The Design of Phase II Clinical Trials Testing Cancer Therapeutics: Consensus Recommendations from the Clinical Trial Design Task Force of the National Cancer Institute Investigational Drug Steering Committee

Lesley Seymour1, S. Percy Ivy2, Daniel Sargent3, David Spriggs4, Laurence Baker5, Larry Rubinstein2, Mark J. Ratain6, Michael Le Blanc7, David Stewart6, John Crowley8, Susan Groshen9, Jeffrey S. Humphrey10, Pamela West11, and Donald Berry8

Abstract

The optimal design of phase II studies continues to be the subject of vigorous debate, especially studies of newer molecularly targeted agents. The observations that many new therapeutics “fail” in definitive phase III studies, coupled with the numbers of new agents to be tested as well as the increasing costs and complexity of clinical trials, further emphasize the critical importance of robust and efficient phase II design. The Clinical Trial Design Task Force (CTD-TF) of the National Cancer Institute (NCI) Investigational Drug Steering Committee (IDSC) has published a series of discussion papers on phase II trial design in *Clinical Cancer Research*. The IDSC has developed formal recommendations about aspects of phase II trial design that are the subject of frequent debate, such as endpoints (response versus progression-free survival), randomization (single-arm designs versus randomization), inclusion of biomarkers, biomarker-based patient enrichment strategies, and statistical design (e.g., two-stage designs versus multiple-group adaptive designs). Although these recommendations in general encourage the use of progression-free survival as the primary endpoint, randomization, inclusion of biomarkers, and incorporation of newer designs, we acknowledge that objective response as an endpoint and single-arm designs remain relevant in certain situations. The design of any clinical trial should always be carefully evaluated and justified based on characteristic specific to the situation. *Clin Cancer Res*; 16(6); 1764–9. ©2010 AACR.

Background

Many new drugs targeting molecular pathways are ready for clinical development, necessitating the use of efficient trial designs to quickly and accurately identify promising agents, while also identifying those for which all further development should be stopped. Although the development of some drugs is discontinued after phase I, the major drug development decision is generally made on the basis of phase II results. Although traditional oncology trial designs using the endpoint of response and a single arm design seem to have done this task reasonably well for cytotoxic agents, the same does not seem to be true for newer agents in which high rates of tumor shrinkage may not be expected, nor for combinations of agents (such as a new drug combined with standard treatments). Certainly, success rates for phase III trials seem to be decreasing (1). This decrease has led to considerable scientific discussion, debating the advantages and disadvantages of using response versus progression (2) or other imaging endpoints (3), single arm versus randomized designs (4), patient enrichment and biomarker endpoints (5), and optimal statistical designs, such as adaptive design or phase I–II designs.

The Investigational Drug Steering Committee (IDSC) of the National Cancer Institute Cancer Therapy and Evaluation Program (NCI CTEP) appointed a Clinical Trial Design Task Force to advise on the design of early (phase I and II) clinical trials (Table 1). In keeping with its broad mandate, Task Force members include IDSC members, as well as external representation from academia and the
Although the most common approach is from randomization, ideally, however, the selection would be efficient. We recognized that there are circumstances in which such a “proof of concept” approach might be reasonable, such as seeking a signal about the selection of tumor types for further study (e.g., when not readily apparent from preclinical or phase I studies) or for biomarker-based studies to validate a proposed mechanism of action. Ideally, however, these concepts would be embedded, possibly adaptively, in a single phase II trial. Thus, we used single-arm versus randomized studies as our primary categorization of phase II trials (Fig. 3).

**Selection of the appropriate primary endpoint.** The advantages and limitations of objective tumor response as the primary endpoint in phase II trials, and alternative endpoints such as biomarker-based studies to validate a proposed mechanism of action. Although the Task Force accepted that response-based endpoints are still relevant for some agents (when tumor shrinkage and clinically relevant response rates are expected) and some trials, the recommendations emphasize the need to consider the inclusion of a progression-free survival primary endpoint as more informative (8, 9). Overall survival is not recommended as an endpoint, as subsequent therapy may confound conclusions, and progression is usually substantially earlier, thus shortening the duration of the trial and follow-up.

