

## On clinical trials with a high placebo response rate



George Y.H. Chi <sup>a,4,\*</sup>, Yihan Li <sup>b,1</sup>, Yanning Liu <sup>a,2</sup>, David Lewin <sup>a,3</sup>, Pilar Lim <sup>a,2</sup>

<sup>a</sup> Janssen Research & Development, L.L.C, USA

<sup>b</sup> AbbVie, Pharmaceutical Research & Development, L.L.C, USA

### ARTICLE INFO

#### Article history:

Received 27 August 2015

Accepted 26 October 2015

Available online 18 November 2015

#### Keywords:

Adjusted treatment effect

Combination test

Consistency test

Doubly randomized delayed start design

Enrichment design

Joint test

Monotonicity condition

Placebo response

Sequential parallel design

### ABSTRACT

The basic problem that causes the frequent failure of a standard randomized parallel placebo-controlled clinical trial with a high placebo response rate is the underestimation of the treatment effect by the observed relative treatment difference. A two-period sequential parallel enrichment design has been proposed where the first period is a standard parallel design and at the end of the first period, the placebo non-responders are identified and re-randomized in the second period. Based on such a design, available methods have primarily focused on testing either the first period treatment null hypothesis or the global null hypothesis defined as the joint period 1 and period 2 treatment effect null hypothesis by a test statistic which is either derived from a combined statistic or defined directly as a weighted z-score where the weights are functions of some population and design parameters satisfying certain power optimality criterion. However, in some cases, it is not clear what their combined statistics are estimating and in others, the combined statistics are estimating the apparent treatment effect; but generally, there is no discussion of the need to provide a proper assessment of the treatment effect for the intended study population. It should be clear that an appropriate assessment of the treatment effect for the intended study population is critical for the benefit/risk analysis as well as the proper dosage recommendation. Any benefit/risk analysis and dosage recommendation that are based on an apparent treatment effect from a standard parallel design such as the first period of a sequential parallel enrichment design tend to underestimate the benefit/risk ratio which in turn may lead to overdosing recommendation. It is the purpose of this paper to introduce the concept of an adjusted treatment effect which is derived by adjusting the apparent treatment effect from the first period of a sequential parallel enrichment design with information from the second period subject to a consistency condition. The adjustment properly compensates for the high placebo response rate. It is proposed that this adjusted treatment effect should be used to assess the treatment effect for the intended study population and should be the basis for the benefit/risk analysis and the dosage recommendation.

© 2015 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. Introduction

The basic reason for the failure of many standard randomized parallel placebo-controlled clinical trials with high placebo response rate is that the observed relative treatment difference only provides an estimate of an apparent treatment effect since the treatment effect has been diminished by the presence of a

substantial proportion of placebo responders in the population. The full treatment effect cannot be directly estimated by the relative treatment difference. An appropriate assessment of the full treatment effect is critical for making a risk/benefit analysis and dosage recommendation. The primary purpose of this paper is to propose a method for adjusting the apparent treatment effect to account for the high placebo response rate within the framework of a doubly randomized delayed start (DRDS) design as discussed in Liu et al. [1] which improves upon the earlier sequential parallel design (SPD) of Fava et al. [2].

## 2. Background

### 2.1. The sequential enrichment design

The problem of a high placebo response rate in clinical trials

\* Corresponding author. 920 US Highway 202S, Raritan, NJ 08869, USA.

E-mail addresses: [chionroad@gmail.com](mailto:chionroad@gmail.com) (G.Y.H. Chi), [yihan.li@abbvie.com](mailto:yihan.li@abbvie.com) (Y. Li), [yliu@its.jnj.com](mailto:yliu@its.jnj.com) (Y. Liu), [Lewin@StatSpeaking.com](mailto:Lewin@StatSpeaking.com) (D. Lewin), [plim@its.jnj.com](mailto:plim@its.jnj.com) (P. Lim).

<sup>1</sup> AP9A-2, 1 North Waukegan Rd, North Chicago, IL 60064, USA.

<sup>2</sup> Trenton-Harbourton Road, Titusville, NJ 08568, USA.

<sup>3</sup> 1434 W Belden Ave, Chicago, IL 60614, USA.

<sup>4</sup> Current address: 12436 Over Ridge Rd, Potomac, MD 20854, USA.

occurs in several therapeutic areas, but it is most often observed in trials involving subjects with psychiatric disorders. In these populations of subjects with psychiatric disorders, the placebo response rate has been estimated to vary from 30% to 50%. Trials in these therapeutic areas often failed because in a standard randomized parallel placebo-controlled trial, the observed relative treatment difference only provides an estimate of an *apparent* treatment effect which does not reflect the full treatment effect due to the dilution resulting from the presence of a substantial proportion of placebo responders. This problem has been known for quite some time. Temple [3] had suggested an enrichment design whereby subjects responding to placebo in a run-in period are excluded from a second period during which placebo non-responders are re-randomized to treatment and placebo in a parallel design. The purpose of Temple's enrichment design is merely to show that the treatment is effective in some subpopulation and in this case in the subpopulation of placebo non-responders. However, one problem with this enrichment design is that the claim of treatment effectiveness cannot be readily extended to the entire intended study population. Another problem with this design is that if the treatment is to be indicated for the enriched subpopulation, then in actual clinical practice, a patient has to be given placebo first to verify his/her placebo response status before the treatment can be prescribed; however, this would entail an ethical dilemma.

Fava et al. [2] proposed a SPD design where subjects are randomized to a treatment group and two placebo groups in the first period. At the end of the first period, the non-responders in one placebo group will be given treatment in the second period, while the non-responders in the other placebo group will continue with placebo in the second period. The subjects in the treatment group in the first period will continue on the treatment in the second period. It should be noted that in the original proposed SPD design, the randomization in Period 2 refers to the original randomization conducted at the beginning of the first period. The lack of a re-randomization in the second period poses potential imbalance in key covariates between the two placebo non-responder groups at the end of the second period if there is a differential placebo dropout rate between the two placebo arms. Such imbalance may introduce bias and cause difficulty in the statistical inference. Liu et al. [1] proposed a doubly randomized delayed start (DRDS) design which was presented earlier at the 2010 BASS Conference. This DRDS design involves randomizing the subjects to treatment and placebo in the first period and then re-randomizing the

placebo non-responders identified at the end of the first period based on some pre-specified response threshold to treatment and placebo in the second period. The terms “delayed start” were used for the obvious application of this design to trials involving progressive diseases. A simple diagram of such a design is depicted in Fig. 1.

Chen et al. [4] considered a SPD design with re-randomization in the second period which they termed a SPD-ReR design. Now, the original SPD design has since also been revised to include re-randomization in the second period. In this paper, the DRDS design may refer to a SPD ReR design or a SPD design with re-randomization if found appropriate, and for convenience, some of the terminologies and notations used in Liu et al. [1] are adopted. The DRDS design has been accepted by the regulatory agencies as an innovative design. However, the regulatory agencies have raised issues with various proposed methods of analysis. In order to address these issues, a new statistical methodology is proposed here that includes the DRDS design and a statistical approach for this design that differs from the currently available methods.

## 2.2. Some key issues associated with the current methods for a DRDS design

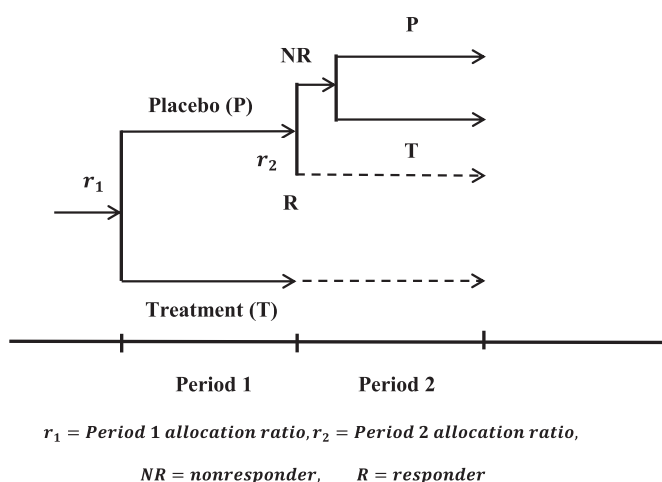
There are a few important conceptual and technical issues related to the problem of a high placebo response rate in a DRDS design that have not been mentioned nor discussed by the previous authors. These basic issues need to be satisfactorily resolved before a DRDS design can be applied to phase 3 trials to obtain the evidence of effectiveness required. These issues will now be discussed and they will be addressed in the new approach to be proposed in Section 4.

### 2.2.1. Issue 1

The customary view considers the standard randomized parallel double blind placebo-controlled design as the design of choice because the relative treatment difference from such a design reflects the net treatment effect over and beyond what is expected of a placebo which should be minimal for this view to be valid. In a study population that has a substantial proportion of placebo responders, the relative treatment difference is only an apparent treatment difference, because it ignores the mitigating effect of the presence of a high placebo response rate on this treatment difference. This is the primary reason why many such trials have failed in the past. In a DRDS design, this same problem is present in the first period. Therefore, clearly the apparent treatment effect from the first period would be underestimating the full treatment effect. Another problem inherent in the above view is that even if perchance the apparent treatment effect shows the treatment is superior to placebo, any dosage recommendation based on an apparent dose–response relationship would likely lead to overdosing. Hence, for these two reasons alone, an appropriate assessment of the treatment effect adjusting for high placebo response rate is needed.

### 2.2.2. Issue 2

A problem that is born of the above view is present in the current proposed methods of analysis of a DRDS design. These methods variously proposed to estimate the apparent treatment effect of Period 1 by a combined statistic, which is defined as a weighted combination of the apparent treatment effect of Period 1 and the enriched treatment effect of Period 2 under some assumptions. For example, in Huang and Tamura [5], a score test is derived under the constancy assumption which requires that the enriched treatment effect of Period 2 be equal to the apparent treatment effect of Period 1, while for binary outcome, in Tamura



**Fig. 1.** A basic DRDS design for assessing treatment effect in trials with a high placebo response rate.

and Huang [8], the combined statistic is derived under the monotonicity condition which assumes that each placebo responder is also a treatment responder. In each instance, the assumption may be invalid or unnecessarily stringent. Furthermore, the combined statistic is used to derive a combination test for testing either the apparent treatment null hypothesis of Period 1, or a global null hypothesis which is defined as the joint apparent treatment null of Period 1 and the enriched treatment null of Period 2. Even if these assumptions are appropriate, the rejection of these null hypotheses by these combination tests would not have solved the problem discussed under Issue 1 above.

### 2.2.3. Issue 3

A problem that arises as a result of the two issues discussed above is that the weights used in the combined statistics are functions not only of the population parameters, but also some DRDS design parameters, in particular the placebo to treatment allocation ratios in Period 1 and Period 2. One can place more weight on Period 2 treatment effect estimate in the combined statistic by simply increasing the allocation ratio in Period 1. Such bias is present even when the allocation ratio in Period 1 is equal to 2 as is the case in most of the DRDS designs used in these earlier papers. Such potential bias causes concern over these combined statistics and is interpreted as biasing the estimate of the apparent treatment effect of Period 1. Such misleading use of the Period 2 result and a misleading interpretation of the purpose of the second period of a DRDS design is unfortunate and should be corrected.

### 2.2.4. Issue 4

Assuming for the moment that a combined statistic with weights that are independent of the allocation ratios has been defined. Then, one needs to know what this combined statistic is estimating and how to interpret it. Is the combined statistic estimating a treatment effect for the intended study population? Does the treatment effect represent an appropriate assessment of the full treatment effect in the intended study population? Does the treatment effect adjust for the presence of placebo responders in the intended study population? Interpretability of the estimate of a combined statistic is crucial in its acceptability as an estimate of the full treatment effect for the intended study population. Such interpretation is lacking for the combined statistics in most of the current available methods, except for those cases where the combined statistics are meant to estimate the apparent treatment effect of Period 1 as discussed in Issue 2 above.

### 2.2.5. Issue 5

Assuming that a combined statistic is estimating the true treatment effect for the intended study population as discussed in Issue 4, one problem that may arise is that it is possible for the combined statistic to show a positive combined treatment effect, yet the estimate of the apparent treatment effect from Period 1 may be negative. This kind of inconsistency is not a desirable outcome, since it suggests that the treatment effect may be substantially worse than placebo among the placebo responders. This issue is also not addressed relative to the combined statistics in the current available methods in addition to their problems as discussed above although it is related to the monotonicity condition introduced in Tamura et al. [8].

### 2.2.6. Issue 6

In all of the currently available methods, Period 2 of a DRDS design is simply viewed as a trial independent of Period 1. However, realistically, the probability structure underlying Period 2 in a DRDS design is conditional in nature. The sample cohorts in Period 2 represent placebo non-responders in Period 1 who are re-

randomized in Period 2 into treatment and placebo groups. Therefore, the distributions of the response variables for these cohorts in Period 1 and Period 2 are singly truncated bivariate normal distributions where the Period 1 placebo responses of these cohorts have been truncated at some pre-specified threshold. Hence, the distributions of these cohorts in Period 2 are conditional distributions with the condition specified by the truncation of their placebo response in Period 1 at some threshold. Thus, the treatment effect at the end of Period 2 will be conditional in nature which has some interesting and useful properties that are not available or apparent under the unconditional probability structure.

To address the above issues, a new approach is proposed in this paper. The probability structure underlying a DRDS design is first developed in Section 3. Then, in Section 4, the key concept of an adjusted treatment effect will be defined as a specific weighted treatment effects from Period 1 and Period 2 where the weights are independent of the allocation ratios and any design parameters. This adjusted treatment effect can be interpreted as an adjustment of the apparent treatment effect from Period 1 by appropriately accounting for the presence of placebo responders in the intended study population. Period 2 of a DRDS design provides the information needed to make this adjustment possible. Therefore, this adjusted treatment effect provides an appropriate assessment of the full treatment effect for the intended study population. Then, in Section 5, a new combined statistic can be derived directly from the definition of the adjusted treatment effect so that it will provide an unbiased estimate of the adjusted treatment effect. The combination test derived from this combined statistic will then be used to test the adjusted treatment null hypothesis. In addition to this combination test, a new consistency measure is introduced in Section 6, which can be viewed as a natural generalization of the monotonicity condition for a continuous outcome. A consistency null hypothesis is defined from this consistency measure and a consistency test is derived to test the consistency of the treatment effects from the two periods which is now a condition needed for excluding the situation where the adjusted treatment effect is positive while the apparent treatment effect of Period 1 is negative. Finally, in Section 7, a joint test, which is defined as the simultaneous testing of both the adjusted treatment null by the combination test and the consistency null by the consistency test, is proposed for demonstrating that a treatment is effective for the intended study population. It is shown that this joint test controls the type I error strongly under most of the scenarios encountered in practice. In addition, it is shown that if a particular application scenario appears to fall in certain range that suggests potential inflation in type I error may be expected, then one can control the expected inflation of this type I error by increasing the allocation ratio  $r_1$  to a level  $>2$ . It should be noted that since the weights used to define the combined statistic is independent of the allocation ratios, a DRDS design is free to choose any allocation ratios in Period 1 and Period 2 as long as they satisfy certain inequalities that are usually met in any practical application. Once the joint null has been rejected, then the estimated adjusted treatment effect derived from the combined statistic should represent an appropriate assessment of the full treatment effect for the intended study population. In Section 8, a simulated DRDS designed trial is presented for illustration. A summary discussion concludes the paper in Section 9.

## 3. The DRDS design and its underlying probability structure

Before introducing the adjusted treatment effect, it is important to first discuss the probability structure underlying a DRDS design. The previous authors have essentially adopted the view that the two periods in a DRDS design may be considered as two

independent trials. In this section, a trial using the basic DRDS design is described and the probability structure behind this design is discussed which forms the basis for the proposed methodology. It will become clear that this underlying probability structure is crucial in establishing the needed properties for the proposed test statistics. In addition, it will be relevant at the study design stage.

