



Published in final edited form as:

*Stat Med.* 2011 December 30; 30(30): . doi:10.1002/sim.4412.

## Estimation of Treatment Effect for the Sequential Parallel Design

Roy N. Tamura<sup>1</sup>, Xiaohong Huang<sup>2</sup>, and Dennis D. Boos<sup>3</sup>

<sup>1</sup>Lilly Research Laboratories, Eli Lilly and Company, Lilly Corporate Center, Indianapolis, IN 46285

<sup>2</sup>Sanofi-Aventis, 200 Crossing Blvd, Bridgewater, NJ 08807

<sup>3</sup>Department of Statistics, North Carolina State University, Raleigh, NC 27695

### Summary

The sequential parallel clinical trial is a novel clinical trial design being used in psychiatric diseases which are known to have potentially high placebo response rates. The design consists of an initial parallel trial of placebo versus drug augmented by a second parallel trial of placebo versus drug in the placebo non-responders from the initial trial. Statistical research in the design has focused on hypothesis tests. However, an equally important output from any clinical trial is the estimate of treatment effect and variability around that estimate. In the sequential parallel trial, the most important treatment effect is the effect in the overall population. This effect can be estimated by considering only the first phase of the trial but this ignores useful information from the second phase of the trial. We develop estimates of treatment effect which incorporate data from both phases of the trial. Our simulations and a real data example suggest that there can be substantial gains in precision by incorporating data from both phases. The potential gains appear to be greatest in moderate sized trials which would typically be the case in Phase II trials.

### 1. Introduction

In psychiatry, randomized, double-blind, placebo controlled clinical trials are necessary to determine the efficacy of a new treatment. However, even for drugs which are known to be effective, such trials have a high failure rate. Khan et al. [1] reviewed the data from nine antidepressants approved by the United States Food and Drug Administration between 1985 and 2000. For these antidepressants, there were 51 randomized, double-blind, placebo-controlled trials and 92 treatment arms with an eventually approved dose; however of these 92 arms, only 45 showed statistically significant separation compared to placebo. Further analysis of the data suggests that high placebo response is the major contributor to the problem of high type II error.

In order to address the placebo response, Fava et al. [2] proposed a novel design aimed at increasing the efficiency of placebo controlled psychiatric clinical trials. The basic idea is that in addition to an initial phase of a standard parallel design, there is a second phase in which patients initially randomized to placebo and who have not responded are randomized to drug or placebo. The inference on treatment is based on a combined analysis of both the initial and second phases of the study. Fava et al. [2,3] called this design the sequential parallel design.

The sequential parallel design is an example of an enrichment design [4] in which the enrichment is the population of placebo non-responders. In contrast to other enrichment designs, however, the primary population of interest in the sequential parallel design is the overall population of depressed patients. This is because in clinical practice, it would be unethical to initiate patients on placebo and wait to observe their response. The null hypothesis of the sequential parallel design is that there is no treatment effect in either the

overall population or the subpopulation of placebo non-responders. If it can be assumed that placebo responders are also drug responders, then a significant effect in the subpopulation implies a significant effect in the overall population. Fava et al. [2] proposed a weighted sum of the observed treatment differences over the two phases to test the null hypothesis of no treatment effect in either the overall or the subpopulation of placebo non-responders. Huang and Tamura [5] proposed a score statistic under the assumption of equal treatment effect in the two phases of the study. Ivanova et al. [6] extended the score statistic under the more general assumption that the ratio of treatment effect in the two phases is known. Whichever statistic is chosen, the efficiency of the design can be substantially increased over the traditional parallel design.

In addition to testing of a statistical hypothesis, a second goal of a clinical trial is to yield an estimate of the treatment effect and an associated confidence interval. Although the treatment effect in the population of placebo non-responders may be interesting, the more important treatment effect is the effect in the overall population. How should one estimate the treatment effect in the overall population from a sequential parallel design? One obvious possibility to estimate the treatment effect in the overall population by only considering the data from the first phase of the trial. The usual estimated treatment effect from the first phase will be unbiased and the confidence intervals will have correct coverage if the study is large enough. This approach, however, fails to use any data from the second phase of the study and hence may be inefficient. Standard confidence intervals using only the first phase of the design may result in inconsistencies with the statistical test since the test uses information from both phases. The goal of this manuscript is to explore various options for estimating the treatment effect in the overall population while incorporating information from both phases of the study. We consider binary endpoints and measure the treatment effect as the difference in response rates. In Section 2, we describe the various methods under consideration and in Section 3, we compare the finite sample properties of these estimates and their approximate 95% confidence intervals by simulation. In Section 4, we apply the methods to a recently presented sequential parallel trial of L-methylfolate in the augmentation treatment of depression. We summarize our findings and recommendations in Section 5.

## 2. Description of Design and Estimators

### 2.1 Sequential Parallel Design

Assume that the clinical trial consists of a single drug and a placebo. In the sequential parallel design, patients are typically randomized into three groups. The basic idea is that there are two phases of treatment with the duration of each phase sufficiently long for the drug to elicit a response. In trials of major depressive disorder, a four to six week duration for each phase would be reasonable. The three treatment groups are characterized by the choice of placebo or drug in each phase. The first group receives placebo in both phases of the study, the second group receives placebo in the first phase and drug in the second phase, and the third group receive drug in both phases. Patients are randomized to the three groups

according to an  $a : a : (1-2a)$  ratio where  $0 < a < \frac{1}{2}$ . Typically,  $a$  is chosen to initially allocate more patients to placebo yet be easy to implement. For example,  $a=1/3$  leads to 1:1:1, and  $a=3/8$  leads to 3:3:2. All patients in all three groups typically continue through both phases and the blind is maintained throughout the study. When assessing the treatment effect, information is used from all randomized patients in the first phase but only from the first phase placebo non-responders in the second phase. Define  $p_1 = P(\text{drug response}_{\text{first phase}})$ , and  $q_1 = P(\text{placebo response}_{\text{first phase}})$  as the probabilities of being a drug responder and a placebo responder, in the first phase; define  $p_2 = P(\text{drug response}_{\text{second phase}} | \text{placebo non-responder}_{\text{first phase}})$  and  $q_2 = P(\text{placebo response}_{\text{second phase}} | \text{placebo non-}$

*responder<sub>first phase</sub>*) as the conditional probabilities of being a drug responder and a placebo responder in the second phase given that the patients are placebo non-responders in the first phase; the sequential parallel design is shown in Table 1.

