

Adaptive Design for Confirmatory Clinical Trials

Cyrus R. Mehta

Cytel Corporation and Harvard School of Public Health

Email: mehta@cytel.com

3 March 2009

Abstract

This paper discusses the benefits and limitations of adaptive sample size re-estimation for late stage confirmatory clinical trials. Comparisons are made with more traditional fixed sample and group sequential designs. It is seen that the real benefit of the adaptive approach arises through the ability to invest sample size resources into the trial in stages. The trial starts with a small up-front sample size commitment. Additional sample size resources are committed to the trial only if promising results are obtained at an interim analysis. This strategy is seen to be more advantageous than the fixed sample or group sequential approaches in certain settings. The discussion is illustrated through a case study of an actual adaptive trial.

1 The Setting

Consider a two-arm trial to determine if there is an efficacy gain for an experimental drug relative to the industry standard treatment for negative symptoms schizophrenia. The primary endpoint is the improvement from baseline to week 26 in the Negative Symptoms Assessment (NSA), a 16-item clinician-rated instrument for measuring the negative symptomatology of schizophrenia. Let μ_t denote the difference between the mean NSA at baseline and the mean NSA at week 26 for the treatment arm and let μ_c denote the corresponding difference of means for the control arm. Denote the efficacy gain by $\delta = \mu_t - \mu_c$. The trial will be designed to test the null hypothesis $H_0: \delta = 0$ versus the one-sided alternative hypothesis that $\delta > 0$. It is expected, from limited data on related studies, that $\delta \geq 2$ and σ , the between-subject standard deviation, is believed to be about 7.5. In the discussion that follows we shall focus our attention on adaptive sample size adjustments due to uncertainty surrounding the true value of δ . Even though the statistical methods discussed here are applicable when there is uncertainty about either δ or σ , the adaptive approach requires careful justification primarily when δ is involved. Adaptive sample size adjustments relating to uncertainty about σ are fairly routine and non-controversial.

We shall consider fixed-sample, group sequential and adaptive design options for this study. There are advantages and disadvantages to each option with no single approach dominating over the others. We are interested, however, in exploring whether the adaptive methodology can add value to the better established fixed sample and group sequential approaches to trial design. We will see that an adaptive design alleviates to some extent the problem of “overruns” encountered by group sequential designs when the primary endpoint is observed after a lengthy follow-up period as is the case here. Additionally, we will see that an adaptive design may, in certain settings, have a more favorable risk versus benefit trade-off. This case study relates to an actual trial. For reasons of confidentiality, however, the identity of the trial sponsor and the name of the product being tested are not disclosed.

2 Fixed Sample Design

Since it is believed a priori that $\delta \geq 2$, we first create Plan 1, a single-look design with 80% power to detect $\delta = 2$ using a one-sided level 0.025 test, given $\sigma = 7.5$. With these design parameters we can show that Plan 1 will be fully powered if a total of 442 subjects are enrolled (221/arm). There is, however, considerable uncertainty about the true value of δ , and to a lesser extent about σ . Nevertheless it is believed that even if the true value of δ were as low as 1.6 on the NSA scale, that would constitute a clinically meaningful effect. We therefore also create Plan 2, having 80% power to detect $\delta = 1.6$ using a one-sided level-0.025 test, given $\sigma = 7.5$. Plan 2 requires a total sample size of 690 subjects.

We have now proposed two design options. Under Plan 1 we would enroll 441 subjects and hope that the study is adequately powered, which it will be if $\delta = 2$ and $\sigma = 7.5$. If, however $\delta = 1.6$ the power drops from 80% to 61%. There is thus a risk of launching an underpowered study for an effective drug under Plan 1. Under Plan 2 we will enroll 690 subjects, thereby ensuring 80% power at the smallest clinically meaningful value, $\delta = 1.6$, and rising to 93% power at $\delta = 2$. The operating characteristics of Plan 1 and Plan 2 are displayed side by side in Table 1 for values of δ between 1.6 and 2.0.

