

When should the sequential parallel comparison design be used in clinical trials?

Clin. Invest. (2013) 3(9), 823–833

Randomized clinical trials in many conditions such as mental illness, pain, addiction, allergy and asthma are subject to moderate-to-high rates of placebo response. This reduces the effect size for the comparison of the active treatment to the placebo group, hence reducing the efficiency of studies. Since attempts to resolve this problem using a placebo lead-in design have not been successful, in 2003 the sequential parallel comparison design (SPCD) was developed. In the SPCD, the study is conducted in two treatment stages, with the efficacy analysis including (a) all subjects in stage 1 and (b) stage 2 data from stage 1 placebo nonresponders. The relevant data from the two stages are pooled to compute an overall p-value. This article reviews the situations in which the SPCD is preferable to the standard parallel design as well as the circumstances in which SPCD is not a good choice.

Keywords: placebo response • randomized trial • sequential parallel comparison design

High rates of placebo response in clinical trials, even of drugs previously approved by the US FDA for a particular condition, have often failed to demonstrate a significant difference of the active treatment from placebo [1]. For example, Laughren reviewed randomized trials for new treatments for depression and schizophrenia that were considered by the FDA between 1987 and 1999 and that were ultimately approved [1]. Laughren found many cases in which neither the investigational drug nor the already-approved drug differed significantly from placebo [1].

The causes of placebo response have been commented upon by many authors, such as Fava *et al.* [2]. Some of these causes relate to actual improvement in the condition of the patient (e.g., from regression to the mean for fluctuating illnesses, or improvement from expectancy that may lead to physiologically active processes that can be similar to those elicited by active treatment), while others have nothing to do with the patients but are results of trial design or implementation flaws (e.g., patients being initially evaluated as more ill than they actually are, so that if later evaluations are more accurate, patients appear to have improved even if they were receiving only placebo).

There are far-ranging public health consequences associated with the false-negative statistical results associated with high placebo response. At a minimum, failed trials delay the introduction of new therapies to patients by a number of years and raise the cost of development. Worse still, failed trials can cause companies to abandon development of effective treatments. In fact, some companies have entirely abandoned their development efforts in certain fields, such as depression, based on a judgment that the risk of an effective treatment failing to show efficacy is simply too high.

What designs have attempted to compensate for high placebo response rates?

Khan *et al.* [3] and Mallinckrodt *et al.* [4] have prepared analyses supportive of the intuitively logical conclusion that lower placebo response is associated with

Lee Baer¹ & Anastasia Ivanova^{*2}

¹Department of Psychiatry, Massachusetts General Hospital & Harvard Medical School, Boston, MA, USA

²Department of Biostatistics, The University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

*Author for correspondence:
E-mail: aivanova@bios.unc.edu

greater drug–placebo difference. This concept has led to development of many trial design features; for example, recruiting patients with more severe illnesses and who are less likely to spontaneously improve, or randomizing more patients to placebo and thereby mitigating expectation of improvement.

Another common practice in conditions with a moderate-to-high placebo response is the use of a placebo lead-in design (also referred to as ‘placebo run-in’ or ‘placebo wash-out’ design), in which all patients first receive placebo, and then only those patients who did not respond to placebo are randomized to the placebo versus drug arms for the efficacy analysis treatment phase (i.e., stage) [5]. This is done in an attempt to reduce the placebo response rate in the treatment phase, and thereby increase the drug versus placebo difference and reduce the risk of false-negative findings with a given sample size.

The placebo lead-in design is considered an enrichment design [6] because the sample randomized in the trial (i.e., only the nonresponders to placebo during the ‘lead-in’) is a subsample of the full sample recruited. This subsample is expected to have a greater treatment effect compared with placebo than the total sample. However, in 2011 Khin *et al.* [7], citing a 2001 article by Faries *et al.* on major depressive disorder (MDD) [8] and meta-analyses by Trivedi and Rush of 1994 [5] and Lee *et al.* of 2004 [9], noted: “It has been shown ... that using a placebo run-in has not been a successful strategy in lowering placebo response rate, nor has it increased drug–placebo differences in MDD trials.” This indicates that, on average, the use of the placebo lead-in approach added time and cost to trials (e.g., the number of subjects recruited who signed informed consents was larger than the ‘n’ used in the trial) but yielded little or no benefit.

These failures of the standard placebo lead-in design may have resulted from the short duration of the placebo lead-in period or from investigators not being blinded during this period. Some techniques have attempted to overcome these deficiencies (e.g., use of longer lead-in periods, or of double blinding during the lead-in). Regarding longer lead-ins, Chen *et al.* noted, with respect to a review of both depression and schizophrenia trials, that trials with a longer placebo lead-in phase have a smaller placebo response rate [10]. However, no data were provided indicating that a lower placebo response rate led to a larger drug versus placebo difference.

Mallinckrodt *et al.* [11] and Mallinckrodt *et al.* [12] examined the use of double-blind placebo lead-ins in 11 depression trials. They found that, while the effect size was slightly larger in the patients who did not respond to placebo during the lead-in (the ‘qualified subset’) as compared with the ‘all-patient’ group, the difference was

small (0.360 vs 0.422 using the Hamilton Depression Rating Scale [HDRS]), and the use of the ‘qualified subset’ provided very little advantage in terms of correctly identifying drug as superior to placebo [11].

Although it is counter-intuitive, use of a placebo lead-in, even if it does successfully reduce the placebo response rate, will not necessarily result in a larger drug versus placebo difference because placebo non-responders (the subsample resulting from a placebo lead-in) may be less responsive to both drug and placebo. In addition, even if the sponsor of a trial uses a long placebo lead-in or a double-blind placebo lead-in, and decides to assume that this will result in both some reduction of the placebo response rate and some increase in the drug versus placebo difference, the placebo lead-in may nonetheless increase the overall duration and cost of the trial because of the need to recruit and obtain consent from more subjects than the ‘n’ of the trial.