The total sample size of the trial is  $n$ . In the placebo-placebo treated group,  $n_{11}$  is the observed number of non-responders in the first phase and responders in the second phase,  $n_{12}$  is the observed number of non-responders in both phases, and  $n_{13}$  is the observed number of responders in the first phase. In the placebo-drug treated group,  $n_{21}$  is the observed number of non-responders in first phase and responders in the second phase,  $n_{22}$  is the observed number of non-responders in both phases, and  $n_{23}$  is the observed number of responders in the first phase. In the drug-drug treated group,  $n_{31}$  and  $n_{32}$  are the observed number of responders and non-responders, respectively, in the first phase.

### 2.2 Maximum Likelihood Estimator

Consider the case where the treatment effect is potentially different in the two phases of the study. The joint likelihood for  $n_{11}, n_{12}, n_{13}, n_{21}, n_{22}, n_{23}, n_{31}$  and  $n_{32}$  is

$$L(p_1, q_1, p_2, q_2) \propto p_1^{n_{31}}(1 - p_1)^{n(1-2a)-n_{31}} q_1^{n_{13}+n_{23}}(1 - q_1)^{2na-(n_{13}+n_{23})} q_2^{n_{11}}(1 - q_2)^{n_{12}} p_2^{n_{21}}(1 - p_2)^{n_{22}}$$

The MLE can be obtained by setting the first derivatives of  $\log L(p_1, q_1, p_2, q_2)$  to 0 and solving, which results in:

$$\hat{p}_1 = \frac{n_{31}}{n(1-2a)}, \quad \hat{q}_1 = \frac{n_{13}+n_{23}}{2na}, \quad \hat{p}_2 = \frac{n_{21}}{n_{21}+n_{22}}, \quad \hat{q}_2 = \frac{n_{11}}{n_{11}+n_{12}}$$

Define

$$\theta = \begin{pmatrix} p_1 \\ q_1 \\ p_2 \\ q_2 \end{pmatrix}$$

then the information matrix of  $\theta$  is  $I(\theta) = \frac{1}{n} E \left\{ -\frac{\partial^2}{\partial \theta \partial \theta^T} \log L(p_1, q_1, p_2, q_2) \right\}$ ;

$$I(\theta) = \begin{pmatrix} \frac{1-2a}{p_1(1-p_1)} & 0 & 0 & 0 \\ 0 & \frac{2a}{q_1(1-q_1)} & 0 & 0 \\ 0 & 0 & \frac{a(1-q_1)}{p_2(1-p_2)} & 0 \\ 0 & 0 & 0 & \frac{a(1-q_1)}{q_2(1-q_2)} \end{pmatrix}$$

Thus, we expect from the general theory that the MLEs  $\hat{\theta}_{MLE} = (\hat{p}_1 \hat{q}_1 \hat{p}_2 \hat{q}_2)^T$  are asymptotically normally distributed with mean  $\theta$  and covariance matrix

$$\sum_{\hat{\theta}_{MLE}} = \frac{1}{n} I(\theta)^{-1} = \frac{1}{n} \begin{pmatrix} \frac{p_1(1-p_1)}{1-2a} & 0 & 0 & 0 \\ 0 & \frac{q_1(1-q_1)}{2a} & 0 & 0 \\ 0 & 0 & \frac{p_2(1-p_2)}{a(1-q_1)} & 0 \\ 0 & 0 & 0 & \frac{q_2(1-q_2)}{a(1-q_1)} \end{pmatrix}$$

If we define  $\Delta = p_1 - q_1$ , then  $\hat{\Delta}_{MLE} = p_1\hat{ - } - q_1\hat{ }$ . This estimator utilizes only data from the first phase and we will refer to it as the Phase 1 estimator.

### 2.3 Constrained Maximum-Likelihood Estimator

In order to utilize data from the second phase, one needs to make assumptions about the relationship between  $p_1, q_1, p_2,$  and  $q_2$ . One possibility is to make the assumption of equal treatment effect in the two stages, as was proposed by Huang and Tamura [5] for inferential testing. Define  $p_1 - q_1 = p_2 - q_2 = \Delta$ . Then the joint likelihood function for  $n_{11}, n_{12}, n_{13}, n_{21}, n_{22}, n_{23}, n_{31}$  and  $n_{32}$  can be written as:

$$L(\Delta, q_1, q_2) \propto (q_1 + \Delta)^{n_{31}} (1 - (q_1 + \Delta))^{n(1-2a) - n_{31}} q_1^{n_{13} + n_{23}} (1 - q_1)^{2na - (n_{13} + n_{23})} q_2^{n_{11}} (1 - q_2)^{n_{12}} (q_2 + \Delta)^{n_{21}} (1 - (q_2 + \Delta))^{n_{22}}$$

