Published in final edited form as:

Clin Trials. 2013 April ; 10(2): . doi:10.1177/1740774512468806.

An analysis of adaptive design variations on the sequential parallel comparison design for clinical trials

Michael Y. Mi^a and Rebecca A. Betensky^{b,*}

^aHarvard Medical School, Harvard University, Boston, MA 02115, USA

^bHarvard School of Public Health, Harvard University, Boston, MA 02115, USA

Abstract

Background—Currently, a growing placebo response rate has been observed in clinical trials for antidepressant drugs, a phenomenon that has made it increasingly difficult to demonstrate efficacy. The sequential parallel comparison design (SPCD) is a clinical trial design that was proposed to address this issue. The SPCD theoretically has the potential to reduce the sample size requirement for a clinical trial and to simultaneously enrich the study population to be less responsive to the placebo.

Purpose—Because the basic SPCD design already reduces the placebo response by removing placebo responders between the first and second phases of a trial, the purpose of this study was to examine whether we can further improve the efficiency of the basic SPCD and if we can do so when the projected underlying drug and placebo response rates differ considerably from the actual ones.

Methods—Three adaptive designs that used interim analyses to readjust the length of study duration for individual patients were tested to reduce the sample size requirement or increase the statistical power of the SPCD. Various simulations of clinical trials using the SPCD with interim analyses were conducted to test these designs through calculations of empirical power.

Results—From the simulations, we found that the adaptive designs can recover unnecessary resources spent in the traditional SPCD trial format with overestimated initial sample sizes and provide moderate gains in power. Under the first design, results showed up to a 25% reduction in person-days, with most power losses below 5%. In the second design, results showed up to a 8% reduction in person-days with negligible loss of power. In the third design using sample size re-estimation, up to 25% power was recovered from underestimated sample size scenarios.

Limitations—Given the numerous possible test parameters that could have been chosen for the simulations, the study's results are limited to situations described by the parameters that were used, and may not generalize to all possible scenarios. Furthermore, drop-out of patients is not considered in this study.

Conclusions—It is possible to make an already complex design such as the SPCD adaptive, and thus more efficient, potentially overcoming the problem of placebo response at lower cost. Ultimately, such a design may expedite the approval of future effective treatments.

Correspondence to: Rebecca A. Betensky, Ph.D., Professor of Biostatistics, Department of Biostatistics, Harvard School of Public Health, 655 Huntington Avenue, Building II Room 421, Boston, Massachusetts 02115, USA, Phone: 617.432.2821, betensky@hsph.harvard.edu.

Keywords

clinical trial; adaptive design; sequential parallel comparison design; power; SPCD; placebo response

Introduction

Demonstrating the efficacy of psychoactive drugs has become increasingly difficult because of the considerable rise in patients' response to placebo [1–6]. Possible contributing factors include the cumulative effects of nonspecific interventions, changes in the patient populations being recruited into the studies, use of an inactive versus active placebo, and regression toward the mean [2,3,7]. Other contributing factors may arise from the actual study itself, such as amount of patient interaction with physicians and effective control of blinding [2]. Methods have been proposed to minimize these aforementioned effects, such as standardizing diagnostic procedures and patient-doctor interactions, using a placebo lead-in trial before the start of a clinical trial, requiring multiple measurements to screen for eligibility, and enriching the study population for chronic disease sufferers [2,3,7].

Motivated by the rising placebo response problem, Fava et al. [2] proposed the sequential parallel comparison design (SPCD), which aims to reduce both the placebo response rate and the sample size requirement. In the basic SPCD format, patients are randomly allocated to one of three groups, drug/placebo, placebo/drug, and placebo/placebo, which follow a distribution of $1-2a$, a , and a respectively with a being a number between 0 and 0.5. At the completion of phase I, patients are designated responders or non-responders based on pre-determined study criteria. Responders and drug non-responders exit the double-blind study and either continue with open-label therapy or discontinue treatment altogether. Placebo non-responders enter phase II and receive either drug or placebo. For example, a phase I responder to placebo in the placebo/drug group would exit the study after phase I while a phase I non-responder to placebo in the placebo/drug group will go on to drug in phase II. The value for a can be optimized to maximize the power of the test statistic if the underlying drug and placebo response rates are known. Last, the original linear combination SPCD weighs the difference detected in each phase differently using the parameter w —a value between 0 and 1—which is the weight placed on the phase I difference; consequently, $1-w$ is the weight placed on the phase II difference. Like a , w can also be optimized if the underlying response rates are known. The advantages of the SPCD are that it reduces the required number of patients in the trial by pooling two phases of data for analysis and that the patients in the second phase may have a reduced placebo response, allowing for observation of a larger treatment effect.