Merlo-Pich *et al.* have proposed an alternative approach to address the problem of high placebo response, called the band-pass methodology [13]. Using this method, all data for patients treated with drug and with placebo are excluded from the efficacy analysis for those sites that had an unusually high or low placebo response. This method can, on a *post hoc* basis, detect a significant drug–placebo difference in some trials that failed. However, because this method can cause many subjects not to be utilized, it could be expensive in practice because of the larger ‘gross’ sample size that may be required. Therefore, to our knowledge it has not been used prospectively in a trial.

Why was the SPCD developed & how is it conducted?

Since attempts to resolve the problem of placebo response using a placebo lead-in design and other approaches has not been successful, a design called the sequential parallel comparison design (SPCD) was developed at Massachusetts General Hospital (MGH; MA, USA) in 2003 (see ‘Financial and competing interests disclosure’ for information regarding patents owned by MGH and the availability of licenses for use of SPCD in clinical trials) [2,14]. Figures 1 & 2 illustrate two of the SPCD formats for a two-arm randomized clinical trial. In the SPCD, the study is conducted in two treatment stages that can be of equal or unequal duration. The efficacy analysis includes (a) all subjects in the stage 1 and (b) stage 2 data from only subjects who did not respond to placebo in the stage 1. Patients who receive active drug in stage 1, as well as placebo responders from stage 1, usually participate in stage 2 to maintain blinding and test additional hypotheses, but their data are not included in the efficacy analysis. When using SPCD, the relevant data from the two stages (mean responses and variances, treatment effects

and test statistics, among others, as appropriate, based upon the analytical method being utilized) are pooled to compute an overall p-value. This approach (i.e., utilizing data from some subjects in both stage 1 and 2) is akin to what occurs in a crossover trial, although a crossover trial is different because in such a trial, the populations in both stages are the same.

With SPCD, more data is generated from any given sample size (in comparison with a conventional single stage design or a placebo lead-in design) and, therefore, efficiency is generally increased. For treatments that have an effect, SPCD generally provides greater power for any specified sample size compared with a single stage design or, said another way, greater efficiency allows use of a smaller sample size without reducing power.

If placebo response is moderate to high, additional benefit from SPCD can be achieved if:

- The placebo response rate is lower in stage 2 among subjects who were placebo nonresponders in stage 1 (as was the case in all of the trials described in Table 1);
- This also leads to an improvement in the drug versus placebo difference (as occurred in three out of four trials described in Table 1; however, such an increase should not be expected to occur universally).

SPCD is generally suitable for trials using placebo control as monotherapy or adjunct therapy, whether placebo response is expected to be high, moderate, or low, except in the following circumstances: when a normal treatment period is long and therefore using two treatment periods is not feasible, or when, in monotherapy trials, it is not desirable for some patients to receive 'placebo only' during two treatment periods (see 'When should the SPCD not be used?').

SPCD would often be suitable for trials in chronic illnesses that are not yet curable, such as asthma, arthritis, hypertension, epilepsy, migraine headache and pain; that is, illnesses for which crossover designs are sometimes considered, but frequently investigators do not use crossover design because of a concern as to the risk of a 'carry over' effect. Using SPCD, there is no such concern because no patients switch from active treatment to placebo (a circumstance that leads to uncertainty as to whether later responsiveness is a delayed result of the active treatment), and no placebo responders switch to active treatment (a circumstance that leads to uncertainty as to whether continued response would have occurred even in the absence of active treatment). Using SPCD, only those subjects who did not respond to placebo are switched from placebo to active treatment, and therefore the assumption of no carry-over effect from placebo is reasonable.

SPCD is already being used in many psychiatry trials (e.g., depression, schizophrenia, bipolar disorder), where the benefits of utilizing data from subjects in two

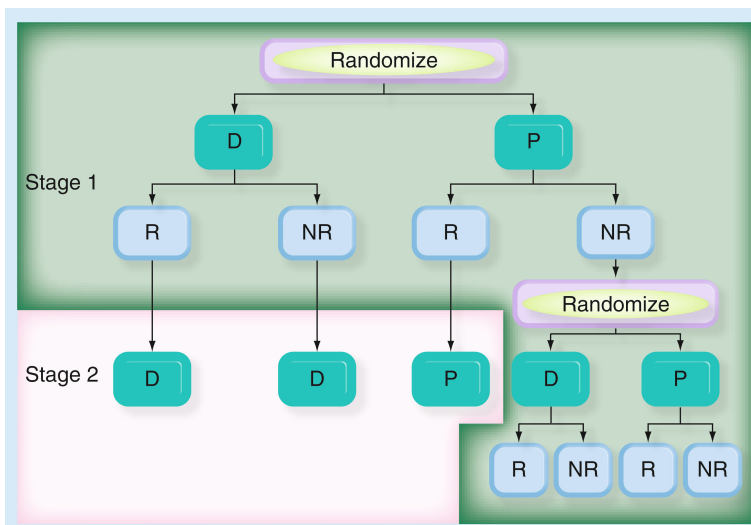


Figure 1. Sequential parallel comparison design re-randomization format. Efficacy analysis is based only upon results of the eight groups in blue boxes. D: Medical device or other type of therapy; P: Placebo; R: Responders; NR: Nonresponders.

treatment stages, and of enriching stage 2 with patients likely to have a reduced placebo response, are especially evident.