**Randomization, blinding, and crossover.** The Task Force agreed that randomization was generally required to evaluate the efficacy of combinations of agents (e.g., for approved drugs and investigational agents). Randomization is usually essential for a phase II trial in which progression-free survival is the most appropriate endpoint. Nonetheless, single-arm designs are still appropriate for the evaluation of a monotherapy or when a well-defined historical control database is available (10). As for any trial the design and the selected null and alternative hypotheses must be carefully justified. If a randomized design is selected, blinding of the agents (against placebo, other doses of the same agent, or other active agents) should be considered. When the primary endpoint is progression-free survival or response based, designs that allow crossover after progression maintain the integrity of the study and can provide additional data that could inform the future development of the agent.

**Biomarkers.** Biomarkers are of considerable interest in the setting of the phase II study, but they present significant challenges in their incorporation, measurement, and interpretation. In most instances, the biomarker is...
not clinically validated as a predictive marker (of efficacy) early in the development of a new agent (i.e., at the time of the phase II trial). The IDSC’s recommendations about biomarker use in early clinical trials are detailed in the current Focus Series (11–16). Because of these limitations, these recommendations encourage the prospective inclusion of molecular markers in phase II trials to evaluate predictive markers, but discourage prospective patient selection on the basis of a biomarker (unless already clinically validated), except in the setting of an appropriate (and explicit) adaptive design. Phase II trials including patients with a specific biomarker but with multiple histological subtypes were considered of particular interest and may be a more efficient screening tool, especially when combined with an adaptive design.

**Statistical designs.** Improved efficiencies in clinical trial design with associated shortening of development times for effective agents are highly desirable. Numerous designs have been proposed, including randomized selection designs (pick-the-winner), adaptive designs (17), randomized discontinuation designs (18), and other randomized designs (19). Prospectively specified adaptive designs are of particular interest in the context of phase II studies of molecularly targeted agents in which biomarker identification and validation may be emergent during the conduct of the trial, limiting the ability to select patients or identify optimal doses and/or schedules at the trial outset. Such adaptive designs are also particularly useful for trials including patients with a range of histologic subtypes but with biomarkers of interest. Adaptive designs in such settings should be efficient and may result in improved precision. Despite the multiplicity of new designs that have been proposed, their inclusion in new trials has in general been modest at best (20, 21). Reasons postulated include requirements for statistical support as well as concerns about robustness, accrual, and cost.

The Task Force is strongly supportive of designs that improve efficiency and shorten development time, such as adaptive designs, but recognized the need to continue to formally evaluate these designs to encourage wider acceptance and implementation. An ongoing initiative is the creation of a database to allow the formal testing, *in silico* of newer designs, in order to validate their use in future trials.

Interestingly, although formulated prior to the publication of the editorial, these recommendations are congruent with a review of phase II trials published in the Journal of Clinical Oncology (22), as well as with other reviews and recommendations (23).

### Consensus Recommendations

**Choosing the appropriate primary endpoint**

The first and critical decision point for the design of a phase II trial is based on the choice of the most appropriate primary endpoint, which should be tailored to the disease and drug(s) under investigation.

- Response-based endpoints such as those defined by Response Evaluation Criteria In Solid Tumors (RECIST), are standard, especially in early phase II trials. Other qualified biomarkers, such as molecular imaging or tumor markers, may be appropriate in select circumstances. Response-based endpoints are appropriate primary endpoints if unambiguous and clinically relevant direct antitumor activity (such as tumor shrinkage) is hypothesized.
- If a response-based endpoint is not appropriate, especially in later phase II trials, progression-free survival is recommended as the primary endpoint. Other biomarker endpoints (such as tumor burden, tumor markers, novel imaging, tumor response, molecular biomarkers) and patient-reported outcomes are always encouraged as secondary endpoints, especially in the context of studies that aim to qualify such endpoints. It is acknowledged that once qualified, these biomarker endpoints will become appropriate primary endpoints.

