Published in final edited form as:

Clin Trials. 2015 April ; 12(2): 128–138. doi:10.1177/1740774514562029.

Suspension of Accrual in Phase II Cancer Clinical Trials

Yimei Li^{1,2}, Rosemarie Mick¹, and Daniel F. Heitjan¹

¹Department of Biostatistics & Epidemiology, University of Pennsylvania Perelman School of Medicine

²Division of Oncology, The Children's Hospital of Philadelphia

Abstract

Purpose—Phase II cancer clinical trial designs commonly incorporate an interim analysis for lack of efficacy. To strictly and ethically implement such designs, one should suspend accrual in cases where pending patient outcomes can affect early termination decisions. This paper aims to evaluate various options for accrual suspension and illustrate how the suspension strategy affects operating characteristics of the trial.

Methods—We define a strict suspension strategy for determining whether one should continue, suspend, or restart accrual at any point within the trial. The strategy is compared to a naïve implementation of suspension and a strategy of no suspension. We evaluate the methods' operating characteristics by simulation.

Results—The suspension strategy has little effect on type I error, power, and early termination probability. Methods that involve stricter suspension policies generally lead to smaller but longer trials. Differences across strategies are substantial when the ratio of enrollment rate to outcome availability rate is high.

Conclusions—The suspension strategy is most relevant in trials that accrue rapidly and require lengthy observation of each subject. The choice of suspension strategy involves a tradeoff between the cost of implementing a potentially complex suspension algorithm in real time vs. the cost of enrolling more patients and exposing them to a potentially toxic and ineffective treatment regimen.

Keywords

Accrual; cancer; clinical trial; phase II; suspension

Introduction

One conducts a phase II clinical trial to obtain a preliminary evaluation of the efficacy of an experimental therapy. A common design in oncology is a single-arm trial with the primary endpoint clinical response, defined in terms of the degree of tumor shrinkage.¹ Designs

Correspondence: Yimei Li PhD, Colket Translational Research Building 4032, 3501 Civic Center Boulevard, Philadelphia, PA 19104, USA. yimeili@mail.med.upenn.edu.

Declaration of interest:

All authors declare that they have no conflict of interest.

typically incorporate one or more pre-planned interim analyses that allow early termination when evidence of activity is lacking.²⁻⁷

A standard two-stage trial proceeds to the second stage if the number of responses in an initial n_1 subjects exceeds a designated critical value r_1 . The second stage involves enrolling a further $n-n_1$ patients; if the number of responses in the total n subjects exceeds a critical value r the regimen proceeds to further testing. One selects n_1 , r_1 , n and r to satisfy desired statistical properties. For example, the commonly used Simon optimal design minimizes the expected total sample size under a null hypothesis response rate, subject to constraints on type I and type II error probabilities.⁶

Strict implementation of such a design may require suspension of accrual when the interim analysis sample size has been reached but the outcomes are not yet available on all enrolled patients.^{1,8} For example, consider a design with $n_1=17$, $r_1=3$, $n=37$ and $r=10$; that is, a design that specifies an interim analysis after the first 17 patients have been observed, rejecting the regimen if there are 3 or fewer responses. If after 16 patients there are exactly 3 responses, then the trial will continue only if patient 17 is also a responder. But suppose that after enrolling patient 17, whose outcome will not be known for, say, three months, potential subject 18 appears and consents to participate. Should patient 18 be turned away on ethical grounds? And what if we enroll patient 18 but later observe that patient 17 does not respond, and the interim analysis therefore suggests that the regimen is inactive and the trial should have been terminated? And if moreover patient 18 is observed to respond, does this not cast doubt on our decision to terminate after only 3 responses in the first 17 patients? An unambiguous strategy for the avoidance of such conundrums is to suspend accrual after the enrollment of patient 17, restarting accrual only if patient 17 responds. Indeed, as we demonstrate below, a careful implementation of such a two-stage design could require multiple suspensions.

Suspending trials in this way avoids the ethical difficulties and ambiguities described above but creates a further complex of potential problems. Most importantly, it complicates trial operation, as there must be real-time monitoring of study data to know whether and when to suspend; this not only adds to the work load of trial staff, but through its complexity invites errors in protocol execution. Secondly, frequent suspension of enrollment may dampen the enthusiasm of participating investigators and stall study momentum. Thirdly, the occurrence of multiple temporary stops could be informative about the unfolding outcomes of the ongoing trial, defeating measures to control the release of trial results. And finally, the occurrence of suspensions will lengthen the trial.

In this article we undertake a detailed study of accrual suspension in two-stage phase II cancer trials. We compare a range of suspension strategies: A naïve strategy that suspends only after the first stage is fully enrolled, a strict strategy that terminates the trial whenever data are sufficient to justify a conclusion and suspends accrual whenever the results of pending patients could determine the conclusion, and a strategy of continuous enrollment without suspension. We use simulations to evaluate the performance of these methods with respect to type I error, power, probability of early termination, trial duration, total

enrollment, probability of suspension(s), and the number of patients exposed unnecessarily to an ineffective treatment.