The purpose of this article is to review when use of SPCD is appropriate and when it is not. The next section provides a further explanation of the biostatistical approach used with SPCD and a brief mention of the various specific analytical methods that have been proposed by various authors. The authors next provide a review of the four SPCD trials for which results have been published, as well as information about one additional SPCD trial for which summary results have been

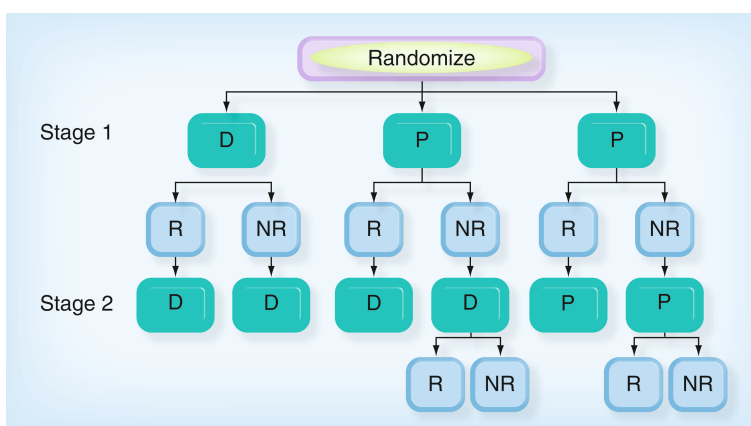


Figure 2. Sequential parallel comparison design single randomization format. Efficacy analysis is based only upon results of the ten groups in blue boxes. D: Medical device or other type of therapy; P: Placebo; R: Responders; NR: Nonresponders.

Table 1. Results of published clinical trials using sequential parallel comparison design.

Identifier (n)	Authors	Experimental conditions tested in SPCD	Response rate active drug (%)		Response rate placebo (%)		One-sided p-values	Ref.
			Stage 1	Stage 2	Stage 1	Stage 2		
NCT00555997 (120)	Papakostas <i>et al.</i>	Ziprasidone monotherapy (max 160 mg/day) versus placebo monotherapy	44.8 (13/29)	23.8 (5/21)	31.9 (29/91)	28.0 (7/25)	Stage 1 p = 0.15 Stage 2 p = 0.74 SPCD p = 0.21	[21,102]
NCT00321152 (148)	Papakostas <i>et al.</i> , study 1	Adjunctive L-methylfolate (7.5 mg/day) versus adjunctive placebo	19.4 (7/36)	17.1% (6/35)	28.5 (32/112)	9.1 (3/33)	Stage 1 p = 0.91 Stage 2 p = 0.27 SPCD p = 0.52	[22,103]
NCT00683852 (221)	Fava <i>et al.</i>	Adjunctive aripiprazole (2 mg/day) versus adjunctive placebo	18.5 (10/54)	18.0 (11/61)	17.4 (29/167)	7.9 (5/63)	Stage 1 p = 0.50 Stage 2 p = 0.08 SPCD p = 0.09	[23,104]
NCT00955955 (75)	Papakostas <i>et al.</i> , study 2	Adjunctive L-methylfolate (15 mg/day) versus adjunctive placebo	36.8 (7/19)	27.7 (5/18)	19.6 (11/56)	9.5 (2/21)	Stage 1 p = 0.12 Stage 2 p = 0.14 SPCD p = 0.02	[22,105]

SPCD: Sequential parallel comparison design.

reported, and information about eight other SPCD trials that are ongoing or for which results are not yet reported. The results of the published SPCD trials will then be discussed and, finally, the article will offer comments as to the circumstances in which the use of SPCD is suitable and beneficial, as well as when its use is not appropriate.

What are the statistical considerations in SPCD designs?

Since patient populations in stage 1 and 2 of an SPCD trial are different (the stage 2 population is a subset of the stage 1 population), one cannot assume that the treatment responses and therefore treatment differences in the two stages are the same. Therefore one cannot use analysis methods of crossover trials. The null hypothesis in the SPCD trial is that the treatment is not effective in 'all comers' (the stage 1 population of patients) and it is not effective in the population of placebo nonresponders (the stage 2 population). If the sample size is large enough, the null hypothesis is rejected if the treatment is effective in at least one of these populations. In other words, the treatment does not have to be effective in both populations to reject the null hypothesis.

Statistical analysis of SPCD data pools the inference from first and second stages to form a single test statistic. Methods for analyzing data from SPCD trials with both continuous and binary end points have been developed by many different researchers. These methods are efficient, address the correlation of observations from subjects included in both stages and preserve the type I error rate well.

Fava *et al.* proposed a linear combination test for binary outcome [2]. The test statistic is based on the linear combination of treatment effects from the two stages. The

weight (w between 0 and 1) to form the linear combination needs to be specified by the user in advance. The optimal w (i.e., the w that yields the highest power) depends on response rates; however the w of 0.6, where the first stage difference is weighted slightly higher than the second stage difference, works well in most cases. Huang and Tamura [15] and Ivanova, Qaqish and Schoenfeld [16] proposed the score test for binary outcomes. The Ivanova approach [16], like the linear combination test of Fava [2], requires the user to specify in advance the test parameter, r , which is an estimate of the ratio of the difference between active treatment and placebo in stage 2 in comparison with stage 1. The power is maximized when this test parameter is equal to the true ratio of the differences between treatment and placebo in the two stages of the trial. As it may be difficult to estimate the ratio of the treatment differences precisely before the trial, a recommended robust parameter to use in the test is $r = 1$, that is, assuming that treatment differences will be the same in both stages. Over a range of outcomes, the score test yields somewhat high power compared with the linear combination test in Fava *et al.* [2]. Simulations find that the score test preserves type I error well even for relatively small sample sizes [15,16].

Chen *et al.* proposed a linear combination test for continuous outcomes [10]. As in the linear combination test for binary outcomes, a weighting of $w = 0.6$ is recommended. Other methods of analyzing SPCD continuous outcome data have been proposed [17–19]. Simulations show that all methods preserve type I error rate for sample sizes that are usually used in SPCD. Tamura *et al.* described how to estimate the treatment effect in an SPCD trial [20].

