

A Comparison of Phase II Study Strategies

Sally Hunsberger, Yingdong Zhao, and Richard Simon

Abstract The traditional oncology drug development paradigm of single arm phase II studies followed by a randomized phase III study has limitations for modern oncology drug development. Interpretation of single arm phase II study results is difficult when a new drug is used in combination with other agents or when progression-free survival is used as the endpoint rather than tumor shrinkage. Randomized phase II studies are more informative for these objectives but increase both the number of patients and time required to determine the value of a new experimental agent. In this article, we compare different phase II study strategies to determine the most efficient drug development path in terms of number of patients and length of time to conclusion of drug efficacy on overall survival. (Clin Cancer Res 2009;15(19):5950-5)

The clinical development of oncology drugs has traditionally involved three distinct phases, each with its own goal and characteristic design. In phase I, the maximum tolerated dose of the drug is determined, the underlying assumption being that higher doses, although more toxic to normal tissue, are more effective for eradicating tumors. Phase II studies attempt to determine whether an antitumor effect in a particular diagnostic category is sufficient to warrant conducting a phase III clinical trial. An antitumor effect has traditionally been evaluated by using an endpoint such as tumor shrinkage. Phase II studies are typically single arm studies with 15-40 patients per diagnostic category. Phase III clinical trials are generally large randomized, controlled studies, with the endpoint being a direct measure of patient benefit, such as survival.

The classic paradigm described above has several limitations for modern oncology drug development that arise in the phase II setting. First, successful development of agents that extend survival in patients with cancer has led to the need to study combinations of agents. This makes the design of phase II studies more complex (1) and means that objective responses in single arm phase II studies of combination regimens containing a new drug do not necessarily represent evidence of antitumor activity for the new drug. To interpret the phase II study, one needs a comparison of the activity of the combination containing the new drug to the activity of the regimen given at maximum tolerated doses without the new drug. Such a comparison, if based on prospective randomization, would require a much larger sample size than the traditional single arm phase II trial. The limitations of using historical control

information for estimating the activity of the control regimen are well documented (2), and even if such information is used, larger sample sizes are required because a comparison is involved (3, 4).

The traditional paradigm is also problematic for the development of drugs that may inhibit tumor growth without shrinking tumors. A design based on tumor shrinkage may indicate that a potentially active drug is inactive. As a solution, investigators are beginning to use progression-free survival (PFS) (defined as time from entry on study to documented progression or death) as an endpoint in phase II studies. It is, however, very difficult to reliably determine whether a new drug extends PFS in a single arm phase II trial. Whereas tumors rarely shrink spontaneously, PFS times often vary widely among patients.

As an example of how traditional drug development has not worked, consider advanced pancreatic cancer. From 2004 to 2006, three negative, randomized phase III clinical trials were reported (5-7). In the clinical trials, the addition of Oxaliplatin, Cisplatin, or Irinotecan to Gemcitabine was studied. All three studies followed single arm phase II studies with promising evidence of activity for the combinations (8-10). From these three negative studies it is clear that single arm phase II studies of combination regimens in this population of patients are unreliable. It appears that the response endpoint can be influenced merely by the selection of the patients. Thus, there is a strong need for randomized phase II studies rather than single arm phase II studies in this disease.

In this article, we consider the role of phase II studies in modern oncology drug development. We consider single arm phase II studies, randomized phase II studies, designs that integrate phases II and III into the same study, and skipping phase II altogether. To understand the impact of phase II studies on drug development, we calculate $E[T]$, the expected time from the beginning of phase II to a final conclusion on overall survival (OS), and $E[N]$, the expected number of patients needed in both phase II and phase III.

The outline of the article is as follows. In section 2 we provide a literature review. Section 3 discusses different phase II designs along with details of simulation studies that we performed to

Authors' Affiliation: Biometric Research Branch, Division of Cancer Treatment and Diagnosis, National Cancer Institute, Bethesda, Maryland
Received 12/10/08; revised 5/5/09; accepted 5/29/09; published OnlineFirst 9/29/09.