Tamura and Huang examined the efficiency of the design with binary endpoints while incorporating the possibility of dropout before phase II and with continuous endpoints [8]. In general, the reduction in sample size over the conventional design was between 15–30%. However, the SPCD does increase the length of the trial for each subject by extending it to two phases. Depending on accrual, and treatment period, it may or may not increase the overall length of the trial.

Although the SPCD is more efficient than conventional designs, it is highly dependent on *a priori* estimates of drug and placebo responses, which may not always be accurate. In such cases, the SPCD could potentially lose much of its advantage over conventional designs. In the context of group sequential clinical trials, adaptive designs have been proposed to give investigators and sponsors the option to stop a study early for futility or to change certain design features mid-trial [9]. The purpose of the current study is to incorporate features of

existing adaptive designs into the SPCD to make it more robust and provide additional gains in power or savings in resources. In particular, interim analysis was used to implement three adaptations. Because adaptive designs can be too complex to solve analytically for power or type I error rate, the results for this study were empirically calculated using simulations.

Simulations

To conduct the appropriate simulations for the SPCD, a model was created using the programming language R. The basic model generated patient information and clinical trial outcome based on the following set of parameters taken from the SPCD: p_1 , the response rate of patients to drug during phase I; p_2 , the response rate of patients to drug during phase II; q_1 , the response rate of patients to placebo—or a current standard treatment—during phase I; q_2 , the response rate of patients to placebo during phase II; n , the number of patients recruited; a , the distribution of patients in the SPCD groups; and w , the weight applied to the test statistics from the two phases of the trial. Using these parameters, a list of patients was randomly generated and assigned to the three arms of the design: placebo-placebo, placebo-drug, and drug-placebo, i.e., each patient was assigned at the beginning of the trial both phase I and phase II treatments, but, with respect to phase II, only phase I placebo non-responders are included in the efficacy analysis. The patients' responses to drug or placebo in each phase were then generated. As in Fava et al. [2], patient responses were assumed to be binary. In addition, each patient was assigned an accrual day uniformly between 1 and 400. Each phase lasted 30 days.

The previously mentioned study by Tamura and Huang [8] suggested that the values of a and w that optimize power for a wide range of drug and placebo response rates are 0.2857 and 0.7–0.9 respectively. However, in one of the few clinical trials that uses the SPCD, Fava et al. [10] selected 0.375 and 0.5 as the pre-randomization values of a and w respectively. This study opted to use 0.2857 and 0.8 for a and w , given that they appear optimal when nothing is known about the drug or placebo response, information that Fava et al. may have had when they designed their trial. The sample size, n , was 225 in all simulations, since that was the number of patients in the above clinical trial [10]. It was chosen to mimic the real life limitations on patient recruitment.

After patient data were generated, estimates of the response rates were tabulated. Because the SPCD involves two phases for analysis, a modified traditional two proportions test statistic was used, namely, $\hat{z} = \hat{h}/\hat{s}$, where:

$$\hat{h} = w(\hat{p}_1 - \hat{q}_1) + (1 - w)(\hat{p}_2 - \hat{q}_2)$$

$$\hat{s}^2 = \frac{2(2a - 1)(w - 1)^2(-\hat{p}_2 + \hat{p}_2^2 - \hat{q}_2 + \hat{q}_2^2) + (\hat{q}_1 - 1)((\hat{q}_1 - 1)\hat{q}_1 + 2a(\hat{p}_1^2 - \hat{q}_1^2 - \hat{p}_1 + \hat{q}_1))w^2}{2a(2a - 1)(\hat{q}_1 - 1)n}$$

Those quantities with hats on them denote their estimates using the simulated data. The power of the test is $\Phi(z - 1.96) + \Phi(-z - 1.96)$, where Φ is the cumulative distribution of the normal distribution. For each set of parameters, 10,000 trials were simulated, from which the empirical power was calculated.

In this study, interim analyses were carried out to implement three different adaptive design variations of the basic SPCD. The interim evaluations were conducted at day 200 of the simulated trials unless otherwise specified. A total of 153 sets of parameters with varying

values of p_1 , p_2 , q_1 , and q_2 were tested to observe the influence of these parameters on the power. These parameters were based on those examined by Fava et al. [2]. Another 11 sets of parameters were used to measure type I error rates for the interim analyses. Subsets of the results are provided in this paper to illustrate general and representative trends.