The power of the test for the SPCD also depends on the allocation ratio between the new treatment and placebo in stage 1. If treatment effects were known, one

could compute the optimal allocation ratio in stage 1. Since actual response rates are not known beforehand, we recommend, if treatment differences are expected to be approximately equal in both stages, a stage 1 placebo:drug allocation ratio of 2:1. If the treatment difference in stage 2 is expected to be twice or more than the difference in stage 1, theoretically a very high percentage of patients should be assigned to placebo in stage 1 to maximize the sample size in stage 2, but in practice a placebo:drug ratio greater than 3:1 is not recommended. In the unlikely scenario where the treatment difference in stage 1 is expected to be much higher than the difference in stage 2, the best placebo:drug allocation ratio in stage 1 is 1:1.

If the placebo:drug allocation ratio and the test parameter (w or r) are chosen within a fairly wide range of values, the SPCD is more powerful than a conventional single stage design, even if the treatment difference in stage 2 is the same as in stage 1 or slightly lower. For any given number of enrolled patients, SPCD is generally more powerful than a placebo lead-in unless the lead-in period causes the treatment difference following the lead-in to be substantially larger (at least twofold) than the treatment difference would have been in the absence of the lead-in period – that is, unless the hoped for benefit of the lead-in period (i.e., an increased drug vs placebo difference) is achieved to the maximum degree.

How are power & sample size computed for the SPCD?

A sample size and power ‘calculator’ is available for free at www.rctlogic.com using a typical ‘click through’ license. The calculator uses the methodology in [16] for binary outcomes and the methodology in [10] for continuous outcomes, and shows, based on input assumptions specified by the user, the ‘n’ and power resulting from use of SPCD in comparison to a conventional single stage design.

What completed trials have used the SPCD design?

We performed an online search of the medical literature through pubmed.gov for all published studies that have used the SPCD (using search terms ‘sequential’, ‘parallel’, ‘comparison’ and ‘design’). The information extracted with respect to these studies included the population treated, the outcome measure, the criterion for success in stage 1 and stage 2, whether or not subjects were re-randomized after stage 1, and the placebo and active drug response rates in stage 1 and 2, along with the final combined one-sided p-value for each trial.

Our literature search found the four published studies using SPCD that have appeared in the psychiatric literature. These are summarized in Table 2. The four studies all involved patients with major depressive disorder, likely due to the high placebo response rate in this

condition. The sample sizes ranged from 75 to 224, and the number of sites ranged from six to 22. All of these studies were funded by corporations. As illustrated in Table 2, all studies were two-armed studies comparing drug with placebo, and used the single randomization format of SPCD as illustrated in Figure 2. In all four studies the allocation ratio for placebo:drug was 3:1. Table 2 includes additional design characteristics of each of these studies.

The first SPCD protocol developed, which involved ziprasidone monotherapy, used two 6-week stages. The other three studies used two 30-day stages. The ziprasidone study used a continuous outcome measure as the primary efficacy measure, while the other three used either a dichotomous improvement variable, continuous outcomes, or both dichotomous and continuous outcomes as the primary efficacy measure in the combined analysis of the two stages (Table 2).

In order to identify subjects who were placebo nonresponders and should be included in the efficacy analysis of stage 2, the ziprasidone study used a single criterion: <50% improvement in the 17-item HDRS (HDRS-17), a gold standard scale for depression research. The other three studies, with protocols developed subsequent to the ziprasidone study, used dual criteria: <50% improvement in a well-recognized scale, and an illness rating at the end of stage 1 (i.e., at the beginning of stage 2) of at least some prespecified numeric level. For illnesses such as depression that involve a ‘flooring effect’ (i.e., the ability to detect a signal is diminished when the level of illness severity is low), the refinement of using dual criteria can be beneficial (see below). While all four of the published studies examined patients with major depressive disorder, three also required that the patients had not responded to previous trials of antidepressant medications, and tested the addition of the study medication to the patient’s current antidepressant medication. The ziprasidone study, however, did not require failure of previous antidepressant drug trials, and tested ziprasidone as monotherapy for depression.

Table 1 illustrates the response rates for subjects in each stage of the four published studies, as well as p-values for the active treatment versus placebo for each stage and the combined SPCD p-value. p-values for each individual stage were computed using the Fisher’s exact test and the SPCD p-values were computed using the score test with test parameter $r = 1$. All p-values are one-sided and therefore are equal to about a half of p-values reported in corresponding publications of the trials. One-sided p-values are shown in Table 1 because a stage-specific two-sided p-value is not meaningful when the observed treatment effect in that stage is negative (e.g., stage 2 of a trial in [21], or stage 1 of study 1 in [22]).

Table 2. Design details of clinical trials with published results that used sequential parallel comparison design.

Authors (identifier)	Subject diagnosis	Study length	Experimental conditions tested in SPCD	Primary outcome measure	Type of outcome measure	Nonresponse criterion to enter efficacy analysis in stage 2	Ref.
Papakostas <i>et al.</i> (NCT00555997)	Major depressive disorder	12 weeks, divided into two 6-week stages	Ziprasidone monotherapy (max 160 mg/day) versus placebo monotherapy	Change in the total HDRS-17 score	Continuous	<50% improvement in HDRS-17 compared with baseline	[21,102]
Papakostas <i>et al.</i> , study 1 (NCT00321152)	SSRI-resistant major depressive disorder	60 days, divided into two 30-day stages	Adjunctive L-methylfolate 7.5 mg/day versus adjunctive placebo	HDRS-17 response defined as improvement of 50% or more and change in total HDRS-17	Dichotomous and continuous	<50% improvement in HDRS-17 compared with baseline and HDRS-17 >7 at end of stage 1	[22,103]
Fava <i>et al.</i> (NCT00683852)	Major depressive disorder with inadequate antidepressant therapy response	60 days, divided into two 30-day stages	Adjunctive aripiprazole 2 mg/day versus adjunctive placebo	MADRAS response defined as improvement of 50% or more	Dichotomous	<50% improvement in MADRAS compared with baseline and MADRAS >16 at end of stage 1	[23,104]
Papakostas <i>et al.</i> , study 2 (NCT00955955)	SSRI-resistant major depressive disorder	60 days, divided into two 30-day stages	Adjunctive L-methylfolate 15 mg/day versus adjunctive placebo	HDRS-17 response defined as improvement of 50% or more and change in total HDRS-17	Dichotomous and continuous	<50% improvement in HDRS-17 compared with baseline and HDRS-17 >7 at end of stage 1	[22,105]

All trials used single randomization format and 3:1 allocation ratio between placebo and drug in stage 1.