Requests for reprints: Sally Hunsberger, Biometric Research Branch, 6130 Executive Boulevard, EPN-8120, MSC 7434, National Cancer Institute, Bethesda, MD 20892. Phone: 301-402-0637; Fax: 301-402-0560; E-mail: sallyh@ctep.nci.nih.gov.

© 2009 American Association for Cancer Research.
doi:10.1158/1078-0432.CCR-08-3205

evaluate the impact of phase II strategies on drug development. Section 4 gives the results of the simulation studies. A discussion of the results is presented in section 5.

Literature Review

Rubinstein and colleagues (11) discuss the challenges of drug development with molecularly targeted agents. They describe the pitfalls of single arm studies and recommend the use of randomized phase II studies in which type I error rates are relaxed from the traditional 0.05 to 0.20. These issues were also described by Simon and colleagues (12), for therapeutic vaccine studies, and by Ratain and colleagues (13). Ratain and colleagues (14) used a "randomized discontinuation design" in which patients are initially treated with the same experimental agent; patients with stable disease after a specified time are then randomized to either continue receiving the experimental agent or receive placebo.

Inoue and colleagues (15) presented a Bayesian phase II/III design in which patients are randomized to an experimental arm or a standard arm; the decision to stop the study early or continue the study is made repeatedly based on simultaneous hypotheses tests of survival and response rates. They compare the efficiency of the design to two independent studies, with the first study being a single arm study based on response rates and the second study being a randomized study with survival as the endpoint. In a simulation patterned after a non-small cell lung cancer study, they found that the phase II/III design used fewer patients and took less time to complete.

Buaer and colleagues (16) and Proschan and Hunsberger (17) have developed adaptive designs that are very flexible and allow the primary endpoint to be analyzed during the study and to be used to determine whether the study should continue. In these designs, the sample size can also be readjusted. The framework of the adaptive design allows one to maintain the type I error rate by adjusting the critical value at the end of the study.

Parmar and colleagues (18) advocate integrated phase II/III designs and give an in depth discussion on the motivation for these designs. Goldman and colleagues (19) consider a phase III study design with an interim analysis that can stop for futility based on a composite endpoint of either PFS or overall survival. They show that including a futility analysis can save time and reduce the number of patients accrued to a phase III trial under the null hypothesis. The criteria they use for stopping is more conservative than the one we consider.

Methods

In this article, we evaluate the impact of different phase II study strategies on drug development and use $E[T]$ and $E[N]$ to evaluate the impact. We first discuss the single arm phase II study design. In this type of study, PFS is the primary endpoint and is used as an early indicator of activity. PFS data from patients given the experimental treatment are compared to the historical experience. If this single arm study is promising, a randomized phase III study based on OS is organized. Although it is possible to plan such studies by using specific historical controls and taking into account the number of such controls, this is rarely done. Usually, historical data (often from small studies) are used to specify a null comparison level of activity.

For PFS data, this comparison level may represent PFS at a landmark time, or median PFS for an exponential distribution.

Because historical controls may not be prognostically comparable to patients accrued to the phase II trial, the specified null level of PFS may not be correct. When a null PFS rate is specified that is larger than the true rate for the population under study, the benefit of the new treatment will be underestimated, thus reducing the probability of finding activity in the phase II study and continuing on to the phase III study. Ultimately, this reduces the probability of finding a significant benefit on OS. Conversely, a treatment that has no benefit on PFS is more likely to appear active when the null rate that is specified is smaller than the true rate for the population under study. This will result in continuation to a phase III study with probability greater than the specified type I error, thus increasing $E[T]$ and $E[N]$. We study the effect of over- or underspecifying the null PFS rates in a single arm phase II study.

The problem of incorrectly specifying the null PFS rate in a single arm phase II study can be alleviated by performing a randomized phase II study comparing the new treatment to the control regimen by using PFS as an endpoint. If the new treatment appears to be better than the control based on PFS, then a phase III trial comparing the new treatment to the control regimen by using OS as an endpoint is organized.

