Novel Designs and End Points for Phase II Clinical Trials

Alex A. Adjei,1 Michaele Christian,2 and Percy Ivy2

Abstract The large number of negative phase III trials in oncology over the last several years has renewed interest in refining phase II oncology clinical trials to maximize the chances of success in phase III testing. More efficient phase II study designs will improve our ability to identify promising agents for testing while accurately identifying nonefficacious agents. Recognizing that new paradigms of phase II trial designs need to be developed, the Clinical Trial Design Task Force of the National Cancer Institute (NCI) Investigational Drug Steering Committee has tackled the question of improving efficiency of phase II clinical trials. In this issue of CCR Focus, four of the major topics discussed are presented. First, the task force recommended that alternate phase II end points should be studied. Second, depending on the characteristics of the specific trial and study population, historical controls or a randomized design may be more appropriate. Third, rational incorporation of biomarkers into phase II trials should be encouraged. Last, novel imaging modalities will be critical in evaluating the clinical benefit of new cytostatic agents.

Recent advances in molecular biology have led to the development of a myriad of anticancer agents that specifically target aberrant pathways and other proteins that are relatively specific for tumor cells. The number of agents available for testing dictates that a more efficient system aimed at quickly and accurately identifying promising agents for phase III testing be developed. This will allow investigators to efficiently discard nonefficacious agents and devote time to the development of the more promising agents, thus conserving financial and human resources. Only about 5% of investigational agents tested in phase III trials make it onto the market in oncology. This is in contrast to the 20% or so success rate in other areas of medicine such as cardiovascular disease and the average of 11% for all disease sites (Fig. 1). A barrier to achieving this goal is the lack of good surrogates of true patient benefit, which in oncology is improvement in overall survival. In addition to identifying good surrogates, which could be imaging biomarkers or biochemical, genetic, or molecular biology biomarkers, novel approaches to phase II study design need to be tested. In this issue of CCR Focus, four subjects relevant to more efficient phase II study design, namely, alternate end points, randomized designs, biomarkers, and imaging tools are discussed.

Alternate End Points for Screening Phase II Studies

A major challenge in drug development is the efficient design of phase II trials to identify active compounds for phase III testing. The inability to accurately predict effective agents results in a disproportionately high incidence of negative phase III trials (Table 1). End points in phase II studies that accurately predict phase III success are needed. Traditionally, objective response rates have been used, based on the idea that tumor shrinkage predicted for prolonged survival. Accumulating evidence suggests, however, that the effect of novel agents inhibiting angiogenic and proliferative signaling, among others, may not be accurately captured by traditional radiographic evaluation. An example is sorafenib in the treatment of hepatocellular carcinoma. Sorafenib showed a doubling of the time-to-progression rate compared with a placebo, with an objective response rate of 2% (1). Alternative phase II end points that may accurately predict the true efficacy of novel noncytotoxic agents are currently being explored. In this series, Dhani et al. (2) review alternate end points proposed for phase II trials. The focus is on using conventional imaging techniques to acquire data that are then analyzed in innovative ways. Current phase II designs are considered, and an analysis using data from phase III trials in pancreatic cancer and colorectal cancer is done. These investigators conclude that continued research to validate alternate phase II end points is essential.

Improving Predictive Power of Phase II Studies

As discussed previously, a number of novel agents possess antiproliferative effects and do not cause significant tumor shrinkage to the point where classical response criteria are met. For example, sorafenib does cause tumor shrinkage as measured by waterfall plots; on the other hand, these decreases in unidimensional [Response Evaluation Criteria In Solid Tumors (RECIST)] or bidimensional (WHO) size do not quantitatively meet accepted response end points. Sunitinib, another multitargeted kinase inhibitor with vascular endothelial growth factor receptor 2 inhibitory activity, shows a 40% measurable response rate in renal cell cancer that is associated with an improvement in progression-free survival.
These data indicate that the clinically effective novel agents do elicit changes in tumor size to a variable extent. Conventional imaging end points (RECIST or WHO), however, that dichotomize the continuous variable of tumor burden to “response” and “no response” fail to accurately capture patient benefit and predict eventual phase III success. Insignificant decrements in tumor size caused by blocking proliferative signals, but wherein very little viable tumor is left in the residual mass, are particularly complicated to evaluate. In these cases, the tumor mass may be composed predominantly of necrotic tissue, a thin rim of viable tissue with central necrosis or inflammatory and fibrous connective tissue (3, 4).