HDRS-17: 17-point Hamilton Depression Rating Scale; MADRAS: Montgomery Asberg Depression Rating Scale; SPCD: Sequential parallel comparison design;

SSRI: Selective serotonin reuptake inhibitor.

As expected, placebo response in stage 1 was moderate in the three adjunctive therapy trials (ranging from 17.4 to 28.5%), but was somewhat higher in the ziprasidone monotherapy trial (31.9%). The observed placebo response rate was either greater than, or approximately equal to, the observed rate of active drug in stage 1 in two of the studies ([23] and study 1 in [22]; stage 1 is essentially comparable to a conventional single stage design). Even though observed placebo response rates in stage 1 were moderate, in all four trials the placebo response rate in stage 2 was lower still (<10% in all of the trials other than the ziprasidone study, which was 28%). The authors believe it is likely that the reduction in the placebo response rate of stage 2 of the ziprasidone study would have been greater, and the drug versus placebo difference more favorable, if dual criteria for the identification of placebo nonresponders entering the efficacy analysis in stage 2 had been used, as was done in the three published subsequent trials.

In three of the trials the 'signal' (i.e., drug–placebo difference) improved in stage 2 and the p-value achieved is lower compared with the p-value that otherwise would have been achieved in a conventional single-stage trial with the same sample size, an equal placebo:drug allocation and observed response rates equal to those observed in the trials.

In two of the studies the drug, or dose of the drug, showed essentially no efficacy. A benefit of SPCD is that, with two populations studied, we are able to obtain the estimates of treatment effect in both overall and enriched population of placebo nonresponders to help in planning future studies with this drug, or to help decide that further investment of time and money is not warranted.

What other trials have used the SPCD design?

An additional nine SPCD trials were not included in our literature review because articles relating to them

have not yet been published. These trials are outlined in [Table 3](#), with corresponding clinicaltrials.gov 'NCT' identifier codes.

These nine trials address schizophrenia, bipolar disorder, depression and irritable bowel syndrome, and include diverse conditions, therapies, formats and

approaches: monotherapy and adjunct therapy, use of active comparator, use of multiple doses, use of binary and continuous outcomes, and use of drug therapies and a medical device. Of the nine trials set forth in [Table 3](#), three have NIH funding, three have nonprofit funding and three have corporate funding. In four

Table 3. Details of the nine additional Sequential Parallel Comparison Design trials.

Identifier	Name of trial	Sponsor	Funding	Principal investigators	Patients (n)	Sites (n)	Indication	Ref.
NCT01337609	Study of probiotic ganeden BC30 for IBS and major depressive disorder	Massachusetts General Hospital (MA, USA)	Ganeden Biotech (OH, USA)	Maurizio Fava	5 [†]	2	IBS and depression	[106]
NCT01318434	A paroxetine- and placebo-controlled study of 50 and 100 mg/day of EB-1010 among outpatients with major depressive disorder	Euthymics Bioscience Inc. (MA, USA)	Euthymics Bioscience Inc.	Pierre Tran	300	43	Depression	[107]
NCT01298427	Efficacy and tolerability of riluzole and biomarker of treatment response in treatment-resistant depression	NIMH (MD, USA)	NIMH	Carlos Zarate, Sanjay Matthew, Maurizio Fava, Gerard Sanacora	275	4	Depression	[108]
NCT01665950	Simvastatin augmentation of lithium treatment in bipolar depression	Massachusetts General Hospital	NIMH	Roy Perlis	30	1	Bipolar disorder	[109]
NCT01561742	Minocycline augmentation in schizophrenia	University of Texas Health Science Center (TX, USA)	Stanley Medical Research Institute (MD, USA)	Mujeeb Shad	120	1	Schizophrenia	[110]
NCT01724476	A trial of folate with B-12 in patients with schizophrenia with residual symptoms in Ethiopia	Massachusetts General Hospital	Stanley Medical Research Institute	David C Henderson	240	1	Schizophrenia	[111]
NCT01784666	Adjunctive isradipine for the treatment of bipolar depression	Massachusetts General Hospital	Stanley Medical Research Institute	Roy Perlis	30	1	Bipolar disorder	[112]
NCT01654796	Trial of low field magnetic stimulation augmentation of antidepressant therapy in treatment-resistant depression	NIMH	NIMH	Maurizio Fava	80	6	Treatment-resistant depression	[113]
NCT01500200	A study to evaluate ALKS 5461 in subjects with major depressive disorder	Alkermes, PLC (Dublin, Ireland)	Alkermes, PLC	Richard Leigh-Pemberton	130	27	Treatment-resistant depression	[114]

[†]The original planned sample size was 32.

IBS: Irritable bowel syndrome; NIMH: National Institute of Mental Health.

of the nine trials, MGH is the sponsor; in six of the nine trials, an MGH employee is either the primary investigator or one of the primary investigators.

Discussion of completed SPCD trials

The SPCD has been used in four published studies in depression to date, and is also being used in several ongoing studies that address a variety of illnesses. In the five reported studies (four published and one not yet published) the SPCD has appeared to accomplish its goal of increasing the efficiency of the design in three ways:

- By utilizing response data from some subjects twice (in stage 1 and 2);
- By enriching the data analysis with placebo nonresponders, hence decreasing the placebo response rate in stage 2 in all cases and improving the drug placebo difference in three cases;
- By permitting additional experiment-within-an-experiment analysis to be accomplished with subjects in groups other than placebo nonresponders, providing important information in planning future studies.