Interim Analysis I

Rationale and Method

In the first interim analysis, the estimate for the placebo response rate q_1 at the interim time was compared to a predetermined threshold, c , and if the estimated rate were less than c , then an adaptive design would be implemented. The rationale for this design was that if the placebo response were low enough, then phase II would not be necessary to detect a drug's efficacy over placebo [2]. Furthermore, because drug efficacy might be detected as early as four weeks into some trials [3], not all patients would need to complete the SPCD trial. Therefore, if the evaluation criterion were met, patients not yet enrolled and those who had not yet started phase II would not continue with phase II of the trial. Furthermore, patients not yet enrolled would be reassigned treatment groups so as to achieve a drug to placebo allocation ratio of as close as possible to 1:1, rather than the 1–2 a :2 a ratio afforded by the basic original SPCD. Consequently, at the end of the restructured adaptive study, there were some patients who enrolled in two phases of SPCD and some patients who participated only in phase I.

By taking advantage of independence between the two patient groups, the final test statistic was calculated using the combined data of the two groups. Those enrolled up to the interim analysis period contributed to a partial SPCD test statistic, while those who enrolled after the interim analysis time—if the condition was met—contributed to a conventional two proportion test statistic. Using a variance spending method proposed by Fisher [11] to preserve the overall type I error rate in the presence of an interim analysis with potential adaptation, we weighted the test statistics from before and after the interim analysis so that the combined null distribution would still have mean zero and variance one.

For each set of simulation parameters, empirical power and savings in person-days deriving from early termination of phase II were recorded. Percent savings was based on how many person-days were saved compared with the expected total (13,500 person-days, derived from 225 total patients and 60 days of treatment). The interim analysis time was day 200, the midpoint of the accrual period, which lasted 400 days, and the cut-off criterion, c , was 0.3.

Results and Conclusions

Despite the implementation of interim analysis 1 with the Fisher variance spending method [11] to control for type I error rate, the SPCD and interim analysis 1 design still exceeded the an error rate of 0.05. After additional simulations, it was thought that this inflation of type I error rate could be contributed to a slow convergence of the SPCD to normality as larger sample sizes resulted in better type I error rates. Nevertheless, to control for type I error with small sample sizes, a threshold of 0.045 was used instead of 0.05 to obtain a final error rate below 0.05, as reflected in Figure 1A.

The power differences between the basic SPCD and the interim design were generally 1% or less (Table 1). Of the 153 sets of parameters tested, few had power losses greater than 1%. In general, power losses were high in cases that had large optimal a values and low optimal w values, which meant that more patients went on to phase II, and more weight was placed on phase II.

Consequently, the elimination of phase II by the interim design could reduce more power. On the other hand, the gains in savings were significant in those settings. The simulated results showed up to 25% reduction in total person-days with the savings being dependent on the time at which the interim analysis was conducted and c , the cut-off threshold.

To probe the effects of the placebo response threshold and the interim evaluation time on power and savings, a range of values of each was examined. A threshold of 0.3 can provide substantial savings with minimum to negligible decreases in power (Supplemental Figure 1). Interim analysis conducted as early as day 80 (i.e., at 20% of planned accrual) had little impact on the power. Results from simulations using other parameters support this conclusion. With 225 patients, enough patients will have enrolled by day 80 to give a good estimate of q_I . In most cases where q_I is less than 0.35, the sample size required is less than 225.

In conclusion, this interim analysis design is useful when the sample size exceeds actual power requirements, and recovery of some of the planned costs is desirable. Furthermore, it has the advantage of subjecting fewer patients to prolonged exposure to either placebo or experimental treatment. For the most part, this design adequately shortens the duration of the trial without unacceptable sacrifice to power. Because a few cases resulted in up to 9% absolute reduction in power, this risk should be considered when evaluating whether to pursue this adaptive design.

Interim Analysis 2

Rationale and Methods

Building on the previous design, the second proposed interim analysis evaluated whether the estimate of $p_I - q_I$, the phase I treatment effect, at the interim time exceeded a threshold, c . We note that this threshold, c , is not the same as that used for interim analysis 1. The rationale for this design was that if the difference between drug and placebo detected in phase I were large enough, then the data from phase II would not be necessary. Consequently, all patients who had not started phase II and all patients who had not yet enrolled would not enter phase II. Patients not yet enrolled would be reassigned to achieve as close to an overall allocation ratio of 1:1 as possible. The two proportions test was substituted for the SPCD test statistic for these patients. Empirical power and savings in person-days were recorded for various sets of test parameters.