Table 1: Operating Characteristics of Plan 1 and Plan 2

δ	Plan 1		Plan 2	
	Sample Size	Power	Sample Size	Power
1.6	441	61%	690	80%
1.7	441	66%	690	85%
1.8	441	71%	690	88%
1.9	441	76%	690	91%
2.0	441	80%	690	93%

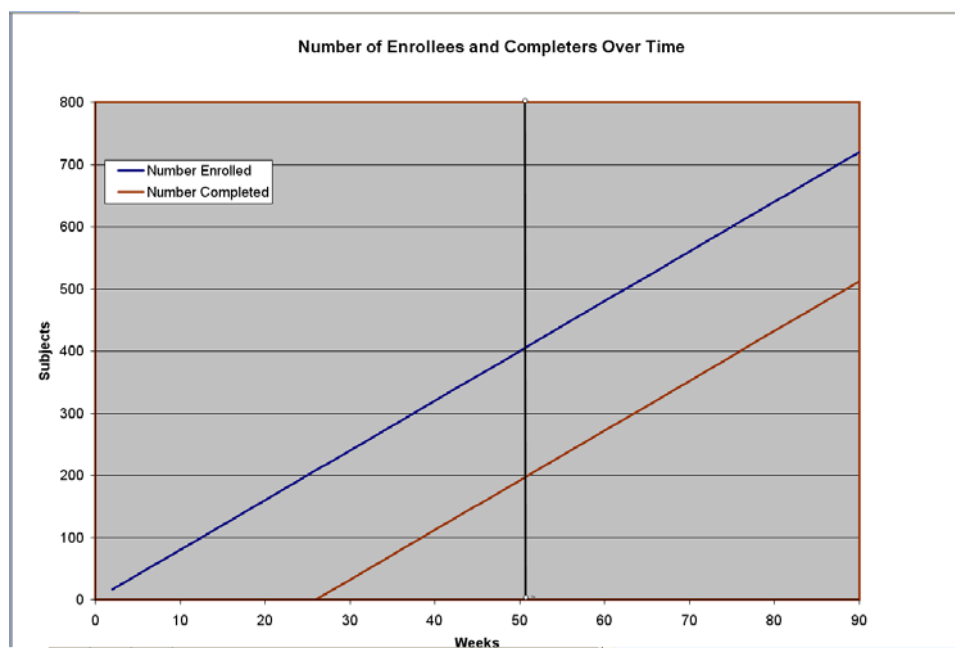
If resources were plentiful, Plan 2 would clearly be the preferred option. The sponsor must, however, allocate scarce resources over a number of studies and in any case is not in favor of

designing an overpowered trial. This leads naturally to considering a design that might be more flexible with respect to sample size than either of the above two single-look fixed sample designs. We will consider two types of flexible designs; group sequential and adaptive.

3 Group Sequential Design

When sample size flexibility is desired for late-stage trials, it is usually appropriate to first explore the group sequential option. Let us then construct a group sequential design with one interim look and 80% power to detect $\delta = 1.6$ such that if in fact $\delta = 2$, the trial will stop early. While this would appear to be an attractive option, it is important to consider not just the saving in study duration but also the saving in the actual number of subjects randomized to the study. Since the efficacy endpoint for this trial will only be observed at week 26, the actual saving in sample size will be affected by the enrollment rate. In the current study it is anticipated that subjects will enroll at an average rate of 8 per week. The number of subjects enrolled and the number of completers over time are displayed graphically in Figure 1