The trial testing L-methylfolate ([22], study 2) found a two-sided p-value below 0.05. For this trial the p-value was less than half of the p-value that would have resulted in a conventional single stage trial with the same sample size, equal patient allocation to placebo and drug, and the same observed response rates. An even larger reduction of p-value was provided by SPCD in the aripiprazole trial [21] and in another study testing L-methylfolate ([22], study 1). Although the results of the Alkermes Plc trial are not yet published, on May 31 2013 the company reported in a press release that this trial had a p-value of 0.026 [101].

Since the two stage approach of SPCD can provide some evidence of effectiveness in either stage, very useful information is obtained that can indicate that further development is warranted, even if a p-value below 0.05 is not achieved. This was in fact the case with the first study of L-methylfolate: a dose of 7.5 mg/day proved ineffective; however, in that study the drug nonresponders of stage 1 received a higher dose of 15 mg/day in stage 2, and several of the subjects then responded. Therefore a second trial was conducted with the higher dose (15 mg/day), which showed significant efficacy with an 'n' of only 75.

When should the SPCD be used?

The SPCD can be used to study chronic conditions that are not yet curable, such as asthma, arthritis, hypertension, epilepsy, migraine headaches, diabetes,

schizophrenia and cystic fibrosis, as well as many illnesses for which treatment may generate remission, in some cases lasting remission, such as for depression. This design can increase the efficiency of randomized clinical trials, including proof-of-concept/Phase II and registration trials, particularly in the presence of high placebo response, but also in cases of low placebo response. In Table 3, the Alkermes and Euthymics trials are examples of Phase II and III trials, respectively.

We believe that SPCD can also be especially advantageous in the following situations:

- Trials in pediatric and adolescent populations, since they are difficult to recruit for and there is often a desire to generate adequate power while minimizing the exposure of young people to new compounds;
- Trials for orphan diseases, where recruitment is often very difficult;
- 'Postmarketing commitment' trials (certain Phase IV trials), when the sponsor may want to fulfill its commitment to conduct an adequately powered trial for the smallest possible investment;
- Dose-response studies, since SPCD allows a sponsor to compare a new treatment at several alternative doses with placebo in the population of patients in stage 1, and to compare some of the doses in placebo nonresponders in stage 2.

To date, there have not been examples of SPCD trials illustrating the first three cases above.

Although some researchers may have initially questioned the generalizability of the results of SPCD trials, this trial design can actually have some benefits in comparison with the results of a typical parallel trial. First, typical trials often use placebo lead-ins. This is a widely accepted design, even in Phase III trials. Arguably, results of a placebo lead-in trial are less generalizable than those of an SPCD trial, because SPCD trials do give weight to the results of all-comers. Second, in clinical practice, patients that would be characterized as 'placebo responders' often respond spontaneously, before visiting a doctor, and therefore are not treated. The more resistant patients (placebo nonresponders) seek treatment. Therefore, a design that focuses on placebo nonresponders can in fact address the population that new treatments seek to help.

Natural history of disease can contribute to placebo response, and can be used to determine the relative contribution of placebo and drug to individual outcomes. An SPCD trial is typically longer than placebo lead-in or a single stage trial, with some patients

staying on placebo or drug therapy for the duration of the SPCD trial, thereby allowing the sponsor and the FDA to obtain valuable data on response over time.

A key feature of assessing safety is understanding the impact of a treatment over an extended time period. With SPCD, each patient is treated longer than in a standard trial design, with, as noted above, some patients being treated with active treatment for the entire trial while others are treated with placebo. For this reason, the SPCD design has the potential to provide a more realistic assessment of safety, since drugs are typically prescribed in clinical practice for a longer period than the length of a normal trial.

When should the SPCD not be used?

There are circumstances when use of SPCD would not be appropriate; for example:

- SPCD should not be used for condition that can result in considerable deterioration or for which a complete cure is possible; for example, infectious diseases or acute illnesses.
- While the two stages can be of different lengths, each stage in an SPCD trial is a true treatment stage, and not a 'lead-in'. Therefore, each stage should be long enough to allow a treatment response. If the aggregate length of the two stages of an SPCD trial would be unacceptably long, then SPCD is not suitable. This would typically involve a circumstance in which a conventional single-stage trial would have a long treatment duration for each patient (e.g., several months or a year), and adding a second stage would either, first, be very expensive; for example, if there would be a need to very significantly increase the number of visits and each visit is costly; or second, lead to an unacceptably high number of dropouts among the subjects who are included in the efficacy analysis of stage 2 (i.e., among the placebo nonresponders of stage 1). As to the first example, a trial sponsor considering use of SPCD should weigh the additional cost of each patient remaining in the trial longer than 'usual' (i.e., for two treatment periods) versus the cost saving from being able to reduce sample size without reducing power, or the benefit of risk reduction by gaining greater power for any specified sample size. As to the second example, of course, conventional single stage trials experience dropouts. The key question is, will there be many additional dropouts during stage 2 of an SPCD trial. It may be the case that patients who complete stage 1 are 'persistent', and therefore will generally remain in stage 2. In fact, in the SPCD trials reported to date, only approximately 10% of the placebo nonresponders

from stage 1 have dropped out during stage 2. Therefore, to date, the expected power of SPCD trials has not been adversely affected by high dropout rates in stage 2.

- Use of SPCD would not be suitable if it would be inappropriate for any subject to be treated with placebo for the full length of the trial (i.e., during both stages). This concern, however, is typically obviated if the therapy under investigation is adjunct therapy.

Future perspective

SPCD was proposed by Fava *et al.* more than 10 years ago [2]. Since then, many researchers from academia, industry and the FDA have reviewed SPCD in detail and published their findings extensively.

Starting several years ago, adoption of SPCD began to gain momentum, and now results of five trials have been reported. We are unaware of any material problems that have been associated with these trials. With respect to the four published SPCD trials, SPCD has reduced the observed placebo response in all cases, and in all but one case has increased observed treatment–placebo difference in stage 2 compared with stage 1.