This design had greater potential for improved efficiency over basic SPCD than the first proposed adaptive design. However, by directly probing the effect size of drug over placebo, this design also had greater potential to inflate the type I error rate. We tried to control for this by using a similar variance spending method as in interim analysis 1. Percent savings was based on how many person-days were saved compared with the maximum (13,500 person-days). The interim time was day 200, and the threshold for $p_I - q_I$ was 0.3.

Results and Conclusions

This adaptive design had similar issues with type I error rate as interim analysis 1, but the same method of using a threshold of 0.045 successfully controlled for the final error rate (Figure 1B). Savings in person-days and changes in power largely depend on the timing of the interim analysis and the threshold at which a change would be made in the design after the interim analysis. Therefore, a range of values for each variable was examined. With a lower cut-off threshold, there was greater savings in person-days, but there was also an associated drop in power compared to the SPCD design (Supplemental Figure 2). A cut-off of 0.25 was optimal, which incidentally was the largest difference between p_I and q_I that Fava et al. examined [2].

An earlier interim analysis time was associated an increase in savings of person-days, but did not have much impact on power. The sample size with which interim analysis 2 was simulated was not affected by the ideal SPCD sample size beyond providing greater power whenever the former exceeded the latter. However, a smaller SPCD sample size requirement was correlated with a larger difference between p_1 and q_1 , which resulted in a greater likelihood of meeting the cut-off threshold of interim analysis 2 and greater savings. Because values of a and w determined the impact of terminating phase II, as the results from interim analysis 1 suggested, a separate set of parameters was tested with varying values of the threshold and interim evaluation time to investigate the effects of the use of suboptimal values for a and w .

Similar trends were shown for other parameters as well. Most notably, this adaptive design demonstrated gains in power over the basic SPCD for sets of parameters with suboptimal estimates of a and w . This result suggests that this design can effectively recover power when the estimates of a and w are incorrect prior to the start of a trial. In the context of this analysis, it is logical that when more weight is placed on phase I data, as indicated by a high w and low a , there would be less of an impact on the power when phase II data are discarded than when less weight is placed on phase I. Therefore, our second proposed adaptive design may not provide as much improvement relative to the basic SPCD as would be desired.

Interim Analysis 3

Rationale and Methods

The third proposed interim analysis incorporated sample size re-estimation into the basic SPCD. Because the SPCD drastically reduces sample size requirements compared to the conventional design, it is rare in practice for the SPCD to overestimate¹ the required sample size. However, if the design values of the drug and placebo response parameters are inaccurate, it becomes necessary to increase the sample size to demonstrate efficacy. Wittes and Brittain [12] introduced a sample-size re-estimation method at an interim analysis, which they termed an internal pilot study. In their setting of normal data, this involved re-estimation of the variance of the data, a so-called nuisance parameter. Other methods have been suggested for studies with binary outcomes [13, 14]. In this study, we followed Fisher's [12] method for increasing sample size while controlling for type I error.

At the interim time, estimates of p_1 , q_1 , p_2 , and q_2 were used to re-calculate the sample size. If the new sample size exceeded the initial estimate, new patients were recruited to reach the target size. As with any sample-size re-estimation, it may not always be feasible to increase the sample size to the required level, and an upper bound must be placed. If the new sample size exceeds the upper bound, then one can either discontinue the study out of futility and not test the hypotheses or continue with the planned trial to measure secondary outcomes of the study. If the planned sample size exceeds the new sample size, one can stop early and recover some of the cost of the study. However, Wittes and Brittain cautioned that stopping a trial early may have a negative impact on the type I error rate and should be carefully evaluated [12]. In this study, the simulated trial was stopped at the interim analysis time if the re-estimated sample size was smaller than the number of patients already recruited. If the re-estimated sample size exceeded the upper bound, then more patients were recruited until the minimum of the new sample size and the upper bound was reached.

¹We recognize that the calculation of sample size is based on estimated parameters. Here and throughout the rest of the manuscript, we refer to this process as sample size "estimation" or "re-estimation" if it is conducted at the interim analysis.

The same sets of parameters were analyzed with interim analysis 3 as were used with interim analyses 1 and 2. Likewise with the previous two analyses, we started by looking at the type I error rate of this design and moved on to empirical power analyses.