In summary, failure to elicit a measurable response by RECIST or WHO is disassociated from conventional activity measurements but remains associated with molecular targeted effects and often results in decreases in tumor size when measured by waterfall plots. As such, investigators should increasingly give attention to continuous measures of tumor burden as a phase II trial end point. The development of better markers of effective therapy must be prospectively validated because tumor shrinkage does not always correlate with improved outcome. Timed end points such as progression-free survival or overall survival may be more appropriate than response rates in assessing treatment benefit for anticancer agents, including

### Table 1. Examples of promising phase II studies of novel agents that led to negative phase III studies

<table>
<thead>
<tr>
<th>Agent</th>
<th>Target</th>
<th>Comparator</th>
<th>Disease</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enzastaurin</td>
<td>PKC</td>
<td>—</td>
<td>Glioblastoma</td>
<td>No benefit (23)*</td>
</tr>
<tr>
<td>Vatalanib</td>
<td>VEGFR</td>
<td>FOLFOX</td>
<td>Colon cancer</td>
<td>No benefit (24, 25)*</td>
</tr>
<tr>
<td>Bevacizumab</td>
<td>VEGF</td>
<td>GC</td>
<td>Pancreas</td>
<td>No benefit (26)*</td>
</tr>
<tr>
<td>Sorafenib</td>
<td>MKI</td>
<td>PC</td>
<td>NSCLC</td>
<td>No benefit (27)*</td>
</tr>
<tr>
<td>Timifarnib</td>
<td>FT</td>
<td>—</td>
<td>AML</td>
<td>No benefit (28)*</td>
</tr>
<tr>
<td>Lonicarnib</td>
<td>FT</td>
<td>PC</td>
<td>NSCLC</td>
<td>No benefit (29)*</td>
</tr>
<tr>
<td>ISIS 3521</td>
<td>PKC</td>
<td>PC</td>
<td>NSCLC</td>
<td>No benefit (30)*</td>
</tr>
<tr>
<td>Bexarotene</td>
<td>RXR</td>
<td>VC</td>
<td>NSCLC</td>
<td>No benefit (31)*</td>
</tr>
<tr>
<td>Bevacizumab</td>
<td>VEGFR</td>
<td>Erlotinib</td>
<td>NSCLC</td>
<td>No benefit*†</td>
</tr>
<tr>
<td>PF3512676-</td>
<td>TLR9</td>
<td>PC</td>
<td>NSCLC</td>
<td>No benefit/toxic (32)</td>
</tr>
</tbody>
</table>

**NOTE:** Where Comparator is indicated, the compound is combined with the standard regimen or agent and compared with the regimen/agent alone.

**Abbreviations:** PKC, protein kinase C; VEGFR, vascular endothelial growth factor receptor; VEGF, vascular endothelial growth factor; MKI, multikinase inhibitor; FT, farnesyl transferase; RXR, retinoid X receptor; TLR9, toll-like receptor 9; FOLFOX, 5-fluorouracil/leucovorin/oxaliplatin; GC, gemcitabine/cisplatin; PC, paclitaxel/carboplatin; VC, vinorelbine/cisplatin; NSCLC, non–small cell lung cancer; AML, acute myeloid leukemia.