Define

$$\theta_C = \begin{pmatrix} \Delta \\ q_1 \\ q_2 \end{pmatrix}$$

The first derivatives of  $\log L(\Delta, q_1, q_2)$  are:

$$\frac{\partial \log L(\Delta, q_1, q_2)}{\partial \theta_C} = \begin{pmatrix} \frac{n_{31}}{q_1 + \Delta} - \frac{n(1-2a) - n_{31}}{1 - (q_1 + \Delta)} + \frac{n_{21}}{q_2 + \Delta} - \frac{n_{22}}{1 - (q_2 + \Delta)} \\ \frac{n_{31}}{q_1 + \Delta} - \frac{n(1-2a) - n_{31}}{1 - (q_1 + \Delta)} + \frac{n_{13} + n_{23}}{q_1} - \frac{2na - (n_{13} + n_{23})}{1 - q_1} \\ \frac{n_{11}}{q_2} - \frac{n_{12}}{1 - q_2} + \frac{n_{21}}{q_2 + \Delta} - \frac{n_{22}}{1 - (q_2 + \Delta)} \end{pmatrix}$$

There is no closed form solution for the MLE of  $\theta_C$ . The MLE for these parameters have to be obtained via numerical methods. The information matrix of  $\theta$  is

$$I(\theta_C) = \frac{1}{n} E \left\{ - \frac{\partial^2}{\partial \theta_C \partial \theta_C^T} \log L(\Delta, q_1, q_2) \right\} :$$

$$\begin{pmatrix} \frac{1-2a}{(q_1 + \Delta)(1 - (q_1 + \Delta))} + \frac{a(1-q_1)}{(q_2 + \Delta)(1 - (q_2 + \Delta))} & \frac{1-2a}{(q_1 + \Delta)(1 - (q_1 + \Delta))} & \frac{a(1-q_1)}{(q_2 + \Delta)(1 - (q_2 + \Delta))} \\ \frac{1-2a}{(q_1 + \Delta)(1 - (q_1 + \Delta))} & \frac{1-2a}{(q_1 + \Delta)(1 - (q_1 + \Delta))} + \frac{2a}{q_1(1-q_1)} & 0 \\ \frac{a(1-q_1)}{(q_2 + \Delta)(1 - (q_2 + \Delta))} & 0 & \frac{a(1-q_1)}{(q_2 + \Delta)(1 - (q_2 + \Delta))} + \frac{a(1-q_1)}{q_2(1-q_2)} \end{pmatrix}$$

The constrained MLE (constraint of equal treatment effect)  $\theta_{CMLE} = (\Delta, \hat{q}_1, \hat{q}_2)^T$  are approximately normally distributed with mean  $\theta_C$  and asymptotic covariance matrix

$$\sum_{\hat{\theta}_{CMLE}} = \frac{1}{n} I(\theta_C)^{-1}.$$

### 2.4 Alternative Treatment Estimator

The assumption of equal treatment effect in the two phases may be clinically debatable because the two populations are different. The overall population however can be envisioned to consist of two subpopulations, the subpopulation of placebo non-responders and the subpopulation of placebo responders. Thus the overall treatment effect  $\Delta$  can be written as:

$$\Delta = q_1 \Delta_{RES} + (1 - q_1) \Delta_{NRES}$$

where  $\Delta_{RES}$  and  $\Delta_{NRES}$  represent the treatment effect in responders and non-responders, respectively. Suppose we assume that within the population of placebo responders, all patients also respond to drug. Then  $\Delta_{RES} = 0$ . This assumption is referred to as the monotonicity assumption in the causal effects literature and needs to be examined on a case by case basis when considering a sequential parallel design. Without this assumption, the inference from the design becomes problematic because a significant test statistic could arise solely from the population of placebo non-responders.

The quantity  $p_2 - q_2$  represents the treatment effect in placebo non-responders after being treated by placebo. If one assumes that the initial placebo treatment does not affect the treatment effect in placebo non-responders, then  $\Delta_{NRES} = p_2 - q_2$ . We call this assumption the constancy of treatment effects. Under monotonicity and constancy assumptions, an alternative estimator of the overall treatment effect is:

$$\hat{\Delta}_2 = (1 - \hat{q}_1)(\hat{p}_2 - \hat{q}_2)$$

### 2.5 Linear Combination Estimator

The assumptions of monotonicity and constancy of treatment effect in placebo non-responders would be difficult to verify in practice. Thus, rather than relying on  $\hat{\Delta}_2$  alone, we consider linear combinations of the two estimators  $\hat{\Delta}_{MLE}$  and  $\hat{\Delta}_2$ . Serfling [7, p.127] shows that the optimal (minimal variance) linear combination of two estimators that are jointly asymptotically normal is a function of the elements of the asymptotic covariance matrix.

If the covariance matrix of  $(\hat{\Delta}_{MLE}, \hat{\Delta}_2)$  is denoted:

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

then the linear combination  $\hat{\Delta}_O = w \hat{\Delta}_{MLE} + (1 - w) \hat{\Delta}_2$  with minimum variance occurs when

$$w = \frac{\sum_{22} - \sum_{12}}{\sum_{11} - 2\sum_{12} + \sum_{22}}$$

At this value of  $w$ , the asymptotic variance of the linear combination times  $n$  is

$$(\sum_{11}\sum_{22} - \sum_{12}^2) / (\sum_{11} - 2\sum_{12} + \sum_{22}) \quad (1)$$

For the two estimators  $\hat{\Delta}_{MLE}$  and  $\hat{\Delta}_2$ , the delta method yields the asymptotic covariance matrix

$$\sum \frac{1}{n} \begin{pmatrix} \frac{p_1(1-p_1)}{1-2a} + \frac{q_1(1-q_1)}{2a} & \frac{(p_2-q_2)q_1(1-q_1)}{2a} \\ \frac{(p_2-q_2)q_1(1-q_1)}{2a} & \frac{(p_2-q_2)^2q_1(1-q_1)}{2a} + \frac{(1-q_1)p_2(1-p_2)}{a} + \frac{(1-q_1)q_2(1-q_2)}{a} \end{pmatrix} \quad (2)$$

Thus  $\hat{\Delta}_O$  is asymptotically normal with mean  $w(p_1 - q_1) + (1 - w)(p_2 - q_2)$  and variance given by (1) with the values of  $\Sigma_{ij}$  taken from (2). When  $p_1 - q_1 = (1 - q_1)(p_2 - q_2)$ , the asymptotic mean of  $\hat{\Delta}_O$  is  $p_1 - q_1$  and  $\hat{\Delta}_O$  is asymptotically unbiased.

