Published in final edited form as:

J Biopharm Stat. 2009 ; 19(3): 556–562. doi:10.1080/10543400902802474.

Current Issues in Oncology Drug Development, with a Focus on Phase II Trials

Daniel J. Sargent and

Division of Biomedical Statistics and Informatics, Mayo Clinic, Rochester, MN, 55905

Jeremy M.G. Taylor

Department of Biostatistics, University of Michigan, Ann Arbor, MI, 48019

Abstract

In this commentary we discuss several challenges that are of current relevance to the design of clinical trials in oncology. We argue that the compartmentalization of trials into the three standard phases, with non overlapping aims, is not necessary and in fact may slow the clinical development of agents. Combined phase I/II trials and/or phase I trials that at minimum collect efficacy data and more optimally include a preliminary measure of efficacy in dosing determination should be more widely utilized. Similarly, we posit that randomized phase II trials should be used more frequently, as opposed to the traditional historical single arm phase II trial that usually does not have a valid comparison group. The use of non binary endpoints is a simple modification that can improve the efficiency of early phase trials. The heterogeneity in scientific goals and contexts in early phase oncology trials is considerable, and the potential to improve the design to match these goals is great. Our overall premise is that the potential benefits associated with the oncology clinical trial community moving away from the one size fits all paradigm of trial design are great, and that more flexible and efficient designs tailored to match the goals of each study are currently available and being used successfully.

Keywords

Efficient designs; flexible adaptive designs; summarization of information; randomized phase II; phase I/II

The standard paradigm for the evaluation of new agents for the treatment of cancer is a phase I trial to establish the dose, a phase II trial to obtain preliminary evidence of efficacy; and a phase III trial to compare with a standard therapy. In this approach phases I and II are for gathering information, phase III is for a definitive comparison. These three phases of development represent an idealized scenario that was established more than 30 years ago; the compartmentalization of trials into the three phases with narrowly defined goals and data collection is a convenient simplifying approach. More recently, it has been recognized that it is not necessarily desirable to force a uniform approach that ignores the scientific context of the disease and the experimental agent. The current interest in adaptive designs, seamless Phase II/III studies, Phase I trials with an expansion phase, and Phase I/II trials suggest that the clinical trial community is exploring a variety of approaches beyond the simple compartmentalization of phase I, II, and III.

In practice there are many scientific and logistical reasons why the path of a single phase I, followed by a single phase II then a phase III trial is not exactly followed. The particular context may be better addressed using a different approach, such as if the agent under study is expected to be minimally toxic. Accumulating data from these and other studies may suggest deviations from the approach; issues related to funding and availability of patients also play important roles. Together, these factors suggest the need for greater flexibility and innovation in the design of these studies, particularly the early phase studies.

The typical phase I trials have been small studies that assess only toxicity with the goal of establishing the maximum tolerated dose. These studies have generally been too small to establish the dose reliably, especially since in the idealized paradigm there is no further investigation of the choice of dose in phase II and phase III. It is interesting to note that in other diseases randomized phase II trials are used to choose the dose for definitive testing in phase III. There has also been a surprising reluctance to collect efficacy data in phase I studies (Booth et al 2008 a). In our opinion, since the number of available patients is frequently limited, each patient should be recognized as a valuable resource and it would be wise to gather and utilize as much relevant information as possible from each patient, suggesting the need to collect efficacy data in phase I studies, as well as toxicity data. A phase I study is also one of the few opportunities to collect dose response data for efficacy as well toxicity, so it would seem wise to take advantage of this opportunity. The standard design (3+3) for phase I designs is routinely used, despite overwhelming evidence that more statistical model based designs are preferred (O'Quigley and Zohar Br J Can 2006, Zohar & Chevret, JBC 2008, Rogatko, JCO 2007, Onar et al JBS 2009). It does seem strange to us that investigators continue to use the same design irrespective of the details of the scientific context or the objectives.

For phase II studies the most prevalent design historically has been the single-arm study of the new agent(s), with a binary variable, tumor response, as the outcome measure. The trials may also include early stopping possibilities. The size of these trials is usually justified using hypothesis tests, comparing the tumor response rate to an assumed known historical response rate. Just like for phase I trials, many choices about the design are kept simple. For example, tumor response is a graded variable, yet it is reduced to binary in most trials despite the fact that this results in a loss of efficiency. The clinical trial community (Booth et al 2008) as well as the statistical community recognize this as a problem. While it is convenient to justify the size of the trial using hypothesis testing, after the trial is complete the preferred analysis would usually be summarization of the data in terms of estimates and measures of uncertainty. Also, it is quite likely that there will be other similar phase II studies of a new agent, and any decision to proceed to a phase III study would usually be based on the information about efficacy and other factors gained from all relevant studies, not just be based on a hypothesis test from a single phase II study with the binary measure of tumor response as the endpoint.