*No benefit over standard therapy.
standard cytotoxics. For such end points, the general consensus is that historical controls may be unreliable. For combination therapies, a real potential for bias using time-to-event end points is observed. The use of placebos and/or blinding may help avoid bias. Thus, a control arm in the evaluation of such agents may be important in determining the true effect of novel agents. To prevent an erroneous determination of progression-free survival, which could be dependent on the frequency of patient evaluation, progression-free survival rate at a specified time point may be more reliable. It should be noted that trials that are underpowered for a small difference are not necessarily underpowered for large effects. In advanced disease settings with a median progression-free survival of 3 months, looking for a substantial improvement in progression-free survival before embarking on phase III trials is appropriate. The randomized design is particularly relevant wherein a novel agent is combined with standard chemotherapy. Determining the additional effect of the novel agent is difficult, and a randomized design incorporating a standard chemotherapy arm helps resolve some of these problems.

In recent years, two types of randomized designs have been used in phase II oncology trials, namely, randomized phase II trials with a standard chemotherapy control arm and randomized discontinuation designs (5, 6). Rubinstein et al. (7) discuss the merits and disadvantages of historical controls and various randomized designs. They conclude that randomized designs need to be considered in appropriate instances. A recent article on this subject is also recommended to the interested reader (8).

**Imaging Tools for Phase II Treatment Trials in Oncology**

Tumor shrinkage (objective response) as measured by anatomic imaging to determine therapeutic efficacy is increasingly unreliable as an end point for a number of novel agents. Standard anatomic criteria are based on changes in tumor size. A number of novel agents such as sorafenib (discussed previously) cause tumor stasis or lead to necrosis with minimal viable tumor tissue but no overall change in tumor size. These changes may remain undetected using standard anatomic criteria. As a result of these limitations, standard anatomic response may not correlate with clinical outcome (9). The choice of relevant imaging modalities is crucial regardless of the end point used. Because imaging is used to define treatment failures, the imaging modality used represents a true surrogate of patient benefit, which ultimately is survival.

Among the newer imaging approaches available, fluorodeoxyglucose positron emission tomography has emerged as the most used functional imaging modality and has been validated in certain diseases such as gastrointestinal stromal tumors. Shankar et al. (10) discuss fluorodeoxyglucose positron emission tomography and outline problems and opportunities using data from a number of studies. It is becoming increasingly clear that novel positron emission tomography imaging agents and other functional imaging modalities, including digital contrast-enhanced magnetic resonance imaging and magnetic resonance spectroscopy among others, will be the focus of future research. As the proportion of cytotoxic drugs in oncology therapy diminishes, objective response rates measured by tumor shrinkage will become less reliable as a clinically meaningful end point and may be replaced with functional imaging modalities.

**Biomarkers in Phase II Trials**

Most of the new, approved anticancer agents, as well as most of those in clinical testing, have distinct targeting capabilities against malignant cells. Several critical issues in drug development need to be addressed as these novel agents proceed in clinical evaluation. Agents have to show inhibition of the intended target, and inhibition should lead to clinical benefit. If the drug target is not ubiquitous, patients need to be selected according to the presence or absence of specific tumor-related molecular signatures to enhance clinical benefit (11). Correlative biomarkers increase in importance in this setting. In phase I clinical trials, biomarkers are important for proof of target inhibition at achievable drug concentrations. In phase II clinical trials, biomarker assays potentially provide an early marker of agent activity, if a strong correlation between assay results and clinical outcome can be established. Alternately, biomarkers can be used to select patients with a molecular profile characteristic of those who are likely to respond to a particular regimen. There are a number of challenges inherent in incorporating biomarkers into clinical trials. Often, the true target of the agent in human tumors is unknown, there may be multiple targets such as a number of the protein kinase inhibitors in development, making the identification of a robust biomarker difficult. In cases in which the drug target is known and used as a surrogate end point, inhibition of the drug target may be necessary but not sufficient for tumor shrinkage or patient benefit. This has been shown for instance in the evaluation of MAP/ERK kinase inhibitors (12, 13). A number of times, the assays to be used are neither standardized nor validated, and different results are obtained by different investigators. These and other challenges faced in incorporating biomarkers into early clinical trials have been clearly outlined by McShane et al. (14) in this CCR Focus.