Methods

Design framework

A typical phase II cancer trial aims to test the null hypothesis $H_0: p \leq p_0$ against the alternative $H_1: p > p_1$, where p is the regimen's unknown true response rate, p_0 is the maximum uninteresting response rate, and p_1 is the target response rate one hopes the regimen to achieve. The trial proceeds to stage 2 only if more than r_1 of the first n_1 patients exhibit clinical responses. The second stage enrolls up to a sample size of n and deems the treatment a success if the total number of responses exceeds r .

Design characteristics include type I error, the probability of falsely accepting a regimen with a true response rate p_0 ; power, the probability of correctly accepting a regimen with a true response rate p_1 ; probability of early termination $PET(p)$, the probability of terminating the trial at the first stage under true response rate p ; and expected sample size $EN(p)$, the average number of subjects to be enrolled in a hypothetical series of replications of the study assuming true response rate p : $EN = n_1 + (1 - PET) \times (n - n_1)$. Typically we specify design parameters to control the type I error rate at level α (often 5% or 10%) with power at least $1 - \beta$ (often 80% or 90%). The Simon optimal design, for example, is the design with the smallest possible $EN(p_0)$ that satisfies the designated type I error rate and power inequalities.

Suspension of accrual

The implementation of such a design may require suspension of accrual while awaiting outcomes of the enrolled subjects. The need to suspend may occur as we approach the end of either stage 1 or stage 2. We let X denote the number of responses observed currently; N the number of available outcomes (subjects whose response status has been observed); N_P the number of pending outcomes (subjects whose response status has not yet been determined); N_L the number of subjects left to be enrolled, which equals the targeted sample size (i.e., n_1 for stage 1 and n for stage 2) minus the number of enrolled subjects; and R the stopping criterion (equal to r_1 for stage 1 and r for stage 2).

We illustrate suspension in an example of Simon's optimal design for $p_0=0.2$, $p_1=0.4$, $\alpha=\beta=0.1$, which yields $n_1=17$, $r_1=3$, $n=37$, $r=10$. Assume we have enrolled 16 subjects, among whom 15 already have an outcome, 3 of them responses, and 1 is pending, and that at this moment subject 17 is identified; that is, $N=15$, $N_P=1$, $N_L=1$, and $R=3$.

Naïve implementation with suspension (abbreviated as NaïveS)

In this strategy, we continue enrolling until we have n_1 subjects, at which point we suspend until all have provided outcome data. Depending on the interim analysis findings, we either terminate the trial early or re-open accrual and proceed to stage 2. Subjects who arrive during the suspension are turned away. With this approach, in the example scenario,

regardless of X we enroll #17 and then suspend until the outcomes of subjects 16 and 17 are revealed.

Strict implementation with suspension (StrictS)

NaïveS is unethical under some scenarios, because it permits accrual of subjects when there is no hope that the trial will continue, and may suspend accrual when it is already established that the trial should continue. A strict strategy would adjust decisions depending on the current value of X . For example, consider these scenarios based on outcomes in the first 15 subjects:

1. With 0 or 1 responses, we terminate the trial before enrolling subject 17, because there is no hope of achieving 4 or more responses.
2. With 2 responses, we suspend temporarily and re-open only if subject 16 responds, in which case we need the result of subject 17 to determine whether to proceed; if subject 16 does not respond, we terminate the trial without enrolling subject 17.
3. With 3 responses, we enroll subject 17 without suspension. This is because if subject 16 is a response, the trial will continue to the second stage, whereas if subject 16 is not a response, we need the result of subject 17 to determine whether to proceed. After enrolling subject 17, we suspend accrual until his result becomes known.
4. With 4 or more responses, we enroll subject 17 and proceed to stage 2 without suspension, because we have already reached the passing criterion regardless of the outcomes of subjects 16 and 17.

Implementation of this strict strategy demands timely and careful collection and examination of the accruing data as the close of either stage approaches.

We summarize the process into a general decision algorithm that one should apply in real time in any two-stage trial if the objective is to adhere strictly to the stopping rule with minimal potential exposure of patients to an ineffective therapy and maximal efficiency once passage of the interim criterion is established (Table 1):

1. If $X \geq R - N_P - N_L$, it is impossible to satisfy the continuation criterion; we terminate the trial.
2. If $R - N_P - N_L < X \leq R - N_L$, it is still possible, but not certain, that the trial will satisfy the continuation criterion; as our decision will hinge on the outcomes of the pending subjects, we suspend accrual and re-evaluate when the next outcome becomes available.
3. If $R - N_L < X \leq R$, it is possible that the trial will satisfy the continuation criterion, depending only on the outcomes of the as-yet unenrolled subjects; therefore we continue accrual, re-evaluating when either the next outcome becomes available or the next patient presents for accrual, whichever occurs first.