Although the randomized phase II study alleviates the need to specify a null rate, it does require more patients than a single arm study. Therefore, in order to address the increase in sample size, we consider integrating a randomized phase II into a phase III study. With this approach, accrual to a randomized phase II study is designed to continue on into a phase III study if a specified criterion is met. The endpoint used for the phase II evaluation will differ from that used for the phase III analysis (as in the single arm study and sequence of studies), but data from patients accrued during the phase II study are used in the phase III study. Goldman and colleagues (19) have described these designs as a phase III study with an interim futility analysis with an intermediate endpoint. Finally, we consider a strategy of skipping the phase II study and performing a single randomized study with survival as the endpoint, and including an interim futility analysis based on survival.

In this article, we wish to evaluate the strategies by comparing the total number of patients ($E[N]$), both in phase II and phase III, and the total time until completion ($E[T]$) under null and alternative hypotheses, by using parameters from the pancreatic cancer example for illustration. Appendix 1 provides equations for the calculation of $E[N]$ and $E[T]$. In the calculation of $E[N]$ and $E[T]$, the sample size and length of accrual for a phase III study are included, and the same phase III study design is used for all strategies. The sample size and length of accrual of the phase III study are based on a design that has a primary endpoint of OS and 90% power for a two-sided 0.05 level test.

The pancreatic cancer literature suggests that the median OS rate is 6 mo. For the sample size calculations of the phase III study, an improvement in OS to 7.8 mo is used (hazard ratio of 1.3). Although this improvement appears small, it is likely that this improvement would be of interest because this study is for an advanced disease population, and even small OS improvements would be interesting because the drug could then be studied in earlier stages of disease. Assuming an accrual rate of 15 patients per month with a minimum follow up of 6 mo would require 46.1 mo of accrual or 692 patients.

For the two strategies that have independent phase II and phase III studies (i.e., the single arm study and randomized phase II), the phase II primary endpoint will be PFS, and the study will be designed to have 90% power for a one-sided 0.1 level test. We continue with the pancreatic example to design the phase II studies based on PFS. The literature suggests that the median PFS for pancreatic cancer is between 2 and 4 mo; thus, for the single arm study, we specify 3 mo as the null PFS rate, and for the randomized study we based the sample size calculation on a control arm median PFS rate of 3 mo. We power both studies to detect an improvement in the median PFS to 4.5 mo (hazard ratio of 1.5).

For the integrated phase II/III study design, patients will be accrued until time t_1 . At t_1 , accrual will be suspended and patients will be followed for a minimum time, f_1 . After $t_1 + f_1$, a comparison of the treated versus control groups based on PFS will be performed. If the P value for PFS in this interim analysis is not less than a specified threshold, α_1 , accrual will terminate and no claims for the new treatment will be made. Otherwise, accrual will resume until a total of M patients are accrued. After accruing M patients, follow-up will continue for an additional minimum time, f_o . At the end of the study, OS will be evaluated on all M patients. The total sample size M is that of the phase III study.

The strategy of skipping the phase II study and performing an interim futility analysis on OS requires a specification of t_1 (the time of the interim analysis) and α_1 , the criteria for continuing. That is, if the P value for the comparison of OS is less than α_1 , the study will continue.

For the integrated phase II/III study and for the phase III study with a futility analysis, we determined t_1 and α_1 so that the overall study power (probability of concluding a benefit on OS when starting from phase II) will be maintained at 81%. Note, this 81% is the power for the strategy of a randomized phase II study with 90% power for PFS, followed by a randomized phase III study with 90% power for OS. For the integrated phase II/III study and the futility design, we evaluate $E[N]$ and $E[T]$ for different α_1 values, but we always adjusted t_1 to maintain 81% power.

