patients who experience stable disease are then randomized in a placebo-controlled double-blind manner to continuing or discontinuing therapy. The primary end point is the fraction of patients in each randomized group who maintain stable disease at an arbitrary postrandomization time point. Time to event measures, such as progression, or overall changes in tumor burden, rather than a dichotomous ‘‘stable or better’’ versus ‘‘progression,’’ could also be used as end points (23).

The design was first proposed by Kopec et al. (24) and has since been statistically analyzed for efficiency and biases by several statisticians (25–27). In oncology, the design has been used to determine that the putative antiangiogenic agent carboxyaminoimidazole is inactive in renal cancer (28) and that sorafenib is active (29). The latter is the first successful adaptation with an active drug and thus the specifics are worthy.
Sorafenib was originally identified as a RAF kinase inhibitor with predominantly growth-inhibitory activity in preclinical and phase I trials (30–32). A phase II program focused on colon cancer, in which activated RAF was thought to be important, was designed using the RDT approach. Because RAF is an integrating kinase in multiple growth factor receptor kinase cascades, the trial allowed entry of any refractory solid tumor patient in addition to colon cancer (29). A total of 502 patients were enrolled, of which 202 were renal cancer patients, largely due to investigator observations of tumor shrinkage in this population. Of these, 34% experienced bidimensional tumor shrinkage of 25% or more (a nonstandard protocol defined end point) by 12 weeks of treatment and continued on therapy. Sixty-nine patients had tumor measurements within 25% of baseline after 12 weeks of which 64 (plus one additional patient not meeting criterion) were randomized. At the 12-week postrandomization primary end point, 16 of 32 patients on sorafenib maintained stable disease, whereas only 6 of 33 patients randomized to placebo maintained stable disease \((P = 0.0077)\). Median progression-free survival from the randomization point in the sorafenib group (24 weeks) was also greater than that of the placebo group (6 weeks; \(P = 0.0087\)). Interestingly, the vascular endothelial growth factor receptor– and platelet-derived growth factor receptor–inhibitory properties of sorafenib, which are likely the more important mechanisms for its effects against renal cancer, became evident only after the RDT trial was designed (33).

**Advantages and Disadvantages of the Randomized Discontinuation Trial Design**

The sorafenib study results confirm prior statistical analyses of the RDT design, which have concluded that the overall trial size can be smaller than standard upfront randomization, the overall number of patients exposed to placebo is less than in a standard upfront randomization, and the trial is most useful when the responding subpopulation represents a relatively small proportion of the entire investigated population (24–27). For growth-inhibitory agents, “responding” implies antitumor effect and not necessarily “objective radiologic response.” These statistical analyses also emphasized that a lack of “carryover” effect of the agent following discontinuation is critical for successful application of the design. Although some investigators have emphasized the possible overall lower number of patients enrolled in a RDT trial compared with an upfront randomization, these differences are relatively modest and manifest only in a few situations (24). Nevertheless, the increasing reluctance of cancer patients and their physicians to accept the use of placebos in oncology trials does mean that the lower number of patients exposed to a placebo arm in the RDT design is a distinct advantage.

Some critics have argued that discontinuation of an “active” agent raises ethical issues. In the context of phase II trials, in which the true clinical value of a drug or regimen remains unknown, the ethics of discontinuation are inherently no more problematic in this design than in an upfront randomization scheme, in which some patients do not receive the new drug at all. More importantly, randomization of patients with stable disease should meet the ethical criteria of equipoise because it is not clear whether lack of growth is due to a drug effect or simple selection of an indolent population. Patient acceptability of this trial design is highlighted by the very rapid accrual not only in the RDT of the active agent sorafenib but also in the RDT of the inactive agent carboxyaminoimidazole (28, 29).

Importantly, trial size and number of patients randomized is highly dependent on the prerandomization, or “run-in,” duration (27). If this duration is too long, some responding patients will progress and will not be randomized, thus increasing the overall trial size. If this duration is too short, there will be insufficient enrichment of the responding population and the statistical behavior of the trials will become similar to an upfront randomization trial. The planned prerandomization duration must thus be considered carefully in the context of the studied population and expected drug effect.

Although some of the above limitations may preclude the use of the RDT in many situations, it should certainly...
be considered when the population most likely to benefit from a growth-inhibitory agent is not clearly recognizable because the most important advantage of the RDT trial design is its ability to "select" the subpopulation most likely to benefit from the intervention. The trial design is particularly powerful when the drug effect is great (i.e., near complete inhibition of growth) and the responding population constitutes <50% of the overall enrolled population (25, 26). In fact, if the responding population is 30% or less and if the drug has absolutely no effect in the nonresponding population, it is highly unlikely that the power of an upfront randomized trial with a growth-inhibitory agent would be sufficient with a reasonable number of patients (18, 34).

For example, under the assumption that trastuzumab has no effect in HER2-negative breast cancer and given the observation that ~30% of breast cancer patients are HER2 positive (35), >23,000 unselected patients would need to be enrolled to provide 90% power to detect what is now known to be the survival benefit provided by this drug in the HER2-positive population (18). Some have argued that clinical situations where drug effect is great in a small subpopulation of enrolled patients are rare. The trastuzumab example, however, shows that a highly selected population can benefit dramatically. However, as in the above examples for the epidermal growth factor receptor inhibitors and sorafenib, it is far more common for the population most likely to benefit to be unknown at the time late phase II trials are initiated.