Table 2 gives the values for the asymptotic variance ( $\times n$ ) of  $\hat{\Delta}_{MLE}$  and the optimal linear combination for various values of  $p_1, q_1, p_2, q_2$  and  $a$  when the assumptions underlying  $\hat{\Delta}_2$  hold. Clearly, substantial reductions in variance are possible when the data from the second phase are combined with the first phase data. The issue of course is that the optimal weight is a function of the unknown parameters. One obvious approach would be to plug in the maximum likelihood estimators for  $p_1, q_1, p_2,$  and  $q_2$  into  $\hat{\Delta}_O$ , and we denote this plug-in estimator as  $\hat{\Delta}_{EST}$ . The danger however of using weights with random components and ignoring this randomness, has been discussed by Shuster [8] in the context of meta-analyses.

An alternative approach to the plug-in estimator is to consider weights that are only a function of the known allocation ratio  $a$ . In examining (1) we noticed that the covariance tends to be much smaller in magnitude than the variance quantities. For example, if  $n = 1, a = .25, p_1 = .6, q_1 = .5, p_2 = .5,$  and  $q_2 = .3,$  the covariance matrix is

$$\begin{pmatrix} .98 & .1 \\ .1 & .94 \end{pmatrix} .$$

Suppose we ignore the covariance in determining the weight. We also somewhat arbitrarily

assume  $q_1 = .4$  and  $p_1(1 - p_1) = p_2(1 - p_2) = q_2(1 - q_2) = .2$ . Then  $w_a = \frac{.24(1 - 2a)}{.36 - .52a}$ . This allocation based weight is compared to the optimal weight in Table 2. It appears that this estimator does a reasonable job as long as the unknown parameters are not too close to zero. For these values of underlying parameters,  $w_a$  tends to overweight  $\hat{\Delta}_{MLE}$  which may not be unreasonable since  $\hat{\Delta}_{MLE}$  is always unbiased. We call this linear combination estimator  $\hat{\Delta}_{WA}$ . Note in Table 2 that  $Var(\hat{\Delta}_{WA})$  is close to  $Var(\hat{\Delta}_O)$ . Both  $\hat{\Delta}_{EST}$  and  $\hat{\Delta}_{WA}$  are asymptotically unbiased when  $p_1 - q_1 = (1 - q_1)(p_2 - q_2)$ .

The monotonicity and constancy of treatment effect assumptions imply that  $p_1 - q_1 = (1 - q_1)(p_2 - q_2)$ . If this does not hold, the linear combination estimator will be biased for estimating  $p_1 - q_1$ . The magnitude of the bias compared to the variance depends on the underlying parameters and the sample size  $n$ . In the following section, we give results from our simulation study which looks at the various point estimates and their approximate 95% confidence intervals.

### 3. Simulation Study

The estimators,  $\hat{\Delta}_{MLE}$ ,  $\hat{\Delta}_{CMLE}$ ,  $\hat{\Delta}_O$ ,  $\hat{\Delta}_{EST}$ , and  $\hat{\Delta}_{WA}$  (hereafter referred to as MLE, CMLE, O, EST, and WA respectively) were examined in a simulation study. This section presents the results of a small representative sample from that study. The oracle estimator,  $\hat{\Delta}_O$ , is not an estimator one can use in practice since it uses the true values of  $p_1$ ,  $q_1$ ,  $p_2$ , and  $q_2$  in both the weight and the variances of the estimator. However we examined it in the simulations as a benchmark for the other linear combination estimators.

We consider total sample sizes from  $n = 50$  to  $n = 400$ . At each sample size and parameter configuration, 1000 realizations were simulated. For all estimators excluding  $\hat{\Delta}_O$ , we used Wald type confidence intervals obtained by the underlying asymptotic normality of the estimators. The Newton-Raphson method using initial values,  $q_1$ ,  $q_2$  and

$\frac{1}{2}(\hat{p}_1 - \hat{q}_1 + \hat{p}_2 - \hat{q}_2)$  was used to calculate CMLE. If the method failed to converge, a small set of alternative initial values were tested. If nonconvergence was still an issue, then a grid search over allowable values of  $q_1$ ,  $q_2$  and  $\Delta$  was used to determine CMLE. Table 3 shows the performance of the estimators in terms of bias, variance, mean squared error (MSE), coverage probability of the nominal 95% confidence interval, and the average length of the 95% confidence interval. At each sample size, we show the range of S.E. across the five estimators for each performance measure. In Table 3, we show two cases for which  $p_1 - q_1 = (1 - q_1)(p_2 - q_2)$ . These cases represent cases in which the assumptions underlying the linear combination estimators are true and the estimators are asymptotically unbiased. In Table 3, the allocation ratio  $a$  is equal to 0.35 which is close to the case where the initial allocation is 1:1:1 for the sequential parallel design. We also examined other  $a$  ranging from 0.20 to 0.40 and the results were qualitatively similar. The results show that in all cases, the MSE of the linear combination estimators are smaller than the MLE and CMLE estimators and the lengths of the 95% confidence intervals are smaller than the corresponding lengths for the MLE and CMLE. This indicates that the gains in using a linear combination estimator can be substantial, especially at moderate sample sizes from  $n = 100$  to  $n = 200$ . At low sample sizes, the confidence intervals for all of the estimators excluding the benchmark O tend to be liberal, especially for the EST estimator.