**Other Issues**

In addition to the four topics addressed in separate articles in this CCR Focus, other issues that need to be addressed in improving phase II trial designs are discussed here.

**Acclual.** One of the most important reasons for negative clinical trials is lack of accrual (15–17). The subject of poor accrual to clinical trials (it is estimated that only about 5% of the eligible population is enrolled into cancer clinical trials in the United States) has been extensively discussed. Complex, poorly designed studies, increasingly complicated regulatory policies, and financial hardship (lack of adequate funding and insurance coverage) are seen as some of the key barriers to clinical trial participation. The interested reader is referred to excellent reviews of this subject in the recent literature (15–17). Clinicians and statisticians need to remember the challenge of poor accrual and ensure that novel phase II clinical trial designs do not include unnecessarily complex study-related procedures for patients because this could lead to poor accrual, resulting in the study question never being
### Table 2. Key features of WHO, RECIST 1.0, and RECIST 1.1 criteria

<table>
<thead>
<tr>
<th>Description of disease</th>
<th>WHO (Miller 1981)</th>
<th>RECIST 1.0 (Therasse 2000)</th>
<th>RECIST 1.1 (Eisenhauer 2009)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measurable, bidimensional</td>
<td>Measurable, unidimensional</td>
<td>Measurable</td>
<td>Measurable</td>
</tr>
<tr>
<td>Measurable, unidimensional</td>
<td>Evaluable/nonmeasurable</td>
<td>Nonmeasurable</td>
<td>Nonmeasurable</td>
</tr>
<tr>
<td>Measurable lesion definition</td>
<td>Uni- and bidimensional</td>
<td>Unidimensional, longest diameter, ≥10 mm (spiral CT) ≥20 mm other modalities</td>
<td>Unidimensional, longest diameter tumor lesions ≥10 mm (CT; skin by calipers) ≥20 mm if CXR</td>
</tr>
<tr>
<td>Measurable node definition</td>
<td>Not specifically noted</td>
<td>All measurable lesions up to 10 total (5 per organ), these are “target lesions”; all other lesions (measurable and non measurable) are nontarget</td>
<td>≥15 mm short axis</td>
</tr>
<tr>
<td>Disease burden to be assessed at baseline</td>
<td>All (not specified)</td>
<td>All measurable lesions up to five total (two per organ), these are target lesions; all other lesions (measurable and non measurable) are nontarget</td>
<td>All measurable lesions up to five total (two per organ), these are target lesions; all other lesions (measurable and non measurable) are nontarget</td>
</tr>
<tr>
<td>Baseline sum</td>
<td>Sum products of diameters bidimensional</td>
<td>Sum longest diameters all measurable lesions</td>
<td>Sum diameters all measurable lesions</td>
</tr>
<tr>
<td></td>
<td>Sum linear diameters unidimensional</td>
<td></td>
<td>(Short axis nodes, longest diameter others)</td>
</tr>
<tr>
<td>CR</td>
<td>Disappearance of all known disease</td>
<td>Disappearance of all known disease</td>
<td>Disappearance of all known disease; to be considered CR, malignant nodes must be &lt;10 mm</td>
</tr>
<tr>
<td>CR confirmation?</td>
<td>Yes, 4 wks</td>
<td>Yes, 4 wks</td>
<td>Yes (if response primary end point); 4 wks</td>
</tr>
<tr>
<td>CR duration</td>
<td>From date first documented until PD/relapse</td>
<td>From date first documented until PD/relapse</td>
<td>From date first documented until PD/relapse</td>
</tr>
<tr>
<td>PR</td>
<td>Measurable disease, 50% decrease in tumor load</td>
<td>Measurable disease (target lesions), 30% decrease in sum of longest diameters; all other disease, no evidence of progression</td>
<td>Measurable disease (target lesions), 30% decrease in sum of longest diameters; all other disease, no evidence of progression</td>
</tr>
<tr>
<td>PR confirmation?</td>
<td>Yes, 4 wks</td>
<td>Yes, 4 wks</td>
<td>Yes (if response primary end point); 4 wks</td>
</tr>
<tr>
<td>PR duration</td>
<td>From date of start of therapy to date of PD</td>
<td>From date first documented until PD/relapse</td>
<td>From date first documented until PD/relapse</td>
</tr>
<tr>
<td>Progression</td>
<td>Measurable disease, ≥25% increase in size of one or more measurable lesions or appearance of new lesions</td>
<td>Measurable disease, 20% increase in sum longest diameters, taking as reference smallest sum in study or appearance of new lesions</td>
<td>20% Increase in sum diameters, with minimum absolute increase of 5 mm, taking as reference smallest sum in study; or appearance of new lesions</td>
</tr>
<tr>
<td>No change/stable</td>
<td>Stable disease or non-PR and non-PD for 4 wks minimum</td>
<td>Non-PR, non-PD; minimum defined by protocol</td>
<td>Non-PR, non-PD; minimum defined by protocol</td>
</tr>
</tbody>
</table>