We evaluated the designs under: (a) no treatment effect on either PFS or OS (global null); (b) treatment effect on PFS and OS (global alternative). When we evaluate the single arm study, we use the equations in Appendix 1 and assume that PFS and OS follow exponential distributions. Because the two studies are independent, $E[T]$ and $E[N]$ can be calculated analytically. In the integrated phase II/III design, data from the same person are used for the PFS and OS analysis. Therefore, the data will be correlated, and the correlation needs to be accounted for when evaluating $E[T]$ and $E[N]$. The correlation structure we assume makes analytic results difficult to calculate, because the PFS data no longer follow an exponential distribution; hence, computer simulations were conducted to evaluate $E[N]$ and $E[T]$. Because the integrated phase II/III design is compared to the separate randomized phase II strategy, and the futility analysis on OS strategy, simulated data were used to evaluate $E[N]$ and $E[T]$ for these designs as well. (Equations for the calculations of $E[N]$ and $E[T]$ are found in Appendix 1).

In the simulations, we generate correlated PFS and OS data as follows. The distribution of OS for the control group was taken as exponential with median, m_o , months. The treatment effect for OS is specified by a parameter, Δ_o , resulting in an exponen-

tial distribution of OS for the treatment group with $m_o \Delta_o$. Provisional PFS times were generated for control and treatment group patients by using exponential distributions with median values of m_p mo and $\Delta_1 m_p$ mo, respectively. For a patient with an overall survival value of Y_o , and a provisional PFS value of Y_1 , the actual PFS time was set as $Y_p = \min(Y_1, Y_o)$. This introduction of correlation between PFS and OS means that PFS times do not have an exponential distribution. If the medians of OS and PFS are very different, then the correlation is very small and PFS will have an approximate exponential distribution. In the simulations, Δ_1 and Δ_o were varied. All simulations are performed with 10,000 replications.

Simulation Results

For the single arm phase II study, specifying the null PFS comparison rate too low becomes a problem when there is no treatment benefit. Appendix Table 1a shows the increase in the expected sample size and expected length of study when the null rate is underspecified by 2 wk and 1 mo. Because the probabilities of continuing to the phase III study increase from the desired level of 0.1 to 0.4 and 0.72, the expected sample size can almost triple (when compared to the correct specification) if the specification is off by 2 wk or can be four times larger if the specification is off by 1 mo. Appendix Table 1b shows that specifying the null rate too high cuts into the probability of concluding a benefit on OS when a benefit exists. The overall probability is expected to be 0.81, but it is reduced to 0.51 or 0.09 for a 2 wk or 1 mo overspecification. It is clear that incorrectly specifying the null for single arm studies can increase $E[N]$ and $E[T]$ when there is no treatment benefit and can reduce the probability of finding a benefit on OS when there is a treatment benefit. Therefore, we consider the integrated phase II/III study that aims to take advantage of randomization as in a randomized phase II study yet potentially shorten the phase III study under a null treatment effect. We also consider the effect on $E[N]$ and $E[T]$ of skipping phase II and moving to a phase III study with OS as the only endpoint.

Appendix Table 2 gives the $E[T]$ and $E[N]$ for the designs under the global null and global alternative. All designs have 81% power and a type I error rate of less than 0.05 (two-sided). As can be seen in Appendix Table 2, the integrated phase II/III design is effective for substantially reducing the development time and number of required patients under the global null when compared to the study with a futility analysis based on OS. The sample size is comparable to that of the sequence of studies under the global null when $\alpha_1 = 0.1$ or 0.2. Futility monitoring on PFS is more effective than futility monitoring on OS in this setting because progression events can be observed sooner. The dramatic savings in time and patients (when comparing the sequence of studies to the integrated design) comes under the global alternative.

In Appendix Table 1, we show $E[T]$ and $E[N]$ under the global null and alternative hypothesis for different patient accrual rates, different hazard ratios, and median times for PFS and OS. In Appendix Table B1, for the integrated design we only show results for $\alpha_1 = 0.1$ and 0.2 because the other values led to larger $E[N]$ and $E[T]$. We also only show designs with $f_1 = 0$ because setting f_1 larger than 0 only changed the results minimally. Therefore, because, practically, it is much easier to continue accrual, we feel that this is the design of most interest.

The results are qualitatively very similar to those shown in Appendix Table 2.