Results and Conclusions

Similar to the previous interim analyses, this one also resulted in inflation of type I error rate. However, this time, a new threshold of 0.04 was needed to obtain a final type I error rate of less than 0.05 (Figure 1C). Table 3 shows the power analysis results for some of the parameters tested. Of note, we have included relevant sample size calculations other than the optimal sample size for comparison. In particular, because the specific α and w values used in a trial significantly influence the sample needed for the SPCD to achieve a certain power, this sample size should be the more appropriate sample size used for comparison to that derived from sample size re-estimation. Consequently, the examples chosen for this analysis illustrate both the power gained from sample size re-estimation and the possible problem with not choosing the optimal α and w values.

The sample size calculated from the α and w values selected prior to the trial beginning can vary widely from the optimal one (Table 3). It can be as highly accurate as within one patient of optimal or as inaccurate as needing more than 200 patients. If the optimal α and w values are not used, then the SPCD can have less power than that produced by the conventional two proportions test. In cases where the required sample size is smaller than the one estimated prior to the trial beginning, our sample size re-estimation can unnecessarily inflate the sample size need because we always recruit up to the interim analysis time. Nevertheless, in nearly all cases, the median sample size from sample size re-estimation is still smaller than the conventional sample size requirement. On the other hand, in cases where the starting sample size is underestimated, sample size re-estimation can recover up to 24.5% power (Table 3). The amount of power recovered is comparable to previous study of sample size re-estimation for standard trials [15], which suggests that the SPCD could also benefit from sample size re-estimation.

We probed the effects of the lower and upper bound on the outcome of sample size re-estimation. In this study, the time of interim analysis fixes both the amount of information available for sample size re-estimation and the lower bound for re-estimation. To illustrate the range of potential power increases from the sample size re-estimation, we chose parameters associated with an underestimate of the required sample size (Supplement Figure 3). With an increase in the upper bound for the re-estimated sample size, there was an associated increase in the power. For the example shown, the power increase ranged from about 11% to 19%. The median new sample size did not vary much with the increase in upper bound, but the mean new sample size increased considerably. This result suggests that for a majority of time, the re-estimated sample size was outside the bounds established, which resulted in the new sample size being capped at the bounds.

These results show that as the interim time increased, which was associated with an increase in the lower bound, the power increase ranged from 10% to 21%. For interim analysis times greater than 240, the increases in power started to plateau (Supplemental Figure 3). It is likely that not enough patient information was collected before that point to give accurate estimates of the drug and placebo response rates. After that point, the additional information helped increase the power, but not to the same degree as before. Investigators conducting clinical trials will need to decide whether the additional gains in power were worth monitoring the patients for longer periods.

Discussion

Overall, the three proposed adaptive design variations on the basic SPCD either reduced the resource requirements of the SPCD or increased the power without substantially increasing the type I error rate. In the first two designs, the main goal was to reduce the total person-days required for the trial by reducing the length of the trial for some of the patients. On average, the first interim analysis design achieved savings of 5% to 25% without substantial changes in power for some of the parameters tested. Additional savings reduced power. In general, this design worked best when the true response rates were low. The second interim analysis design achieved smaller savings of 2% to 8% with negligible changes in power, which was mostly due to the fact that the second interim analysis criterion was harder to satisfy than that of the first. Unlike the first interim analysis design, it achieved savings for a larger set of parameters, and it worked best when the initial estimates of the true response rates were very inaccurate. The third interim analysis design successfully extended a simpler sample size re-estimation method to the basic SPCD and achieved up to a 25% increase in power when the initial sample size was underestimated due to unknown p_1 , p_2 , q_1 , and q_2 . In general, given the same sample size, the SPCD had much better power than the conventional two-arm design [15], which suggests that the SPCD is generally a very efficient design that also can benefit from sample size re-estimation.

While it is possible that the basic original SPCD can drastically reduce the sample size requirements in certain cases, it relies on accurate estimates of the underlying drug and placebo response rates, which consequently determine a and w . As the result from interim analysis 3 shows, inaccurate response rates can dramatically affect the sample size requirement to the extent that the basic SPCD is somewhat inefficient. Consequently, it is important to collect preliminary data on the response rates of the patients to make reasonable estimates when using the SPCD. Alternatively, an adaptive design, such as those that we have proposed, could be used in conjunction with the basic SPCD to make the SPCD more robust.