(Continued on the following page)
Table 2. Key features of WHO, RECIST 1.0, and RECIST 1.1 criteria (Cont’d)

<table>
<thead>
<tr>
<th></th>
<th>WHO (Miller 1981)</th>
<th>RECIST 1.0 (Therasse 2000)</th>
<th>RECIST 1.1 (Eisenhauer 2009)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Other</td>
<td>Guidance on reporting results</td>
<td>Appendices on imaging XR, CT, and justification for unidimensional; guidance on reporting results</td>
<td>Guidance on reporting results of phase II and III, PFS end point, identification of new lesion, further clarity on nonmeasurable sites; updated appendices on imaging, including MRI, PET, CT; addition of frequently-asked-question appendix and detailed list of changes from RECIST 1.0 to 1.1; special issue in which RECIST 1.1 published contains other articles supporting changes made</td>
</tr>
<tr>
<td>Special comments on bone metastases</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: PFS, progression-free survival; PET, positron emission tomography; MRI, magnetic resonance imaging; CT, computed tomography; CR, complete remission; XR, X-ray; CXR, chest X-ray; PD, progressive disease.

*By convention, bidimensional measurement is generally used for trials in which response is determined using the WHO criteria.

1 In practice, some groups changed this to 25% increase in sum of products of diameters, but original publication is defined as in table.

answered. For example, subjecting every patient to a tumor biopsy is fraught with difficulty. Surgical and medical complications from procedures remain an issue, especially wherein tissues are not easily accessible, leading to difficulty in obtaining usable paired biopsy samples. The prevalence of these procedural complications is not well documented. The search for patients with easily accessible tissue, on the other hand, leads to slow accrual. Thus, the requirement for tissue biopsy should be considered only when necessary, based on robust preclinical data and not as a hypothesis-generating exercise.

Dose schedule issues in phase II trials. Currently, oncology phase I trials define a dose for phase II testing, which can be inaccurate because of the small numbers of patients tested (18, 19). Finding the appropriate dose and schedule for anticancer agents for testing in phase II studies can be problematic. It may be misguided to use a recommended phase II dose determined in very few patients to a larger and more variable patient population in phase II and subsequent phase III trials. This is the approach taken, however, in standard phase II designs. An inappropriately low dose of an agent can contribute to negative phase III studies. This problem is suspected to have contributed to negative results with gefitinib (20), enzastaurin (19), and vatalanib (21). The alternative is using too high a dose in subsequent studies, resulting in significant toxicities, which may result in inappropriately abandoning an agent. It seems that, with dose reductions, however, toxicities are less likely to lead an effective agent being abandoned. For example, the initial approved dose of docetaxel was 100 mg/m² every 3 weeks, but the currently common used dose in practice is 75 mg/m². The original recommended phase II dose of pemetrexed was 600 mg/m². This dose was reduced to 500 mg/m² based on toxicity encountered in one of the early phase II studies (22). The issue of selecting the appropriate agents for phase II testing is particularly relevant with the current novel agents in testing, which happen to be mostly oral agents. It is difficult to determine the delivered dose with these agents because heterogeneity of transport proteins and metabolic enzymes affect drug exposure. To overcome these problems, rigorous dose-finding studies have to be undertaken after the initial determination of the maximum tolerated dose. At least two dose levels should be tested with associated pharmacokinetic studies to help identify the appropriate dose. Treating suitably sized patient population in this setting can give additional information about biomarkers, thus providing information on the relationship between dose and pharmacokinetic-pharmacodynamic markers. It should be noted that phase II dose-selection designs are relatively common in other areas of medicine such as cardiovascular, endocrine, and infectious diseases.