4. If $X > R$, if the trial is in stage 1 we complete stage 1 accrual and continue to stage 2 without suspension, or resume accrual if previously suspended; if in stage 2 we accept the new regimen and complete stage 2 accrual.

Strict implementation without suspension (StrictWoS)

Application of *StrictS* raises logistical difficulties, as one must enter, check, transmit and analyze accrual and outcome data, and broadcast notices of the imposition and lifting of suspensions — all in real time. An intermediate approach is to apply the above decision rule but without the possibility of suspension (i.e. removing rule 3). That is, we would evaluate the data for possible early termination, but not interrupt recruitment in cases of ambiguity.

Such a design would continue enrolling subjects until we have observed the outcomes of the first n_I enrolled patients, by which time we may already have enrolled many more subjects. For example, if we have enrolled 22 subjects by the time we observe the outcomes of the first 17 enrolled patients, then the 5 additional subjects could have been enrolled incorrectly, depending on the outcomes in the first 17. Thus, such a strategy avoids the complications of suspension, at the risk of unnecessarily exposing more than the minimum number of subjects to an inactive drug. It also invites ambiguity in cases where the early termination criterion is met but subsequent outcomes push the observed response rate above the maximum level for early termination.

Simulations

We conducted simulations to investigate the performance of the three strategies with respect to type I error, power, PET, total trial duration, and EN. Moreover for *StrictS* we evaluated the probability of suspending at least once, the probability of suspending more than once, and the probability of terminating the trial prior to enrolling n_I patients. For *NaïveS* and *StrictWoS*, we evaluated the average number of patients treated incorrectly, where a patient is considered incorrectly treated if he or she was enrolled when the trial should have been suspended (under *StrictS*), but later results suggested the drug was inactive. We simulated results for Simon's optimal designs, but results are generalizable to any two-stage design.

We considered various settings of the design parameters, with p_0 ranging from 0.05 to 0.60, p_I ranging from 0.25 to 0.80, and $p_I - p_0 = 0.15$ and 0.20. Type I error rate is set at 10% and power at 90%.

We assume an enrollment rate of λ patients per month, and that it takes m months to determine a subject's response status. Thus, $1/m$ is the rate of outcomes becoming available (the number of outcomes available per month per pending patient), analogous to the enrollment rate. The ratio of enrollment rate and outcome availability rate (i.e., the product λm) determines the relative value of suspension. Therefore we restrict attention to three cases: $\lambda m = 1$ ($\lambda = 1$ and $1/m = 1$), representing slow enrollment and rapid outcome availability; $\lambda m = 6$ ($\lambda = 2$ and $1/m = 1/3$), representing moderate enrollment and moderate outcome availability; and $\lambda m = 24$ ($\lambda = 4$ and $1/m = 1/6$), representing rapid enrollment and slow outcome availability. We simulated 4,000 trials in each scenario.

We performed all computations in R 3.0.2; code is available from the first author.

Results

To illustrate simulation findings we continue with the example of the Simon optimal design with $\alpha=\beta=0.1$, $p_0=0.2$ and $p_I=0.4$, which yields $n_I=17$, $r_I=3$, $n=37$, $r=10$, type I error rate 9.5%, power 90.3%, $PET(p_0)=0.55$, and $EN(p_0)=26.0$. Because we found that type I error rate, power, and PET are robust to suspension strategy (Supplemental Table 1), we henceforth focus on other design properties. For all three strategies, we evaluated total trial duration and average total enrollment EN. For the *StrictS* strategy, we evaluated the probability of suspending accrual and the probability of terminating the trial prior to enrolling n_I patients. For *NaïveS* and *StrictWoS*, we evaluated the average number of patients treated incorrectly.

Low ratio of enrollment to outcome availability

Figure 1 shows the results under a low ratio of enrollment rate λ to outcome availability rate $1/m$: specifically, $\lambda m=1$, with slow enrollment ($\lambda=1$ patient/month) and fast outcome availability ($m=1$ month to evaluate response). Figures 1a and 1b display trial duration and number of patients enrolled against true response rate for the three strategies. Trial duration increases as the true response rate increases, because when the response rate is low the trial is more likely to terminate at stage 1. Trial duration is generally similar across strategies. The number of patients enrolled increases as true response rate increases. The average total enrollment is also similar across designs, and close to the theoretical value $EN(p_0)$ (represented by a solid dot on the curves). These results are unsurprising, because with slow enrollment and fast outcome availability, outcomes are likely to be available before the next eligible patient appears, obviating suspension.

Figure 1c shows additional parameters for *StrictS*. The probability of suspending at least once is high when the true response rate is low (0.94 when $p=p_0=0.2$), and quickly decreases as the true response increases (to 0.18 when $p=p_I=0.4$). The probability of suspending more than once also declines with increasing p , with a value of 0.70 when $p=p_0=0.2$ and 0.12 when $p=p_I=0.4$. There is a moderate probability of terminating the trial prior to enrolling n_I patients, with a value of 0.34 at $p=p_0=0.2$ and 0.02 at $p=p_I=0.4$.