Table 4 shows the same performance measures of the estimators for two cases for which  $p_1 - q_1 \neq (1 - q_1)(p_2 - q_2)$ . In these cases, the linear combination estimators will be biased for the true quantity of interest  $p_1 - q_1$  and this is evident in the bias columns of the table. Negative values for bias indicate cases where the estimator underestimates the true quantity of interest. The bias for the linear combination estimators does not go to zero as  $n$  increases and thus, as  $n$  gets larger, the bias becomes more pronounced in comparison to the variance. This causes the coverage probability to deteriorate as  $n$  increases. At moderate values of  $n$ , the coverage probability of the linear combination estimators tends to be 90-95% as opposed to the nominal 95%. Even though the estimators are biased, the MSE for the linear combination estimators is still smaller than the MSE for the MLE which suggests that as a point estimate, the linear combination estimator is still competitive with the MLE. The second set of parameters in Table 4 is a case in which the constrained maximum likelihood estimator is asymptotically unbiased since  $p_1 - q_1 = p_2 - q_2$ . At small and moderate sample



sizes, failure of convergence for the Newton-Raphson method was more pronounced and 5-20% of the time, a grid search was needed to compute CMLE.

When  $p_1 - q_1 > (1 - q_1)(p_2 - q_2)$ , the bias is a function of how far apart the two estimates are. In general,  $q_2 < q_1$  and thus, as  $q_1$  increases, there is a greater chance that the discrepancy is large. Thus, more caution is needed in using 95% confidence intervals for the linear combination estimators when the placebo response in the first phase is large. Even in cases of large discrepancy, however, we observed that the MSE of the linear combination estimators was lower than the MSE of the MLE.

#### 4. Real Data Example

A sequential parallel clinical trial was conducted in Selective Serotonin Reuptake Inhibitor (SSRI) treated patients with Major Depressive Disorder with the test drug being 15 mg of L-methylfolate (ClinicalTrials.gov Identifier: NCT00955955). In this study, the test drug or placebo was augmented to the SSRI which the patient was currently taking. The allocation ratio was 3:3:2 for the groups identified in Table 1 which corresponds to  $a=0.375$ . The treatment period for both the first phase and the second phase was 30 days. The primary outcome scale was the change in total scores of the 17 item Hamilton Depression Rating Scale. Response was defined as a 50% reduction from baseline (defined as initiation of the phases) in the total score. Figure 1 illustrates the final response data from the trial (Fava et al. [9]). In this example,  $\hat{p}_1 = .368$ ,  $\hat{q}_1 = .196$ ,  $\hat{p}_2 = .278$  and  $\hat{q}_2 = .095$ . The dropout of placebo non-responders adds additional multinomial categories in the sequential parallel design (Tamura and Huang [10]). If  $s$  is defined as the retention rate of placebo non-responders, then the covariance matrix (1) is modified with the last two terms of the asymptotic variance of  $\Delta_2$  divided by  $s$ . The constrained MLE of  $s$  has a closed form solution as the empirical retention rate in placebo non-responders and the constrained MLE of the remaining parameters may be obtained by similar optimization methods as before. With these adjustments, the estimated weight for the plug-in estimator is 0.379, resulting in  $\hat{\Delta}_{EST} = 0.156$  (s.e.=0.078). The weighted estimator  $\hat{\Delta}_{WA} = 0.156$  (s.e.=0.078) and the constrained MLE  $\hat{\Delta}_{CMLE} = 0.177$  (s.e.=0.086). Figure 2 shows the point estimates and the associated 95% confidence limits of the difference in response rates in the overall population. Although all 4 point estimators are very similar, there is a large reduction in the length of the 95% confidence interval for the linear combination estimators and the constrained MLE. The reported p-value for the drug effect was 0.0399 (Fava et al. [9]) which is consistent with the confidence limits which utilize the second phase information.

#### 5. Discussion

Our simulation studies suggest there are substantial gains possible in utilizing information from the second phase of the design in estimating the treatment effect and associated confidence intervals. The linear combination estimators and associated confidence intervals are simple to calculate compared to the constrained maximum likelihood estimator. There appears to be little difference between the two linear combination estimators considered here, however at small sample sizes, we prefer the estimator which weights information only based on the allocation ratio. There are large gains for moderate sample size studies which would be the case in Phase II or proof of concept studies. As the sample size becomes large, bias can cause the coverage probability of the confidence intervals from the linear combination estimators to break down if  $p_1 - q_1 > (1 - q_1)(p_2 - q_2)$ . This is also true of the CMLE if the treatment effects are not equal in the two phases. If the sequential parallel design is used in a large pivotal Phase III trials where correct coverage is important, it seems reasonable to use the MLE confidence interval widths with a linear combination estimator for the point estimate.



The logic we have used to construct the estimators can be extended to continuous data. In such a case, the important factor is the joint distribution of the estimators of treatment effect from the first phase and the second phase. Once the joint distribution is known, then analogous linear combination estimators to what we have constructed for binary data, can be obtained. The monotonicity assumption in the continuous data case assumes that the quantitative response in placebo responders would be exactly the same whether the patient was administered drug or placebo. Some hint of whether this assumption is tenable can be obtained by examining the data in the second phase from placebo responders. If placebo responders when switched to drug show a significant difference compared to placebo responders who remained on placebo, then the assumption of monotonicity would be questionable.

One of the primary goals of Phase II studies is to get information on the treatment effect in order to design the pivotal Phase III studies. In such studies, exact coverage of the confidence interval is less important. However, analytical methods which reduce the uncertainty of the point estimate are important whether the information is used informally or more formally as was suggested by Chuang-Stein [11] for the design of pivotal Phase III studies. Another important usage of the linear combination estimator and confidence intervals would be in graphical displays involving a number of clinical trials. As an example, suppose a sponsor utilized a sequential parallel design in a Phase II study and subsequently used the traditional parallel design in two Phase III studies. In an integrated summary of efficacy, it would be important to use a linear combination estimator and confidence interval for the sequential parallel design (assuming that there was not a large difference between  $p_1\hat{p}_1 - q_1\hat{q}_1$  and  $(1 - q_1)(p_2\hat{p}_2 - q_2\hat{q}_2)$ ), when graphically representing the results from all three studies. Our results show that if the sequential parallel design is used as a Phase II study, it is important that the treatment effect and variability around that effect be calculated using data from both phases of the study.