Consider a trial with a DRDS design as shown in Fig. 1. Let  $\Omega = \Omega_1$  denote the intended study population, and assume that there is a subpopulation of placebo responders  $\Omega_R$  even though this subpopulation can't be characterized prior to the start of the trial. Let  $\Omega_{NR}$  denote the placebo non-responder subpopulation. Let  $T$  denote an experimental treatment and  $P$  the placebo. In Period 1,  $n_1$  subjects are randomly assigned to  $T$  and  $P$  in a placebo-to-treatment allocation ratio of  $r_1 \geq 1$  with  $n_{1,T}$  subjects assigned to treatment  $T$  and  $n_{1,P} = r_1 n_{1,T}$  subjects assigned to placebo  $P$ , where  $n_1 = n_{1,P} + n_{1,T}$ . Let  $X_1$  denote a continuous clinical response variable of interest,  $X_{1,T}$  and  $X_{1,P}$  the response variables under the treatment  $T$  and the placebo  $P$  respectively. Let  $X_{1,P} \sim N(\mu_{1,P}, \sigma_{1,P}^2)$  and  $X_{1,T} \sim N(\mu_{1,T}, \sigma_{1,T}^2)$  be normally distributed with the mean and variance  $(\mu_{1,P}, \sigma_{1,P}^2)$  and  $(\mu_{1,T}, \sigma_{1,T}^2)$  respectively. For simplicity and without much loss in generality, it will be assumed that  $\sigma_{1,P}^2 = \sigma_{1,T}^2 = \sigma_1^2$ . Let  $\Delta_1 = \mu_{1,T} - \mu_{1,P}$  denote the relative treatment difference in Period 1.

Let  $\{x_{1,P,i}, i = 1, 2, \dots, n_{1,P}\}$  and  $\{x_{1,T,j}, j = 1, 2, \dots, n_{1,T}\}$  denote the observed sample responses from the placebo and treatment groups respectively. Then,  $\hat{\Delta}_1 = (\hat{\mu}_{1,T} - \hat{\mu}_{1,P}) \sim N(\Delta_1, \sigma_1^2/n_{1,T}R_1)$ , where  $\hat{\mu}_{1,P} = 1/n_{1,P} \sum_{i=1}^{n_{1,P}} x_{1,P,i}$ ,  $\hat{\mu}_{1,T} = 1/n_{1,T} \sum_{j=1}^{n_{1,T}} x_{1,T,j}$ , and  $R_1 = r_1/(1 + r_1) = n_{1,P}/(n_{1,P} + n_{1,T})$  is the fraction of placebo subjects among the entire sample of  $n_1$  subjects.

When the variances  $\sigma_1^2$  and  $\sigma_2^2$  for  $\hat{\Delta}_1$  and  $\hat{\Delta}_2$  from Period 1 and Period 2 are considered unknown as is usually the case, then one may estimate these unknown variances by their respective pooled sample variances given by  $\hat{\sigma}_1^2 = ((n_{1,T} - 1) \hat{S}_{1,T}^2 + (n_{1,P} - 1) \hat{S}_{1,P}^2) / (n_{1,T} + n_{1,P} - 2)$  and  $\hat{\sigma}_2^2 = ((n_{2,T} - 1) \hat{S}_{2,T}^2 + (n_{2,P} - 1) \hat{S}_{2,P}^2) / (n_{2,T} + n_{2,P} - 2)$  where

$$\hat{S}_{1,T}^2 = \frac{1}{(n_{1,T} - 1)} \sum_{i=1}^{n_{1,T}} (X_{1,T,i} - \bar{X}_{1,T})^2,$$

$$\hat{S}_{1,P}^2 = \frac{1}{(n_{1,P} - 1)} \sum_{i=1}^{n_{1,P}} (X_{1,P,i} - \bar{X}_{1,P})^2$$

$$\hat{S}_{2,T}^2 = \frac{1}{(n_{2,T} - 1)} \sum_{i=1}^{n_{2,T}} (X_{2,T,i} - \bar{X}_{2,T})^2,$$

$$\hat{S}_{2,P}^2 = \frac{1}{(n_{2,P} - 1)} \sum_{i=1}^{n_{2,P}} (X_{2,P,i} - \bar{X}_{2,P})^2$$

At the end of Period 1, a pre-specified criterion will be applied to determine the response status of each placebo subject who completed the trial. This criterion may be translated into a threshold  $c$  in the range of the response variable  $X_1$ . At the end of Period 1, placebo subjects who are identified as responders, that is, if  $X_{1,P} > c$ , and along with the placebo dropouts will be excluded from the second period of the study. Those placebo subjects classified as non-responders, that is,  $X_{1,P} < c$ , will be re-randomized to treatment and placebo at the start of Period 2 in a placebo-to-treatment allocation ratio of  $r_2 \geq 1$ . For practical consideration,  $r_2$  is set to the value 1 in the present paper as is the case in most applications for obvious reason. It will also be assumed that the proportion of placebo non-responders among the placebo dropouts

in Period 1 is similar to their population proportion. For simplicity, it is assumed here that there were no placebo dropouts. Let  $\tau = (c - \mu_{1,P})/\sigma_{1,P}$  be the placebo response threshold standardized relative to the placebo response distribution in Period 1. Let  $n_2$  equal the number of placebo non-responders who completed Period 1 of the study and  $\gamma = \Phi(\tau) = \Phi((c - \mu_{1,P})/\sigma_{1,P})$  denote the population proportion of placebo non-responders in  $\Omega = \Omega_1$ . Then, the ratio  $\hat{\gamma} = n_2/n_{1,P}$  should be a consistent estimate of the parameter  $\Phi(\tau)$  in the absence of placebo dropouts, or under the above assumption if placebo dropouts are present.

At the start of Period 2, the  $n_2$  placebo non-responders from Period 1 will be re-randomized to treatment and placebo under equal allocation  $r_2 = 1$ . Then, it follows that  $n_{2,T} = n_{2,P} = n_2/(1 + r_2) = \gamma n_{1,P}/(1 + r_2) = \gamma n_{1,T} r_1 / (1 + r_2) = n_{1,T} \gamma R_{1,2}$ , where  $R_{1,2} = r_1 / (1 + r_2)$ .

Now without loss in generality and for obvious reason, consider relabeling the entire placebo sample in Period 1 as follows:

$$\{X_{1,P,i}, i = 1, 2, \dots, n_{2,T}, n_{2,T} + 1, n_{2,T} + 2, \dots, n_2, n_2 + 1, n_2 + 2, \dots, n_{1,P}\}$$

where the first  $n_{2,T}$  placebo subjects  $\{X_{1,P,i}, i = 1, 2, \dots, n_{2,T}\}$  are placebo non-responders that have been re-randomized in Period 2 to treatment, and the next set of  $n_{2,P}$  placebo subjects  $\{X_{1,P,i}, i = n_{2,T} + 1, n_{2,T} + 2, \dots, n_2\}$  are placebo non-responders that have been re-randomized in Period 2 to placebo, while the remainder of the placebo sample  $\{X_{1,P,i}, i = n_2 + 1, n_2 + 2, \dots, n_{1,P}\}$  are the placebo subjects who were placebo responders (or placebo dropouts if any, although it is assumed none here) in Period 1. Note that under equal allocation in Period 2,  $n_{2,P} = n_{2,T} = (n_{2,P} + n_{2,T})/2 = n_2/2 = \gamma n_{1,P}/2$ .

Assuming that the randomization in Period 1 holds, the placebo sample should be representative of the population  $\Omega = \Omega_1$ . If the entire placebo sample at the end of Period 1 were re-randomized in Period 2 to treatment, then the pair of response variables  $(X_{1,P}, X_{2,T})$  should follow a bivariate normal distribution  $(X_{1,P}, X_{2,T}) \sim N(\mu_{12,T}, \Sigma_{12,T})$ , where  $\mu_{12,T} = \begin{pmatrix} \mu_{1,P} \\ \mu_{2,T} \end{pmatrix}$  and  $\Sigma_{12,T} = \begin{pmatrix} \sigma_{1,P}^2 & \rho_T \sigma_{1,P} \sigma_{2,T} \\ \rho_T \sigma_{1,P} \sigma_{2,T} & \sigma_{2,T}^2 \end{pmatrix}$  upon assuming that  $\sigma_{1,P}^2 = \sigma_{1,T}^2 = \sigma_1^2$ ,  $\sigma_{2,P}^2 = \sigma_{2,T}^2 = \sigma_2^2$ , and  $\rho_T$  is the correlation  $corr(X_{1,P}, X_{2,T})$ , where  $X_{2,T}$  is the response variable in Period 2 under the treatment T. Similarly, if the entire placebo sample at the end of Period 1 were re-randomized in Period 2 to placebo, then the pair of response variables  $(X_{1,P}, X_{2,P})$  should follow a bivariate normal distribution  $(X_{1,P}, X_{2,P}) \sim N(\mu_{12,P}, \Sigma_{12,P})$ , where  $\mu_{12,P} = \begin{pmatrix} \mu_{1,P} \\ \mu_{2,P} \end{pmatrix}$  and  $\Sigma_{12,P} = \begin{pmatrix} \sigma_{1,P}^2 & \rho_P \sigma_{1,P} \sigma_{2,P} \\ \rho_P \sigma_{1,P} \sigma_{2,P} & \sigma_{2,P}^2 \end{pmatrix}$  upon assuming that  $\sigma_{1,P}^2 = \sigma_{1,T}^2 = \sigma_1^2$ ,  $\sigma_{2,P}^2 = \sigma_{2,T}^2 = \sigma_2^2$ , and  $\rho_P$  is the correlation  $corr(X_{1,P}, X_{2,P})$ , where  $X_{2,P}$  is the response variable in Period 2 under the placebo P. Indeed, in this case, one may even assume that  $\sigma_{1,P}^2 = \sigma_{2,P}^2$ ,  $\sigma_{1,T}^2 = \sigma_{2,T}^2$  and hence  $\sigma_1^2 = \sigma_2^2$ . It should be pointed out that if the treatment is not effective, then it is likely that  $\rho_P = \rho_T$ , i.e.,  $\rho_P - \rho_T = 0$ . Otherwise, if the treatment is more effective than placebo, then one should expect that  $\rho_P \geq \rho_T$ , i.e.,  $\rho_P - \rho_T \geq 0$ .

### 3.1. Truncated distributions of the two placebo non-responder cohorts in period 2

However, in a DRDS design, since only the placebo non-responders at the end of Period 1 are re-randomized to placebo

and treatment in Period 2. Therefore, for the cohort of placebo non-responders who were re-randomized to treatment in Period 2 denoted by  $(P \rightarrow T)$ , the sample pairs  $\{(X_{1,P,i}, X_{2,T,i}), i = 1, 2, \dots, n_{2,T}\}$  would follow a singly truncated bivariate normal distribution

$$((X_{1,P}|X_{1,P} < c), (X_{2,T}|X_{1,P} < c)) \sim N(\mu_{12,T|X_{1,P} < c}, \Sigma_{12,T|X_{1,P} < c})$$

where

$$\mu_{12,T|X_{1,P} < c} = \begin{pmatrix} \mu_{1,P} - \sigma_{1,P} \left( \frac{\varphi(\tau)}{\Phi(\tau)} \right) \\ \mu_{2,T} - \rho_T \sigma_{2,T} \left( \frac{\varphi(\tau)}{\Phi(\tau)} \right) \end{pmatrix}$$

$$\Sigma_{(12,T|X_{1,P} < c)} = \begin{pmatrix} \text{var}(X_{1,P}|X_{1,P} < c) & \text{cov}(X_{1,P}, X_{2,T}|X_{1,P} < c) \\ \text{cov}(X_{1,P}, X_{2,T}|X_{1,P} < c) & \text{var}(X_{2,T}|X_{1,P} < c) \end{pmatrix}$$

$$= \begin{pmatrix} \sigma_{1,P}^2 & \rho_{(T|X_{1,P} < c)} \sigma_{(1,P|X_{1,P} < c)} \sigma_{(2,T|X_{1,P} < c)} \\ \rho_{(T|X_{1,P} < c)} \sigma_{(1,P|X_{1,P} < c)} \sigma_{(2,T|X_{1,P} < c)} & \sigma_{(2,T|X_{1,P} < c)}^2 \end{pmatrix}$$

where the elements of the variance-covariance matrix are given by

$$\sigma_{1,P|X_{1,P} < c}^2 = \text{var}(X_{1,P}|X_{1,P} < c) = \left[ 1 - \tau \frac{\varphi(\tau)}{\Phi(\tau)} - \left( \frac{\varphi(\tau)}{\Phi(\tau)} \right)^2 \right] \sigma_{1,P}^2$$

$$\sigma_{2,T|X_{1,P} < c}^2 = \text{var}(X_{2,T}|X_{1,P} < c)$$

$$= \left( \rho_T^2 \left[ 1 - \tau \frac{\varphi(\tau)}{\Phi(\tau)} - \left( \frac{\varphi(\tau)}{\Phi(\tau)} \right)^2 \right] \sigma_{1,P}^2 + (1 - \rho_T^2) \right) \sigma_{2,T}^2$$

$$\sigma_{2,P|X_{1,P} < c}^2 = \text{var}(X_{2,P}|X_{1,P} < c)$$

$$= \left( \rho_P^2 \left[ 1 - \tau \frac{\varphi(\tau)}{\Phi(\tau)} - \left( \frac{\varphi(\tau)}{\Phi(\tau)} \right)^2 \right] \sigma_{1,P}^2 + (1 - \rho_P^2) \right) \sigma_{2,P}^2$$

$$\text{cov}(X_{1,P}, X_{2,T}|X_{1,P} < c) = \rho_T \left[ 1 - \tau \frac{\varphi(\tau)}{\Phi(\tau)} - \left( \frac{\varphi(\tau)}{\Phi(\tau)} \right)^2 \right] \sigma_{1,P} \sigma_{2,T}$$

and  $\rho_{T|X_{1,P} < c}$  is the correlation for the truncated  $(P \rightarrow T)$  cohort given by

$$\rho_{T|X_{1,P} < c} = \frac{\text{cov}(X_{1,P}, X_{2,T}|X_{1,P} < c)}{\sqrt{\text{var}(X_{1,P}|X_{1,P} < c)} \sqrt{\text{var}(X_{2,T}|X_{1,P} < c)}}$$

$$= \frac{\rho_T}{\sqrt{\left( \rho_T^2 \sigma_{1,P}^2 + \left[ 1 - \tau \frac{\varphi(\tau)}{\Phi(\tau)} - \left( \frac{\varphi(\tau)}{\Phi(\tau)} \right)^2 \right] \right)}}$$

Now in practice, the variances  $\text{var}(X_{1,P}|X_{1,P} < c)$ ,  $\text{var}(X_{2,T}|X_{1,P} < c)$  and the  $\text{cov}(X_{1,P}, X_{2,T}|X_{1,P} < c)$  may be estimated by their respective sample variances and the sample covariance given by

$$S_{X_{1,P}|X_{1,P} < c}^2 = \frac{1}{n_{2,T} - 1} \sum_{i=1}^{n_{2,T}} (X_{1,P,i} - \widehat{X}_{(1,P|X_{1,P} < c)})^2$$

$$S_{X_{2,T}|X_{1,P} < c}^2 = \frac{1}{n_{2,T} - 1} \sum_{i=1}^{n_{2,T}} (X_{2,T,i} - \widehat{X}_{(2,T|X_{1,P} < c)})^2 \text{ and}$$

$$S_{(X_{1,P}, X_{2,T}|X_{1,P} < c)} = \frac{1}{n_{2,T} - 1} \sum_{i=1}^{n_{2,T}} X_{1,P,i}^* X_{2,T,i}^*, \text{ where}$$

$$X_{1,P,i}^* = X_{1,P,i} - \widehat{X}_{(1,P|X_{1,P} < c)}, X_{2,T,i}^* = X_{2,T,i} - \widehat{X}_{(2,T|X_{1,P} < c)} \text{ and}$$

$$\widehat{X}_{1,P|X_{1,P} < c} = \frac{1}{n_{2,T}} \sum_{i=1}^{n_{2,T}} (X_{1,P,i}|X_{1,P} < c) \text{ and}$$

$$\widehat{X}_{2,T|X_{1,P} < c} = \frac{1}{n_{2,T}} \sum_{i=1}^{n_{2,T}} (X_{2,T,i}|X_{1,P} < c).$$

The sample correlation is given by  $\widehat{\rho}_{T|X_{1,P} < c} = S_{(X_{1,P}, X_{2,T}|X_{1,P} < c)} / \sqrt{S_{X_{1,P}|X_{1,P} < c}^2 S_{X_{2,T}|X_{1,P} < c}^2}$ .