Figure 1d shows the number of patients treated incorrectly for *NaïveS* and *StrictWoS*. In this scenario on average few patients are treated incorrectly, with a maximum of 0.9 for *NaïveS* and 1.1 for *StrictWoS*.

Moderate ratio of enrollment to outcome availability rate

Figure 2 shows the results under a moderate ratio of enrollment rate λ to outcome availability rate $1/m$: specifically, $\lambda m=6$, with intermediate enrollment ($\lambda=2$ patients/month) and outcome availability ($m=3$ months to evaluate response). As shown in Figure 2a, trial durations are similar across designs when the true response rate is low, because all designs terminate at the first stage. However when the response rate is high, *NaïveS* leads to longer trials because it always suspends enrollment after reaching n_I patients, whereas the other two strategies do not suspend if many responses are observed in the beginning. *StrictWoS* gives the shortest trial duration because it never suspends. For the same reason, as shown in

Figure 2b, *StrictWoS* leads to the largest number of patients enrolled, as it continues accrual while awaiting pending outcomes. In contrast, *StrictS* leads to the smallest mean total enrollment because it never enrolls unnecessarily (e.g., it stops the trial before enrolling n_I subjects such as when $X/N=0/14$; indeed, for this reason its $EN(p_0)$ can even be less than that of the theoretical Simon bound). Differences in patient numbers disappear when the true response rate is high, because in such cases the trial commonly proceeds to stage 2 without hesitation and ultimately enrolls the full n subjects.

Figure 2c shows that with *StrictS* the trial almost always suspends at least once when p is low (e.g., with probability 0.995 when $p=p_0=0.2$), with the probability becoming smaller as the true response increases (0.57 when $p=p_I=0.4$). In fact the trial almost always suspends multiple times when the true response proportion is low (with probability 0.992 at $p=p_0=0.2$) but less often at higher response rates (0.49 at $p=p_I=0.4$). Again there is a moderate probability of terminating the trial prior to enrolling n_I patients: 0.34 at $p=p_0=0.2$ and 0.02 at $p=p_I=0.4$. Note that these values are the same as those observed in the first scenario, because this property does not depend on enrollment and outcome availability rates (similar to PET in the Supplementary Table). Figure 2d shows that in the scenario of moderate enrollment and outcome availability, *NaïveS* and *StrictWoS* could incorrectly treat a few patients when the true response rate is low; average numbers are 2.3 for *NaïveS* and 5.5 for *StrictWoS* at $p=p_0=0.2$, declining to 0.2 for *NaïveS* and 0.5 for *StrictWoS* at $p=p_I=0.4$.

High ratio of enrollment to outcome availability

Figure 3 shows the results under a high ratio of enrollment rate to outcome availability rate: specifically, $\lambda m=24$, with rapid enrollment ($\lambda=4$ patients/month) and slow outcome availability ($m=6$ months to evaluate response). Differences in trial durations are greatest in this scenario, because during the trial many outcomes will be pending as new patients become eligible, so that one often needs to consider suspension. When p is small to moderate, *StrictS* leads to the longest trial duration as it suspends accrual multiple times (Figure 3a); *StrictWoS* has the shortest trial duration as it never suspends; and *NaïveS* is intermediate, suspending exactly once. When p is high, *StrictS* has a shorter trial duration than *NaïveS*, because it avoids suspension when there are many early responses (enough to exceed the futility boundary). In terms of mean enrollment (Figure 3b), again *StrictS* is best, whereas *StrictWoS* almost always enrolls the maximum sample size even if the interim analysis reveals futility.

As shown in Figure 3c, under *StrictS* the trial almost always suspends at least once (with minimum probability 0.998 across all response rates), and almost always suspends multiple times when p is small to moderate (with a probability of 1 at $p=p_0=0.2$ and 0.998 at $p=p_I=0.4$). Again there is a moderate probability of terminating the trial prior to enrolling n_I patients, with a value of 0.34 at $p=p_0=0.2$ and 0.02 at $p=p_I=0.4$. Figure 3d reveals that in the scenario of fast enrollment and slow outcome availability, a few patients could be treated incorrectly under *NaïveS* and many more under *StrictWoS*; the average number is 1.8 for *NaïveS* and 12.2 for *StrictWoS* at $p=p_0=0.2$, and 0.2 for *NaïveS* and 1.1 for *StrictWoS* at $p=p_I=0.4$.

Other design settings

Tables 2-4 present mean trial duration and enrollment for three other Simon optimal designs with $\alpha=\beta=0.1$: $p_0=0.1$, $p_1=0.3$ (i.e., a disease that is more difficult to treat); $p_0=0.5$, $p_1=0.7$ (a disease that is amenable to treatment); and $p_0=0.2$, $p_1=0.35$ (the same p_0 as our example but a less optimistic alternative response rate). Results are similar to those shown above, with negligible differences across strategies in the scenario of slow enrollment and fast outcome availability, modest differences in the scenario of moderate enrollment and outcome availability, and substantial differences in the scenario of fast enrollment and slow outcome availability. Typically *StrictS* enrolls the fewest patients but gives the longest trials; *StrictWoS* has the shortest trials but enrolls the most patients; and *NaïveS* falls in between.