We believe that, as more SPCD trials are reported, adoption of SPCD will continue to accelerate. Potentially, SPCD will become the standard approach used in psychiatric trials, with substantial adoption also in trials related to a wide range of other illnesses.

We believe that such widespread adoption would provide significant public health benefits by accelerating the approval of new therapies, reducing the cost of developing such therapies, and helping to avoid wasted expenditures and patient exposure with respect to therapies that are in fact ineffective.

Financial & competing interests disclosure

L Baer has consulted for Lilly and Pfizer Pharmaceuticals. A Ivanova is a consultant for, among other companies, RCT Logic, LLC, which is the exclusive licensee of patents owned by Massachusetts General Hospital (MGH) relating to the use of SPCD in clinical trials. The authors thank M Bowman, Chief Executive Officer, RCT Logic, for assistance in the production of this manuscript. RCT Logic offers licenses under MGH's issued patents to those who wish to conduct a clinical trial using any SPCD format. Such licenses involve payment of a one-time license fee but no ongoing royalty payments. Significant license fee reductions are available for academic institutions. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Executive summary

Background

- Randomized clinical trials in many medical conditions are subject to moderate-to-high rates of placebo response, reducing the efficiency of the studies and increasing the chances of false-negative statistical results.

What designs have attempted to compensate for high placebo response rates?

- A common practice in conditions with a moderate-to-high placebo response is to use a placebo lead-in design, but meta-analyses of placebo lead-in studies in depression indicate that there is no increase in the drug versus placebo difference from typical single-blind placebo lead-ins, and little increase from longer or double-blind placebo lead-ins.

Why was the sequential parallel comparison design developed & how is it conducted?

- The sequential parallel comparison design (SPCD) was developed because attempts to resolve the problem of placebo response using a placebo lead-in design have not been successful.
- In the SPCD, the study is conducted in two treatment stages, with the efficacy analysis including (a) all subjects in stage 1 and (b) stage 2 data from only subjects who did not respond to placebo in stage 1, and relevant data from the two stages are pooled to compute an overall p-value.
- Since data from some subjects are utilized from both stage 1 and 2, use of SPCD with a given sample size can generally provide, for treatments that have at least some efficacy, a reduction in p-value compared with a conventional single stage parallel design, even when placebo response rate is low.

What completed trials have used the SPCD design?

- This article provides information on 13 SPCD trials. The four studies with published results all involve patients with major depressive disorder, likely due to the high placebo response rate in this condition. The nine other trials address schizophrenia, bipolar disorder and depression, and test diverse conditions, therapies, formats and approaches.

When should the SPCD be used?

- SPCD is appropriate for pivotal and earlier stage trials, particularly when high placebo response rate is expected.
- SPCD is appropriate for trials in pediatric and adolescent population, since these trials are difficult to recruit for and there is often a desire to generate adequate power while minimizing the exposure of young people to new compounds.
- SPCD is appropriate for trials for orphan diseases, where recruitment is often very difficult.
- SPCD is appropriate for 'postmarketing commitment' trials (certain Phase IV trials), when the sponsor may want to fulfill its commitment to conduct an adequately powered trial for the smallest possible investment.
- SPCD is appropriate for dose-response studies, since SPCD allows a sponsor to compare a new treatment at several alternative doses to placebo in the population of patients in stage 1, and to compare a selected optimal dose in placebo nonresponders in stage 2.

When should the SPCD not be used?

- The SPCD is not appropriate for acute illnesses.
- The SPCD is not appropriate when a normal treatment period is long and therefore using two treatment periods is not feasible or would be very costly.
- The SPCD is not appropriate when the percentage of dropouts after stage 1 among placebo nonresponders would be unacceptably high.
- The SPCD is not appropriate when we do not want some patients to receive 'placebo only' during two treatment periods.

References

Papers of special note have been highlighted as:

- of interest
 - of considerable interest
- Laughren TP. The scientific and ethical basis for placebo-controlled trials in depression and schizophrenia: an FDA perspective. *Eur. Psychiatry* 16, 418–423 (2001).
 - Fava M, Evins AE, Dorer DJ, Schoenfeld DA. The problem of placebo response in clinical trials for psychiatric disorders: culprits, possible remedies, and a novel study design approach. *Psychother. Psychosom.* 72, 115–127 (2003).
 - First report of the rationale and development of the sequential parallel comparison design (SPCD) as an effective method to deal with placebo response.
 - Khan A, Detke M, Khan SRF, Mallinckrodt C. Placebo response and antidepressant clinical trial outcome. *J. Nerv. Ment. Dis.* 191(4), 211–218 (2003).
 - Mallinckrodt CH, Zhang L, Prucka WR, Millen BA. Signal detection and placebo response in schizophrenia: parallels with depression. *Psychopharmacol. Bull.* 43(1), 53–72 (2010).
 - Trivedi MH, Rush AJ. Does a placebo run-in or placebo treatment cells affect the efficacy of antidepressant medications? *Neuropsychopharmacology* 11, 33–43 (1994).
 - Fedorov VV, Liu T. Enrichment design. In: *Wiley Encyclopedia of Clinical Trials*
 - D'Agostino RB, Sullivan LM, Massaro JM (Eds). John Wiley & Sons, Inc., Hoboken, NJ, USA, 1–8 (2007).
 - Khin NA, Chen YF, Yang Y, Yang P, Laughren TP. Exploratory analyses of efficacy data from major depressive disorder trials submitted to the US Food and Drug Administration in support of new drug applications. *J. Clin. Psychiatry* 72(4), 464–472 (2011).
 - Faries DE, Heiligenstein JH, Tollefson GD, Potter WZ. The double-blind variable placebo lead-in period: results from two antidepressant clinical trials. *J. Clin. Psychopharmacol.* 21(6), 561–568 (2001).