A common assumption in traditional phase II studies is that the response rate of the standard therapy is known from historical data. The accuracy with which the historical response rate is known is frequently quite limited. This uncertainty is typically not considered in designing the study, unless a Bayesian approach is used (Thall and Simon 1994, Taylor et al

2006). Changes in the standards of care in terms of therapy, disease assessment, and ancillary and supportive care make the historical data less useful for comparison purposes. Furthermore the response rate is usually very dependent on the disease severity of the population of patients. For any planned Phase II study there is no guarantee that the population of patients on that trial will be similar to that in the historical group. This problem with a lack of a valid comparison group can be overcome by randomizing patients to a standard treatment arm.

There are many additional reasons why patients in a single-arm Phase II study may not be comparable to those in some hypothetical historical group. Phase II trials involving new agents are typically undertaken in large academic medical centers, where the patient population may vary in many ways from those in a subsequent phase III trial (patients more mobile, more heavily pretreated, better socio-economic status, better supportive care). For new agents there is a natural enthusiasm amongst the investigators for the new agent and a desire for it to “look good”. This enthusiasm may manifest itself in various ways, such as setting the historical response rate at a low value (Thall and Estey 2005, Vickers et al 2007) or only enrolling patients who look in some sense ‘promising’. These aspects cause problems in an uncontrolled phase II study, but not in a randomized Phase II study.

A number of papers have discussed the pro and cons of randomized phase II studies (Rubenstein et al 2005, Wieand 2005, Booth et al 2008) and some (Liu et al 1999, Redman and Crowley 2007) have argued strongly against randomized phase II studies. The main argument against such trials is that randomized phase II studies are mistaken for phase III studies but they are substantially underpowered, thus a hypothesis tests at the end of the study may result in too many false negatives. Alternatively the alpha level can be increased to give more power, but at the expense of more false positives. If the study was the only Phase II trial to be conducted and a single hypothesis test was to be used as the basis to decide whether to start a Phase III trial, then this would be a valid concern. However, this is not the real situation. As indicated above there are typically a number of similar Phase II trials, and since the main goal of the phase II trial is to provide information on efficacy, the preferred analysis to summarize this information is to give estimates and uncertainty intervals, rather than p-values. With this analysis the amount of information provided by the data will be appropriately reflected in the width of the confidence interval, which is something the p-value does not provide.

Phase III studies are large randomized comparative experiments typically with time to event outcomes (e.g. overall survival) endpoints. These trials are designed to provide definitive information to guide clinical practice. Hypothesis testing is a valid approach for this setting. However, there has been a high failure rate amongst phase III trials. There are a number of factors that may be contributing to this; one is clearly that the quality of the data and evidence from phases I and II is not sufficient (Ratain and Karrison 2007).

The usual statistical tradeoff between bias and variance remains relevant in oncology clinical trials. In the large sample size setting of a phase III study, bias is the major concern, leading to the imperative of having randomized studies to ensure valid comparison groups. Phase I and II studies are smaller; for these variance will also be important. For a given

sample size, variance can be reduced by efficient analysis of the data; clearly this is an area where statisticians can have a major impact. Many of the articles in this issue are concerned with developing more efficient methodology for early phase trials.

In the setting of cancer clinical trial research in 2009, many of these challenges are directly relevant. The prior experience of having an inadequate number of agents to test, thus the need to minimize type II error, has been replaced in many settings with the case of too many drugs to test in a timely manner given limited patient resources. Novel agents with non-cytotoxic (and thus non response-inducing) mechanisms of action have reached the stage of clinical testing. Patient populations are being described not only by traditional standards of disease site and pathology, but also by status on one or more biomarkers (e.g Her2 status in breast cancer, Romond, NEJM 2005; KRAS mutation status in colon cancer Amado, JCO 2008). Novel endpoints, such as functional imaging results or levels of circulating tumor cells, are in development and testing. Each of these issues is addressed, in full or in part, by a collection of 5 papers on phase II studies in this special issue, the papers of Cheung, Jung, Chen, Sun, and Kocherginsky.

In 4 of these 5 papers, the principle of randomization in the phase II setting is either advocated for, or taken for granted as a necessary component. This is consistent with recent literature on the topic, which clearly has led to an increase in the use of randomization in the phase II setting (Rubinstein, JCO 2006, Ratain and Sargent, EJC, 2009). This need for randomization is driven by several factors: 1) the testing of agents in new patient populations defined by biomarkers, where historical data is not available, 2) the use of novel endpoints, which lack historical data, 3) the increased use of progression free survival as a phase II endpoint, which is more heavily influenced by factors beyond therapy than the traditional endpoint of tumor response, and 4) the high rate of phase III clinical trials that fail to achieve success on their primary endpoint, which is a tremendous cost to the clinical trials system. We agree in principle with the tenor of these papers, both for the reasons previously listed as well as the fact that the ongoing rapid evolution in standards of care in terms of therapy, disease assessment, ancillary and supportive care, as well as the differences in patient populations between phase II trials frequently done at a limited number of academic sites and phase III trials done in a very large number of centers make comparisons with historical controls intrinsically unreliable.