Examples of phase II cancer trials

We used four published phase II cancer clinical trials as examples to illustrate the implications of our results in implementing suspension strategies. The first is a study of rituximab for nodular lymphocyte-predominant Hodgkin lymphoma.⁹ This study used Simon's two-stage optimal design with $p_0=0.2$ and $p_1=0.4$. The trial was open from March 1999 to September 2006 and enrolled 39 patients. Therefore the enrollment was very slow with $\lambda=0.43$ patients per month. Patients received rituximab once per week for four consecutive weeks and response was determined at three months after the fourth dose of rituximab. That is, response was determined 4 months post enrollment and the ratio of enrollment to outcome availability rate was $\lambda m=1.72$, which is a scenario of low ratio of enrollment to outcome availability. As shown in Figure 1, trial duration and total patients enrolled would be similar across the three suspension strategies and in this case we could adopt the logistically simplest strategy *StrictWoS*.

The second example is a phase II trial of bendamustine for relapsed or refractory T cell lymphoma.¹⁰ Recruitment occurred at 21 centers in France leading to much faster enrollment. The study enrolled 60 patients in 20 months; that is, $\lambda=3$. Patients received six 37week cycles of bendamustine, and response was assessed after three cycles of treatment (i.e. $m=2.25$ months). Therefore the ratio of enrollment to outcome availability rate was moderate ($\lambda m=6.75$). According to Figure 2, we could infer that the differences across suspension strategies would be modest. With an expected response rate of 0.2 or higher, *StrictS* may be preferred as it has reasonable trial duration and minimizes the number of patients enrolled.

The third example is a trial evaluating lenalidomide for mantle cell lymphoma.¹¹ The study recruited 134 patients at 45 sites worldwide from January 2009 to July 2012 and thus had a fast enrollment rate of $\lambda=3.12$. Patient response was determined following six 28-day cycles. Due to the delayed outcome availability ($m=6$ months) the ratio of enrollment to outcome availability is high ($\lambda m=18.72$). As shown in Figure 3, in this scenario the differences across the three suspension strategies are substantial. When the response rate is between 0.15 and 0.3 as expected for this study, *NaïveS* would be preferred because *StrictS* has long trial duration and *StrictWoS* yields a large number of patients enrolled.

The final example is a study to evaluate trastuzumab emtansine and pertuzumab for patients with HER2-positive metastatic breast cancer.¹² The study enrolled 64 patients in 11 months at 17 participating sites, which is a very rapid enrollment rate of $\lambda=5.82$. Tumor assessment was performed at screening and every 6 weeks until disease progression or initiation of another anti-cancer therapy. Thus the first outcome assessment became available shortly after enrollment ($m=1.5$ months). This scenario could be considered as a moderate-to-high ratio of enrollment to outcome availability ($\lambda m=8.73$), and we could expect the results from the three suspension strategies would be similar to what were shown in Figure 2, with somewhat larger differences across strategies. With an expected response rate in the range of 0.3 to 0.5, the *StrictWoS* may be preferred as it has the shortest trial duration and a comparable number of patients enrolled to *NaïveS* and *StrictS*.

Discussion

We have presented a range of approaches to suspension in phase II cancer trials, using simulation to evaluate their long-run statistical properties. The *StrictS* strategy—which adheres precisely to the characteristics of the interim analysis design, terminates the trial as soon as justified, and never exposes patients unnecessarily—will lead to longer studies but the smallest possible sample sizes. We suspect that some aspects of *StrictS* are applied in many trials, but rarely in the rigorous and comprehensive way specified here. *StrictWoS*, which never suspends enrollment, is the simplest to implement and gives the shortest trials but is most likely to expose subjects unnecessarily. *NaïveS*, which suspends automatically after n_I are enrolled but otherwise ignores pending results, may be the most commonly implemented practice, and is intermediate between the other two with respect to its statistical properties.

Our results suggest that type I error, power and PET are insensitive to suspension, but trial duration, the number of enrolled patients, and the number of patients enrolled incorrectly can vary substantially across strategies. As trial efficiency and patient safety have an inverse relationship, no strategy can provide best performance with respect to both. Thus an investigator will have to balance these competing imperatives in the light of the full design situation.

If the salient concern is to protect patients from an inactive regimen (e.g., if there is an effective standard of care), then one should apply *StrictS*. As illustrated in Figures 1a and 2a, this strategy does not overly increase trial duration, provided the ratio of the enrollment rate to the outcome availability rate is not large. But one needs to keep in mind that this strategy could make the trial much longer if enrollment is rapid, it takes a long time to ascertain response, and the true response rate is close to the null rate. Also this strategy is likely to require more than one suspension, with the attendant operational complications.