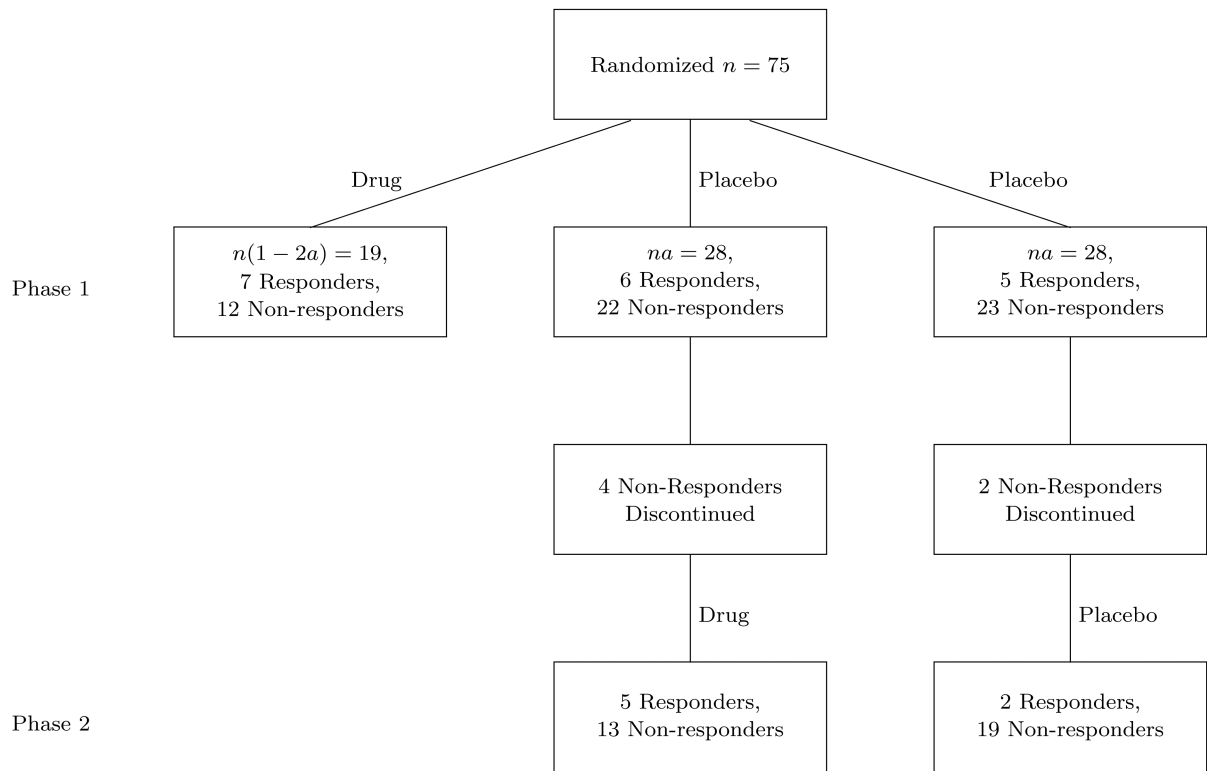
## Acknowledgments

The authors thank Dr. Maurizio Fava and Dr. David Schoenfeld for sharing the L-methylfolate results with us prior to their presentation. Dennis Boos was supported by NIH grant P01 CA142538-01.