Two papers deal with the explicit issue of having more treatments available for testing than can be adequately evaluated using standard criteria. Chueng proposes conducting multiple arm screening trials with frequent monitoring to drop arms that are not promising, or to allow re-evaluation of the trial's sample size. The concept of multiple arm screening trials is not new, being proposed at least as far back as 1985 (Simon & Thall), with many subsequent variations (Thall, Simon, Ellenberg, 1988, Sargent and Goldberg, 2001). Cheung's suggested approach of using a sequential probability ratio test, with very frequent between arm comparisons and adaptive sample size estimation, may hold promise in settings where the administrative burden of such frequent analyses is acceptable; a comparison of this approach with a more standard group sequential testing approach would be insightful. Chen and Beckman propose a considerably more theoretical approach to the multitude of drug available for testing issue, using a cost-effectiveness model to evaluate different possible

strategies of drug development at a trial level, agent level, or franchise-level (i.e. considering multiple agents under development by a single sponsor with constrained resources). Chen and Beckman conclude, based on their modeling, that more, smaller trials are more cost-effective than fewer larger trials. A limitation of this approach is that the only costs considered in the phase III trial that would follow a phase II trial are based on the sample size of the trials; the loss of potential patient benefit, and revenue of the sponsor, from an effective agent that improperly terminated after phase II testing is not considered. As a concept, however, a strategic, model-based study of drug development processes is highly appropriate, and similar thinking has led to the field of clinical trials simulation, which is becoming increasingly popular (Holford, 2000).

Two other papers in this issue address issues in the more usual randomized phase II trial. Sun et al extend the work of Rubinstein et al (JCO 2006) to apply the usual randomized phase II design to allow dual endpoints; in this particular situation, both tumor response and early disease progression. Sun et al advocate high alpha and beta error rates of 0.20, which we feel are appropriate in the randomized phase II setting. Sun et al do propose performing interim monitoring at one time point of the trial on the tumor response variable, although early progression may be more indicative of harm and could be considered as appropriate to monitor. In addition, it is unclear if a simpler approach of assessing progression free survival, which captures perhaps the more biologically meaningful patient-level effect of therapy (delaying disease worsening), might be adequate. Jung and George propose an alternative estimation approach in the randomized phase II trial based on the UMVUE; a more direct comparison of this approach to standard approaches would be enlightening. In addition Jung and George assume the primary comparison of interest in the randomized phase II trial is an individual comparison of each arm to historical controls, which in our opinion, does not take full advantage of the randomization.

The final phase II design paper of Kocherginsky et al propose a design to address the same fundamental premise of Sun et al, that is, that we must go beyond the traditional endpoint of tumor response. Kocherginsky propose a dual endpoint trial with both tumor response and disease stabilization as endpoints; similar to Sun et al, but propose this in a single arm phase II trial.

As a whole, this set of papers presents potentially useful solutions to some of the issues facing the cancer clinical trials and drug development arena. The need to strategically consider drug development as an integrated program as opposed to a collection of isolated studies, the benefits in many cases of randomization earlier in the drug development process, and the need for new endpoints all are challenging the standard paradigms. Continued research is clearly warranted as the statistical and clinical communities strive to bring the most effective therapies to patients, in an as quick and cost-effective manner as possible.

References

- Amado RG, Wolf M, Peeters M, Van Cutsem E, Siena S, Freeman DJ, Juan T, Sikorski R, Suggs S, Radinsky R, Patterson SD, Chang DD. Wild-type KRAS is required for panitumumab efficacy in patients with metastatic colorectal cancer. *J Clin Oncol*. 2008; 26(10):1626–1634. [PubMed: 18316791]