Discussion

Traditional oncology drug development strategies are problematic when using PFS for assessing drug activity or when adding a new agent to an active regimen. We explored the impact of different phase II study strategies on drug development by evaluating the expected number of patients and the expected length of time for phases II and III.

We showed that although the single arm phase II study may appear to speed up drug development, even minimal prognostic bias in comparison to historical controls can have major impact on producing misleading results that either lead to futile phase III trials or, more importantly, result in missing active agents. Our investigation showed that the integrated phase II/III design worked well under the global null. That is, $E[N]$ and $E[T]$ were no larger than those of a randomized phase II study and were considerably smaller than when the phase II study was skipped. The integrated phase II/III study performed better than the separate randomized phase II study under the global alternative and did not increase $E[T]$ and $E[N]$ when compared to skipping the phase II component.

There are several practical issues that must be addressed when using an integrated phase II/III design. When an integrated phase II/III design is proposed, institutions must have sufficient accrual for a phase III study, and there must be enough funding to support a phase III study. This could have implications for studies that are funded by grants or for small companies that typically only develop agents through phase II.

In some situations, it may be difficult for physicians to enter patients into a randomized phase III study after a positive randomized phase II study with an endpoint of PFS. The difficulty may also exist for the integrated phase II/III study because continuation of the study would imply that PFS results were promising for the new treatment. The randomized phase II or the integrated phase II/III would only be appropriate in situations in which PFS was not an accepted measure of patient benefit or a validated surrogate of OS. In such settings, a positive randomized phase II study with PFS as the endpoint will often not be sufficient to change clinical practice, and a sequence of randomized studies or the integrated phase II/III design would both be appropriate in an ethical sense.

The combined phase II/III design and the separate trial designs are reasonable only if it is expected that improvement of PFS is a necessary condition, although not a sufficient condition, for improvement in OS. If improvement in PFS is not a necessary condition for improvement in OS, then clearly the study with a futility analysis based on OS would be preferable. A future area of research for the integrated approach would be to include a futility analysis on OS along with the PFS analysis. This would improve performance for the trials in which improvement in PFS is not accompanied by an improvement in OS. This approach can also be used with endpoints other than PFS such as molecular biomarkers or new imaging diagnostics. However, one needs to critically examine the relationship of any proposed intermediate endpoint and OS.

After the parameters of the integrated phase II/III design have been chosen, the design and monitoring plan should be clearly described in the protocol. The protocol would specify the number of progression events that would be needed for the PFS analysis and the α_1 for stopping the study. The total number of events for the OS analysis would also be specified. After the PFS analysis has been performed, typical interim Data Safety Monitoring Committee monitoring based on OS (for efficacy) would be specified in the protocol. The protocol should indicate clearly that early stopping of accrual because of a treatment effect on PFS is not a part of the analysis plan.

The integrated phase II/III design has some properties that may be of value when accelerated or when provisional approval of a new drug is of interest. Although a phase III study is required after accelerated approval has been obtained, it has not always been performed. This design would ensure that a randomized phase III trial based on OS was in place at the time that accelerated approval was obtained and would provide a well-powered, well-designed randomized phase II study with PFS as the basis for the provisional claim. If accelerated approval were of interest, α_1 would generally be set no greater than 0.05.

We have provided a web-based computer program that calculates the expected sample size, expected study duration, and the power for the randomized designs studied in this article.¹ The calculations are exact for the randomized phase II strategy and are approximations for the integrated phase II/III study and the single study with a futility analysis strategy. Although some of the calculations are approximations, the approximation of the savings in sample size or time that could be obtained by using the integrated phase II/III approach would be adequate to decide whether the design should be considered further. When designing an integrated phase II/III study, we recommend evaluating various sets of parameters. For example, the accrual rate should be varied along with the medians of PFS and OS and the size of the treatment effect on PFS and OS. If an integrated approach looks promising, a simulation study that reflects the best assumptions on the distribution of PFS and OS and accounts for the correlation in the integrated and single study designs should be performed.