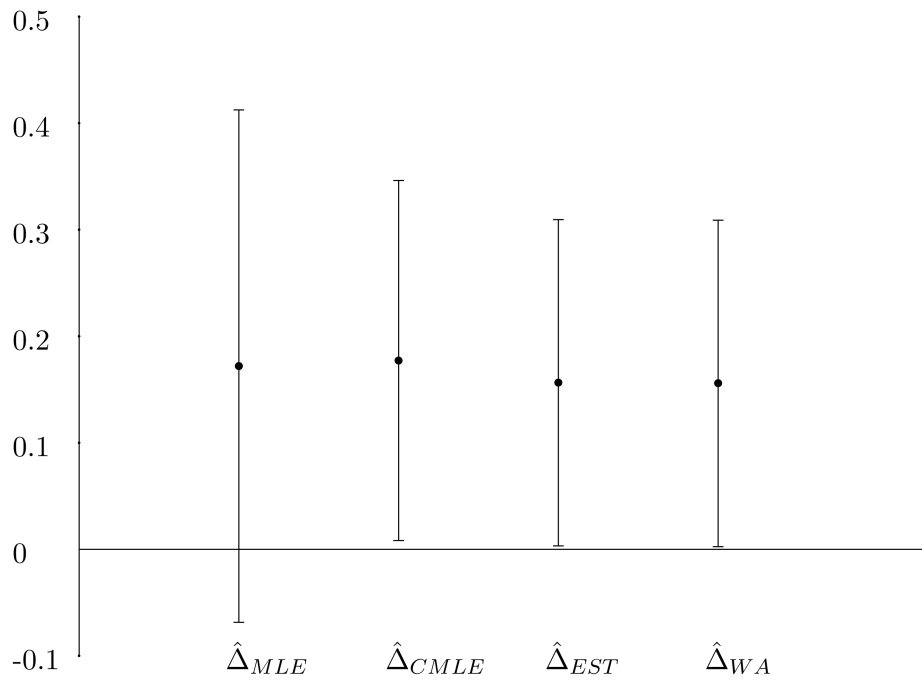
## References

1. Khan A, Khan SR, Walens G, Kolts R, Giller EL. Frequency of positive studies among fixed and flexible dose antidepressant clinical trials: an analysis of the Food and Drug Administration summary basis of approval reports. *Neuropsychopharmacology*. 2003; 28:552–557.10.1038/sj.npp.1300059 [PubMed: 12629536]
2. 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. *Psychotherapy and Psychosomatics*. 2003; 72:115–127.10.1159/000069738 [PubMed: 12707478]
3. Fava M, Evins AE, Dorer DJ, Schoenfeld DA. Erratum. *Psychotherapy and Psychosomatics*. 2004; 73:123.10.1159/000076725
4. Fedorov, VV.; Liu, T. *Wiley Encyclopedia of Clinical Trials*. New York: Wiley; 2007. Enrichment design.
5. Huang X, Tamura RN. Comparison of Test Statistics for the Sequential Parallel Design. *Statistics in Biopharmaceutical Research*. 2010; 2(1):42–50.10.1198/sbr.2010.08015
6. Ivanova A, Qaqish B, Schoenfeld DA. Sample Size and Power Calculations for the Sequential Parallel Comparison Design. *Statistics in Medicine*. 2010 Submitted to.
7. Serfling, RJ. *Approximation Theorems of Mathematical Statistics*. New York: Wiley; 1980.
8. Shuster JJ. Empirical vs natural weighting in random effects meta-analysis. *Statistics in Medicine*. 2010; 29(12):1259–65.10.1002/sim.3607 [PubMed: 19475538]

9. Fava, M.; Shelton, R.; Zajecka, J.; Rickels, K.; Clam, A.; Bace, L.; Schoenfeld, D.; Nelson, E.; Baeber, J.; Lydracd, B.; Migeoulon, D.; Alpect, J.; Zisook, S.; Papakogtas, G. L-methylfolate Augmentation of Selective Serotonin Reuptake Inhibitors (SSRIs) for SSRI-Resistant Major Depressive Disorder: Results of Two Randomized Double Blind Trials. Poster Presented at 2010 ACNP Annual Meeting; 2010.
10. Tamura RN, Huang X. An examination of the efficiency of the sequential parallel design in psychiatric clinical trials. *Clinical Trials: Journal of the Society of Clinical Trials*. 2007; 4:309–317.10.1177/1740774507081217
11. Chuang-Stein C. Sample size and the probability of a successful trial. *Pharmaceutical Statistics*. 2006; 5:305–309.10.1002/pst.232 [PubMed: 17128428]



**Figure 1. Response data from sequential parallel clinical trial of 15mg L-methylfolate**



**Figure 2.** Point estimates and 95% confidence intervals for the treatment-placebo response rates for the L-methylfolate example.

**Table 1**

Sequential parallel trial design.

Treatment	Response		Frequency	Probability
	First phase	Second phase		
Placebo	Placebo	No	$n_{11}$	$(1 - q_1)q_2$
	<i>na</i>	No	$n_{12}$	$(1 - q_1)(1 - q_2)$
		Yes	$n_{13}$	$q_1$
Drug	Drug	Yes	$n_{21}$	$(1 - q_1)p_2$
	<i>na</i>	No	$n_{22}$	$(1 - q_1)(1 - p_2)$
		Yes	$n_{23}$	$q_1$
Drug	Drug	Yes	$n_{31}$	$p_1$
	$n(1 - 2a)$	No	$n_{32}$	$1 - p_1$

**Table 2**

Asymptotic variances ( $\times n$ ) of the Phase I estimator,  $\hat{\Delta}_{MLE}$ , the optimal linear combination estimator,  $\hat{\Delta}_{\hat{O}}$  and the allocation weighted estimator  $\hat{\Delta}_{WA}$ .

$a$	$p_1$	$q_1$	$p_2$	$q_2$	$w^*$	$w_a^{**}$	$Var(\hat{\Delta}_{MLE})$	$Var(\hat{\Delta}_{\hat{O}})$	$Var(\hat{\Delta}_{WA})$
.25	.6	.5	.5	.3	.488	.522	0.980	0.530	0.532
	.4	.3	.35	.1	.505		0.900	0.506	0.507
.30	.6	.5	.5	.3	.429	.471	1.017	0.483	0.486
	.4	.3	.35	.1	.439		0.950	0.466	0.468
.35	.6	.5	.5	.3	.356	.404	1.157	0.458	0.462
	.4	.3	.35	.1	.361		1.100	0.445	0.448

\* Optimal weight

\*\* Allocation based weight

**Table 3**

Performance of estimators when  $p_1 - q_1 = (1 - q_1)(p_2 - q_2)$ . Coverage is the empirical coverage of the nominal 95% confidence interval and length is the average length of the 95% confidence interval. Allocation  $\alpha = 0.35$ .