Similarly, for the cohort of placebo non-responders who are re-randomized to placebo, denoted by,  $(P \rightarrow P)$  in Period 2, the sample pairs  $\{(X_{1,P,n_{2,T}+i}, X_{2,P,i}), i = 1, 2, \dots, n_{2,P}\}$  also follows a singly truncated bivariate normal distribution with

$$((X_{1,P}|X_{1,P} < c), (X_{2,P}|X_{1,P} < c)) \sim N(\mu_{12,P|X_{1,P} < c}, \Sigma_{12,P|X_{1,P} < c})$$

where

$$\mu_{12,P|X_{1,P} < c} = \begin{pmatrix} \mu_{1,P} - \sigma_{1,P} \left( \frac{\varphi(\tau)}{\Phi(\tau)} \right) \\ \mu_{2,P} - \rho_P \sigma_{2,P} \left( \frac{\varphi(\tau)}{\Phi(\tau)} \right) \end{pmatrix}$$

$$\Sigma_{12,P|X_{1,P} < c} = \begin{pmatrix} \text{var}(X_{1,P}|X_{1,P} < c) & \text{cov}(X_{1,P}, X_{2,P}|X_{1,P} < c) \\ \text{cov}(X_{1,P}, X_{2,P}|X_{1,P} < c) & \text{var}(X_{2,P}|X_{1,P} < c) \end{pmatrix}$$

the expressions for the elements of the above variance-covariance matrix  $\Sigma_{12,P|X_{1,P} < c}$  are similar to the previous expressions derived for the  $(P \rightarrow T)$  cohort and will not be repeated here.

Now, with the underlying conditional probability structure for a DRDS design as described above, the Period 2 expected treatment effect is now given by the conditional (truncated) mean difference

$$(\Delta_2|X_{1,P} < c) = \mu_{2,T|X_{1,P} < c} - \mu_{2,P|X_{1,P} < c} = \left[ \mu_{2,T} - \rho_T \sigma_{2,T} \left( \frac{\varphi(\tau)}{\Phi(\tau)} \right) \right]$$

$$- \left[ \mu_{2,P} - \rho_P \sigma_{2,P} \left( \frac{\varphi(\tau)}{\Phi(\tau)} \right) \right]$$

$$= (\mu_{2,T} - \mu_{2,P}) + (\rho_P \sigma_{2,P} - \rho_T \sigma_{2,T}) \left( \frac{\varphi(\tau)}{\Phi(\tau)} \right) \tag{1}$$

which may be estimated by the observed mean difference given by

$$(\widehat{\Delta}_2|X_{1,P} < c) = \widehat{\mu}_{2,T|X_{1,P} < c} - \widehat{\mu}_{2,P|X_{1,P} < c}$$

where

$$\widehat{\mu}_{2,T} | X_{1,P} < c = \widehat{X}_{2,T} | X_{1,P} < c = \frac{1}{n_{2,T}} \sum_{i=1}^{n_{2,T}=n_{2,P}} (X_{2,T,i} | X_{1,P} < c)$$

and

$$\widehat{\mu}_{2,P} | X_{1,P} < c = \widehat{X}_{2,P} | X_{1,P} < c = \frac{1}{n_{2,P}} \sum_{i=1}^{n_{2,P}=n_{2,T}} (X_{2,P,i} | X_{1,P} < c)$$

Thus,

$$\begin{aligned} E(\widehat{\Delta}_2 | X_{1,P} < c) &= E(\widehat{\mu}_{2,T} | X_{1,P} < c - \widehat{\mu}_{2,P} | X_{1,P} < c) = (\Delta_2 | X_{1,P} < c) \\ &= (\mu_{2,T} - \mu_{2,P}) + (\rho_P \sigma_{2,P} - \rho_T \sigma_{2,T}) \frac{\varphi(\tau)}{\Phi(\tau)} \end{aligned}$$

Note that in the above expression for  $E(\widehat{\Delta}_2 | X_{1,P} < c)$  or Eq. (1), if the duration of Period 1 is relatively short, then the first term  $(\mu_{2,T} - \mu_{2,P}) = (\mu_{1,T} - \mu_{1,P})$  which is the apparent treatment effect from Period 1, and hence the increase in the expected treatment effect in Period 2 would come from the second term  $(\rho_P \sigma_{2,P} - \rho_T \sigma_{2,T}) \varphi(\tau) / \Phi(\tau)$  which is 0 when there is no treatment effect and should be positive when the treatment is effective, since in that case, one expects that  $(\rho_P \sigma_{2,P} - \rho_T \sigma_{2,T}) > 0$ . Eq. (1) will be important as will be seen later.

Some of the above expressions are well-known (see e.g., Johnson and Kotz [9], Gajjar and Subrahmaniam [10], Rosenbaum [11], Shah and Parikh [12] and Tallis [13]) and others can be derived from them.

### 3.2. The joint distribution of $(\widehat{\Delta}_1, (\widehat{\Delta}_2 | X_{1,P} < c))$

Now with the above derivation of the expressions for the various distribution parameters for the conditional distributions as a function of the distribution parameters of their underlying unconditional distributions for the two cohorts ( $P \rightarrow T$ ) and ( $P \rightarrow P$ ) in Period 2, one can establish the following lemma within the framework of a DRDS design.

**Lemma.** For a DRDS design, the treatment effect estimates  $\widehat{\Delta}_1$  and  $(\widehat{\Delta}_2 | X_{1,P} < c)$  from Period 1 and Period 2 follow an asymptotically normal bivariate distribution  $(\widehat{\Delta}_1, (\widehat{\Delta}_2 | X_{1,P} < c)) \sim \Phi(\mu_{12}, \Sigma_{12})$ , where the means are given by

$$\begin{aligned} \mu_{12} &= \begin{pmatrix} \Delta_1 \\ (\Delta_2 | X_{1,P} < c) \end{pmatrix} = \begin{pmatrix} \mu_{1,T} - \mu_{1,P} \\ \mu_{2,T} | X_{1,P} < c - \mu_{2,P} | X_{1,P} < c \end{pmatrix} \\ &= \begin{pmatrix} \mu_{1,T} - \mu_{1,P} \\ (\mu_{2,T} - \mu_{2,P}) + (\rho_P \sigma_{2,P} - \rho_T \sigma_{2,T}) \frac{\varphi(\tau)}{\Phi(\tau)} \end{pmatrix} \end{aligned}$$

and the variance-covariance matrix is given by

$$\Sigma_{12} = \begin{pmatrix} \text{var}(\widehat{\Delta}_1) & \text{cov}(\widehat{\Delta}_1, (\widehat{\Delta}_2 | X_{1,P} < c)) \\ \text{cov}(\widehat{\Delta}_1, (\widehat{\Delta}_2 | X_{1,P} < c)) & \text{var}(\widehat{\Delta}_2 | X_{1,P} < c) \end{pmatrix}$$

where

$$\text{var}(\widehat{\Delta}_1) = \frac{\sigma_1^2}{n_{1,T} R_1}, \text{ assuming that } \sigma_{1,T}^2 = \sigma_{1,P}^2 = \sigma_1^2 \quad (2)$$

$$\begin{aligned} \text{var}(\widehat{\Delta}_2 | X_{1,P} < c) &= \text{var}(\widehat{\mu}_{2,T} | X_{1,P} < c - \widehat{\mu}_{2,P} | X_{1,P} < c) \\ &= \frac{1}{n_{2,T}} (\text{var}(X_{2,T} | X_{1,P} < c) + \text{var}(X_{2,P} | X_{1,P} < c)) \end{aligned} \quad (3)$$

where

$$\begin{aligned} \text{var}(X_{2,T} | X_{1,P} < c) &= \left( \rho_T^2 \left[ 1 - \tau \frac{\varphi(\tau)}{\Phi(\tau)} - \left( \frac{\varphi(\tau)}{\Phi(\tau)} \right)^2 \right] \sigma_{1,P}^2 \right. \\ &\quad \left. + (1 - \rho_T^2) \right) \sigma_{2,T}^2 \end{aligned} \quad (4)$$

$$\begin{aligned} \text{var}(X_{2,P} | X_{1,P} < c) &= \left( \rho_P^2 \left[ 1 - \tau \frac{\varphi(\tau)}{\Phi(\tau)} - \left( \frac{\varphi(\tau)}{\Phi(\tau)} \right)^2 \right] \sigma_{1,P}^2 \right. \\ &\quad \left. + (1 - \rho_P^2) \right) \sigma_{2,P}^2 \end{aligned} \quad (5)$$

and

$$\begin{aligned} \text{cov}(\widehat{\Delta}_1, (\widehat{\Delta}_2 | X_{1,P} < c)) &= \text{cov}((\widehat{\mu}_{1,T} - \widehat{\mu}_{1,P}), (\widehat{\mu}_{2,T} | X_{1,P} < c \\ &\quad - \widehat{\mu}_{2,P} | X_{1,P} < c)) \\ &= \frac{1}{n_{1,P}} (\text{cov}(X_{1,P}, X_{2,P} | X_{1,P} < c) \\ &\quad - \text{cov}(X_{1,P}, X_{2,T} | X_{1,P} < c)) \rightarrow 0 \end{aligned} \quad (6)$$

asymptotically where the covariance terms  $\text{cov}(X_{1,P}, X_{2,P} | X_{1,P} < c)$  and  $\text{cov}(X_{1,P}, X_{2,T} | X_{1,P} < c)$  are as given previously.

The proof of this lemma will be omitted since these expressions can be directly derived from the preceding conditional distribution parameters for the two cohorts ( $P \rightarrow P$ ) and ( $P \rightarrow T$ ).

Note that for the conditional (truncated) variances and covariance, one can use their sample variance and covariance as estimates.

### 3.3. An example of a DRDS design

Table 1 displays a summary of the data from a very small completed phase II study based on a DRDS design as described in Fig. 1. Using the conditional probability structure described above, the data from Period 1 of this study will be used later as the basis for illustrating the proposed method with a simulated trial using a DRDS design. In addition, selected power and sample size calculations for the combination and consistency tests will also be based on the data from this table.

## 4. The adjusted treatment effect

### 4.1. The reason for adjusting the apparent treatment effect $\Delta_1$

In a trial with high placebo response rate, the first problem encountered is the inability to characterize the subpopulation of placebo responders  $\Omega_R$ . Therefore, if a traditional randomized parallel design is used, such as the first period of a DRDS design, then the high placebo response rate in the intended study population  $\Omega = \Omega_1$  would obviously reduce the treatment effect because it is measured as a relative difference  $\Delta_1 = \mu_{1,T} - \mu_{1,P}$  between the

**Table 1**  
Hypothetical Distributions of a HDRS<sub>17</sub> Subscale Score based on an Early Phase 2 Major Depressive Disorder Trial using a DRDS Design with Parameter Values:  $r_1 = 2, \pi = 0.60, \gamma = 0.40, r_2 = 1$ .

Period 1					
$\mu_{1T}$	$\sigma_{1T}$	$\mu_{1P}$	$\sigma_{1P}$	$\Delta_1$	$\sigma_1$
3.30	2.44	3.00	2.40	0.30	2.42
Period 2					
$\mu_{2T}$	$\sigma_{2T}$	$\mu_{2P}$	$\sigma_{2P}$	$\Delta_2$	$\sigma_2$
3.90	1.95	2.80	2.00	1.10	1.98

treatment and placebo groups, a problem that is all too familiar in an active control trial. If placebo responders are present in substantial proportion, then this relative difference will become much smaller. This reduced treatment effect termed the *apparent treatment effect* in a parallel design is the reason why many such trials had failed in the past.

To further elaborate on this problem, assume for the moment that one is able to characterize the placebo responders  $\Omega_R$  and the placebo non-responders  $\Omega_{NR}$  relative to a response variable  $X \sim N(\mu, \sigma^2)$  and a response threshold  $c$ , where larger values of the response variable  $X$  represent better outcomes. Let  $\tau = (c - \mu)/\sigma$ , then  $\alpha_{NR} = \Phi(\tau)$  would be the proportion of placebo non-responders in  $\Omega = \Omega_1$ . Let  $X_{R,T} \sim N(\mu_{R,T}, \sigma_{R,T}^2)$  and  $X_{R,P} \sim N(\mu_{R,P}, \sigma_{R,P}^2)$  denote the response distribution for treatment  $T$  and placebo  $P$  respectively in  $\Omega_R$ , and  $X_{NR,T} \sim N(\mu_{NR,T}, \sigma_{NR,T}^2)$  and  $X_{NR,P} \sim N(\mu_{NR,P}, \sigma_{NR,P}^2)$  denote the response distribution for treatment  $T$  and placebo  $P$  respectively in  $\Omega_{NR}$ . Furthermore, let  $\Delta_R = \mu_{R,T} - \mu_{R,P}$  and  $\Delta_{NR} = \mu_{NR,T} - \mu_{NR,P}$  denote the respective treatment effects in  $\Omega_R$  and  $\Omega_{NR}$ . Under homogeneity, the apparent treatment effect  $\Delta_1$  in Period 1 of a DRDS design can be defined as a simple weighted average of  $\Delta_R$  and  $\Delta_{NR}$  given by  $\Delta_1 = \alpha_R \Delta_R + \alpha_{NR} \Delta_{NR}$ . Clearly, when the proportion of placebo responders  $\alpha_R$  is low, then the apparent treatment effect  $\Delta_1$  is close to  $\Delta_{NR}$  and the impact of  $\Delta_R$  would be small. On the other hand, when the placebo response rate  $\alpha_R$  is relatively high, then the impact of  $\Delta_R$  would be great on the apparent treatment effect  $\Delta_1$ . In this latter case, the apparent treatment effect  $\Delta_R$  due to the placebo response in  $\Omega_R$  results in the apparent treatment effect  $\Delta_1$ . Therefore, this suggests that one should adjust the weights  $\alpha_R$  and  $\alpha_{NR}$  in  $\Delta_1 = \alpha_R \Delta_R + \alpha_{NR} \Delta_{NR}$  in an objective manner to account for the high placebo response rate in  $\Omega_R$  which is reflected in the apparent treatment effect  $\Delta_R$ . In the next section, an adjusted treatment effect is defined which represents an adjustment of the weights in  $\Delta_1 = \alpha_R \Delta_R + \alpha_{NR} \Delta_{NR}$  to account for the impact of the presence of placebo responders in  $\Omega = \Omega_1$ .

4.2. An adjusted treatment effect

Recall that for simplicity and without loss in generality, one may assume that  $\sigma_{1,P}^2 = \sigma_{1,T}^2$  which is also suggested by the first period data in the example given in Table 1. Denote this common variance by  $\sigma_1^2$ , and hence  $\sigma_{\Delta_1}^2 = \sigma_1^2 / (n_{1,T}R_1)$ . Similarly, one may assume that in Period 2, the conditional variances are equal, i.e.,  $\sigma_{2,T|X_{1,P} < c}^2 = \text{var}(X_{2,T}|X_{1,P} < c) = \sigma_{2,P|X_{1,P} < c}^2 = \text{var}(X_{2,P}|X_{1,P} < c) = \sigma_2^2$ , which is also suggested by the data in the example given in Table 1, although it was not assumed to be so in the earlier expression for  $\sigma_{\Delta_2|X_{1,P} < c}^2$ , and hence here one has  $\sigma_{\Delta_2|X_{1,P} < c}^2 = \sigma_2^2 / (n_{2,T}R_2)$ . If

one were to combine the treatment effect estimate  $\hat{\Delta}_1$  from Period 1 and  $(\hat{\Delta}_2|X_{1,P} < c) = (\hat{\mu}_{2,T}|X_{1,P} < c - \hat{\mu}_{2,P}|X_{1,P} < c)$  from Period 2

using weights defined through their inverse variances following the method of weighted least square [14], then the least square estimator of the treatment effect is given by

$$\hat{\Delta} = \alpha_1 \hat{\Delta}_1 + \alpha_2 (\hat{\Delta}_2|X_{1,P} < c) \tag{7}$$

where the coefficients  $\alpha_1$  and  $\alpha_2$  are given in general by

$$\alpha_1 = 1 - \alpha_2$$

where

$$\alpha_2 = \frac{\frac{\sigma_1^2}{n_{1,T}R_1} - \text{cov}(\hat{\Delta}_1, (\hat{\Delta}_2|X_{1,P} < c))}{\frac{\sigma_1^2}{n_{1,T}R_1} + \frac{\sigma_2^2}{n_{2,T}R_2} - 2\text{cov}(\hat{\Delta}_1, (\hat{\Delta}_2|X_{1,P} < c))} \tag{8}$$

Now, since  $\text{cov}(\hat{\Delta}_1, (\hat{\Delta}_2|X_{1,P} < c)) \rightarrow 0$  asymptotically as noted earlier, hence under large sample,  $\alpha_2$  may be approximately given by

$$\alpha_2 = \frac{\frac{\sigma_1^2}{n_{1,T}R_1}}{\frac{\sigma_1^2}{n_{1,T}R_1} + \frac{\sigma_2^2}{n_{2,T}R_2}} = \frac{\frac{n_{1,T}\gamma R_{12}R_2}{\sigma_2^2}}{\frac{n_{1,T}R_1}{\sigma_1^2} + \frac{n_{1,T}\gamma R_{12}R_2}{\sigma_2^2}} = \frac{1}{1 + \left(\frac{\sigma_2}{\sigma_1}\right)^2 \frac{1}{\gamma} \left(\frac{R_1}{R_{12}R_2}\right)}$$

where under a DRDS design,  $n_{2,T} = n_{1,T}\gamma R_{12}$  and  $\gamma = \Phi(\tau)$  is the population proportion of placebo non-responders which can be consistently estimated by the fraction of placebo subjects remained at the end of Period 1 who are placebo non-responders exclusive of the placebo responders and placebo dropouts and under the previous assumptions.