Figure 1: Impact of Enrollment Rate and Length of Follow-Up on Trial Completion

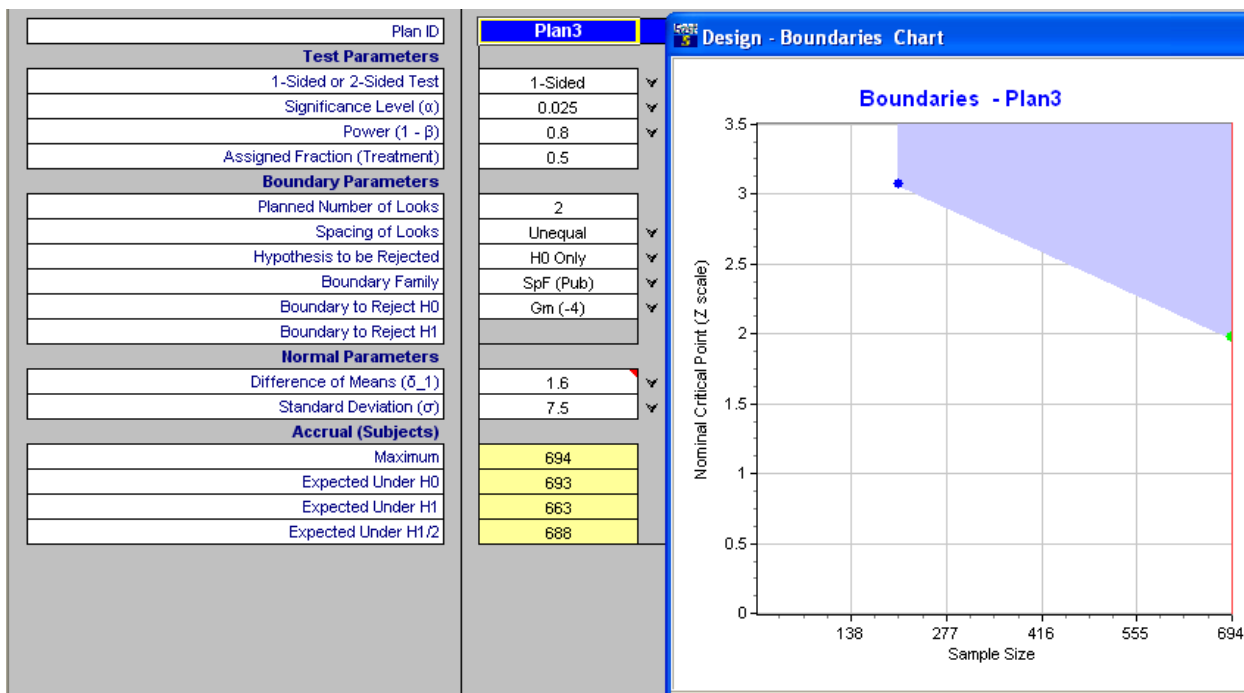


Observe that there is a 26-week horizontal separation between the two parallel lines depicting, respectively, the graph for enrollment and the graph for study completion. This 26-week gap must be taken into consideration when evaluating the savings achieved by utilizing a group sequential design.

The two major design parameters to be specified for a two-look group sequential design are the timing of the interim analysis and the amount of type-1 error to be spent. We will assume that

data must be available for at least 200 completers before the trial can be terminated for efficacy so that an adequate safety profile may be developed for the study drugs. Therefore a suitable time point for the interim analysis is week 52, when we will have enrolled 416 subjects with data on 208 completers. Next we must decide on the amount of type-1 error to spend (see Lan and DeMets, 1983) for the early stopping boundary. It is generally held that the type-1 error should be spent conservatively in the early stages of a trial so as to ensure that results based on premature termination will be compelling and have the capacity to change medical practice (see Pocock, 2005). Suppose then that we use the $\gamma(-4)$ error spending function proposed by Hwang, Shih and DeCani (1990) to obtain the early stopping boundary. The boundary thus produced resembles the conservative O'Brien-Fleming (1979) boundary. The corresponding group sequential design, having a sample size of 694, is displayed in Figure 2 as Plan 3.

Figure 2: Group Sequential Design Denoted as Plan 3



In Plan 3 the nominal critical point for early stopping is 3.067 standard deviations. The one sided p-value corresponding to this early stopping boundary is $1 - \Phi(3.067) = 0.0011$ which, if met, would indeed be compelling enough to justify premature termination. Both Plan 2 and Plan 3 have 80% power to detect $\delta = 1.6$ with a one-sided level-0.025 test. Their sample size commitments too are almost the same. However, under Plan 2 there is no possibility of early stopping whereas under Plan 3, it is possible to stop early and thereby save on sample size. Figure 2 shows that the expected sample size if in truth $\delta = 1.6$, is 663 subjects, a saving of 61 subjects compared to the maximum sample size of 694. The saving will be even more if the true value of δ is greater than 1.6. These expected savings in sample size are discussed next along with the problem of “overruns”.