With the number and type of new drugs that are being developed today, it may be necessary to use new types of designs in the phase II and phase III settings. We suggest that investigators explore the efficiency of integrated phase II/III designs.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Appendix 1

We now provide equations to calculate the expected sample size ($E[N]$) and expected length of study ($E[T]$). In this section we will use the following notational convention: a subscript of "1" will refer to parameters in the PFS analysis (or for the futility analysis if there is no phase II study), and a subscript of "o" will refer to parameters in the final phase III OS analysis. $E[N]$ and $E[T]$ for the different phase II strategies:

¹ <http://linus.nci.nih.gov/brb>

Single arm phase II and Randomized phase II

$$E[N] = n_1 + (M)P\{\text{continuing}\}$$

$$E[T] = t_1 + f_1 + (M/a + f_o)P\{\text{continuing}\}$$

Integrated II/III and Futility based on OS

$$E[N] = n_1 + (M - n_1)P\{\text{continuing}\}$$

$$E[T] = t_1 + f_1 + (M/a - t_1 + f_o)P\{\text{continuing}\},$$

where a is the accrual rate, M is the final sample size for the phase III study of OS, n₁ is the sample size based on PFS or at the futility analysis, t₁ is the length of accrual based on the PFS or futility analysis, and f₁ or f_o is the minimum follow-up (f₁ = 0 for the futility analysis).

For all designs (assuming the null rate is correctly specified in the single arm study), the P{continuing} is α₁ under the global null hypothesis and is (1 - β₁) under the global alternative hypothesis. Here, (1 - β₁) is the probability of concluding a treatment benefit on PFS or the probability of not stopping the study for futility.

Standard power calculations (ref. 20; or Southwest Oncology Group Statistical Tools)² that assume an exponential distribution for PFS and OS can be used to calculate n₁, f₁, M, f_o, and P{continuing} for the studies with phase II and phase III being independent. For the integrated phase II/III design, power calculations that assume independent exponential distributions for PFS and OS can give good approximations, but simulations would need to be performed to account for a specified correlation structure (as discussed in the article). Similarly, for the single study with a futility analysis based on OS, assuming an exponential distribution for OS and treating the two analyses as being independent can give a good approximation, but simulations would need to be performed to find the true E[N] and E[T].

The P{continuing} is different for the single arm study when the null rate is incorrectly specified. In this case, P{continuing} = P{rejecting specified null true median}, which is the probability of rejecting the specified null rate; this probability is calculated

by using a distribution with the true median. Specifying the null rate too low gives a P{continuing} that is higher than the desired level of α₁ under the global null. Specifying a null rate too high gives a P{continuing} that is lower than the desired level of (1 - β₁) under the global alternative.

For the studies with an independent phase II and phase III, the probability of concluding a positive treatment effect on OS is Power_{II/III} = P{continuing}(1 - β_o). For the integrated phase II/III study and the single study with a futility analysis based on OS this is a lower bound for Power_{II/III}.

A web-based computer program that calculates expected sample size, expected study duration, and Power_{II/III} for the randomized designs studied in this article can be found at the National Cancer Institute Biometric Research Branch.¹ Independent exponential distributions are assumed for PFS and OS in the calculation of P{continuing}. Also, the futility analysis based on OS is assumed to be independent of the final OS analysis. Therefore, for the integrated design and for the single study with a futility analysis based on OS, the web-based program provides approximations.

Appendix 2

Appendix Table 1. Simulation results comparing drug development with different phase II components. The simulations vary accrual, hazard ratios, and control medians. The PFS and OS are correlated and generated according to two exponentials: Y₁, with median 3 mo and a treatment effect hazard ratio of 1.5; and Y₂, with median 6 mo and a treatment hazard ratio of 1.3. Progression is the min(Y₁, Y₂), and survival is Y₂. All designs have 81% power to conclude there is a positive effect on OS under the global alternative while having probability of less than 0.05 (when using a two-sided test) of concluding a positive effect on OS under the global null. E[N] is the expected sample size, and E[T] is the expected study time. All time is in months.