One important issue that arose from the simulation is that the SPCD had difficulty achieving a type I error rate of 0.05. Additional analysis suggests that the SPCD does converge toward normality with a much larger sample size. Given that a large sample size is not feasible in reality, this paper proposed to correct this issue by setting a lower alpha threshold. Despite setting a stricter cut-off to control for type I error rates, however, the three interim analysis designs were still able to provide substantial gains in person-time savings or improvements in power.

Nevertheless, in practice, this means that meticulous simulations must be conducted prior to the start of a clinical trial to ensure that sponsors can have a more accurate assessment of what the expected type I error rate is.

In future work, we would like to perform more extensive simulations to better optimize the SPCD for given parameter spaces. We would also like to examine how we can optimize a and w to maximize power. Preliminary data show that a simple optimization function of w at the interim analysis would inflate the type I error rate. More research is needed to determine how to correct this and how to optimize the SPCD test statistic. We believe this is feasible given that other test statistics have been developed for the SPCD that do not utilize w . [16–17] We also explored a previously published method [15] for sample size re-estimation that allows blinding of response rates at interim analysis. It too suffered from inflation of type I error rate, an issue we would like to correct in the future.

In summary, the results from the three interim analysis designs provide evidence that the basic SPCD can be improved by incorporating aspects of adaptive designs. However, because the SPCD strongly depends on good estimates of patients' actual response rates to drug and placebo, more testing and simulations are needed to determine how exactly these adaptations are to be implemented, as well as what the loss in power in certain cases would be when translated into additional patients. In reality, situation specific simulations that incorporate known data would be conducted prior to the start of a clinical trial to understand how uncertainty may impact the outcome and how to optimize it. Furthermore, a metric needs to be developed so that the savings in sample size and trial length can be evaluated under real life economic constraints. The costs of recruiting a patient and the costs of keeping a patient on a trial for a certain period of time must be considered when weighing the advantages and disadvantages of the savings.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We appreciate the helpful suggestions of Dr. David Schoenfeld and Mr. Matthias Bowman.

Grant support: NIH CA075971; Harvard NeuroDiscovery Center; Harvard Clinical and Translational Science Center (NIH #UL1 RR 025758)

References

1. Quitkin FM. Placebos, drug effects, and study design: A clinician's guide. *Am J Psychiatry*. 1999; 156:829–836. [PubMed: 10360119]
2. Fava M, Evins AE, Dorner DJ, Schoenfeld DA. The problem of the placebo response in clinical trials for psychiatric disorders: Culprits, possible remedies, and a novel study design approach. *Psychother Psychosom*. 2003; 72:115–127. [PubMed: 12707478]
3. Walsh BT, Seidman SN, Sysko R, Gould M. Placebo response in studies of major depression: Variable, substantial, and growing. *JAMA*. 2002; 287:1840–1847. [PubMed: 11939870]
4. Khan A, Leventhal RM, Khan SR, Brown WA. Severity of depression and response to antidepressants and placebo: An analysis of the food and drug administration database. *J Clin Psychopharmacol*. 2002; 22:40–45. [PubMed: 11799341]
5. Kirsch I, Deacon BJ, Huedo-Medina TB, Scoboria A, Moore TJ, Johnson BT. Initial severity and antidepressant benefits: A meta-analysis of data submitted to the food and drug administration. *PLoS Med*. 2008; 5:e45. [PubMed: 18303940]
6. Fournier JC, DeRubeis RJ, Hollon SD, et al. Antidepressant drug effects and depression severity: A patient-level meta-analysis. *JAMA*. 2010; 303:47–53. [PubMed: 20051569]
7. Thomson R. Side effects and placebo amplification. *Br J Psychiatry*. 1982; 140:64–68. [PubMed: 7037102]
8. Tamura RN, Huang X. An examination of the efficiency of the sequential parallel design in psychiatric clinical trials. *Clin Trials*. 2007; 4:309–317. [PubMed: 17848492]
9. Chang M, Chow SC, Pong A. Adaptive design in clinical research: Issues, opportunities, and recommendations. *J Biopharm Stat*. 2006; 16:299–309. discussion 311–2. [PubMed: 16724486]
10. Fava, M.; Mischoulon, D.; Iosifescu, D., et al. A double-blind, placebo-controlled study of aripiprazole adjunctive to antidepressant therapy (ADT) among depressed outpatients with inadequate response to prior ADT (ADAPT-A study) [abstract]. NCDEU 2010 meeting; June 14 – 17, 2010; Boca Raton, FL. p. 59abstract session II-21.
11. Fisher LD. Self-designing clinical trials. *Statist. Med*. 1998; 17:1551–1562.
12. Wittes J, Brittain E. The role of internal pilot studies in increasing the efficiency of clinical trials. *Stat Med*. 1990; 9:65–71. discussion 71–2. [PubMed: 2345839]