3.1 The Problem of Overruns

Care must be taken when estimating the actual sample size savings of a group sequential design. Even if the early stopping boundary is crossed at week 52 on the basis of the data from the 208 completers, we must still take into account the additional 208 randomized subjects who enrolled between week 26 and week 52 for whom the week 26 endpoint will not yet have been attained. These additional 208 subjects are referred to as the “overruns”. When the overruns are accounted for, the saving in sample size due to early stopping is only $694 - 416 = 278$ subjects, rather than $694 - 208 = 486$ subjects. The power and expected sample size values of the group sequential Plan 3 for different choices of δ are displayed in Table 2. The table shows the impact of overruns on the expected sample size. For comparison we have also included corresponding power and sample size values for the fixed sample Plan 2 in Table 2.

Table 2: Operating Characteristics of Plan3 (Group Sequential) and Plan2 (Fixed Sample)

δ	Plan3 (Group Sequential)				Plan2 (Fixed Sample)	
	Probability of Early Stopping	Expected Sample Size		Power	SampSiz	Power
		No Overruns	With Overruns			
1.6	6.6%	662	676	80%	690	80%
1.7	7.9%	656	672	84%	690	85%
1.8	9.3%	649	668	88%	690	88%
1.9	11.0%	640	663	91%	690	91%
2.0	13.0%	631	658	94%	690	94%

It is seen from Table 2 that Plan 3 offers a modest benefit relative to Plan 2. After accounting for the overruns, the expected sample sizes under Plan 3 range between 658 and 676 for corresponding values of δ between 2 and 1.6, as compared to a fixed sample size of 690 under Plan2. In terms of power, Plan 2 and Plan 3 are practically identical. For the current trial a group sequential design with conservative error spending offers no substantial advantage over a conventional single look design with a fixed sample size. One is still faced with the dilemma of committing excessive sample size resources up front in order to ensure adequate power at $\delta = 1.6$, with limited prospects of saving on sample size in the event that $\delta = 2$.

Although in general group sequential designs do offer savings in expected sample size, their actual benefit may be diminished if a study enrolls subjects very rapidly but the primary endpoint can only be observed after a lengthy follow-up. In the current example we assumed that subjects are enrolled at the rate of 8 per week and the endpoint is observed after 26 weeks of follow-up for each subject. This resulted in 208 additional subjects being on-study who were not yet followed for 26 weeks at the time of the interim analysis. The efficiency loss due to an overrun of this magnitude was difficult to overcome. If instead the enrollment rate were to be halved to 4 subjects per week, and the endpoint were to be observed after only 12 weeks instead of 26 weeks, there would only be an overrun of 48 subjects, and the resulting operating characteristics of the two group sequential designs would be more favorable relative to the

corresponding fixed sample design. The accrual rate and the duration of follow-up are thus two extremely important design parameters for a group sequential trial.

We next consider adopting an adaptive design for this study. This is a radically different approach to trial design in which the difficulties encountered by group sequential designs – rapid accrual, delayed endpoint, and large up-front commitment of patient resources – can to some extent be mitigated.

4 Adaptive Design

To motivate the adaptive design let us recall that although the actual value of δ is unknown, the investigators believe that $\delta \geq 2$. For this reason Plan 1 was constructed to have 80% power to detect $\delta = 2$. Plan 2 on the other hand was constructed to have 80% power to detect $\delta = 1.6$, the smallest clinically meaningful treatment effect. If there were no resource constraints one would of course prefer to design the study for 80% power at $\delta = 1.6$ since that would imply even more power at $\delta = 2$. However, as we saw in Table 1, this conservative strategy carries as its price a substantially larger up-front sample size commitment which is, moreover, unnecessary if in truth $\delta = 2$. Plan 3 was therefore constructed as a group sequential alternative to Plan 2. Plan 3 also has 80% power to detect $\delta = 1.6$ but there is a possibility of early stopping. We have seen, however, that due to the overruns problem, the expected sample size savings realized by Plan 3 is small while the up-front sample size commitment is large.