Appendix Table 1a. E[N] and E[T] for drug development using a single arm phase II study when the experimental agent provides no benefit. The phase II study design is based on a null median PFS rate of 3 mo, i.e., an assumption that control patients would have a median PFS of 3 mo. If that assumption is correct, the single arm phase II trial will have 90% power to detect a 1.5 mo improvement (hazard ratio of 1.5), with a

² <http://www.swogstat.org/statoolsout.html>

Accrual per month	Hazard Ratios		Medians		Designs	α ₁	Global Null		Global Alternative	
	Y1	OS	Y1	OS			E[N]	E[T]	E[N]	E[T]
30	1.5	1.3	3	6	Futility based on OS	0.2	513	17.1	719	24.0
						0.5	504	16.8	699	23.3
						0.1	325	14.44	935	39.6
						0.1	350	11.7	713	23.8
15	1.5	1.3	3	12	Integrated II/III (f ₁ = 0)	0.2	336	11.2	704	23.5
						0.2	524	34.9	725	48.3
						0.5	500	33.3	670	46.6
						0.1	297	24.0	906	74.2
15	1.3	1.2	3	6	Integrated II/III (f ₁ = 0)	0.1	292	19.4	708	47.2
						0.2	302	20.2	705	47.0
						0.2	765	51.0	1253	83.6
						0.5	819	54.6	1210	80.7
15	1.3	1.2	3	6	Futility based on OS	0.1	629	45.4	1702	121.9
						0.1	586	39.1	1249	83.3
						0.2	547	36.5	1227	81.8
						0.2	765	51.0	1253	83.6

one-sided 0.1 level test and a minimum follow-up of 3 mo. This will require 69 patients. The phase III randomized study will require 692 patients, with a minimum follow-up of 6 mo, and will have 90% power to detect a hazard ratio of

True median PFS rate for the population included in the study (months)	Probability of continuing to the phase III study	E[N]	E[T]
3*	0.1	138	12.8
3.5	0.4	346	28.5
4	0.72	567	45.1

*The truth agrees with assumption.

1.3, assuming a median OS rate of 6 mo in the control arm. A two-sided 0.05 level test will be used. (Assumptions are based on the pancreatic cancer example).

Appendix Table 1b. The probability of correctly concluding a treatment benefit on overall survival when the specified null rate is too large and the experimental agent provides a benefit

True median PFS rate for the population included in the study (months)	Probability of continuing to the phase III study	Probability of concluding an overall survival benefit
3*	0.9	0.81
2.5	0.59	0.53
2	0.1	0.09

*The truth agrees with assumption.

on PFS that results in a hazard ratio of 1.5 (based on the true control PFS rate). The phase II study design is specified in Appendix Table 1a. The phase III randomized study is assumed to have 90% power.

Appendix Table 2. Simulation results based on the pancreatic cancer example. Accrual of 15 patients/mo; data were generated according to two exponentials: Y_1 , with a median of 3 mo and a treatment effect hazard ratio of 1.5; Y_2 , with a median of 6 mo and a treatment hazard ratio of 1.3. Progression was the min

Designs	α_1	t_1	Global Null		Global Alternative	
			E[N]	E[T]	E[N]	E[T]
Futility based on overall survival	0.2	24.0	427	28.5	649	43.2
	0.5	11.9	433	28.9	627	41.8
Sequence of Phase II and Phase III	0.1	15.1	296	23.3	849	65.0
Integrated II/III with ($f_1 = 0$)	0.05	20.4	325	21.7	646	43.1
	0.1	16.7	294	19.6	644	42.9
	0.2	12.3	287	19.2	634	42.3
	0.5	6.1	391	26.0	625	41.7
Integrated II/III with ($f_1 = 3$)	0.05	18.3	295	22.7	644	46.0
	0.1	14.7	268	20.9	640	45.7
	0.2	10.8	268	20.9	633	45.2
	0.5	4.2	378	28.2	623	44.5

(Y_1, Y_2), and survival was Y_2 . All designs have 81% power to conclude there is a positive effect on OS under the global alternative while having probability of less than 0.05 of concluding a positive effect on OS under the global null. E[N] is the expected sample size, and E[T] is the expected study time. All time is in months.