13. Gould AL. Interim analyses for monitoring clinical trials that do not materially affect the type I error rate. *Stat Med.* 1992; 11:55–66. [PubMed: 1557576]
14. Herson J, Wittes J. The use of interim analysis for sample size adjustment. *Drug Information Journal.* 1993; 27:753–760.
15. Shih WJ, Zhao PL. Design for sample size re-estimation with interim data for double-blind clinical trials with binary outcomes. *Stat Med.* 1997; 16:1913–1923. [PubMed: 9304763]
16. Ivanova A, Qaqish B, Schoenfeld DA. Optimality, sample size, and power calculations for the sequential parallel comparison design. *Stat. Med.* 2011; 30:2793–2803. [PubMed: 21805484]
17. Tamura RN, Huang X, Boos DD. Estimation of treatment effect for the sequential parallel design. *Stat. Med.* 2011; 30:3496–3506. [PubMed: 22139795]

Figure 1. Type I error rates of SPCD and the three interim analysis designs

The n , a , and w values shown were the optimal ones that will achieve 90% power with the minimum number of patients for equal values of p and q (C1: $p_1=q_1=p_2=q_2=0.5$, C2: $p_1=q_1=p_2=q_2=0.4$, C3: $p_1=q_1=p_2=q_2=0.3$, C4: $p_1=q_1=p_2=q_2=0.2$). Each simulation used 225, 0.2857, and 0.8 for n , a , and w respectively. Interim time point was day 200. Cut-off for interim analysis 1 was 0.3. Cut-off for interim analysis 2 was 0.3. Upper bound sample size for interim analysis 3 was 900, and the lower bound was number of patients recruited at interim analysis.

Table 1

Performances of SPCD and interim analysis 1 design

The n , α , and w values shown were the optimal ones that will achieve 90% power with the minimum number of patients for the values of p and q listed. The simulation for this interim analysis design actually used 225, 0.2857, and 0.8 respectively for all trials. Percent savings was based on how many person-days were saved compared to the expected total (13,500 person-days). Interim time was day 200, and the cut-off threshold, c , was 0.3.

p_1	p_2	q_1	q_2	n	α	w	SPCD			Interim 1			Difference	Savings (%)
							Error	Power	Error	Power	Error	Power		
0.5	0.4	0.35	0.25	322	0.29	0.7		0.735		0.723		0.723	-0.012	5.29%
0.5	0.4	0.35	0.15	179	0.4	0.36		0.840		0.828		0.828	-0.013	5.50%
0.5	0.4	0.3	0.25	199	0.27	0.77		0.918		0.908		0.908	-0.01	12.17%
0.45	0.25	0.3	0.05	168	0.4	0.29		0.838		0.792		0.792	-0.046	12.57%
0.5	0.4	0.25	0.25	130	0.25	0.81		0.983		0.980		0.980	-0.003	19.96%
0.4	0.3	0.25	0.1	188	0.36	0.42		0.851		0.778		0.778	-0.073	20.16%
0.35	0.25	0.2	0.05	148	0.4	0.29		0.893		0.806		0.806	-0.087	24.03%
0.45	0.35	0.2	0.2	119	0.24	0.8		0.992		0.987		0.987	-0.006	24.11%
0.35	0.25	0.15	0.05	112	0.3	0.54		0.986		0.966		0.966	-0.02	24.91%
0.4	0.3	0.15	0.15	106	0.23	0.8		0.996		0.993		0.993	-0.003	24.92%
0.35	0.25	0.1	0.05	77	0.24	0.68		1.000		0.999		0.999	-0.001	24.97%
0.35	0.25	0.1	0.1	89	0.22	0.8		0.999		0.998		0.998	-0.001	24.99%

Table 2
Performances of SPCD and interim analysis 2 design

The n , α , and w values shown were the optimal ones that will achieve 90% power with the minimum number of patients for the values of p and q listed. The simulation for this interim analysis design actually used 225, 0.2857, and 0.8 respectively for all trials. Percent savings was based on how many person-days were saved compared to the expected total (13,500 person-days). Interim time was at day 200, and the cut-off threshold, c , was 0.3.

