Randomized Phase II Trials

Time for a New Era in Clinical Trial Design

Sumithra J. Mandrekar, PhD, and Daniel J. Sargent, PhD

Abstract: The classic single-arm oncology phase II trial designs for evaluating an experimental regimen/agent are limited by multiple sources of bias arising from the inability to separate trial effects (such as patient selection, trial eligibility, imaging techniques and assessment schedule, and treatment locations) from treatment effect on clinical outcomes. Changes in patient population based on biologic subsetting, newer imaging technologies, the use of alternative end points, constrained resources, and the multitude of promising therapies for a given disease make randomized phase II designs, with a concurrent control arm where necessary, attractive. In this brief report, we discuss the salient features of the randomized designs for phase II trials, which when properly applied under the constraints of their underlying inference framework can assure optimal use of limited phase III financial and patient resources.

Key Words: False-positive, False-negative, Phase II, Randomized, Screening, Selection.

Phase II clinical trials are designed to identify promising experimental therapies that can then be tested further in a definitive phase III trial. With the recent explosion of molecularly targeted agents in oncology, frequently resulting in multiple agents aiming at the same target, coupled with the growth in combination therapies for cancer treatment, there is a definite need to evaluate the efficacy of multiple regimens quickly and concurrently. The spiraling costs of phase III trials demand that the failure rate of such trials be reduced. Changes in patient population based on biologic subsetting and evolution in imaging technologies make comparison against historical controls inaccurate. Moreover, with new biologic agents, end points such as tumor response are no longer useful because many of these agents are cytostatic; the subsequent appropriate use of alternative end points such as progression-free survival in a phase II setting imply a greater need for concurrent control. The classic single-arm phase II trial designs for evaluating each experimental regimen/agent

Department of Health Sciences Research, Mayo Clinic, Rochester, Minnesota.
Disclosure: The authors declare no conflicts of interest.
Address for correspondence: Sumithra J. Mandrekar, PhD, Department of Health Sciences Research, Mayo Clinic, 200 1st Street SW, Rochester, MN 55905. E-mail: mandrekar.sumithra@mayo.edu

individually are limited by outcome-trial effect confounding arising from the inability to separate trial effects (such as patient selection, trial eligibility, imaging techniques and assessment schedule, and treatment locations) from treatment effect on clinical outcomes. Designs with randomization (using stratification or dynamic allocation where necessary) to experimental regimens/agents, using a concurrent control arm when necessary, offer an attractive proposition by assuring better patient comparability and reducing outcome-trial effect confounding.[1–4] Moreover, randomized phase II designs greatly enhance the potential for biomarker discovery, which is an important first step toward the aim of personalized medicine.

TYPES OF RANDOMIZED PHASE II DESIGNS

The design of a clinical trial is largely driven by three statistical parameters: (1) α, the type I error or probability of a false-positive result, (2) β, the type II error or probability of a false-negative result, and (3) δ, the targeted difference or targeted effect size. The sample size is determined to detect δ with a significance level of α and power of $(1 - \beta) \times 100\%$. The randomized phase II designs are differentiated by the choice of the values for these statistical parameters, which is dictated by the inference framework of the design.

Randomized phase II designs fall into one of the following three categories: (1) randomization to parallel noncomparative single-arm experimental regimens each with independent decision rule; (2) randomized selection (or pick the winner) designs for selecting the most promising experimental regimen among several similar experimental regimens; and (3) randomized screening design for comparing an experimental regimen to standard of care.[3–7] We review the salient features of each of these designs below.

Randomization to Parallel Noncomparative Regimens

The first class of randomized designs includes randomization to two or more experimental treatment arms in which the randomization is primarily for the purpose of reducing various types of bias, including patient selection bias and controlling for known or unknown baseline imbalances across the arms. Each individual treatment arm within the randomized phase II design is structured as an independent phase II study with determination of "promising activity" based on a comparison against historical control with appropriate thresholds for α (typically 0.1) and β (typically 0.1).

The arms have independent decision rules including rules for early termination for lack of efficacy. Such a design would be useful in the concurrent evaluation of two or more experimental regimes, with no direct comparison, such that each regimen that meets the success criteria has the potential to be tested further in a larger trial.

Although this is not a common design in most disease settings, they are nevertheless attractive in early phase II situations where there is a reliable early end point to demonstrate success (such as tumor response) and that success on that end point is directly attributable to the experimental regimen in question (i.e., a single-agent trial). Success in such a trial for the most part still dictates the need for a more thorough evaluation of those regimens that show promise in a phase IIb setting, with a direct comparison of safety and efficacy outcomes between the randomized arms.

Randomized Selection Designs

The second class of randomized phase II designs was first introduced by Simon et al.[5] with the aim of choosing the most promising experimental regimen from among similar ones using a ranking and selection approach. The experimental regimen(s) selected as the most promising is then compared with the standard of care in a subsequent larger phase III trial. Scenarios in which such a design would be useful include comparing different modes of drug administration or dosing schedules or comparing different combination regimens, all of which have a new experimental agent added to a common core regimen.