References

- Korn EL, Simon R. Using the tolerable-dose diagram in the design of phase I combination chemotherapy trials. *J Clin Oncol* 1993;11:794-801, PubMed.
- Green S, Benedetti J, Crowley J. *Clinical trials in oncology*. Cambridge: Chapman & Hall; 1997.
- Dixon DO, Simon R. Sample size considerations for studies comparing survival curves using historical controls. *J Clin Epidemiol* 1988;41:1209-13, PubMed.
- Makuch RW, Simon RM. Sample size considerations for nonrandomized comparative studies. *J Chronic Dis* 1980;33:175-81, PubMed.
- Louvet C, Labianca R, Hammel P, et al. Gemcitabine in combination with oxaliplatin compared with gemcitabine alone in locally advanced or metastatic pancreatic cancer: results of a GERCOR and GISCAD phase III trial. *J Clin Oncol* 2005;23:3509-16, PubMed.
- Heinemann V, Quietzsch D, Gieseler F, et al. Randomized phase III trial of gemcitabine plus cisplatin compared with gemcitabine alone in advanced pancreatic cancer. *J Clin Oncol* 2006;24:3946-52, PubMed.
- Rocha Lima CM, Green MR, Rotche R, et al. Irinotecan plus gemcitabine results in no survival advantage compared with gemcitabine monotherapy in patients with locally advanced or metastatic pancreatic cancer despite increased tumor response rate. *J Clin Oncol* 2004;22:3776-83, PubMed.
- Louvet C, Andre T, Lledo G, et al. Gemcitabine combined with oxaliplatin in advanced pancreatic adenocarcinoma: final results of a GERCOR multicenter phase II study. *J Clin Oncol* 2002;20:1512-8, PubMed.
- Philip PA, Zalupski MM, Vaitkevicius VK, et al. Phase II study of gemcitabine and cisplatin in the treatment of patients with advanced pancreatic carcinoma. *Cancer* 2001;92:569-77, PubMed.
- Wagener DJ, Verdonk HE, Dirix LY, et al. Phase II trial of CPT-11 in patients with advanced pancreatic cancer, an EORTC early clinical trials group study. *Ann Oncol* 1995;6:129-32, PubMed.
- 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;23:7199-206, PubMed.
- Simon RM, Steinberg SM, Hamilton M, et al. Clinical trial designs for the early clinical development of therapeutic cancer vaccines. *J Clin Oncol* 2001;19:1848-54, PubMed.
- Ratain MJ, Stadler WM. Clinical trial designs for cytostatic agents. *J Clin Oncol* 2001;19:3154-5, PubMed.
- Ratain MJ, Eisen T, Stadler WM, et al. Phase II placebo-controlled randomized discontinuation trial of sorafenib in patients with metastatic renal cell carcinoma. *J Clin Oncol* 2006;24:2505-12, PubMed.
- Inoue LY, Thall PF, Berry DA. Seamlessly expanding a randomized phase II trial to phase III. *Biometrics* 2002;58:823-31, PubMed.
- Bauer P, Kohne K. Evaluation of experiments with adaptive interim analyses. *Biometrics* 1994;50:1029-41, PubMed.
- Proschan MA, Hunsberger SA. Designed extension of studies based on conditional power. *Biometrics* 1995;51:1315-24, PubMed.
- Barthel FMS, Sydes M, et al. Speeding up the evaluation of new agents in cancer. *J Natl Cancer Inst* 2008;100:1204-14, PubMed.
- Goldman B, LeBlanc M, Crowley J. Interim futility analysis with intermediate endpoints. *Clin Trials* 2008;5:14-22.
- Rubinstein LV, Gail MH, Santner TJ. Planning the duration of a comparative clinical trial with loss to follow-up and a period of continued observation. *J Chronic Dis* 1981;34:469-79. PubMed doi:10.1016/0021-9681(81)90007-2.