Now, in a DRDS design, for practical reasons, the following restrictions on the allocation ratios are expected  $1 \leq r_2 \leq r_1$ . Hence, based on this restriction, the ratio  $(R_1/(R_{12}R_2))$  in the above expression for  $\alpha_2$  achieves its maximum value of 2 which is the value actually attained under the case of equal allocations, when  $r_1 = r_2 = 1$ .

Therefore, one can define

$$\alpha_2 = \frac{1}{1 + \left(\frac{\sigma_2}{\sigma_1}\right)^2 \frac{2}{\gamma}} \text{ and } \alpha_1 = 1 - \alpha_2 \tag{9}$$

which will minimize the weight placed on  $\Delta_2|X_{1,P} < c$ , the expected treatment effect from Period 2.

The coefficients in Eq. (9) are the weights that will be used to define the adjusted treatment effect in the following definition.

**Definition 1:** Under a DRDS design, the *adjusted treatment effect* is defined as the convex combination

$$\hat{\Delta} = \alpha_1 \Delta_1 + \alpha_2 (\Delta_2|X_{1,P} < c) \tag{10}$$

where the coefficients  $\alpha_1 = \alpha_1(\gamma, \sigma_1, \sigma_2)$  and  $\alpha_2 = \alpha_2(\gamma, \sigma_1, \sigma_2)$  are as defined in Eq. (9).

Note that the weights as defined above assumes that  $cov(\widehat{\Delta}_1, (\widehat{\Delta}_2 | X_{1,p} < c)) \rightarrow 0$  and it would be valid in a DRDS design with the conditional probability structure discussed above under large sample as shown in the Lemma. However, for small samples, the weights may not be appropriate and the combined statistics as defined may not be valid and should be interpreted with caution, particularly when the covariance  $cov(\widehat{\Delta}_1, (\widehat{\Delta}_2 | X_{1,p} < c))$  is negative suggesting that the Period 1 apparent treatment effect and the Period 2 treatment effect are not consistent. This inconsistency will be discussed later under a consistency condition to be introduced.

In addition, the weights defined in Eq. (9) for the adjusted treatment effect as defined in Eq. (10) are dependent on the population parameters  $\gamma$ ,  $\sigma_1$  and  $\sigma_2$ , but they are independent of any design parameters particularly the allocation ratios  $r_1$  and  $r_2$ . This is important because if the weights are dependent on the allocation ratios, then one can easily bias the results in favor of the treatment by increasing the allocation ratio  $r_1$  and thus placing greater and greater weights on the Period 2 results. In fact, when the weights are dependent on the allocation ratios, the combined statistic will provide an estimate that is biased in favor of the treatment *even when*  $r_1 = 2$  and  $r_2 = 1$  which are the allocation ratios used in the SPD design of Fava et al. [3] and the DRDS design of Liu et al. [1].

**Remark 1:** It is important to emphasize again that the adjusted treatment effect is independent of the allocation ratios in the class of DRDS designs that are subject to the restriction  $1 \leq r_2 \leq r_1$ . More importantly, the coefficient  $\alpha_2$  represents the smallest possible weight assigned to  $\Delta_2$  under a DRDS design subject to the above restriction and  $\alpha_2$  is actually attained under the case of a DRDS design with equal allocation. Also, with  $\alpha_2$  so defined, the actual DRDS design can still assume allocation ratios other than equal allocation provided the allocation ratios satisfy the above restriction. Therefore, with the weights  $\alpha_1$  and  $\alpha_2$  as defined in Eq. (9), there is no possibility for a DRDS design that is subject to the above allocation ratio restriction to introduce bias into the adjusted treatment effect by over-weighting the treatment effect ( $\Delta_2 | X_{1,p} < c$ ) from the enriched subpopulation of the placebo non-responders from Period 2 by increasing the allocation ratio  $r_1$  in favor of placebo in Period 1 and thereby overweighting the Period 2 results. Even though the coefficient  $\alpha_2$  is the weight actually attained under equal allocations  $r_1 = r_2 = 1$  which does not involve overweighting the Period 2 results, it would be an unlikely configuration to be adopted in practical applications. Thus, if a given DRDS design adopts an allocation ratio  $r_1 > 1$ , it will only improve the precision of estimates, but will not affect the estimate of the adjusted treatment effect as defined in Eq. (10) above.

The weights used in the current combined statistics are implicitly dependent on the allocation ratios, although they are not noted as such. However, Tamura et al. [8] discussed the combined statistic with a view to estimating the treatment effect. But the authors' combined statistic is actually defined as an estimate of the apparent treatment effect  $\Delta_1$  which is not solving the basic problem at hand. Furthermore, the authors prefer weights that are dependent on the allocation ratios which are clearly not appropriate. Therefore, with any allocation  $r_1 > 1$ , these combined statistics would tend to bias the results in favor of the treatment by placing more weight on ( $\Delta_2 | X_{1,p} < c$ ) from Period 2.

**Remark 2:** The weights defined in Eq. (9) for the adjusted treatment effect as defined in Eq. (10) are independent of the allocation ratios  $r_1$  and  $r_2$  as long as they satisfy the constraint  $1 \leq r_2 \leq r_1$ . This property allows one to freely choose a DRDS design with any allocation ratios  $r_1$  and  $r_2$  as long as they satisfy the constraint  $1 \leq r_2 \leq r_1$ . This flexibility has a very interesting, unintended and useful property in assuring type I error control of the joint test which will be discussed in Section 7.2.

**Note:** It is important to point out that the combined statistic as given in Eq. (7) will not necessarily retain the efficiency property of a least square estimator in light of the weights as defined in Eq. (9) unless it is a DRDS design with equal allocation ratios. But this may be the trade-off that one has to consider if one wishes to be able to define an adjusted treatment effect where the weights are independent of the DRDS design parameters, particularly, the allocation ratios, so that the adjusted treatment effect is not biased in favor of the treatment by placing more weights on the enriched treatment effect from Period 2. This latter seems to be a more important issue than optimal efficiency consideration, because an appropriate definition of adjusted treatment effect is critical and would allow a proper assessment of the treatment effect for the intended study population.

#### 4.3. Interpretation of the adjusted treatment effect

As noted earlier, if one were able to characterize the subpopulation  $\Omega_R$  of placebo responders and the subpopulation  $\Omega_{NR}$  of placebo non-responders, then for the overall study population  $\Omega$  in Period 1 of a DRDS design, the overall apparent treatment effect  $\Delta_1$  can be expressed as

$$\Delta_1 = \alpha_R \Delta_R + \alpha_{NR} \Delta_{NR} \quad (11)$$

Then, the adjusted treatment effect given by Eq. (10) becomes

$$\begin{aligned} \Delta &= \alpha_1 \Delta_1 + \alpha_2 (\Delta_2 | X_{1,p} < c) \\ &= \alpha_1 [\alpha_R \Delta_R + \alpha_{NR} \Delta_{NR}] + \alpha_2 (\Delta_2 | X_{1,p} < c) \end{aligned}$$

under the assumption that the distribution of the placebo responders/non-responders among the placebo dropouts, if any, is the same as its population distribution, which implies that  $(\Delta_2 | X_{1,p} < c) \equiv \Delta_{NR}$ . Hence, it follows that

$$\begin{aligned} \Delta &\equiv \alpha_1 \alpha_R \Delta_R + (\alpha_2 + \alpha_1 \alpha_{NR}) \Delta_{NR} \\ &\equiv (1 - \alpha_2) \alpha_R \Delta_R + (\alpha_2 + (1 - \alpha_2) \alpha_{NR}) \Delta_{NR}, \text{ since } \alpha_1 = (1 - \alpha_2) \\ &\equiv (\alpha_R - \alpha_2 \alpha_R) \Delta_R + (\alpha_{NR} + \alpha_2 (1 - \alpha_{NR})) \Delta_{NR} \\ &\equiv (\alpha_R - \alpha_2 \alpha_R) \Delta_R + (\alpha_{NR} + \alpha_2 \alpha_R) \Delta_{NR}, \text{ since } \alpha_R = 1 - \alpha_{NR} \end{aligned} \quad (12)$$

Upon comparing Eq. (11) and Eq. (12), one notes that the adjusted treatment effect  $\Delta$  as defined in Eq. (10) can be viewed as a weighted average of  $\Delta_R$  and  $\Delta_{NR}$  as in Eq. (11) for  $\Delta_1$  except the weights now have been changed in the following manner: The weight for  $\Delta_R$  has been decreased by the fractional amount  $\alpha_2 \alpha_R$  while the weight for  $\Delta_{NR}$  has been increased by the same fractional amount  $\alpha_2 \alpha_R$ . Therefore, Eq. (12) shows that the adjusted treatment effect  $\Delta$  can be viewed as a weighted average of the treatment effect  $\Delta_R$  and  $\Delta_{NR}$  and hence represents a treatment effect for the intended MDD study population  $\Omega = \Omega_1$ . The fraction  $\alpha_2 \alpha_R$  represents the amount of adjustment needed to account for the presence of placebo responders  $\Omega_R$  in  $\Omega = \Omega_1$ .

On the other hand, Eq. (12) can also be rearranged as follows:

$$\begin{aligned} \Delta &\equiv (\alpha_R - \alpha_2 \alpha_R) \Delta_R + (\alpha_{NR} + \alpha_2 \alpha_R) \Delta_{NR} \\ &\equiv (\alpha_R \Delta_R + \alpha_{NR} \Delta_{NR}) + \alpha_2 [\alpha_R (\Delta_{NR} - \Delta_R)] \\ &\equiv \Delta_1 + \alpha_2 [\alpha_R (\Delta_{NR} - \Delta_R)] \end{aligned} \quad (13)$$

Now from Eq. (13), one can see that if there are no placebo responders, i.e.,  $\Omega_R = \emptyset$ , then  $\alpha_R = 0$  and  $\Delta = \Delta_1$ . That is, the adjusted treatment effect  $\Delta$  and the apparent treatment effect  $\Delta_1$  are identical and hence no adjustment is really needed.

Now if  $\Omega_R \neq \emptyset$ , then it is expected that  $\Delta_{NR} > \Delta_R$ . In this case, then  $[\alpha_R (\Delta_{NR} - \Delta_R)]$  represents the total amount of expected



treatment effect  $\Delta_{NR}$  that is not observed due to the placebo response in  $\Omega_R$ . Now, because  $\Delta_{NR} = \Delta_2$ , one can view  $[\alpha_R(\Delta_{NR} - \Delta_R)] = [\alpha_R(\Delta_2 - \Delta_R)]$  as the equivalent amount of treatment effect from Period 2 that has been nullified by the placebo response in  $\Omega_R$ . Then, it follows that  $\alpha_2[\alpha_R(\Delta_2 - \Delta_R)]$  represents the appropriately weighted amount of  $[\alpha_R(\Delta_2 - \Delta_R)]$  from Period 2 that needs to be added to the apparent treatment effect  $\Delta_1$  from Period 1 to account for the presence of placebo responders  $\Omega_R$ . Hence, the quantity  $\alpha_2[\alpha_R(\Delta_{NR} - \Delta_R)]$  represents the appropriate *adjustment* that needs to be made to the apparent treatment effect  $\Delta_1$  to account for the presence of placebo responders.

**5. The combination test**

For a DRDS design, under large sample, consider the adjusted treatment effect  $\Delta = \alpha_1\Delta_1 + \alpha_2(\Delta_2|X_{1,P} < c)$  as given in Definition 1 above. The adjusted treatment null hypothesis and its alternative are defined as follows:

$$H_{0,Adj} : \Delta = \alpha_1\Delta_1 + \alpha_2(\Delta_2|X_{1,P} < c) \leq 0 \quad vs. \quad H_{a,Adj} : \Delta = \alpha_1\Delta_1 + \alpha_2(\Delta_2|X_{1,P} < c) > 0 \quad (14)$$

It should be pointed out that the above adjusted null hypothesis is a stronger null hypothesis than the global null hypothesis defined by  $\{(\Delta_1, (\Delta_2|X_{1,P} < c)) | \Delta_1 \leq 0 \ \& \ (\Delta_2|X_{1,P} < c) \leq 0\}$ , because the parameter space defined by the adjusted null is a half-space in the product space  $\Delta_1 \times (\Delta_2|X_{1,P} < c)$  below a straight line that goes through the origin (0,0) defined by  $\alpha_1\Delta_1 + \alpha_2(\Delta_2|X_{1,P} < c) = 0$  and it covers the global null space which is the third quadrant of the product space  $\Delta_1 \times (\Delta_2|X_{1,P} < c)$  as illustrated in Fig. 2.

Let the estimate of the adjusted treatment effect  $\Delta$  be given by the least square estimator as defined by Eq. (7) with weights defined by Eq. (9):

$$\widehat{\Delta} = \alpha_1\widehat{\Delta}_1 + \alpha_2(\widehat{\Delta}_2|X_{1,P} < c)$$

Then, it follows that

$$E(\widehat{\Delta}) = \alpha_1E(\widehat{\Delta}_1) + \alpha_2E((\widehat{\Delta}_2|X_{1,P} < c)) = \alpha_1\Delta_1 + \alpha_2(\Delta_2|X_{1,P} < c) = \Delta$$

and

$$\begin{aligned} var(\widehat{\Delta}) &= \Sigma_{\widehat{\Delta}}^2 \\ &= \alpha_1^2 var(\widehat{\Delta}_1) + \alpha_2^2 var(\widehat{\Delta}_2|X_{1,P} < c) \\ &\quad + 2\alpha_1\alpha_2 cov(\widehat{\Delta}_1, (\widehat{\Delta}_2|X_{1,P} < c)) \end{aligned}$$

where  $var(\widehat{\Delta}_1)$ ,  $var(\widehat{\Delta}_2|X_{1,P} < c)$  and  $cov(\widehat{\Delta}_1, (\widehat{\Delta}_2|X_{1,P} < c))$  are as given in the earlier lemma.

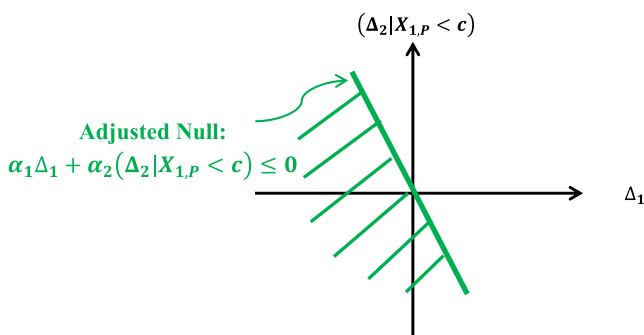


Fig. 2. Region of the parameter space for the adjusted treatment null.

The combination test for testing the adjusted null hypothesis is then given by

$$\widehat{Z} = \frac{(\widehat{\Delta} - \Delta)}{\sqrt{var(\widehat{\Delta})}}$$

where  $\widehat{\Delta}$ ,  $\Delta$  and  $var(\widehat{\Delta})$  are as given above.

**Note:** It is important to point out that the adjusted treatment effect  $\Delta$  and its estimate  $\widehat{\Delta}$  are independent of the allocation ratios  $r_1$  and  $r_2$ , but the variance of  $\widehat{\Delta}$  does depend on the allocation ratios. This is fine, since the variance of  $\widehat{\Delta}$  should take into account the actual allocation ratios in the design. This will not affect the estimate of the adjusted treatment effect, but only its precision.