For these reasons one may prefer to avoid suspension altogether, as in the *StrictWoS* strategy. This is the simplest and also generally the fastest design, but in cases of rapid enrollment and slow evaluation it also enrolls the largest number of patients, and it could lead to the inadvertent treatment of many patients with an ineffective therapy. Simon's optimal design, strictly implemented, gives the smallest possible expected sample size under

the null response rate. But without suspension, as in *StrictWoS*, $EN(p_0)$ can far exceed its nominal optimal value.

Our results show that the magnitude of the differences among the three strategies primarily depends on the ratio of the enrollment rate λ to the outcome availability rate $1/m$. When the ratio is small ($\lambda m=1$), implying slow enrollment and rapid outcome availability, the three strategies perform almost identically; therefore in such cases one can safely adopt the least complicated. When the ratio is intermediate ($\lambda m=6$), implying moderate rates of enrollment and outcome availability, differences among the three strategies are modest. When the ratio is large ($\lambda m=24$), implying rapid enrollment and delayed outcome availability, the differences among the three become substantial. In such situations one must carefully weigh the conflicting interests of protecting patients from an inactive regimen against trial speed and logistical feasibility. Before embarking on a two-stage phase II trial, the investigator could calculate the ratio of expected enrollment rate and outcome availability rate and use our results as a reference to understand the potential differences among various suspension strategies, conducting simulations if necessary to more rigorously evaluate the characteristics of the suspension strategies for the planned trial. The examples illustrate that even within a single disease (lymphoma) there may be substantial differences in the λm parameter, leading to potentially different decisions about whether and how to suspend a trial.

If one chooses a strategy that avoids suspension, it is likely that the realized and planned sample sizes will differ. Because such a trial continues to accrue while awaiting pending results, at an interim or final analysis superfluous patients may have been enrolled. The analysis plan should account for the possibility of these extra patients. Several statistical methods now exist for conducting adjusted analyses when the achieved sample size does not match the target.¹³⁻¹⁷ We have proposed a flexible approach that can be applied with any frequentist multi-stage design.¹⁷ Our method translates the sample sizes and cut points (n_I, r_I, n, r) in a frequentist design to a Bayesian criterion in terms of the posterior distribution of the response rate, applying the same criterion to the realized data as one would to notional ideal data under the planned sample size. Our approach approximately maintains the basic operating characteristics of the original design.

Our analysis assumes the binary endpoint clinical response, typically defined as the outcome of a radiographic exam or biomarker measurement conducted at a protocol-defined time. In recent years there has been increasing interest in time-to-event outcomes such as progression-free survival (PFS), defined as the time from initiation of study therapy to either the first documented evidence of disease progression or death from any cause. The use of such endpoints, where many subjects may be censored, requires different analysis methods.¹⁸⁻²² These methods generally allow continuous enrollment without suspension, and the interim analysis uses all patients enrolled by the time of analysis, treating the patients who have not yet had the event as censored. Analyses are based on the estimated survival curve. We are not aware that trials applying such approaches have considered the need for suspension or methods to implement it.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding:

The National Institutes of Health supported the authors' research under USPHS grant P307CA016520.