Methods of assessing tumor burden. A key feature in the evaluation of drug efficacy in phase II trials has been the method used to objectively assess changes in tumor burden after therapy. In an attempt to identify the most accurate assessment tool after anatomic imaging, the WHO criteria (33) were revamped to produce the RECIST (34) criteria, and improved RECIST criteria (RECIST version 1.1; ref. 35) have recently been introduced. A comparison of the key features of these three methods is shown in Table 2, and their evolution is described by Shankar and colleagues (10) in this issue of CCR Focus. A major point to remember as studies are designed is that WHO or RECIST criteria are simply methods of assessing tumor burden after therapy. The information garnered is then used by investigators to determine the study end point. Thus, an investigator may choose to use objective response rate assessed by RECIST as the primary end point or progression-free survival (which is also assessed by RECIST). In cases wherein progression-free survival has proven to be a more accurate predictor of clinical benefit (1), it behooves us to remember that the determination of disease progression was made by RECIST. Thus, the assessment tool was appropriate, as long as we chose the correct measure of clinical benefit.

Conclusions

As the number of agents with novel mechanisms of action are developed, the need to improve efficiency and the predictive power of phase II trials becomes tantamount. The evaluation of
new approaches and novel designs becomes imperative. This issue of CCR Focus examines a number of novel trial designs, with suggestions for future trial development. Although there is accumulating literature on innovative phase II designs, very few of these have been rigorously tested in clinical trials. It will be helpful to do case studies in which novel designs are used in a prospective trial (for example, Bayesian adaptive designs versus classical designs) to allow for comparative analysis. In addition, study designs such as randomized phase II designs and single arm designs with historical controls need to be prospectively compared so that there will be accumulated data guiding investigators on the designs to use for appropriate situations. Imaging modalities that accurately document clinical effects of agents in tumors, so that radiological response will correlate with survival benefit, need to be explored. A variety of imaging approaches may need to be used for different agents depending on their underlying mechanism of action. For example, digital contrast-enhanced magnetic resonance imaging has been used for inhibitors of angiogenesis. Finally, efforts should be focused on prospective validation of biomarkers in the laboratory (assy validation) and clinically. In the phase II setting, assay validation is particularly important, whereas clinical validation is best done in the phase III studies.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Appendix


Co-Chairs: Michaela Christian, M.D. (NCI CTEP); Alex Adjei, M.D., Ph.D. (Roswell Park); S. Percy Ivy, M.D. (NCI CTEP).

References

15. Vickers AJ. Do we want more cancer patients on clinical trials If so, what are the barriers to greater ac-curals. Trials 2008 Jun 3;9:31.


25. Koehne C, Bajetta E, Lin E, et al. Results of an interim analysis of a multinational randomized, double-blind, phase III study in patients (Pts) with previously treated metastatic colorectal cancer (mCRC) receiving FOLFOX4 and PTK787/ZK 222584 (PTK/ZK) or placebo (CONFIRM 2). J Clin Oncol (Meeting Abstracts) 2006;24:1485.


29. Blumenschein G, Ludwig C, Thomas G. A randomized phase III trial of tipifarnib compared to best supportive care (including hydroxyurea) in the treatment of newly diagnosed acute myeloid leukemia (AML) in patients 70 years or older. Presented at the 2007 American Society of Hematology Annual Meeting; December 8-11, 2007; Atlanta, GA.