5.1. The type I error for the combination test

The type I error for the combination test is given by

$$\alpha = P(\widehat{Z} > c_\alpha | H_{0,Adj}) = P(\widehat{Z}_0 > c_\alpha)$$

where

$$\widehat{Z}_0 = \left( \frac{((\alpha_1\widehat{\Delta}_1 + \alpha_2(\widehat{\Delta}_2|X_{1,P} < c)))}{\sqrt{var(\alpha_1\widehat{\Delta}_1 + \alpha_2(\widehat{\Delta}_2|X_{1,P} < c))}} \right)$$

$$\begin{aligned} var(\alpha_1\widehat{\Delta}_1 + \alpha_2(\widehat{\Delta}_2|X_{1,P} < c)) &= \alpha_1^2(\gamma, \sigma_1, \sigma_2)var(\widehat{\Delta}_1) \\ &\quad + 2\alpha_1\alpha_2cov(\widehat{\Delta}_1, (\widehat{\Delta}_2|X_{1,P} < c)) \\ &\quad + \alpha_2^2(\gamma, \sigma_1, \sigma_2)var((\widehat{\Delta}_2|X_{1,P} < c)) \end{aligned}$$

Note that from the following relationship previously derived,

$$\begin{aligned} cov(\widehat{\Delta}_1, (\widehat{\Delta}_2|X_{1,P} < c)) &= \frac{1}{n_{1P}} (cov(X_{1,P}, X_{2,P} | X_{1,P} < c) \\ &\quad - cov(X_{1,P}, X_{2,T} | X_{1,P} < c)) \end{aligned}$$

which may be estimated by the sample covariance from the two cohorts ( $P \rightarrow P$ ) and ( $P \rightarrow T$ ).

The type I error control for the combination test is illustrated in Table 8

5.2. The power and sample size for the combination test

The power of the combination test at a specified alternative  $(\Delta_1, \Delta_2)$  in the first quadrant is given by

$$\begin{aligned} 1 - \beta &= P(\widehat{Z}_0 > c_\alpha | H_{a,Adj} : (\Delta_1, \Delta_2) \text{ in 1st Quadrant, } \rho_P > \rho_T) \\ &= P\left(\widehat{Z}_a > c_\alpha - \frac{\alpha_1\Delta_1 + \alpha_2(\Delta_2|X_{1,P} < c)}{\Sigma_{\widehat{\Delta},a}} \middle| H_{a,Adj} \right. \\ &\quad \left. : (\Delta_1, \Delta_2) \text{ in 1st Quadrant, } \rho_P > \rho_T \right) \end{aligned}$$

where

$$\widehat{Z}_a = (\alpha_1\widehat{\Delta}_1 + \alpha_2(\widehat{\Delta}_2|X_{1,P} < c)) - (\alpha_1\Delta_1 + \alpha_2(\Delta_2|X_{1,P} < c)) / \Sigma_{\widehat{\Delta},a} \sim N(0, 1) \text{ and } \Sigma_{\widehat{\Delta},a} = \Sigma_{\widehat{\Delta},0}$$

From the above power function, one can derive the sample size formula as follows:

$$n_{1T} = \left( \frac{c_\alpha + c_{1-\beta}}{\alpha_1 \Delta_1 + \alpha_2 (\Delta_2 | X_{1,P} < c)} \right)^2 \left( \alpha_1^2 \frac{\sigma_1^2}{R_1} + \alpha_2^2 \frac{\sigma_2^2}{\gamma R_{12}} \right) + 2 \alpha_1 \alpha_2 \rho_{1,2} \frac{\sigma_1}{\sqrt{R_1}} \frac{\sigma_2}{\sqrt{\gamma R_{12}}} \frac{\sigma_{\Delta_2 | X_{1,P} < c}}{\sqrt{\gamma R_{12}}}$$

**Note:** Alternatively, instead of the power and sample size formulas given in the above equations, one can actually find the power and sample size formulae via the bivariate normal probability integral below:-

$$1 - \beta = \int_{-\infty}^{\infty} \varphi_{\hat{V}_1}(x) \left( 1 - \int_{-\infty}^{\frac{c_\alpha \sqrt{1 + \left(\frac{\alpha_1}{\alpha_2}\right)^2} - \left(\frac{\alpha_1^* U_1 + U_2}{\alpha_2} - \frac{\alpha_1^* x}{\alpha_2}\right) - \rho_{1,2} x}}{\sqrt{1 - \rho_{1,2}^2}} \varphi(y) dy \right) dx$$

where  $\alpha_1^* = \alpha_1 \sigma_1 / \sqrt{n_{1,T} R_1}$ ,  $\alpha_2^* = \alpha_2 \sigma_2 / \sqrt{n_{2,T} R_2}$  and  $\varphi$  and  $\Phi$  represent the standard normal density and cumulative distribution functions.

Table 2, Tables 3 and 8 provide the power and sample size for selected scenarios and DRDS design parameter values based on the HDRS17 Anxiety and Somatization subscale score data given in Table 1.

### 5.3. The monotonicity condition

The rejection region for the adjusted treatment null hypothesis as defined by the combination test is depicted in Fig. 3 below.

Fig. 3 shows that there is still a small area shaded green in Fig. 2 under the rejection region that is situated inside the second quadrant. This suggests that even though the probability is small, the adjusted treatment null may be rejected by the combination test, but the Period 1 treatment effect  $\Delta_1$  may be negative. From Eq. (11), one can see that a negative  $\Delta_1$  suggests that the treatment may perform worse than placebo in the subpopulation  $\Omega_R$ . Now in the subpopulation  $\Omega_R$ , the placebo acts like an active control trial in a non-inferiority trial. In a non-inferiority trial, a treatment is still considered effective if it performs no worse

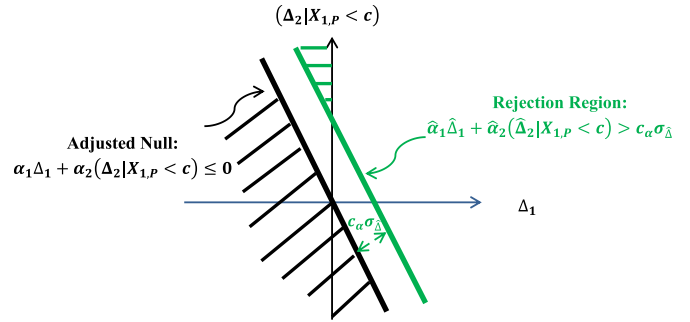


Fig. 3. Rejection region of the combination test.

than placebo by a given non-inferiority margin  $\delta > 0$ . So, what would be an equivalent non-inferiority margin for assessing the effectiveness of a treatment effect in the subpopulation  $\Omega_R$  of placebo responders?

As a condition required for a treatment effectiveness claim to be extendable to the intended study population, Tamura et al. [8] introduced a monotonicity condition for the case under binary outcome. This monotonicity condition simply requires each placebo responder also responds to treatment. Under binary outcome, this monotonicity condition is equivalent to requiring that the treatment be at least as effective as placebo. Now for continuous outcome, this monotonicity condition does not rule out the possibility that the treatment could perform worse than the placebo. Therefore, what should then be the monotonicity condition? Now if one were to require that the treatment should perform at least as effectively as placebo, then this is equivalent to requiring the treatment to show superiority to an active control, and hence would be too stringent. On the other hand, if one were simply to require that each placebo responder also responds to treatment, then under this condition, the treatment can still perform worse than placebo. But then what would be a corresponding non-inferiority margin in this case?

From Eq. (9), one can see that the condition that requires the treatment to be at least as effective as placebo can be stated as the following equivalent condition:

**Table 2**

Selected Powers and Sample Sizes at One-sided  $\alpha = 0.025$  for the Combination Test at the Specified DRDS Design Parameter Values and the Hypothetical Distributions of a HDRS17 Subscale Score under Treatment and Placebo as given in Table 1 (DRDS Design Parameter Values:  $r_1 = 2, r_2 = 1, c = 2.50, \gamma = 0.42$ ),  $\Delta = \alpha_1^* \Delta_1 + \alpha_2^* \Delta_{2|C}$ .

$\mu_{1T}$	$\mu_{1P}$	$\Delta_1$	$\sigma_1$	$\rho_P$	$\rho_T$	$\Delta_{2 C}$	$\sigma_{2 C}$	$\Delta$	$1 - \beta$	$N_1$	$n_{1T}$	$n_{2T}$
3.30	3.00	0.30	2.42	0.80	0.20	1.43	3.18	0.42	80%	960	320	134
									85%	1098	366	154
									90%	1287	429	180
				0.80	0.50	0.88	3.28	0.38	80%	1338	446	187
									85%	1587	529	222
									90%	1893	631	265

**Table 3**

Selected Powers and Sample Sizes at One-sided  $\alpha = 0.025$  for the Combination Test at the Specified DRDS Design Parameter Values and the Hypothetical Distributions of a HDRS17 Subscale Score under Treatment and Placebo as given in Table 1 (DRDS Design Parameter Values:  $r_1 = 2, r_2 = 1, c = 2.75, \gamma = 0.44$ ),  $\Delta = \alpha_1^* \Delta_1 + \alpha_2^* \Delta_{2|C}$ .

$\mu_{1T}$	$\mu_{1P}$	$\Delta_1$	$\sigma_1$	$\rho_P$	$\rho_T$	$\Delta_{2 C}$	$\sigma_{2 C}$	$\Delta$	$1 - \beta$	$N_1$	$n_{1T}$	$n_{2T}$
3.50	3.10	0.40	2.42	0.80	0.20	1.48	3.23	0.52	80%	636	212	93
									85%	720	240	106
									90%	838	280	123
				0.80	0.50	0.96	3.32	0.46	80%	819	273	120
									85%	936	312	137
									90%	1095	365	161

$$\Delta_1 = \alpha_R \Delta_R + \alpha_{NR} \Delta_{NR} > \gamma (\Delta_2 | X_{1,P} < c) \text{ OR } (\Delta_2 | X_{1,P} < c) < \frac{1}{\gamma} \Delta_1 \quad (15)$$

since under the earlier assumptions on the placebo dropouts if any,  $\alpha_{NR} = \gamma = \Phi(\tau)$  and  $\Delta_{NR} = \Delta_2$ . This condition in Eq. (15) is depicted in Fig. 4.

It is clear that this condition is quite stringent and besides this superiority condition is also not required for a non-inferiority trial. Therefore, a less stringent monotonicity condition is needed, a condition that allows the treatment to perform no worse than placebo by a non-inferiority margin. An obvious general monotonicity condition is to require that

$$(\Delta_2 | X_{1,P} < c) < \eta \Delta_1, \text{ for some } \eta > \frac{1}{\gamma} \quad (16)$$

The slope  $\eta$  can be viewed here as the equivalent of a non-inferiority margin  $\delta$ . But how should  $\eta$  be determined? This would be a challenging problem. But even the general monotonicity condition as defined by Eq. (16) is very stringent if the condition is required to be tested as illustrated in Fig. 5.

Note that in the general monotonicity conditions defined by Eq. (16), a constraint is placed on the expected Period 2 treatment effect  $(\Delta_2 | X_{1,P} < c)$ . This constraint is really not necessary because from Eq. (1),

$$\begin{aligned} (\Delta_2 | X_{1,P} < c) &= (\mu_{2,T} - \mu_{2,P}) + (\rho_P \sigma_{2,P} - \rho_T \sigma_{2,T}) \left( \frac{\varphi(\tau)}{\Phi(\tau)} \right) \equiv \Delta_1 \\ &+ (\rho_P \sigma_{2,P} - \rho_T \sigma_{2,T}) \left( \frac{\varphi(\tau)}{\Phi(\tau)} \right) \end{aligned} \quad (17)$$

and it is seen from Eq. (17) that the magnitude of the expected Period 2 treatment effect  $(\Delta_2 | X_{1,P} < c)$  is determined by the magnitude of the Period 1 treatment effect  $\Delta_1$  and the term  $(\rho_P \sigma_{2,P} - \rho_T \sigma_{2,T}) (\varphi(\tau)/\Phi(\tau))$  the magnitude of which in turn is determined by the standard deviations  $\sigma_{2,P} = \sigma_{1,P}$ ,  $\sigma_{2,T} = \sigma_{1,T}$ , the correlations  $\rho_P$  and  $\rho_T$  and the hazard ratio  $(\varphi(\tau)/\Phi(\tau))$ , and it cannot be arbitrarily large.

Therefore, if such constraint imposed by the above condition is not necessary, then one should consider relaxing the condition by letting  $\eta \rightarrow \infty$ . Now as one lets  $\eta \rightarrow \infty$ , the line  $(\Delta_2 | X_{1,P} < c) = \eta \Delta_1 \rightarrow$  the  $(\Delta_2 | X_{1,P} < c) - \Delta_1$  axis. This then naturally leads to the consistency condition which will be introduced in the next section as the condition required for the treatment effectiveness claim to be extendable to the intended study population  $\Omega_1$  in lieu of a general monotonicity condition defined by Eq. (16).

#### 5.4. A measure of consistency

In a DRDS design, what is consistency and why is it necessary? As discussed in Section 5.4, even if the combination test rejects the adjusted null hypothesis, one may still not be able to claim that the treatment is effective for the intended population because the pair of treatment effect  $(\Delta_1, \Delta_2)$  may be located in the second quadrant in the  $\Delta_1 \times (\Delta_2 | X_{1,P} < c)$  parameter space meaning that  $\Delta_1$  could be negative. To remedy this problem, a general monotonicity condition as defined by Eq. (16) can be proposed. But as discussed in Section 5.4, this general monotonicity condition is too stringent. In this section, an alternative consistency test is introduced to test for the consistency between the treatment effects  $\Delta_1$  and  $(\Delta_2 | X_{1,P} < c)$ . However, the consistency test alone does not permit one to conclude that the treatment effects are positive in both periods. It requires the joint rejection of the adjusted null and the consistency null by their respective tests. Therefore, the simultaneous rejection of the adjusted null and the consistency null would be required for one to conclude that the pair of treatment effect

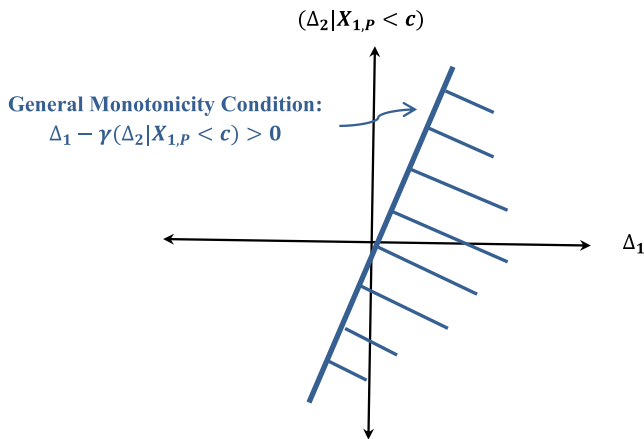


Fig. 4. Region defined by the general monotonicity condition.

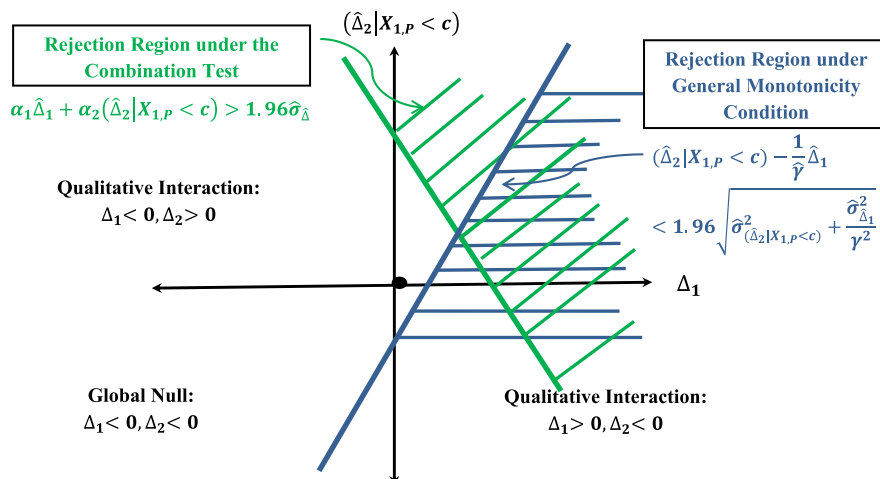


Fig. 5. Rejection region under the combination test and the rejection region under the general monotonicity condition.

$(\Delta_1, (\Delta_2|X_{1,P} < c))$  lies in the first quadrant of the parameter space  $\Delta_1 \times (\Delta_2|X_{1,P} < c)$ .

This consistency test jointly with the combination test may provide sufficient evidence for one to conclude that the pair of treatment effect  $(\Delta_1, (\Delta_2|X_{1,P} < c))$  lies in the first quadrant, that is, both  $\Delta_1$  and  $(\Delta_2|X_{1,P} < c)$  are positive. Once this is established, then the treatment effectiveness claim as represented by the adjusted treatment effect can be extended to the intended study population and the adjusted treatment effect estimates can then be used in the benefit/risk analysis and in proper dosage recommendation.