p_1	p_2	q_1	q_2	n	α	w	SPCD	Interim 2	Difference	Savings (%)
0.6	0.5	0.45	0.3	287	0.32	0.66	0.778	0.777	-0.001	1.74%
0.6	0.5	0.45	0.35	350	0.29	0.75	0.717	0.716	-0.001	1.79%
0.6	0.5	0.45	0.25	224	0.36	0.53	0.817	0.816	-0.002	1.84%
0.6	0.4	0.45	0.15	201	0.38	0.45	0.838	0.837	-0.002	1.84%
0.6	0.4	0.4	0.15	148	0.33	0.59	0.954	0.953	-0.001	3.88%
0.6	0.5	0.4	0.35	215	0.27	0.8	0.898	0.897	-0.001	3.98%
0.6	0.4	0.4	0.25	211	0.27	0.78	0.903	0.903	0	3.98%
0.6	0.4	0.4	0.2	182	0.3	0.69	0.935	0.935	0	4.01%
0.6	0.5	0.4	0.25	160	0.31	0.64	0.948	0.947	0	4.02%
0.6	0.5	0.4	0.3	189	0.29	0.73	0.927	0.927	-0.001	4.17%
0.6	0.5	0.35	0.3	129	0.27	0.77	0.986	0.986	0	7.59%
0.6	0.5	0.35	0.35	141	0.26	0.83	0.976	0.976	0	7.72%
0.6	0.4	0.35	0.2	126	0.28	0.75	0.989	0.988	0	7.72%
0.6	0.4	0.35	0.15	109	0.3	0.66	0.991	0.991	0	7.74%
0.6	0.3	0.35	0.05	95	0.32	0.55	0.995	0.994	-0.001	7.80%
0.6	0.4	0.35	0.25	140	0.26	0.82	0.980	0.979	0	7.90%
0.6	0.5	0.35	0.25	115	0.29	0.71	0.992	0.992	0	8.00%

Table 3

Performances of SPCD and interim analysis 3 design

The n , α , and w values shown were the optimal ones that will achieve 90% power with the minimum number of patients for the values of p and q listed. The simulation for this interim analysis design actually used 225, 0.2857, and 0.8 respectively for all trials. Interim time was at day 200. The lower bound for sample size re-estimation was the number of patients recruited by the interim time, and the upper bound for sample size re-estimation was 4 times the original sample size or 900 in this case. Median interim n was the median re-estimated sample size used for simulation with each set of parameters. The conventional n is the estimated sample size needed with the two proportions test to achieve 90% power.

p_1	p_2	q_1	q_2	α	w	n	Empirical Power			Median SSR n	Conventional n
							SPCD	Interim 3	Difference		
0.75	0.65	0.6	0.5	0.29	0.83	333	0.724	0.956	0.232	320	410
0.5	0.4	0.35	0.25	0.29	0.7	322	0.716	0.961	0.245	335	457
0.8	0.7	0.65	0.55	0.29	0.85	309	0.748	0.963	0.215	293	373
0.45	0.35	0.3	0.2	0.29	0.66	297	0.752	0.969	0.217	315	438
0.75	0.65	0.6	0.45	0.3	0.77	292	0.780	0.965	0.185	283	410
0.65	0.45	0.5	0.2	0.35	0.55	224	0.817	0.974	0.157	254	457
0.6	0.5	0.45	0.25	0.36	0.53	224	0.818	0.977	0.159	262	466
0.7	0.6	0.5	0.45	0.28	0.84	212	0.896	0.982	0.086	204	253
0.55	0.45	0.4	0.2	0.37	0.46	204	0.815	0.976	0.161	260	466
0.6	0.4	0.45	0.15	0.38	0.45	201	0.821	0.976	0.155	258	466
0.8	0.6	0.6	0.45	0.29	0.88	189	0.913	0.980	0.067	189.5	221
0.8	0.7	0.6	0.55	0.29	0.87	188	0.913	0.980	0.067	186	221
0.4	0.3	0.2	0.05	0.36	0.39	99	0.983	0.988	0.005	137	221
0.6	0.3	0.35	0.05	0.32	0.55	95	0.992	0.982	-0.010	110	168
0.45	0.25	0.2	0.05	0.28	0.61	94	0.996	0.984	-0.012	107	148
0.4	0.3	0.15	0.05	0.29	0.56	76	0.999	0.979	-0.020	90	135