Selection designs are designed to make a prioritization between promising "experimental" regimens when there is no a priori data to prefer one regimen over the other. In this design, patients are randomized to two or more "competing" regimens/agents. The final results are then ranked, and the arm with the best observed outcome is selected for further study. The sample size requirements for this design are based on providing a high probability of choosing the best arm as long as the expected outcome in that arm exceeds any other arm by a clinically meaningful margin (e.g., at least 15%). This design does not provide answers concerning the relative merits of similar regimens because it does not test the null hypothesis of equality. This design approach was used by Lustberg et al.[8] to make a selection between two doses of Mitomycin C followed by irinotecan in patients with advanced esophageal and gastroesophageal junction adenocarcinomas. The trial used a two-stage Simon design with individual decision rules for efficacy for each experimental arm with α and β of 0.1. The final results from the two arms were ranked to make a recommendation that the low-dose arm was both well tolerated and efficacious.

The selection of an experimental treatment in a screening design can be based solely on the primary end point or can include other factors when the observed difference in the primary outcome is deemed "small."[5–7] This is a flexible selection design in which other factors such as safety profile, cost, convenience, or quality of life in addition to the primary efficacy measure are taken into consideration in making the selection, much similar to clinical practice.[6,9]

Randomized Screening Designs

The third class of designs is the randomized screening designs for performing a nondefinitive comparison of one or more experimental regimes against the standard of care treatment in a phase II setting. Such an approach was used in the evaluation of two doses of bevacizumab combined with carboplatin and paclitaxel (two experimental arms) versus carboplatin and paclitaxel alone (concurrent control) in previously untreated patients with non-small cell lung cancer.[10] The promising results from this randomized phase II trial for the high-dose bevacizumab arm led to the pivotal phase III trial that established the efficacy of bevacizumab plus carboplatin and paclitaxel in chemotherapy-naive non-small cell lung cancer patients.[11]

Rubinstein et al.[3] formally introduced the paradigm of conducting preliminary and nondefinitive comparisons of experimental regimens to standard of care by carefully selecting the statistical parameters of α, β, and δ such that there is a high chance for identifying nonpromising regimens and taking forward promising regimens for further testing. These designs should not be viewed as a replacement for a definitive phase III trial but rather as a tool to help prioritize experimental regimens using an intermediate end point such as tumor response or progression-free survival for a subsequent more definitive evaluation.

The choice of the three statistical parameters in this design is critical so the sample size is reasonable (typically around 100 patients), and the results are meaningful. Specifically, an overly large false-positive rate ($\alpha > 0.20$) has the risk of increasing the likelihood of negative phase III trials; a high value of β (> 0.20), the false-negative rate, has the risk of terminating further testing of a potentially promising regimen, and a high value of the targeted difference, δ has the risk of rejecting a potentially clinically beneficial regimen. Rubinstein et al.[3] recommend choosing 0.20 for both α and β in a screening setting and a target difference of 20% (or hazard ratio of 1.5). Rubinstein et al.[3] provide a detailed discussion of the possible choice of values for α, β, and δ for a screening design and its impact on the sample size and trial results interpretation.

SUMMARY

Randomized phase II designs are gaining considerable momentum in the current era of constrained resources coupled with the multitude of promising therapies for a given disease. The wealth of opportunities in cancer drug development mandates intelligent clinical trial design. Randomized phase II designs, when properly applied under the constraints of their underlying inference framework, can assure optimal use of limited phase III financial and patient resources.

ACKNOWLEDGMENTS

933

REFERENCES

1. Taylor JMG, Braun TM, Li Z. Comparing an experimental agent to a standard agent: relative merits of a one-arm or randomized two-arm phase II design. *Clin Trials* 2006;3:335–338.
2. Wieand HS. Randomized phase II trials: what does randomization gain? *J Clin Oncol* 2005;23:1794–1795.
3. Rubinstein LV, Korn EL, Freidlin B, et al. Design issues of randomized phase II trials and a proposal for phase II screening trials. *J Clin Oncol* 2005;23:7199–7206.
4. Rubinstein L, Crowley J, Ivy P, et al. Randomized phase II designs. *Clin Cancer Res* 2009;15:1883–1890.
5. Simon R, Wittes RE, Ellenberg SE. Randomized phase II clinical trials. *Cancer Treat Rep* 1985;69:1375–1381.
6. Sargent DJ, Goldberg RM. A flexible design for multiple armed screening trials. *Stat Med* 2001;20:1051–1060.
7. Steinberg SE, Venzon DJ. Early selection in a randomized phase II clinical trial. *Stat Med* 2002;21:1711–1726.
8. Lustberg MB, Bekaii-Saab T, Young D, et al. Phase II randomized study of two regimens of sequentially administered mitomycin C and irinotecan in patients with unresectable esophageal and gastroesophageal adenocarcinoma. *J Thorac Oncol* 2010;5:713–718.
9. Mandrekar SJ, Sargent DJ. Pick the winner designs in phase II cancer clinical trials. *J Thorac Oncol* 2006;1:5–6.
10. Johnson DH, Fehrenbacher L, Novotny WF, et al. Randomized phase II trial comparing bevacizumab plus carboplatin and paclitaxel with carboplatin and paclitaxel alone in previously untreated locally advanced or metastatic non-small-cell lung cancer. *J Clin Oncol* 2004;22:2184–2191.
11. Sandler A, Gray R, Perry MC, et al. Paclitaxel-carboplatin alone or with bevacizumab for non-small-cell lung cancer. *N Engl J Med* 2006;355:2542–2550.