This ability to "select" the responding population nevertheless introduces biases that require the observations to be confirmed in a subsequent phase III trial (24). The ability to be randomized is dependent not only on drug activity but also on toxicity and compliance, an increasingly important issue with oral agents. In other words, the trial may "select" a population of patients willing to tolerate bothersome, low-grade toxicities rather than a population with biological characteristics predicting benefit. It is thus critical that the "learning" that occurs in a randomized discontinuation trial be "confirmed" in a subsequent phase III trial. Whether the phase III trial should be an upfront randomization versus placebo in the patient group identified in the RDT or whether it should be an add on trial of standard therapy with or without the new agent is dependent on the available treatment options, the available standards of care, and information about the clinical ability to combine the new agent with available treatments.

If RDT designs are relatively large, and still require subsequent phase III confirmation, it may reasonably be asked why such an investment is worthwhile. Specifically, an upfront randomization could be conducted and the trial be extended to a more formal phase III trial with larger numbers and more definitive end points if the initial results are promising (5). In such a trial, the initial phase II portion could have a time to progression end point and the phase III could have a survival end point. Such a design, although not unreasonable, once again presupposes that the investigator knows the benefiting population. As stated earlier, this may not always be the case, and the RDT can select the population based on the clinical activity of the drug rather than based on incomplete prior knowledge.

Perhaps, just as important are the extensive time and monetary investments necessary in a phase III trial. These confirmatory studies need to be highly rigorous in their eligibility and outcome measures to assure that the trial has the best ability to truly assess risk/benefit ratio in a clearly defined population. This kind of rigor is expensive not only because of the sheer amount of data that needs to be collected but also because of the need to verify and validate any collected information. In addition, the end points that define true benefit for the patient, such as survival, tend to be much longer than end points in phase II studies, such as time to progression or a fraction of patient progression-free at an arbitrary time point. This also adds significantly to phase III trial expense.

It thus may be more prudent to conduct a shorter RDT trial and focus on collecting clinical and molecular data that may help identify patients most likely to respond and be randomized. Factors to consider include evaluation of molecular drug targets, molecular and imaging diagnostic markers, and pharmacogenetic markers. Such factors would then become putative predictive factors or eligibility criteria in a confirmatory phase III trial. In a "learning trial," multiple testing of numerous factors with the resultant potential for false positives would be acceptable because the intent is to confirm these observations in a subsequent trial.

In addition, RDT trial size and expense could be controlled by allowing generous type I and type II errors. A finding significant at the \(P = 0.1\) level still means that there is only a 10% chance of a false positive, which are reasonable odds for conducting the confirmatory phase III trial. Finally, the larger size of a RDT phase II trial, and its randomized nature, means that lack of positive outcome provides a much higher degree of confidence that the agent is not particularly useful in the studied population than a traditional single-arm phase II study and at much lower expense than a traditional phase III study.

Finally, it may be useful to consider several modifications and additional measures within the RDT design that could further improve its utility. First, it is useful to include a futility analysis such that the trial is ended early if fewer patients than expected are randomized. This was done in a post-hoc manner in the previously noted carboxyaminimidazole trial in renal cancer but could easily be incorporated into the initial statistical plan (28). Such a priori criteria would set the maximum trial size and the minimal fraction of responding patients considered acceptable by the sponsor. It may also be possible to randomize a greater portion of patients by broadening the degree of tumor shrinkage that qualifies for randomization. The sorafenib trial randomized only patients who had between 25% shrinkage and 25% growth on
bidual measurements (29). It is prudent to at least randomize patients who have stable disease by the standard radiologic criteria. However, because “objective partial response” does not always translate into true patient benefit, one could randomize patients who experience tumor shrinkage >30% on unidimensional measurements. One could also consider using several arms with different doses or schedules. Dose or schedule effects could then be assessed not only by the fraction of patients randomized in each cohort but also by the relative benefit in the randomized cohorts. Finally, the use of active treatment controls, in which standard therapy plus investigational therapy is followed by randomization to continuing or discontinuing only the investigative agent, is also possible.

Conclusion

A large number of novel agents and an ever larger number of putative targets are in or will soon be in clinical development. Many of these agents are expected to be growth inhibitory and/or may not necessarily lead to sufficient tumor shrinkages that qualify for objective radiologic response under standard criteria (I). There are clearly insufficient patients and insufficient financial resources to take all these agents to confirmatory phase III trials. It thus becomes ever more important to conduct phase II learning trials that provide findings that are likely to be confirmed in phase III. One such design is the randomized discontinuation trial design. Perhaps, the most important feature of this design is its ability to select the subgroup most likely to benefit from the intervention. This selection then allows the confirmatory trial to either focus on this subgroup or at the very least create a prospective stratification for this subgroup. It is hoped that such an approach will not only speed development of promising agents but also help in reaching the now common goal of individualizing therapy more appropriately than just based on tumor histologic type.

Acknowledgments

We thank Drs. Mark Ratain and Theodore Karrison for helpful discussions.

References

30. Moore M, Hirte HW, Siu L, et al. Phase I study to determine the...


The randomized discontinuation trial: a phase II design to assess growth-inhibitory agents

Walter M. Stadler