n	Estimator	$p_1 = .6, q_1 = .5, p_2 = .5, q_2 = .3$					$p_1 = .4, q_1 = .2, p_2 = .35, q_2 = .1$				
		Bias	Variance	MSE	Coverage	Length	Bias	Variance	MSE	Coverage	Length
50	MLE	.001	.023	.023	.93	0.590	.002	.021	.021	.93	0.557
	CMLE	.032	.016	.017	.93	0.483	.023	.011	.012	.93	0.395
	O	.003	.009	.009	.95	0.374	-.001	.009	.009	.96	0.375
	EST	.012	.009	.010	.92	0.356	-.002	.010	.010	.92	0.358
	WA	.003	.009	.009	.94	0.363	-.001	.009	.009	.94	0.362
Range SE		.003-	.0004-	.0004-	.7-	.0000-	.003-	.0004-	.0004-	.6-	.0000-
		.005	.0010	.0010	.9	.0010	.005	.0010	.0009	.9	.0012
100	MLE	.001	.011	.011	.95	0.416	.004	.010	.010	.94	0.393
	CMLE	.028	.008	.009	.93	0.345	.025	.005	.006	.94	0.283
	O	-.002	.004	.004	.95	0.265	.000	.004	.004	.96	0.265
	EST	.002	.005	.005	.93	0.259	-.001	.004	.004	.94	0.259
	WA	-.001	.004	.004	.95	0.261	.000	.004	.004	.95	0.260
Range SE		.002-	.0002-	.0002-	.7-	.0000-	.002-	.0002-	.0002-	.7-	.0000-
		.003	.0005	.0005	.8	.0004	.003	.0004	.0004	.8	.0005
200	MLE	-.003	.006	.006	.93	0.296	.003	.005	.005	.94	0.279
	CMLE	.028	.004	.005	.91	0.246	.025	.003	.003	.91	0.202
	O	.000	.002	.002	.95	0.188	.001	.002	.002	.94	0.188
	EST	.002	.003	.003	.94	0.185	.001	.002	.002	.94	0.185
	WA	-.001	.002	.002	.94	0.187	.001	.002	.002	.94	0.186
Range SE		.002-	.0001-	.0001-	.7-	.0000-	.002-	.0001-	.0001-	.8-	.0000-
		.002	.0003	.0003	.9	.0002	.002	.0002	.0002	.9	.0002
400	MLE	-.004	.003	.003	.95	0.210	.000	.002	.002	.94	0.198
	CMLE	.026	.002	.003	.91	0.175	.025	.001	.002	.91	0.144
	O	-.002	.001	.001	.95	0.133	.001	.001	.001	.96	0.133
	EST	-.001	.001	.001	.94	0.132	.001	.001	.001	.95	0.132
	Range SE		.002	.0003	.0003	.9	.0002	.002	.0002	.0002	.9



$p_1 = .4, q_1 = .2, p_2 = .35, q_2 = .1$

$p_1 = .6, q_1 = .5, p_2 = .5, q_2 = .3$

n	Estimator	Bias	Variance	MSE	Coverage	Length	Bias	Variance	MSE	Coverage	Length
	WA	-.002	.001	.001	.94	0.133	.001	.001	.001	.96	0.132
	Range SE	.001-	<.0001-	<.0001-	.7-	.0000-	.001-	<.0001-	<.0001-	.6-	.0000-
		.002	.0001	.0001	.9	.0001	.002	.0001	.0001	.9	.0001

**Table 4**

Performance of estimators when  $p_1 = q_1$  (1 -  $q_1$ )/( $p_2 - q_2$ ). Coverage is the empirical coverage of the nominal 95% confidence interval and length is the average length of the 95% confidence interval. Allocation  $\alpha = 0.35$ .

n	Estimator	$p_1 = .4, q_1 = .2, p_2 = .4, q_2 = .1$					$p_1 = .4, q_1 = .2, p_2 = .3, q_2 = .1$				
		Bias	Variance	MSE	Coverage	Length	Bias	Variance	MSE	Coverage	Length
50	MLE	-.002	.022	.022	.92	0.555	.008	.024	.024	.93	0.554
	CMLE	.042	.013	.014	.89	0.402	.002	.011	.011	.92	0.388
	O	.021	.010	.011	.94	0.382	-.023	.009	.009	.95	0.366
	EST	.019	.011	.012	.90	0.366	-.027	.010	.010	.90	0.347
	WA	.021	.010	.011	.92	0.370	-.022	.009	.009	.92	0.353
Range SE		.003-	.0005-	.0005-	.8-	.0000-	.003-	.0004-	.7-	.0000-	
		.005	.0010	.0010	.10	.0013	.005	.0011	.0011	1.0	.0013
100	MLE	.001	.010	.010	.93	0.391	.001	.010	.010	.93	0.391
	CMLE	.043	.006	.008	.89	0.287	.000	.006	.006	.94	0.279
	O	.021	.005	.005	.93	0.270	-.024	.005	.005	.93	0.259
	EST	.020	.005	.006	.92	0.265	-.025	.005	.005	.91	0.253
	WA	.021	.005	.005	.93	0.266	-.024	.005	.005	.92	0.255
Range SE		.002-	.0002-	.0002-	.8-	.0000-	.002-	.0002-	.7-	.0000-	
		.003	.0005	.0004	1.0	.0005	.003	.0005	.0005	.9	.0005
200	MLE	-.002	.005	.005	.93	0.279	-.003	.005	.005	.95	0.279
	CMLE	.045	.003	.005	.85	0.204	.001	.003	.003	.94	0.199
	O	.021	.002	.003	.93	0.191	-.023	.002	.003	.92	0.183
	EST	.021	.003	.003	.92	0.189	-.023	.002	.003	.91	0.182
	WA	.022	.002	.003	.93	0.189	-.023	.002	.003	.91	0.182
Range SE		.002-	.0001-	.0001-	.8-	.0000	.002-	.0001-	.7-	.0000-	
		.002	.0002	.0002	1.1	.0003	.002	.0002	.0002	.9	.0003
400	MLE	-.001	.003	.003	.95	0.198	.000	.002	.002	.94	0.198
	CMLE	.044	.001	.003	.79	0.145	.000	.001	.001	.96	0.141
	O	.021	.001	.002	.92	0.135	-.024	.001	.002	.89	0.130
	EST	.021	.001	.002	.92	0.135	-.025	.001	.002	.88	0.129
	WA	.021	.001	.002	.92	0.135	-.025	.001	.002	.88	0.129

$p_1 = .4, q_1 = .2, p_2 = .3, q_2 = .1$

$p_1 = .4, q_1 = .2, p_2 = .4, q_2 = .1$

n	Estimator	Bias	Variance	MSE	Coverage	Length	Bias	Variance	MSE	Coverage	Length
	WA	.022	.001	.002	.92	0.135	-.024	.001	.002	.89	0.129
	Range SE	.001-	<.0001-	<.0001-	.7-	.0000-	.001-	<.0001-	<.0001-	.7-	.0000-
		.002	.0001	.0001	1.3	.0001	.002	.0001	.0001	1.0	.0001