References

- Schlesselman JJ, Reis IM. Phase II clinical trials in oncology: Strengths and limitations of two-stage designs. *Cancer Invest.* 2006; 24:404–412. [PubMed: 16777694]
- Gehan EA. The determination of the number of patients required in a preliminary and a follow-up trial of a new chemotherapeutic agent. *J Chronic Dis.* 1961; 13:346–353. [PubMed: 13704181]
- Fleming TR. One-sample multiple testing procedure for phase II clinical trials. *Biometrics.* 1982; 38:143–151. [PubMed: 7082756]
- Lee YJ. Phase II trials in cancer: present status and analysis methods. *Drugs Exp Clin Res.* 1986; 12:57–71. [PubMed: 3732056]
- Simon R. Optimal two-stage designs for phase II clinical trials. *Control Clin Trials.* 1989; 10:1–10. [PubMed: 2702835]
- Ye F, Shyr Y. Balanced two-stage designs for phase II clinical trials. *Clin Trials.* 2007; 4:514–524. [PubMed: 17942467]
- Jung SH, Lee T, Kim KM, et al. Admissible two-stage designs for phase II cancer clinical trials. *Stat Med.* 2004; 23:561–569. [PubMed: 14755389]
- Herndon JE. A design alternative for two-stage, phase II, multicenter cancer clinical trials. *Control Clin Trials.* 1998; 19:440–450. [PubMed: 9741865]
- Advani RH, Horning SJ, Hoppe RT, et al. Mature results of a phase II study of rituximab therapy for nodular lymphocyte-predominant Hodgkin lymphoma. *J Clin Oncol.* 2014; 32:912–918. [PubMed: 24516013]
- Damaj G, Gressin R, Bouabdallah K, et al. Results from a prospective, open-label, phase II trial of Bendamustine in refractory or relapsed T-cell lymphomas: the BENTLY trial. *JCO.* 2013; 31:104–110.
- Goy A, Sinha R, Williams ME, et al. Single-agent lenalidomide in patients with mantle-cell lymphoma who relapsed or progressed after or were refractory to bortezomib: phase II MCL-001 (EMERGE) study. *J Clin Oncol.* 2013; 31:3688–3695. [PubMed: 24002500]
- Miller KD, Dieras V, Harbeck N, et al. Phase IIa trial of trastuzumab emtansine with pertuzumab for patients with human epidermal growth factor receptor 2-positive, locally advanced, or metastatic breast cancer. *J Clin Oncol.* 2014; 32:1437–1444. [PubMed: 24733796]
- Green SJ, Dahlberg S. Planned versus attained design in phase II clinical trials. *Stat Med.* 1992; 11:853–862. [PubMed: 1604065]
- Chen TT, Ng T-H. Optimal flexible designs in phase II clinical trials. *Stat Med.* 1998; 17:2301–2312. [PubMed: 9819829]
- Masaki N, Koyama T, Yoshimura I, et al. Optimal two-stage designs allowing flexibility in number of subjects for phase II clinical trials. *J Biopharm Stat.* 2009; 17:721–731. [PubMed: 20183436]
- Koyama T, Chen H. Proper inference from Simon's two-stage designs. *Stat Med.* 2008; 27:3145–3154. [PubMed: 17960777]
- Li Y, Mick R, Heitjan DF. A Bayesian approach for unplanned sample sizes in phase II cancer clinical trials. *Clin Trials.* 2012; 9:293–302. [PubMed: 22523304]
- Cheung YK, Thall PF. Monitoring the rates of composite events with censored data in phase II clinical trials. *Biometrics.* 2002; 58:89–97. [PubMed: 11890331]

19. Rosner GL. Bayesian monitoring of clinical trials with failure-time endpoints. *Biometrics*. 2005; 61:239–245. [PubMed: 15737099]
20. Thall PF, Wooten LH, Tannir NM. Monitoring event times in early phase clinical trials: some practical issues. *Clin Trials*. 2005; 2:467–478. [PubMed: 16422307]
21. Zhao L, Woodworth G. Bayesian decision sequential analysis with survival endpoint in phase II clinical trials. *Stat Med*. 2009; 28:1339–1352. [PubMed: 19226557]
22. Huang B, Talukder E, Thomas N. Optimal two-stage phase II designs with longer-term endpoints. *Stat Biopharm Res*. 2010; 2:51–61.

Figure 1.

Trial characteristics under a low ratio of enrollment rate λ to outcome availability rate $1/m$: $\lambda m = 1$, with slow enrollment ($\lambda = 1$ patient/month) and rapid outcome availability ($m = 1$ month to evaluate the outcome). Simon's optimal design with $n_I = 17$, $r_I = 3$, $n = 37$, $r = 10$ for $p_0 = 0.2$, $p_I = 0.4$, $\alpha = \beta = 0.1$. a. Trial duration; b. number of patients enrolled (the dot represents $EN(p_0)$, the theoretical expected sample size under p_0); c. three other performance probabilities under *StrictS*; d. average number of patients treated incorrectly under *NaiveS* and *StrictWoS*.

Figure 2.

Trial characteristics under a moderate ratio of enrollment rate λ to outcome availability rate $1/m$: $\lambda m=6$, with moderate enrollment ($\lambda=2$ patients/month) and moderate outcome availability ($m=3$ months to evaluate the outcome). Simon's optimal design with $n_I=17$, $r_I=3$, $n=37$, $r=10$ for $p_0=0.2$, $p_I=0.4$, $\alpha=\beta=0.1$. a. Trial duration; b. number of patients enrolled (the dot represents $EN(p_0)$, the theoretical expected sample size under p_0); c. three other performance probabilities for *StrictS*; d. average number of patients treated incorrectly under *NaiveS* and *StrictWoS*.

Figure 3.

Trial characteristics under a high ratio of enrollment rate λ to outcome availability rate $1/m$: $\lambda m = 24$, with rapid enrollment ($\lambda = 4$ patients/month) and slow outcome availability ($m = 6$ months to evaluate the outcome). Simon's optimal design with $n_I = 17$, $r_I = 3$, $n = 37$, $r = 10$ for $p_0 = 0.2$, $p_I = 0.4$, $\alpha = \beta = 0.1$. a. Trial duration; b. number of patients enrolled (the dot represents $EN(p_0)$, the theoretical expected sample size under p_0); c. three other performance probabilities for *StrictS*; d. average number of patients treated incorrectly under *NaiveS* and *StrictWoS*.

Table 1

Decision rule for the *StrictS* strategy: No unnecessary enrollment.