Note that under finite samples, if the pair  $(\Delta_1, (\Delta_2|X_{1,P} < c))$  is inconsistent with  $\Delta_1 < 0$ , then the optimal efficiency of the combined statistic under equal allocation case may be lost. However, as noted earlier, since in practice, equal allocation is unlikely to be used, maintenance of efficiency may be a moot point and is of secondary concern compared to a proper assessment of the treatment effect. But in any case, one should interpret the combined statistic with caution in light of such inconsistency. This suggests that consistency is an important condition needed for the validity and interpretability of the combined statistic.

5.5. The consistency test

Let the consistency measure  $\Gamma$  between  $\Delta_1$  and  $(\Delta_2|X_{1,P} < c)$  be defined as  $\Gamma = \Delta_1(\Delta_2|X_{1,P} < c)$ . Then the consistency null and alternative hypotheses are defined as:

$$H_{0,C} : \Gamma = \Delta_1(\Delta_2|X_{1,P} < c) \leq 0 \text{ vs. } H_{0,C} : \Gamma = \Delta_1(\Delta_2|X_{1,P} < c) > 0 \tag{18}$$

The consistency null hypothesis is depicted by the shaded region in Fig. 6.

Now consider the following statistic:

$$\hat{\Gamma} = \hat{\Delta}_1(\hat{\Delta}_2|X_{1,P} < c) - \widehat{cov}(\hat{\Delta}_1, (\hat{\Delta}_2|X_{1,P} < c))$$

Then, one has

$$E(\hat{\Gamma}) = E(\hat{\Delta}_1(\hat{\Delta}_2|X_{1,P} < c) - cov(\hat{\Delta}_1, (\hat{\Delta}_2|X_{1,P} < c))) = \Delta_1(\Delta_2|X_{1,P} < c)$$

The variance of  $\hat{\Gamma}$  is given approximately asymptotically by

$$var(\hat{\Gamma}) = [var(\hat{\Delta}_1)var(\hat{\Delta}_2|X_{1,P} < c)] + cov^2(\hat{\Delta}_1, (\hat{\Delta}_2|X_{1,P} < c)) + [(\Delta_2|X_{1,P} < c)^2 var(\hat{\Delta}_1)] + [\Delta_1^2 var(\hat{\Delta}_2|X_{1,P} < c)] + [4\Delta_1(\Delta_2|X_{1,P} < c)cov(\hat{\Delta}_1, (\hat{\Delta}_2|X_{1,P} < c))] + \Delta_1^2(\Delta_2|X_{1,P} < c)^2$$

The consistency test for the consistency hypothesis defined by Eq. (18) is then given by

$$\widehat{W} = \frac{\hat{\Gamma} - E(\hat{\Gamma})}{\sqrt{var(\hat{\Gamma})}}$$

where  $\hat{\Gamma}$ ,  $E(\hat{\Gamma})$  and  $var(\hat{\Gamma})$  are given above, with  $cov(\hat{\Delta}_1, (\hat{\Delta}_2|X_{1,P} < c))$  estimated by  $\widehat{cov}(\hat{\Delta}_1, (\hat{\Delta}_2|X_{1,P} < c)) = \frac{1}{n_{1P}}(S_{X_{1,P} < c, X_{2,P}|X_{1,P}} - S_{X_{1,P} < c, X_{2,T}|X_{1,P}})$  where  $S_{X_{1,P} < c, X_{2,P}|X_{1,P}}$  and  $S_{X_{1,P} < c, X_{2,T}|X_{1,P}}$  are the sample covariance estimates for  $cov(\hat{\mu}_{1,P|X_{1,P} < c}, \hat{\mu}_{2,P|X_{1,P} < c})$  and  $cov(\hat{\mu}_{1,P|X_{1,P} < c}, \hat{\mu}_{2,T|X_{1,P} < c})$  for the two cohorts  $(P \rightarrow P)$  and  $(P \rightarrow T)$ , since as previously noted,

$$cov(\hat{\Delta}_1, (\hat{\Delta}_2|X_{1,P} < c)) = \frac{1}{n_{1P}}(cov(X_{1,P}, (X_{2,P}|X_{1,P} < c)) - cov(X_{1,P}, (X_{2,T}|X_{1,P} < c)))$$

5.6. The type I error for the consistency test

The type I error for the consistency test is given under asymptotic normality by

$$\alpha = P(\widehat{W} > c_{\alpha,W} | H_{0,C}) = P\left(\frac{[\hat{\Delta}_1(\hat{\Delta}_2|X_{1,P} < c) - \widehat{cov}(\hat{\Delta}_1, (\hat{\Delta}_2|X_{1,P} < c))] - \Gamma}{\sqrt{var(\hat{\Gamma})}} > c_{\alpha,W} | H_{0,C}\right)$$

where  $var(\hat{\Gamma})$  as derived above and  $\Gamma = \Delta_1(\Delta_2|X_{1,P} < c)$ .

Note that at the boundary of the consistency null, the type I error assumes its maximum at  $(\Delta_1, (\Delta_2|X_{1,P} < c)) = (0,0)$  and  $cov(\hat{\Delta}_1, (\hat{\Delta}_2|X_{1,P} < c)) = 0$ . Therefore, the type I error for the consistency test evaluated at its maximum is given by

$$\alpha = P\left(\frac{\hat{\Delta}_1(\hat{\Delta}_2|X_{1,P} < c) - \widehat{cov}(\hat{\Delta}_1, (\hat{\Delta}_2|X_{1,P} < c))}{\sqrt{var(\hat{\Gamma})|H_{0,C}}} > c_{\alpha,W}\right)$$

where

$$var(\hat{\Gamma} | H_{0,C}) = [var(\hat{\Delta}_1)var(\hat{\Delta}_2|X_{1,P} < c)] + [(\Delta_2|X_{1,P} < c)^2 var(\hat{\Delta}_1)] + [\Delta_1^2 var(\hat{\Delta}_2|X_{1,P} < c)]$$

Analogously, the above type I error can also be evaluated asymptotically via bivariate normal integral as

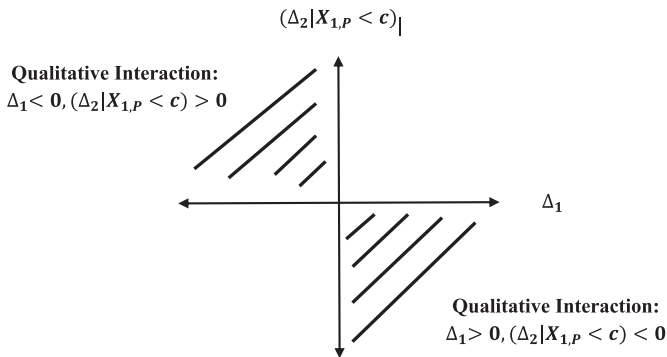


Fig. 6. Region of the parameter space for the consistency null.

$$\begin{aligned} \alpha &= P(\hat{U}_1 \hat{U}_2 > c_{\alpha,W}) = P\left(\hat{U}_2 > \frac{c_{\alpha,W}}{\hat{U}_1} \middle| H_{0,C}\right) \\ &= P\left(\hat{U}_2 > \frac{c_{\alpha,W}}{\hat{U}_1} \middle| (U_1, U_2) = (0, 0), \rho_{1,2} = 0\right) \\ &= \frac{1}{2} + \int_{-\infty}^0 \varphi(z_1) \Phi\left(\frac{c_{\alpha,W}}{z_1}\right) dz_1 \\ &\quad - \int_0^{\infty} \varphi(z_1) \Phi\left(\frac{c_{\alpha,W}}{z_1}\right) dz_1, \text{ with } \hat{U}_1 \\ &= \frac{\hat{\Delta}_1}{\frac{\sigma_1}{\sqrt{n_{1,T}R_1}}}, \hat{U}_2 = \frac{(\hat{\Delta}_2 | X_{1,P} < c)}{\sqrt{\text{var}(X_{2|X_{1,P} < c})}} \\ &\quad \frac{\sigma_2}{\sqrt{n_{2,T}R_2}} \end{aligned}$$

Since  $\widehat{W}_o = \hat{U}_1 \hat{U}_2$  is not normally distributed and has a distribution with heavy tail, its critical values are somewhat larger for the same significance level  $\alpha$  as compared to the critical values from a normal distribution. Critical values for selected levels of significance are given in Table 4.

In light of the proposed procedure of testing both the adjusted treatment null hypothesis by the combination test  $\widehat{Z}_o$  and the consistency null hypothesis by the consistency test  $\widehat{W}_o$ , a rejection of the adjusted treatment null by the test  $\widehat{Z}_o$  implies that  $(\Delta_1, (\Delta_2 | X_{1,P} < c))$  does not lie in the third quadrant which effectively reduces the nominal  $\alpha$  level of the consistency test  $\widehat{W}_o$  by half. Therefore, it is suggested that the type I error rate for the consistency test  $\widehat{W}_o$  be held at the one-sided significance level of  $\alpha = 0.05$  corresponding to a critical value of  $c_{0.05,W} = 1.60$ . This yields an effective significance level of  $\alpha = 0.025$  for the consistency test  $\widehat{W}_o$  under the joint testing procedure. This is the significance level that is used subsequently in generating the various sample size and power calculations for the consistency test  $\widehat{W}_o$ .

Table 8 suggests that the type I error rate for the consistency test is controlled at the one-sided 0.05 level.

The rejection region of the consistency test is depicted in Fig. 7. It shows that the rejection region defined by the consistency test (region in the first and third quadrants) and the combination test (region in the first, second and fourth quadrants defined by the green line) consists of the shaded parabolic region in brown in the first quadrant which represents the intersection of the two rejection regions.

Fig. 8 shows that the rejection region under the combination test and the consistency test is less stringent than the rejection region required by the general monotonicity condition as defined by Eq. (16). The consistency condition here may be viewed as equivalent to a non-inferiority margin in an active control trial

**Table 4**  
Critical values for the consistency test  $\widehat{W}_o$  at selected significance level  $\alpha$ .

$\alpha$	$c_{\alpha,W}$
0.001	5.08
0.005	3.60
0.010	2.98
0.025	2.18
0.050	1.60
0.075	1.26
0.100	1.03

(see the discussion in Section 5.4 where the consistency condition may be viewed as the limiting general monotonicity condition).

5.7. The power and sample size for the consistency test

The Power of the Consistency Test is given by:

$$\begin{aligned} 1 - \beta &= P(\widehat{W}_o > c_{\alpha,W} | H_{a,C}) \\ &= P\left(\frac{\widehat{\Delta}_1(\widehat{\Delta}_2 | X_{1,P} < c) - \widehat{cov}(\widehat{\Delta}_1, (\widehat{\Delta}_2 | X_{1,P} < c))}{\sqrt{\text{var}(\widehat{W}_o)}} > c_{\alpha,W} \middle| H_{a,C}\right) \end{aligned}$$

where  $\text{var}(\widehat{W}_o) = \text{var}(\widehat{\Delta}_1)\text{var}(\widehat{\Delta}_2 | X_{1,P} < c)$ .

Hence, where

$$1 - \beta = P\left(\widehat{W}_a > \frac{c_{\alpha,W} \sqrt{\text{var}(\widehat{W}_o)} - \Delta_1(\Delta_2 | X_{1,P} < c)}{\sqrt{\text{var}(\widehat{W}_a)}} \middle| H_{a,C}\right),$$

where

$$\widehat{W}_a = [\widehat{\Delta}_1(\widehat{\Delta}_2 | X_{1,P} < c) - \widehat{cov}(\widehat{\Delta}_1, (\widehat{\Delta}_2 | X_{1,P} < c)) - \Gamma] / \sqrt{\text{var}(\widehat{W}_a)},$$

and  $\Gamma = \Delta_1(\Delta_2 | X_{1,P} < c)$ .

$$\begin{aligned} \text{var}(\widehat{W}_a) &= \text{var}(\widehat{W}_o) + \text{cov}^2(\widehat{\Delta}_1, (\widehat{\Delta}_2 | X_{1,P} < c)) \\ &\quad + (\Delta_2 | X_{1,P} < c)^2 \text{var}(\widehat{\Delta}_1) + \Delta_1^2 \text{var}(\widehat{\Delta}_2 | X_{1,P} < c) \\ &\quad + 4\Delta_1(\Delta_2 | X_{1,P} < c) \text{cov}(\widehat{\Delta}_1, (\widehat{\Delta}_2 | X_{1,P} < c)) \\ &\quad + \Delta_1^2 (\Delta_2 | X_{1,P} < c)^2 \end{aligned}$$

Note that the power may be evaluated by viewing  $\widehat{\Delta}_1$  and  $(\widehat{\Delta}_2 | X_{1,P} < c)$  as having an asymptotic bivariate normal distribution given by

$$\begin{aligned} 1 - \beta &= P\left(\widehat{V}_2 > \frac{R - U_1 U_2 - U_2 \widehat{V}_1}{U_1 + \widehat{V}_1} \middle| H_{a,C}\right) \\ &= \int_{-\infty}^{\infty} \varphi_{\widehat{V}_1}(x) \int_{\frac{R - U_1 U_2 - U_2 x - \rho_{1,2} x(x + U_1)}{(x + U_1) \sqrt{1 - \rho_{1,2}^2}}}^{\infty} \varphi(y) dy dx \end{aligned}$$

where

$$R = \left(c_{\alpha,W} \sqrt{1 + \rho_{1,2}^2 + U_1^2 + U_2^2 + \rho_{1,2}}\right)$$

where  $U_i = \frac{\Delta_i}{\frac{\sigma_i}{\sqrt{n_{i,T}R_i}}}$ ,  $i = 1, 2$  and  $(\widehat{V}_1, \widehat{V}_2) \sim N(\mu_{1,2}, \Sigma_{1,2}^2)$

where  $\mu_{1,2} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$  and  $\Sigma_{1,2}^2 = \begin{pmatrix} 1 & \rho_{1,2} \\ \rho_{1,2} & 1 \end{pmatrix}$  and  $\rho_{1,2} = \text{corr}(\widehat{\Delta}_1, (\widehat{\Delta}_2 | X_{1,P} < c)) = \rho_{\widehat{\Delta}_1, (\widehat{\Delta}_2 | X_{1,P} < c)}$  as derived earlier.

By substituting the above expressions for  $U_i$ ,  $i = 1, 2$  and noting that  $n_{2,T} = n_{1,T} \gamma R_{12}$ , then one can evaluate the above probability integral for the power at a given sample size,  $n_{1,T}$ .

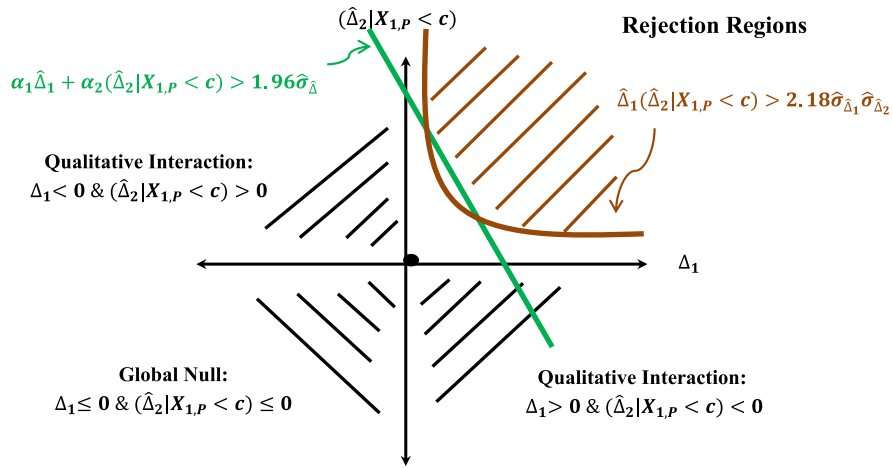


Fig. 7. Rejection regions under the combination and consistency tests.