The above difficulties lead us to consider whether Plan 1, which was intended to detect $\delta = 2$ with 80% power and hence does not have such a large up-front sample size commitment, might be improved so as to provide some insurance against substantial power loss in the event that $\delta = 1.6$. The adaptive approach is suited to this purpose. In this approach we start out with a sample size of 442 subjects as in Plan 1, but take an interim look after data are available on 208 completers. The purpose of the interim look is not to stop the trial early but rather to examine the interim data and continue enrolling past the planned 442 subjects if the interim results are promising enough to warrant the additional investment of sample size. This strategy has the advantage that the sample size is finalized only after a thorough examination of data from the actual study rather than through making a large up-front sample size commitment before any data are available. Furthermore if the sample size may only be increased but never decreased from the originally planned 442 subjects, there is no loss of efficiency due to overruns. The technical problem of avoiding inflating the type-1 error despite increasing the sample size in a data dependent manner has been solved by, among others, Cui, Hung and Wang (1999).

4.1 Selecting the Criteria for an Adaptive Sample Size Increase

The operating characteristics of an adaptive design depend in a complicated way on the criteria for increasing the sample size after observing the interim data. These criteria may combine objective information such as the current estimate of δ or the current conditional

power with assessments of safety and with information available from other clinical trials that was not available at the start of the study. The adaptive approach provides complete flexibility to modify the sample size without having to pre-specify a precise mathematical formula for computing the new sample size based on the interim data. Therefore the full benefit of the flexibility offered by an adaptive design cannot be quantified ahead of time. Nevertheless it is instructive to investigate power and expected sample size by simulating the trial under different values of δ and applying precise pre-specified rules for increasing the sample size on the basis of the observed interim results. This will provide at least some idea, at the design stage, of the trade-off between the fixed sample or group sequential approaches and the adaptive approach.

To this end we create Plan 4, a design with 80% power to detect $\delta = 2$ with a one-sided level-0.025 test, based on a planned enrollment of 442 subjects. Plan 4 specifies, in addition, that there will be one interim analysis after 26 weeks of follow-up data are available on the first 208 subjects enrolled. The purpose of the interim analysis is not to stop the trial early but rather to examine the interim data and decide whether a sample size increase is warranted. If no action were taken at the interim look, Plan 4 would be identical to Plan 1. The timing of the interim look reflects a preference for performing the interim analysis as late as possible but nevertheless while the trial is still enrolling subjects since, once the enrollment sites have closed down, it will be difficult to start them up again. Under the assumption that subjects enroll at the rate of 8 per week we will have enrolled 416 subjects by week 52; 208 of them will have completed the required 26 weeks of follow-up for the primary endpoint, and an additional 208 subjects will comprise the overruns. Only the data from the 208 completers will be used in making the decision to increase the sample size. After this decision is taken, enrollment will continue until the desired sample size is attained. The primary efficacy analysis will be based on the full 26 weeks of follow-up data from all enrolled subjects. It should be noted that, unlike the group sequential setting where the 208 overruns at the time of the interim look played no role in the early stopping decision, here the data from the 208 overruns will be fully utilized in the primary efficacy analysis which will only occur when all enrolled subjects have completed 26 weeks of follow-up. This is one of the advantages of the adaptive approach relative to the group sequential approach for trials with lengthy follow-up.

It remains to specify the criteria for increasing the sample size at the interim look. A well planned trial should pre-specify as far as possible the decision rules to be adopted for increasing the sample size once the interim data are available. Thereby the operating characteristics of the trial can be studied through simulation and if they are unsatisfactory, the rules for sample size adaptation can be modified. It should be stressed, however, that in practice there is flexibility to overrule these pre-specified rules should unexpected results, either internal or external to the trial, be encountered at the time of the interim analysis. Nevertheless a precise formula for increasing the sample size must be pre-specified for purposes of simulation. While there are an infinite number of ways to construct such a formula it must address the following three questions:

- For what range of interim outcomes should a sample size increase be contemplated?
- How should the magnitude of the new sample size be calculated?