**Study design: primary endpoint is tumor response**

**Monotherapy trials.** Single arm designs are acceptable. However, randomization should be encouraged to optimize dose and schedule or to benchmark activity against known active therapies.

### Table 2. Past and present members of the clinical trial design task force

<table>
<thead>
<tr>
<th>Position</th>
<th>Member</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chair</td>
<td>Lesley Seymour</td>
</tr>
<tr>
<td>Co-chair</td>
<td>Donald Berry</td>
</tr>
<tr>
<td>CTEP</td>
<td>S. Percy Ivy</td>
</tr>
<tr>
<td>Imaging</td>
<td>L. Shankar, A. Shields</td>
</tr>
<tr>
<td>Advocate</td>
<td>D. Collyar</td>
</tr>
<tr>
<td>Nonvoting</td>
<td>R. Agarwal, L. Minasian, P. Ujhazy, L. Jensen, P. West</td>
</tr>
</tbody>
</table>
Combination trials. With some exceptions (e.g., availability of a well-validated robust control database), randomization is usually required for trials testing combinations of agents to establish efficacy. An example is standard therapy ± novel agent or combinations of novel agents.

Study design: primary endpoint is progression-free survival

Monotherapy or combination trials

1. With some exceptions (e.g., availability of a robust control database), randomization is required.
2. For randomized trials, blinded designs are encouraged when feasible. Although placebo controlled trials are challenging, they are encouraged whenever possible. Alternatives include dose ranging, randomization versus active controls or other novel agents, and randomized discontinuation and other crossover designs.
3. It may be informative to prospectively incorporate crossover to the standard therapy + novel agent for those patients initially assigned to the standard therapy alone, although careful consideration should be given to the timing of crossover (e.g., only after the primary endpoint has been observed). Such crossover designs increase the access of patients to investigational agents, and also provide additional information about the activity of the study arms.

Patient selection and enrichment strategies

Monotherapy or combination trials

1. A goal of phase (I and) II development should be to define biomarkers predictive of efficacy and/or toxicity. When feasible and appropriate, molecular biomarkers should be explored in order to identify subsets of patients of interest for future study.
2. Enrollment should, in general, not be limited by biomarker status unless there are strong confirmatory and supportive clinical data justifying the enrichment strategy. Adaptive statistical designs may be used to allow modification of enrollment if data suggest a biomarker is predictive.
3. In an unselected trial (i.e., patients not defined by a biomarker), the patient population of primary
interest (i.e., a cohort defined by a biomarker) should be predefined and the study powered accordingly to detect an effect in that subset.

4. Multidisease phase II designs should be considered, especially if the objective is to test a biomarker-focused hypothesis.

**Statistical designs**

Prospective designs that adapt to what is learned during the trial can improve the efficiency of drug development and provide greater precision. Available adaptations include stopping early, continuing longer than anticipated, dropping arms (or doses), adding arms, focusing
on patient subsets, assignment of better performing treatment arms with greater probability, and seamlessly moving from phase I to II or phase II to III during a single trial.

Conclusions

The Task Force formulated recommendations (Fig. 2) for the design of phase II trials of anticancer agents on the basis of consensus gained during a workshop and extensive discussions with members of the IDSC, the Task Force, and external experts. These recommendations were subsequently approved by the IDSC. Although these recommendations in general encouraged the use of progression-free survival as the primary endpoint, the use of randomization, the inclusion of biomarkers, and the use of newer designs, they acknowledge that objective response and single-arm design remain relevant in appropriate circumstances. The design of any clinical trial should always be carefully evaluated and justified.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

The Clinical Trial Design Task Force would like to acknowledge the support of the NCI and the IDSC members and chairs, Amy Gravell, LeeAnn Jensen, the attendees of the original Phase II Workshop, as well as all Task Force members, past and present, especially Drs. Alex Adjei and Michael Christian who originally led the Task Force (Table 2).

Received 12/15/2009; revised 01/08/2010; accepted 01/13/2010; published OnlineFirst 03/09/2010.

References