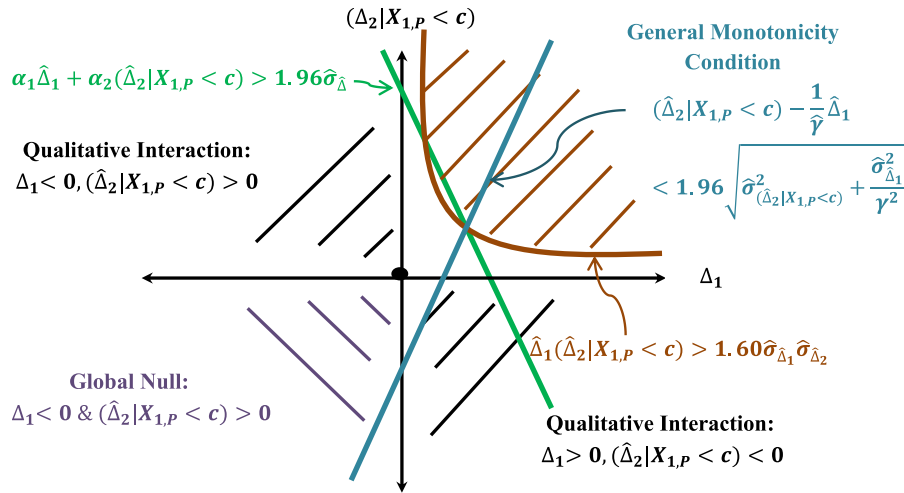


Fig. 8. Rejection regions in the alternative space.

Conversely, to calculate the sample size, one can just solve the above equation implicitly for  $n_{1,T}$  at a given power  $(1 - \beta)$ . Some selected powers and sample sizes are given in Table 5, Tables 6 and 8 based on the example in Table 1.

### 6. The joint test

As mentioned in the preceding section, both the combination test and the joint test are necessary for establishing the effectiveness of a treatment for the intended study population  $\Omega = \Omega_1$  in a DRDS design. A joint test is proposed here for simultaneously testing the adjusted treatment null by the combination test and

the consistency null by the consistency test. Upon the simultaneous rejection of this pair of null hypotheses, one can then derive an estimate of the adjusted treatment effect along with its confidence interval, and an estimate of the consistency of the treatment effects from Period 1 and Period 2 along with its confidence interval. The adjusted treatment effect represents the apparent treatment effect of Period 1 having been adjusted for the presence of high placebo response rate. The consistency condition is viewed as a generalization of the general monotonicity condition and a rejection of the consistency null would permit the extension of the effectiveness of the adjusted treatment effect to the intended study population.

Table 5

Selected Powers and Sample Sizes at One-sided  $\alpha = 0.05$  for the Consistency Test  $\widehat{W}_0$  at the Specified DRDS Design Parameter Values and the Hypothetical Distributions of a HDRS<sub>17</sub> Subscale Score under Treatment and Placebo as given in Table 1 (DRDS Design Parameter Values:  $r_1 = 2, r_2 = 1, c = 2.75, \gamma = 0.44$ )  $\Gamma = \Delta_1^* \Delta_2$ .

$\mu_{1T}$	$\mu_{1P}$	$\Delta_1$	$\sigma_1$	$\rho_P$	$\rho_T$	$\Delta_{2 C}$	$\sigma_{2 C}$	$\Gamma$	$1 - \beta$	$N_1$	$n_{1T}$	$n_{2T}$
3.50	3.10	0.40	2.42	0.80	0.20	1.48	3.23	0.59	80%	825	275	121
									85%	954	318	140
									90%	1083	361	159
				0.80	0.50	0.96	3.32	0.38	80%	1032	344	151
									85%	1176	392	172
									90%	1389	463	204

**Table 6**  
Selected Powers and Sample Sizes at One-sided  $\alpha = 0.05$  for the Consistency Test  $\widehat{W}_o$  at the Specified DRDS Design Parameter Values and the Hypothetical Distributions of a HDRS<sub>17</sub> Subscale Score under Treatment and Placebo as given in Table 1 (DRDS Design Parameter Values:  $r_1 = 2, r_2 = 1, c = 2.50, \gamma = 0.42$ )  $\Gamma = \Delta_1^* \Delta_{2C}$ .

$\mu_{1T}$	$\mu_{1P}$	$\Delta_1$	$\sigma_1$	$\rho_P$	$\rho_T$	$\Delta_{2C}$	$\sigma_{2C}$	$\Gamma$	$1 - \beta$	$N_1$	$n_{1T}$	$n_{2T}$
3.30	3.00	0.30	2.42	0.80	0.20	1.43	3.18	0.43	80%	1128	376	158
									85%	1323	441	185
									90%	1605	535	225
				0.80	0.50	0.88	3.28	0.26	80%	1377	459	193
									85%	1587	529	222
									90%	1893	631	265

Tables 9a, b and c provide the type I error, power and sample size needed for some selected configurations for purpose of illustration. These data can be generated by integrating the combination test and the consistency test through the bivariate normal probability integral since both tests are jointly defined in terms of  $\Delta_1$  and  $(\Delta_1|X_{1,P} < c)$ .

6.1. The joint test ( $\widehat{Z}_o > c_{0.025}, \widehat{W}_o > c_{0.05,W}$ )

Since the test of the adjusted null hypothesis by the combination test alone is deemed not sufficient to establish the effectiveness of the treatment for the intended population in Period 1, it is proposed that a joint testing of the adjusted null hypothesis by the combination test  $\widehat{Z}_o$  at  $\alpha = 0.025$  and the consistency null hypothesis by the consistency test  $\widehat{W}_o$  at  $\alpha = 0.05$  should be performed. When both the adjusted null and the consistency null have been rejected by their respective tests  $\widehat{Z}_o$  and  $\widehat{W}_o$ , then one may conclude that the treatment effect pair  $(\Delta_1, (\Delta_2|X_{1,P} < c))$  is located in the first quadrant and the treatment effects for both Period 1 and Period 2 are positive and consistent. The combination test can then provide an estimate of the adjusted treatment effect and its associated 95% confidence interval given by  $\widehat{\Delta} = \widehat{\alpha}_1 \widehat{\Delta}_1 + \widehat{\alpha}_2 (\widehat{\Delta}_2|X_{1,P} < c)$  where

$$(\widehat{\alpha}_1 \widehat{\Delta}_1 + \widehat{\alpha}_2 (\widehat{\Delta}_2|X_{1,P} < c)) \pm 1.96 \sqrt{\text{var}(\widehat{\alpha}_1 \widehat{\Delta}_1 + \widehat{\alpha}_2 (\widehat{\Delta}_2|X_{1,P} < c))}$$

where

$$\text{var}(\widehat{\alpha}_1 \widehat{\Delta}_1 + \widehat{\alpha}_2 (\widehat{\Delta}_2|X_{1,P} < c)) = \widehat{\alpha}_1^2 \frac{\sigma_1^2}{n_{1,T} R_1} + \widehat{\alpha}_2^2 \frac{\text{var}((\widehat{\Delta}_2|X_{1,P} < c))}{n_{2,T}} + 2\widehat{\alpha}_1 \widehat{\alpha}_2 \text{cov}(\widehat{\Delta}_1, (\widehat{\Delta}_2|X_{1,P} < c))$$

where  $\sigma_1^2$  and  $\text{var}((\widehat{\Delta}_2|X_{1,P} < c))$  may be estimated by the sample variances  $\widehat{\sigma}_1^2$  and  $\widehat{\text{var}}((\widehat{\Delta}_2|X_{1,P} < c))$ , the latter via the sample variances from the two cohorts ( $P \rightarrow P$ ) and ( $P \rightarrow T$ ) for  $\text{var}(X_{2,P}|X_{1,P} < c)$  and  $\text{var}(X_{2,T}|X_{1,P} < c)$ , the covariance  $\text{cov}(\widehat{\Delta}_1, (\widehat{\Delta}_2|X_{1,P} < c))$  can also be estimated by the sample covariance for the two cohorts ( $P \rightarrow P$ ) and ( $P \rightarrow T$ ), and the weights  $\alpha_i$  may be estimated by  $\widehat{\alpha}_i = \alpha_i(\widehat{\gamma}, \widehat{\sigma}_1, \sqrt{\widehat{\text{var}}((\widehat{\Delta}_2|X_{1,P} < c))})$ , for  $i = 1, 2$ .

The magnitude of the consistency measure also provides supportive information for the strength of the consistency.

Fig. 9 provides a graphical description of the estimate of the adjusted treatment effect in relation to the joint test and the general monotonicity condition. It shows that the estimated adjusted treatment effect  $\widehat{\Delta} = \widehat{\alpha}_1 \widehat{\Delta}_1 + \widehat{\alpha}_2 (\widehat{\Delta}_2|X_{1,P} < c)$  appears as the coordinates of the point  $(\widehat{\Delta}_1, \widehat{\Delta}_2)$  which is the intersection of the line  $\widehat{\alpha}_1 \Delta_1 + \widehat{\alpha}_2 (\Delta_2|X_{1,P} < c) = \widehat{\Delta}$  and the 45° diagonal line. The point  $(\widehat{\Delta}_1, (\widehat{\Delta}_2|X_{1,P} < c))$  satisfies the general monotonicity condition  $\Delta_2 < \eta \Delta_1$  as shown in Fig. 9, but if the slope  $\eta$  is smaller, then  $(\widehat{\Delta}_1, (\widehat{\Delta}_2|X_{1,P} < c))$  may very well not satisfy the corresponding monotonicity condition. In addition, if one is required to test the general monotonicity condition, then it would be even more

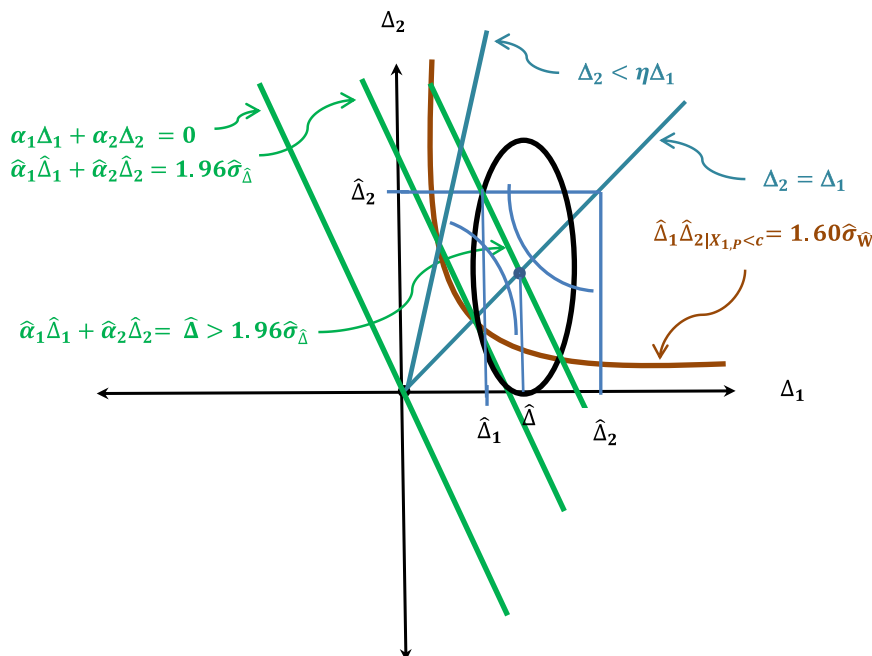


Fig. 9. Estimate of the adjusted treatment effect under the joint test.

stringent. Thus, it is clear from Fig. 9 that the general monotonicity condition as defined by Eq. (16) is unnecessarily restrictive, and the consistency condition should be preferred.

The power of the joint test ( $\widehat{Z}_o > c_{0.025}$ ,  $\widehat{W}_o > c_{0.05,W}$ ) is given in Table 7 and in the last column of Table 8. As expected, the power will be relatively low.

6.2. The type I error control of the joint test

The control of the type I error of the joint test will be investigated in this section.

It suffices to show that the type I error of the joint test is controlled at the positive ( $\Delta_2|X_{1,P}<c$ ) – axis. Let  $(0, (\Delta_2|X_{1,P}<c))$  be a point on the positive ( $\Delta_2|X_{1,P}<c$ ) – axis on the boundary of the joint null.

It is desired to show that

$$\alpha = P\left(\widehat{Z}_a > c_\alpha - \frac{\alpha_2(\Delta_2|X_{1,P}<c)}{\sqrt{V_{Z_a}}}, \widehat{W}_o > c_{\alpha,W}\right) \leq 0.025$$

where  $V_{Z_a} = var(\widehat{\alpha}_1\widehat{\Delta}_1 + \widehat{\alpha}_2(\widehat{\Delta}_2|X_{1,P}<c))|(0, (\Delta_2|X_{1,P}<c))$

First, consider the probability

$$P\left(\widehat{Z}_a > c_\alpha - \frac{\alpha_2(\Delta_2|X_{1,P}<c)}{\sqrt{V_{Z_a}}}\right) \tag{19}$$

where,

$$\begin{aligned} V_{Z_a} &= var\left(\widehat{\alpha}_1\widehat{\Delta}_1 + \widehat{\alpha}_2(\widehat{\Delta}_2|X_{1,P}<c)|H_{0, Adj}: (0, (\Delta_2|X_{1,P}<c))\right) \\ &= \widehat{\alpha}_1^2 var(\widehat{\Delta}_1) + \widehat{\alpha}_2^2 var(\widehat{\Delta}_2|X_{1,P}<c) \\ &\quad + 2\widehat{\alpha}_1\widehat{\alpha}_2 cov(\widehat{\Delta}_1, (\widehat{\Delta}_2|X_{1,P}<c)) \end{aligned}$$

$$\begin{aligned} (\Delta_2|X_{1,P}<c) &= (\mu_{2,T} - \mu_{2,P}) + (\rho_P\sigma_{2,P} - \rho_T\sigma_{2,T})\left(\frac{\varphi(\tau)}{\Phi(\tau)}\right) \cong \Delta_1 \\ &\quad + (\rho_P\sigma_{2,P} - \rho_T\sigma_{2,T})\left(\frac{\varphi(\tau)}{\Phi(\tau)}\right) \end{aligned}$$

Therefore, at the boundary point  $(0, (\Delta_2|X_{1,P}<c))$ , since  $\Delta_1 = 0$ , one has

$$(\Delta_2|X_{1,P}<c) \cong (\rho_P\sigma_{2,P} - \rho_T\sigma_{2,T})\left(\frac{\varphi(\tau)}{\Phi(\tau)}\right) \tag{20}$$

Now without loss in generality, it has been assumed that  $\sigma_{2,P} = \sigma_{2,T} = \sigma_{1,T} = \sigma_{1,P} = \sigma_1$ , therefore Eq. (20) reduces to

$$(\Delta_2|X_{1,P}<c) \cong \sigma_1(\rho_P - \rho_T)\left(\frac{\varphi(\tau)}{\Phi(\tau)}\right) \tag{21}$$

Consider now the variance and covariance terms in the denominator in Eq. (19).

$$var(\widehat{\Delta}_1) = \frac{\sigma_1^2}{n_{1,T}R_1}, \text{ assuming that } \sigma_{1,T}^2 = \sigma_{1,P}^2 = \sigma_1^2$$

$$\begin{aligned} var(\widehat{\Delta}_2|X_{1,P}<c) &= var\left(\widehat{\mu}_{2,T|X_{1,P}<c} - \widehat{\mu}_{2,P|X_{1,P}<c}\right) \\ &= \frac{1}{n_{2,T}} (var(X_{2,T}|X_{1,P}<c) + var(X_{2,P}|X_{1,P}<c)) \\ &= \frac{\sigma_1^2}{n_{2,T}} \left(2 + \left(\rho_P^2 + \rho_T^2\right) \left(\left[1 - \tau \frac{\varphi(\tau)}{\Phi(\tau)} - \left(\frac{\varphi(\tau)}{\Phi(\tau)}\right)^2\right] \sigma_1^2 - 1\right)\right), \end{aligned}$$

which

follows from Eqn. (4) and Eqn. (5), since

Table 7

Selected Powers and Sample Sizes at One-sided  $\alpha = 0.025$  for the Joint Test ( $\widehat{Z}_o, \widehat{W}_o$ ) at the Specified DRDS Design Parameter Values and the Hypothetical Distributions of a HDRS<sub>17</sub> Subscale Score under Treatment and Placebo as given in Table 1 (DRDS Design Parameter Values:  $r_1 = 2, r_2 = 1, c = 2.75, \gamma = 0.44$ ) ( $\Delta, \Gamma$ ) =  $\alpha_1^*\Delta_1 + \alpha_2^*\Delta_2, \Delta_1\Delta_2(c)$ .

$\mu_{1T}$	$\mu_{1P}$	$\Delta_1$	$\sigma_1$	$\rho_P$	$\rho_T$	$\Delta_{2C}$	$\sigma_{2C}$	$(\Delta, \Gamma)$	$1 - \beta$	$N_1$	$n_{1T}$	$n_{2T}$
3.50	3.10	0.40	2.42	0.80	0.20	1.48	3.23	(0.52, 0.59)	80%	951	317	139
									85%	1056	352	155
									90%	1194	398	175
								(0.46, 0.38)	80%	1218	406	179
									85%	1350	450	198
									90%	1524	508	224

Table 8

Powers for the Combination, the Consistency and the Joint Tests at  $c_\alpha = 1.96, c_{0.05,W} = 1.60$  for the Specified DRDS Design Parameter Values and the Hypothetical Distributions of a HDRS<sub>17</sub> Subscale Score (DRDS Design Parameter Values:  $r_1 = 2, r_2 = 1, \gamma = 0.44, c = 2.75$  for First Scenario,  $\gamma = 0.42, c = 2.50$  for Second Scenario).