- What should be the upper limit to the sample size increase?

The answers to these questions might be driven by both clinical and business concerns, and will depend on the importance the investigators place on avoiding a false negative outcome for the current trial.

Range of Interim Outcomes for a Sample Size Increase

It is convenient to partition the sample space of possible interim outcomes into three zones; *unfavorable, promising and favorable*. An adaptive strategy is built on the premise that if the interim outcome lies in either the unfavorable or favorable zones, it is unnecessary to alter the sample size. In one case it would be risky to invest further in what appears to be a failed trial, while in the other case the trial appears slated to succeed anyway, without an additional sample size investment. Thus an adaptive sample size increase is only intended to help studies whose interim results fall in a promising zone, between these two extremes. How might these three zones be identified? One could use the interim estimate $\hat{\delta}$ or its standardized version $z = \hat{\delta}/se(\hat{\delta})$ to partition the sample space into the three zones. Alternatively one could rely on the conditional power or *probability of obtaining a positive outcome at the end of the trial, given the data already observed*. The conditional power approach is favored by most practitioners because it has a meaningful interpretation that is independent of the type of endpoint being measured, and incorporates both the current estimate of treatment effect as well as its standard error. Accordingly for the present trial we pre-specify that a sample size increase will only be contemplated if the conditional power at the interim look lies between 30% and 80%. That is, the unfavorable zone is characterized by conditional power values at most equal to 30%, the promising zone by conditional power values between 30% and 80% and the favorable zone by conditional power values at least equal to 80%.

Computing the Required Sample Size Increase

Just as at the design stage of a trial the sample size is determined by the desired power (80%, say) to detect an anticipated value of δ , so also at the time of the interim analysis the new sample size may be determined by the desired conditional power (also 80%, say) to detect an anticipated value of δ . Now, however, data from the actual trial are available, and may be used to update the anticipated value of δ at which to power the trial. One could, if desired, incorporate prior beliefs, external information and current data into a value of δ at which to power the study. For simplicity however, we shall use the estimate of δ obtained at the interim analysis to recompute the sample size so as to achieve 80% conditional power. It is possible that this calculation could result in a reduction in the total sample size. This is permitted by the statistical methodology of adaptive designs. For the current example, however, we do not wish to decrease the sample size. Therefore if the recomputed sample size constitutes a decrease, the original sample size of 442 subjects will be used.

Specifying an Upper Limit to the Sample Size Increase

Since resources are limited there must be an upper limit to the sample size increase, no matter what sample size is required to attain 80% conditional power. This upper limit is usually restricted to between 50% and 100% of the original sample size and is pre-specified at the start of the trial. Larger sample size increases are undesirable since they could yield statistically significant outcomes that are clinically non-significant. For the current trial we pre-specify an

upper limit of 884 subjects. That is, we are prepared to double our investment in the trial, but only if the interim estimate of conditional power falls in the favorable zone.

Finally, the design specifications of the adaptive Plan 4 are as follows:

1. The initial sample size is 442 subjects, and has 80% power to detect $\delta = 2$ with a one-sided level-0.025 test
2. An interim analysis is performed after data are available on 208 completers with 26 weeks of follow-up data
3. At the interim analysis the conditional power is computed using the estimated value $\hat{\delta}$ as though it were the true value of δ . If the conditional power lies between 30% and 80%, the interim outcome is deemed to be promising
4. If the interim outcome is promising, the sample size is re-computed so as to achieve 80% conditional power at the estimated value, $\hat{\delta}$. The original sample size is then updated to the re-computed sample size, subject to the constraint in item 5 shown below
5. If the re-computed sample size is less than 442, the original sample size of 442 subjects is used. If the re-computed sample size exceeds 884, the sample size is curtailed to 884 subjects

4.2 Operating Characteristics of Adaptive Design

Due to the complex adaptive scheme for re-computing sample size the operating characteristics of Plan 4 can best be evaluated by simulation. Table 3 displays power and expected sample size values for selected values of δ between 1.6 and 2.0, based on 100,000 simulations of Plan 4. For comparative purposes corresponding power and sample size values for Plan 1 are also displayed.