- Booth CM, Calvert AH, Giaccone G, Lobbezoo MW, Eisenhauer EA, Seymour LK. on behalf of the task force on methodology for the development of innovative cancer therapies. Design and conduct of phase II studies of targeted anticancer therapy: Recommendations from the task force on methodology for the development of innovative cancer therapies (MDICT). *European J of Cancer*. 2008; 44:25–29. [PubMed: 17845846]
- Booth CM, Calvert AH, Giaccone G, Lobbezoo MW, Seymour LK, Eisenhauer EA. on behalf of the task force on methodology for the development of innovative cancer therapies. Endpoints and other considerations in phase I studies of targeted anticancer therapies: Recommendations from the task force on methodology for the development of innovative cancer therapies (MDICT). *European J of Cancer*. 2008; 44:19–24. [PubMed: 17890079]
- Cheung YK. Selecting promising treatments in randomized phase II cancer trials with an active control. *JBS*.
- Chen C, Beckman RA. Optimal dose-effective designs of phase II proof of concept trials and associated go-no go decisions. *JBS*.
- Holford NH, Kimko HC, Monteleone JP, Peck CC. Simulation of clinical trials. *Annu Rev Pharmacol Toxicol*. 2000; 40:209–234. [PubMed: 10836134]
- Jung S-H, George SL. Between-arm comparisons in randomized phase II trials. *JBS*.
- Kocherginsky M, Cohen EEW, Karrison T. Design of phase II cancer trials for evaluation of cytostatic/cytotoxic agents. *JBS*.
- Liu PY, LeBlanc M, Desai M. False Positive Rates in Randomized Phase II Designs. *Controlled Clinical Trials*. 1999; 20:343–352. [PubMed: 10440561]
- Onar A, Kocak M, Boyett JM. Continual Reassessment Method vs. Traditional Empirically-Based Design: Modifications Motivated by Phase I Trials in Pediatric Oncology by the Pediatric Brain Tumor Consortium. *JBS*.
- O'Quigley J, Zohar S. Experimental Designs for Phase I phase I/II Dose-Finding Studies. *Br J Cancer*. 2006; 94:609–613. [PubMed: 16434987]
- Ratain MJ, Karrison TG. Testing the Wrong Hypothesis in Phase II Oncology Trials: There is a Better Alternative. *Clin Cancer Research*. 2007; 13(3):781–782.
- Redman M, Crowley J. Small randomized Trials. *Journal of Thoracic Oncology*. 2007; 2(1):1–2. [PubMed: 17410001]
- Rubenstein LV, Korn EL, Freidlin B, Hunsberger S, Ivy SP, Smith MA. Design Issues of Randomized Phase II Trials and a Proposal for Phase II Screening Trials. *J of Clinical Oncology*. 2005; 23–28:7199–7206.
- Rogatko A, Schoeneck D, Jonas W, Tighiouart M, Khuri FR, Porter A. Translation of innovative designs into phase I trials. *J Clin Oncol*. 2007; 25(31):4982–4986. [PubMed: 17971597]
- Romond EH, Perez EA, Bryant J, Suman VJ, Geyer CE Jr, Davidson NE, Tan-Chiu E, Martino S, Paik S, Kaufman PA, Swain SM, Pisansky TM, Fehrenbacher L, Kutteh LA, Vogel VG, Visscher DW, Yothers G, Jenkins RB, Brown AM, Dakhil SR, Mamounas EP, Lingle WL, Klein PM, Ingle JN, Wolmark N. Trastuzumab plus adjuvant chemotherapy for operable HER2-positive breast cancer. *N Engl J Med*. 2005; 353(16):1673–1684. [PubMed: 16236738]
- Ratain MJ, Sargent DJ. Optimising the design of phase II oncology trials: The importance of randomization. *European Journal of Cancer*. 2008 to appear.
- Simon R, Thall PF, Ellenberg SS. New designs for the selection of treatments to be tested in randomized clinical trials. *Stat Med*. 1994; 13(5–7):417–429. [PubMed: 8023026]
- Simon R, Wittes RE, Ellenberg SS. Randomized phase II clinical trials. *Cancer Treat Rep*. 1985; 69(12):1375–1381. [PubMed: 4075313]
- Sargent DJ, Goldberg RM. A flexible design for multiple armed screening trials. *Stat Med*. 2001; 20(7):1051–1060. [PubMed: 11276035]
- Sun LZ, Chen C, Patel K. Optimal two-stage randomized multinomial designs for phase II oncology trials. *JBS*. to appear this issue.
- Taylor JMG, Braun T, Li Z. Comparing an Experimental Agent to a Standard Agent: Relative Merits of a One-Arm or Randomized Two-Arm Phase II Design. *Clinical Trials*. 2006; 3(4):335–348. [PubMed: 17060208]

- Thall PF, Estey EH. Some Ethical Issues in Phase II Trials in Acute Leukemia. *Clinical Advances Hem & Onc.* 2005; 3(12):943–948.
- Vickers AJ, Ballen V, Scher HI. Setting the Bar in Phase II Trials: The Use of Historical Data for Determining “Go/No Go” Decision for Definitive Phase III Testing. *Clin Cancer Research.* 2007; 13(3):792–796.
- Wieand HS. Randomized Phase II Trials: What Does Randomization Gain? *J of Clinical Oncology.* 2005; 23(9):1794–1795.
- Zohar S, Chevret S. Recent developments in adaptive designs for phase I/II dose-finding studies. *J Biopharm Stat.* 2007; 17(6):1071–1083. [PubMed: 18027217]