Conditions	Decisions
$X \leq R - N_P - N_L$	Insufficient activity; terminate the trial.
$R - N_P - N_L < X \leq R - N_L$	Suspend accrual; re-evaluate when the next outcome is available.
$R - N_L < X \leq R$	Continual accrual; re-evaluate when the next outcome becomes available or the next patient presents for enrollment.
$X > R$	If at stage 1, complete stage 1 accrual and continue to stage 2 without suspension, or resume accrual if previously suspended; if at stage 2, accept the new regimen and complete stage 2 accrual.

Note: X =number of responses observed currently; N_P =number of pending results; N_L =number of subjects left to be enrolled = targeted sample size (n_1 for stage 1 and n for stage 2) – number of enrolled subjects; R =the stopping criterion (r_1 for stage 1; r for stage 2).

Table 2

Trial characteristics under Simon's optimal design with $n_I = 12$, $r_I = 1$, $n = 35$, $r = 5$ for $p_0 = 0.1$, $p_I = 0.3$, $\alpha = \beta = 0.1$.

	Trial duration (months) under p_0	Trial duration (months) under p_I	Mean number of patients under the null: $EN(p_0)$	Mean number of patients under the alternative: $EN(p_I)$
Low ratio of enrollment rate ($\lambda=1$) to outcome availability rate ($1/m=1$): $\lambda m=1$				
<i>NaiveS</i>	21.7	34.3	20.1	32.7
<i>StrictS</i>	21.2	34.1	19.8	33.0
<i>StrictWoS</i>	20.4	34.3	20.2	33.1
Moderate ratio of enrollment rate ($\lambda=2$) to outcome availability rate ($1/m=1/3$): $\lambda m=6$				
<i>NaiveS</i>	14.2	21.9	20.1	32.7
<i>StrictS</i>	14.2	20.3	19.8	33.0
<i>StrictWoS</i>	13.0	19.6	24.0	33.6
High ratio of enrollment rate ($\lambda=4$) to outcome availability rate ($1/m=1/6$): $\lambda m=24$				
<i>NaiveS</i>	13.1	19.5	20.1	32.7
<i>StrictS</i>	16.0	20.3	19.8	33.0
<i>StrictWoS</i>	11.1	14.3	34.0	34.9

Table 3

Trial characteristics under Simon's optimal design with $n_I=21$, $r_I=11$, $n=45$, $r=26$ for $p_0=0.5$, $p_I=0.7$, $\alpha=\beta=0.1$.

	Trial duration (months) under p_0	Trial duration (months) under p_I	Mean number of patients under the null: $EN(p_0)$	Mean number of patients under the alternative: $EN(p_I)$
Low ratio of enrollment rate ($\lambda=1$) to outcome availability rate ($1/m=1$): $\lambda m=1$				
<i>NaïveS</i>	29.8	45.0	28.6	43.3
<i>StrictS</i>	27.9	45.3	25.6	43.7
<i>StrictWoS</i>	27.0	44.6	26.7	43.7
Moderate ratio of enrollment rate ($\lambda=2$) to outcome availability rate ($1/m=1/3$): $\lambda m=6$				
<i>NaïveS</i>	18.2	27.3	28.6	43.3
<i>StrictS</i>	23.3	29.1	25.6	43.7
<i>StrictWoS</i>	16.2	24.8	31.3	44.0
High ratio of enrollment rate ($\lambda=4$) to outcome availability rate ($1/m=1/6$): $\lambda m=24$				
<i>NaïveS</i>	15.1	22.4	28.6	43.3
<i>StrictS</i>	34.7	37.8	25.6	43.7
<i>StrictWoS</i>	13.0	16.7	43.7	44.9

Table 4

Trial characteristics under Simon's optimal design with $n_I=27$, $r_I=5$, $n=63$, $r=16$ for $p_0=0.2$, $p_I=0.35$, $\alpha=\beta=0.1$.

	Trial duration (months) under p_0	Trial duration (months) under p_I	Mean number of patients under the null: $EN(p_0)$	Mean number of patients under the alternative: $EN(p_I)$
Low ratio of enrollment rate ($\lambda=1$) to outcome availability rate ($1/m=1$): $\lambda m=1$				
<i>NaïveS</i>	45.8	62.9	44.2	61.2
<i>StrictS</i>	42.6	62.9	41.3	61.0
<i>StrictWoS</i>	42.4	61.7	42.1	61.0
Moderate ratio of enrollment rate ($\lambda=2$) to outcome availability rate ($1/m=1/3$): $\lambda m=6$				
<i>NaïveS</i>	26.6	36.3	44.2	61.2
<i>StrictS</i>	25.9	34.4	41.3	61.1
<i>StrictWoS</i>	24.1	33.4	46.8	61.4
High ratio of enrollment rate ($\lambda=4$) to outcome availability rate ($1/m=1/6$): $\lambda m=24$				
<i>NaïveS</i>	20.5	27.0	44.2	61.2
<i>StrictS</i>	25.5	27.8	41.3	61.0
<i>StrictWoS</i>	16.6	21.1	56.5	62.3