Table 3: Operating Characteristics of Plan 1 (Fixed Sample) and Plan 4 (Adaptive)

Value of δ	Plan 1(Fixed Sample)		Plan 4 (Adaptive)	
	Power	Expected SampleSize	Power	Expected Sample Size
1.6	61%	442	67%	509
1.7	66%	442	72%	507
1.8	71%	442	76%	506
1.9	76%	442	81%	502
2.0	80%	442	84%	499

All Plan 4 results are based on 100,000 simulated trials

The power of the adaptive Plan 4 has increased by 6% at $\delta = 1.6$ and by 4% at $\delta = 2$ compared to Plan 1. These power gains were obtained at the cost of corresponding average

sample size increases of 67 subjects at $\delta = 1.6$ and 57 subjects at $\delta = 2$. The gains in power appear to be fairly modest, especially as they are offset by corresponding sample size increases. However Plan 4 offers a significant benefit in terms of risk reduction, not reflected in Table 3. To see this it is important to note that the sample size under Plan 4 is only increased when the interim results are promising; i.e., when the conditional power at the interim analysis is greater than 30% but less than 80%. This is the very situation in which it is advantageous to increase the sample size and thereby avoid an underpowered trial. When the interim results are unfavorable (conditional power $\leq 30\%$) or favorable (conditional power $\geq 80\%$), a sample size increase is not warranted and hence the sample size is unchanged at 442 subjects for both Plan 1 and Plan 4. But when the interim results are promising (conditional power between 30% and 80%) the sample size is increased under Plan 4 in an attempt to boost the conditional power back to 80%. It is this feature of the adaptive design that makes it more attractive than the simpler fixed sample design.

Table 4 displays the probability of falling into the unfavorable, promising and favorable zones at the interim look, along with the power and expected sample size, conditional on falling into each zone, under both Plan 1 and Plan 4. The table highlights the key advantage of the

Table 4: Operating Characteristics of Plan 1 and Plan 4 Conditional on Interim Outcome

δ	Interim Outcome	Probability of Interim Outcome	Power Conditional on Interim Outcome		Expected Sample Size	
			Plan 1	Plan 4	Plan 1	Plan 4
1.6	Unfavorable	33%	27%	27%	442	442
	Promising	26%	61%	83%	442	696
	Favorable	41%	87%	87%	442	442
1.7	Unfavorable	29%	32%	32%	442	442
	Promising	26%	65%	86%	442	694
	Favorable	45%	90%	90%	442	442
1.8	Unfavorable	26%	36%	36%	442	442
	Promising	26%	69%	89%	442	690
	Favorable	48%	91%	91%	442	442
1.9	Unfavorable	23%	41%	41%	442	442
	Promising	25%	72%	91%	442	689
	Favorable	52%	93%	93%	442	442
2.0	Unfavorable	20%	44%	44%	442	442
	Promising	24%	76%	93%	442	685
	Favorable	56%	95%	95%	442	442

All results are based on 100,000 simulated trials

adaptive Plan 4 compared to the fixed sample Plan 1; i.e., the ability to invest in the trial in stages, with the second stage of the investment being required only if promising results are

obtained at the first stage. This feature of Plan 4 makes it far more attractive as an investment strategy than Plan 1 which has no provision for increasing the sample size if a promising interim outcome is obtained. Suppose, for example that $\delta = 1.6$, the smallest clinically meaningful treatment effect. The trial sponsor only commits the resources needed for 442 subjects at the start of the trial, at which point the chance of success is 61%, as shown in Table 3. The additional sample size commitment is forthcoming only if promising results are obtained at the interim analysis, and in that case the sponsor's risk is substantially reduced because the chance of success jumps to 83%, as shown in Table 4. Similar results are observed for the other values of δ .

The probabilities of entering the unfavorable, promising and favorable zones at the interim analysis, displayed in Table 4, are instructive. Consider again the case $\delta = 1.6$. At this value of δ there is a 26% chance of landing in the promising zone and thereby obtaining a substantial power boost under Plan 4 as compared to Plan 1. That is, 26% of the time the adaptive strategy can rescue a trial that is underpowered at the interim look. The chance of entering the favorable zone is 41%. That is, 41% of the time the sponsor will be lucky and have a well powered trial at the interim look without the need to increase the sample size. The remaining 33% of the time the sponsor will be unlucky and will enter the unfavorable zone from which also there is no sample size increase, and the chance of success is only 27%. These odds improve with larger values of δ .