- 9 Lee S, Walker JR, Jakul L, Sexton K. Does elimination of placebo responders in a placebo run-in increase the treatment effect in randomized clinical trials? A meta-analytic evaluation. *Depress. Anxiety* 19(1), 10–19 (2004).
- 10 Chen YF, Yang Y, Hung H, Wang S. Evaluation of performance of some enrichment designs dealing with high placebo response in psychiatric clinical trials. *Contemp. Clin. Trials* 32, 592–604 (2011).
- **Test statistic to combine results from two stages of the SPCD with continuous outcome.**
- 11 Mallinckrodt CH, Meyers AL, Prakash A, Faries DE, Detke MJ. Simple options for improving signal detection in antidepressant clinical trials. *Psychopharmacol. Bull.* 40(2), 101–114 (2007).
- 12 Mallinckrodt CH, Tamura RN, Tanaka Y. Recent developments in improving signal detection and reducing placebo response in psychiatric clinical trials. *J. Psychiatric Res.* 45, 1202–1207 (2011).
- 13 Merlo-Pich E, Alexander RC, Fava M, Gomeni R. A new population-enrichment strategy to improve efficiency of placebo-controlled clinical trials of antidepressant drugs. *Clin. Psychopharmacol. Ther.* 88(5), 634–642 (2010).
- 14 Fava M, Evins AE, Dorer DJ, Schoenfeld DA. Erratum. *Psychother. Psychosom.* 73, 123 (2004).
- 15 Huang X, Tamura RN. Comparison of test statistics for the sequential parallel design. *Stat. Biopharm. Res.* 2, 42–50 (2010).
- **Review of test statistics used to combine results from two stages of the SPCD with binary outcome.**
- 16 Ivanova A, Qaqish B, Schoenfeld DA. Optimality, sample size and power calculations for the sequential parallel comparison design. *Stat. Med.* 30, 2793–2803 (2011).
- **Discussion of sample size and power calculation considerations for the SPCD with binary outcome.**
- 17 Tamura RN, Huang X. An examination of the efficiency of the sequential parallel design in psychiatric clinical trials. *Clin. Trials* 4, 309–317 (2007).
- **Test statistic to combine results from two stages of the SPCD with continuous outcome.**
- 18 Liu Q, Lim P, Singh J, Lewin D, Schwab B, Kent J. Doubly randomized delayed-start design for enrichment studies with responders or nonresponders. *J. Biopharm. Stat.* 22(4), 737–757 (2012).
- **Test statistic to combine results from two stages of the SPCD with continuous outcome.**
- 19 Doros G, Pencina M, Rybin D, Meisner A, Fava M. A repeated measures model for analysis of continuous outcomes in sequential parallel comparison design studies. *Stat. Med.* 32(16), 2767–2789.
- **Test statistic to combine results from two stages of the SPCD with continuous outcome.**
- 20 Tamura RN, Huang X, Boos DD. Estimation of treatment effect for the sequential parallel design. *Stat. Med.* 30, 3496–3506 (2011).
- 21 Papakostas GI, Vitolo OV, Ishak WW *et al.* A 12-week, randomized, double-blind, placebo controlled, sequential parallel comparison trial of ziprasidone as monotherapy for major depressive disorder. *J. Clin. Psychiatry* 73, 1541–1547 (2012).
- **The first protocol developed using SPCD, and concluded that future studies should employ a severity threshold criterion for entry into stage 2 of the design.**
- 22 Papakostas GI, Shelton RC, Zajecka JM *et al.* L-methylfolate as adjunctive therapy for SSRI-resistant major depression: results of two randomized, double-blind, parallel-sequential trials. *Am. J. Psychiatry* 169, 1267–1274 (2012).
- **Reports two of the four published SPCD studies described in Tables 1 & 2. Demonstrates how results of first study led to successful second study.**
- 23 Fava M, Mischoulon D, Iosifescu D *et al.* A double-blind, placebo-controlled study of Aripiprazole adjunctive to antidepressant therapy among depressed outpatients with inadequate response to prior antidepressant therapy (ADAPT-A Study). *Psychother. Psychosom.* 81, 87–97 (2012).
- **One of four published studies of SPCD discussed in Tables 1 & 2. Contains detailed description and graphical illustration of design.**
- **Websites**
- 101 Alkermes Press Release. Alkermes presents positive results from Phase 2 clinical study of ALKS 5461 in major depressive disorder at 53rd annual NCDEU meeting. <http://investor.alkermes.com/phoenix.zhtml?c=92211&p=RssLanding&cat=news&id=1825817>
- 102 ClinicalTrials Database: NCT00555997. www.clinicaltrials.gov/show/NCT00555997
- 103 ClinicalTrials Database: NCT00321152. www.clinicaltrials.gov/show/NCT00321152
- 104 ClinicalTrials Database: NCT00683852. www.clinicaltrials.gov/show/NCT00683852
- 105 ClinicalTrials Database: NCT00955955. www.clinicaltrials.gov/show/NCT00955955
- 106 ClinicalTrials Database: NCT01337609. www.clinicaltrials.gov/show/NCT01337609
- 107 ClinicalTrials Database: NCT01318434. www.clinicaltrials.gov/show/NCT01318434
- 108 ClinicalTrials Database: NCT01298427. www.clinicaltrials.gov/show/NCT01298427
- 109 ClinicalTrials Database: NCT01665950. www.clinicaltrials.gov/show/NCT01665950
- 110 ClinicalTrials Database: NCT01561742. www.clinicaltrials.gov/show/NCT01561742
- 111 ClinicalTrials Database: NCT01724476. www.clinicaltrials.gov/show/NCT01724476
- 112 ClinicalTrials Database: NCT01784666. www.clinicaltrials.gov/show/NCT01784666
- 113 ClinicalTrials Database: NCT01654796. www.clinicaltrials.gov/show/NCT01654796
- 114 ClinicalTrials Database: NCT01500200. www.clinicaltrials.gov/show/NCT01500200