5 Concluding Remarks

Many small companies with new molecules or technologies under development often rely on outside investors or large pharmaceutical companies for financing their phase 3 trials. The two-stage nature of the investment, with the second installment being obligated only if the interim results have significantly increased the odds of success, might make the adaptive design more attractive to outside investors than a conventional design requiring a fixed investment up-front, even when the two designs have an equivalent unconditional risk profiles. Simulations, performed prior to starting the trial, are necessary to quantify the risks and benefits involved in selecting an adaptive design in preference to a conventional fixed sample or group sequential design, and to enable the sponsor to make an informed decision.

A major additional benefit of the adaptive approach is flexibility. The adaptive methodology controls the type-1 error even if the pre-specified criteria for increasing the sample size are overruled at the interim analysis. This might be desirable for a variety of reasons both internal and external to the current trial. For example, in addition to observing a promising outcome at the interim time analysis, the safety profile for the test drug might turn out to be far superior to what was originally anticipated, and this might make the new drug more competitive in the marketplace. One could therefore justify increasing the sample size by a larger amount than that determined by the pre-specified rules, and thereby further reduce the chances of a false negative outcome. Another possible situation in which one might overrule the pre-specified criteria for sample size change would be if compelling results from other

clinical trials on comparable populations, treated with the same class of drugs became available and caused the sponsor to revise the value of δ at which to power the current study. Ideally one would wish to adhere strictly to the pre-specified criteria for sample size change since the operating characteristics of the design would change if they were overruled. This would certainly be the preference of regulatory authorities. As a practical matter, however, it is not possible to anticipate every contingency under which a sample size change is desirable. It is a strength of the adaptive approach that the validity of the statistical test at the end of the trial is not affected by unanticipated developments arising over the course of the clinical trial that necessitate making changes to the pre-specified criteria for sample size adaptation.

Adaptive trials require very careful up-front planning. An independent interim analysis review committee (IARC) must be appointed with the responsibility to actually implement the adaptive decision rules. A charter listing the members of the IARC, describing their roles and responsibilities, and providing the details of the proposed adaptations must be created. The charter should also discuss the steps that will be taken to ensure that the interim results remain confidential, as premature disclosure of interim results to the trial investigators could compromise the trial. Finally, regulatory approval must be secured in advance through a special protocol assessment (SPA). For this purpose the sponsor is required to submit the protocol, the charter and the simulations backing up the statistical validity of the proposed adaptive approach in good time. Because of all these complexities an adaptive design might not always be the right choice. The more established fixed sample and group sequential designs should always be evaluated alongside an adaptive design. Simulations play a crucial role in understanding the operating characteristics of an adaptive design and deciding whether it is an appropriate choice for the trial under consideration. There should be a tangible, quantifiable benefit arising from the decision to take the adaptive route.

Acknowledgements

The author thanks Dr. Howie Golub for helpful discussions on adaptive designs
Software support for this article was provided by the East (2008) software package.

References

- Cui L, Hung HMJ, Wang S-J (1999). Modification of sample size in group sequential trials. *Biometrics*, 55, 853-857.
- East-5 (2008). *Software for the Design and Interim Monitoring of Flexible Clinical Trials*. Cytel Corporation, Cambridge, MA.
- Hwang IK, Shih WJ, and DeCani JS (1990). Group sequential designs using a family of type I error probability spending functions. *Statistics in Medicine*, 9, 1439-1445.
- Lan KKG, DeMets DL (1983). Discrete sequential boundaries for clinical trials. *Biometrika*, 70, 659-663.

O'Brien PC, Fleming TR (1979). A multiple testing procedure for clinical trials. *Biometrics*, 35, 549-56.

Pocock S (2005). When not to stop a clinical trial. *JAMA*, 204, 17.