$\mu_{1T}$	$\mu_{1P}$	$\Delta_1$	$\sigma_1$	$\rho_P$	$\rho_T$	$\Delta_{2C}$	$\sigma_{2C}$	$N_1$	$P(\widehat{Z}_o > c_{0.025} H_a)$	$P(\widehat{W}_o > c_{0.05,W} H_a)$	$P(\widehat{Z}_o > c_{0.025}, \widehat{W}_o > c_{0.05,W} H_a)$									
3.50	3.50	0.00	2.42	0.80	0.80	1.48	3.23	750	0.025	0.050	0.001									
									3.10	0.81	0.74	0.66								
										0.85	0.79	0.73								
										0.91	0.85	0.82								
									3.00	0.30	2.42	2.42	0.80	0.20	1.43	3.18	750	0.64	0.55	0.41
																		840	0.69	0.59
	990	0.76	0.66	0.57																
	0.80	0.51	0.46	0.26																
		840	0.56	0.50	0.31															
		990	0.63	0.57	0.39															



$$\begin{aligned} \text{var}(X_{2,T}|X_{1,P}<C) &= \left(\rho_T^2 \left[1 - \tau \frac{\varphi(\tau)}{\Phi(\tau)} - \left(\frac{\varphi(\tau)}{\Phi(\tau)}\right)^2\right] \sigma_{1,P}^2 + (1 - \rho_T^2)\right) \sigma_{2,T}^2 \\ \text{var}(X_{2,P}|X_{1,P}<C) &= \left(\rho_P^2 \left[1 - \tau \frac{\varphi(\tau)}{\Phi(\tau)} - \left(\frac{\varphi(\tau)}{\Phi(\tau)}\right)^2\right] \sigma_{1,P}^2 + (1 - \rho_P^2)\right) \sigma_{2,P}^2 \end{aligned}$$

and from the further assumptions that  $\sigma_{2,P} = \sigma_{2,T} = \sigma_{1,T} = \sigma_{1,P} = \sigma_1$ .

Now,

$$\begin{aligned} &\text{cov}(\widehat{\Delta}_1, (\widehat{\Delta}_2|X_{1,P}<C)) \\ &= \frac{1}{n_{1P}} (\text{cov}(X_{1,P}, X_{2,P}|X_{1,P}<C) - \text{cov}(X_{1,P}, X_{2,T}|X_{1,P}<C)), \text{ since} \\ &\widehat{\Delta}_1 = \widehat{\mu}_{1,T} - \widehat{\mu}_{1,P} \text{ and } (\widehat{\Delta}_2|X_{1,P}<C) = \widehat{\mu}_{2,T|X_{1,P}<C} - \widehat{\mu}_{2,P|X_{1,P}<C}. \\ &= \frac{1}{n_{1P}} (\rho_P \sigma_{2,P} - \rho_T \sigma_{2,T}) \left[1 - \tau \left(\frac{\varphi(\tau)}{\Phi(\tau)}\right) - \left(\frac{\varphi(\tau)}{\Phi(\tau)}\right)^2\right] \sigma_{1,P}, \text{ since} \\ &\text{cov}(X_{1,P}, X_{2,P}|X_{1,P}<C) = \rho_P \left[1 - \tau \left(\frac{\varphi(\tau)}{\Phi(\tau)}\right) - \left(\frac{\varphi(\tau)}{\Phi(\tau)}\right)^2\right] \sigma_{1,P} \sigma_{2,P} \text{ and} \\ &\text{cov}(X_{1,P}, X_{2,T}|X_{1,P}<C) = \rho_T \left[1 - \tau \left(\frac{\varphi(\tau)}{\Phi(\tau)}\right) - \left(\frac{\varphi(\tau)}{\Phi(\tau)}\right)^2\right] \sigma_{1,P} \sigma_{2,T}. \\ &= \frac{1}{n_{1T} r_1} (\rho_P \sigma_{2,P} - \rho_T \sigma_{2,T}) \left[1 - \tau \left(\frac{\varphi(\tau)}{\Phi(\tau)}\right) - \left(\frac{\varphi(\tau)}{\Phi(\tau)}\right)^2\right] \sigma_{1,P} \\ &= \frac{\sigma_1^2}{n_{1T} r_1} (\rho_P - \rho_T) \left[1 - \tau \left(\frac{\varphi(\tau)}{\Phi(\tau)}\right) - \left(\frac{\varphi(\tau)}{\Phi(\tau)}\right)^2\right], \text{ under the further} \\ &\text{assumptions that } \sigma_{2,P} = \sigma_{2,T} = \sigma_{1,T} = \sigma_{1,P} = \sigma_1 \end{aligned}$$

Hence,

$$\begin{aligned} \alpha &= P\left(\widehat{Z}_a > c_\alpha - \frac{\alpha_2(\Delta_2|X_{1,P}<C)}{\sqrt{V_{Z_a}}}\middle| (0, (\Delta_2|X_{1,P}<C))\right) \\ &= P\left(\widehat{Z}_a > c_\alpha - \frac{\sqrt{n_{1,T}}(\rho_P - \rho_T) \left(\frac{\varphi(\tau)}{\Phi(\tau)}\right)}{\sqrt{V_{Z_a}}}\right) \end{aligned} \tag{22}$$

where now

$$\begin{aligned} V_{Z_a} &= \left(\frac{\alpha_1}{\alpha_2}\right)^2 \frac{1}{R_1} + \frac{1}{\gamma R_{12}} \left(2 + \left((\rho_T^2 + \rho_P^2) [h(\tau)\sigma_1^2 - 1]\right)\right) \\ &\quad + 2 \left(\frac{\alpha_1}{\alpha_2}\right) \frac{1}{r_1} (\rho_P - \rho_T) h(\tau) \\ &\text{with } h(\tau) = \left[1 - \tau \frac{\varphi(\tau)}{\Phi(\tau)} - \left(\frac{\varphi(\tau)}{\Phi(\tau)}\right)^2\right]. \end{aligned}$$

The power given by Eq. (22) for the combination test is essentially a function of the population parameters  $\rho_P$  and  $\rho_T$  from the two cohorts ( $P \rightarrow P$ ) and ( $P \rightarrow T$ ), the variance  $\sigma_1^2$ , and the standardized response threshold  $\tau$ . The hazard  $(\varphi(\tau)/\Phi(\tau))$  and  $\gamma = \Phi(\tau)$  are in turn influenced by the parameter  $\tau$ . The design parameters may be considered as fixed.

Type I Error rate at a boundary point  $(0, (\Delta_2|X_{1,P}<C))$  on the positive  $(\Delta_2|X_{1,P}<C)$ - axis which is in the consistency null space but in the alternative space of the adjusted treatment null is given by

$$\begin{aligned} &P\left(\widehat{Z}_a > c_\alpha - \frac{\alpha_2(\Delta_2|X_{1,P}<C)}{\sqrt{V_{Z_a}}}\ \& \ \widehat{W}_o > c_{\alpha,W}\middle| (0, (\Delta_2|X_{1,P}<C))\right) \\ &= P\left(\widehat{Z}_a > c_\alpha - \frac{\alpha_2(\Delta_2|X_{1,P}<C)}{\sqrt{V_{Z_a}}}\middle| (0, (\Delta_2|X_{1,P}<C))\right) \\ &\quad \times P\left(\widehat{W}_o > c_{\alpha,W}\middle| (0, (\Delta_2|X_{1,P}<C))\right) \end{aligned}$$

since  $\widehat{Z}_a$  and  $\widehat{W}_o$  are asymptotically independent.

$$\leq P\left(\widehat{Z}_a > c_\alpha - \frac{\sqrt{n_{1,T}}(\rho_P - \rho_T) \left(\frac{\varphi(\tau)}{\Phi(\tau)}\right)}{\sqrt{V_{Z_a}}}\right) \times 0.05,$$

where  $V_{Z_a}$  is as given above.

Table 9a, Tables 9b and 9c provide the type I error rates for the joint test at boundary points on the positive  $(\Delta_2|X_{1,P}<C)$  - axis derived from selected values of the parameters  $\rho_P, \rho_T, \sigma_1 = \sigma_{1,P}, \kappa = \sigma_{1,T}/\sigma_{1,P} = \sigma_{2,T}/\sigma_{2,P}$  and  $\tau$  with the allocation ratios fixed at  $r_1 = 2$  and  $r_2 = 1$  as in the Example given in Table 1.

It can be seen from the first panels of Tables 9a and 9b which are based on the example given in Table 1 that the type I error rates are controlled for various response thresholds. Under the

**Table 9a**

Type I Error Rate for Joint Test at a Boundary Point on the Positive  $(\Delta_2|C)$ -Axis for Selected Parameter Values of  $\rho_T, \rho_P, \sigma_1$  and  $\tau$  (DRDS Design Parameter Values:  $r_1 = 2, r_2 = 1$ ) ( $\Delta, \Gamma = (\alpha_1 \Delta_1 + \alpha_2 (\Delta_2|X_{1,P}<C)), \Delta_1 (\Delta_2|X_{1,P}<C)$ )  $P(\widehat{Z}_a > c_{0.025} - \alpha_2 (\Delta_2|X_{1,P}<C)/\text{std}(\widehat{\Delta}) \ \& \ \widehat{W}_o > c_{0.025,W} | (0, \Delta_2|X_{1,P}<C))$ .

$\rho_P$	$\rho_T$	$\sigma_1$	$\kappa$	$\tau$	$\gamma = \Phi(\tau)$	$\phi(\tau)/\Phi(\tau)$	$\Delta_2 X_{1,P}<C$	Type I error rate
0.80	0.20	2.40	1.0	-0.60	0.274	1.215	1.75	0.0041
				-0.30	0.382	0.998	1.44	0.0047
				0.00	0.500	0.798	1.15	0.0049
				0.30	0.618	0.618	0.89	0.0047
				0.60	0.726	0.459	0.66	0.0040
				0.90	0.834	0.300	0.44	0.0034
		0.5	-0.60	0.274	1.215	2.04	0.0050	
			-0.30	0.382	0.998	1.68	0.0058	
			0.00	0.500	0.798	1.34	0.0061	
			0.30	0.618	0.618	1.04	0.0057	
			0.60	0.726	0.459	0.77	0.0048	
			0.90	0.834	0.300	0.54	0.0041	
	1.20	1.0	-0.60	0.274	1.215	0.87	0.0074	
			-0.30	0.382	0.998	0.72	0.0089	
			0.00	0.500	0.798	0.57	0.0094	
			0.30	0.618	0.618	0.44	0.0086	
			0.60	0.726	0.459	0.33	0.0069	
			0.90	0.834	0.300	0.25	0.0057	
		0.5	-0.60	0.274	1.215	1.02	0.0141	
			-0.30	0.382	0.998	0.84	0.0174	
			0.00	0.500	0.798	0.67	0.0184	
			0.30	0.618	0.618	0.52	0.0166	
			0.60	0.726	0.459	0.39	0.0129	
			0.90	0.834	0.300	0.29	0.0104	

**Table 9b**

Type I Error Rate for Joint Test at a Boundary Point on the Positive ( $\Delta_2|c$ )-Axis for Selected Parameter Values of  $\rho_P, \rho_T, \sigma_1$  and  $\tau$  (DRDS Design Parameter Values:  $r_1 = 2, r_2 = 1, N_1 = 990$ ) ( $\Delta, \Gamma = (\alpha_1 * \Delta_1 + \alpha_2 * (\Delta_2|X_{1,P} < c), \Delta_1 (\Delta_2|X_{1,P} < c)) P(\bar{Z}_a > c_{0.025} - \alpha_2 (\Delta_2|X_{1,P} < c) / \text{std}(\hat{\Delta}) \& \bar{W}_o > c_{0.025,W} | (0, \Delta_2|X_{1,P} < c)$ ).

$\rho_P$	$\rho_T$	$\sigma_1$	$\kappa$	$\tau$	$\gamma = \Phi(\tau)$	$\phi(\tau)/\Phi(\tau)$	$\Delta_2 X_{1,P} < c$	Type I error rate
0.90	0.10	2.40	1.0	-0.60	0.274	1.215	2.33	0.0054
				-0.30	0.382	0.998	1.92	0.0064
				0.00	0.500	0.798	1.53	0.0068
				0.30	0.618	0.618	1.19	0.0063
				0.60	0.726	0.459	0.88	0.0052
				0.90	0.834	0.300	0.63	0.0042
			0.5	-0.60	0.274	1.215	2.48	0.0058
				-0.30	0.382	0.998	2.04	0.0068
				0.00	0.500	0.798	1.63	0.0072
				0.30	0.618	0.618	1.26	0.0067
				0.60	0.726	0.459	0.94	0.0055
				0.90	0.834	0.300	0.63	0.0042
		1.20	1.0	-0.60	0.274	1.215	1.17	0.0124
				-0.30	0.382	0.998	0.96	0.0153
				0.00	0.500	0.798	0.77	0.0161
				0.30	0.618	0.618	0.59	0.0146
				0.60	0.726	0.459	0.44	0.0114
				0.90	0.834	0.300	0.63	0.0269
			0.5	-0.60	0.274	1.215	1.24	0.0236
				-0.30	0.382	0.998	1.02	0.0285
				0.00	0.500	0.798	0.81	0.0296
				0.30	0.618	0.618	0.63	0.0269
				0.60	0.726	0.459	0.47	0.0210
				0.90	0.834	0.300	0.63	0.0269

scenarios in the first panels of these tables, the type I errors are controlled. The lower panels of Tables 9a and 9b show that when the standard deviation  $\sigma_1 = \sigma_{1,P}$  decreases, the type I error rate increases and even more so when the ratio  $\kappa = \sigma_{1,T}/\sigma_{1,P}$  decreases. This is because the variance in the denominator of the test statistic is getting smaller. However, in practical applications, the ratio  $\kappa = \sigma_{1,T}/\sigma_{1,P}$  is not expected to deviate too much from 1 as shown by the example in Table 1. There are some inflation when the correlations  $\rho_P = 0.90, \rho_T = 0.10$  and  $\sigma_1 = \sigma_{1,P} = 0.5$  as shown in the bottom panel of Table 9b. However, interestingly, as Table 9c below shows, the type I error inflation under these scenarios can be controlled if one increases

the allocation ratio  $r_1$ .

As Table 9c illustrates, under these scenarios, the greatest type I error inflation occurs under equal allocation ratios and the type I error starts to decrease as the allocation ratio  $r_1$  increases while holding  $r_2 = 1$ . The reason why the type I error starts to decrease as the allocation ratio  $r_1$  increases is because for a fixed total sample size  $N_1$ , the sample size  $n_{1,T}$  allocated to treatment decreases as  $r_1$  increases. This results in a net decrease in the second term on the right side of the power formula in Eq. (22) and a corresponding reduction in power. This fact holds true across all scenarios. Therefore, from these tables, it appears that the type I error rate of the joint test is controlled at the one-sided

**Table 9c**

Type I Error Rate for Joint Test at a Boundary Point on the Positive ( $\Delta_2|c$ )-Axis for Selected Parameter Values of  $\rho_P, \rho_T, \sigma_1$  and  $\tau$  (DRDS Design Parameter Values:  $r_1 = 1, 2, 3, r_2 = 1, N_1 = 990$ ) ( $\Delta, \Gamma = (\alpha_1 * \Delta_1 + \alpha_2 * (\Delta_2|X_{1,P} < c), \Delta_1 (\Delta_2|X_{1,P} < c)) P(\bar{Z}_a > c_{0.025} - \alpha_2 (\Delta_2|X_{1,P} < c) / \text{std}(\hat{\Delta}) \& \bar{W}_o > c_{0.025,W} | (0, \Delta_2|X_{1,P} < c)$ ).

$\rho_P$	$\rho_T$	$\sigma_1$	$\kappa$	$r_1$	$\tau$	$\gamma = \Phi(\tau)$	$\phi(\tau)/\Phi(\tau)$	$\Delta_2 X_{1,P} < c$	Type I error rate		
0.90	0.10	2.40	0.5	1.0	-0.60	0.274	1.215	2.48	0.0072		
					-0.30	0.382	0.998	2.04	0.0086		
					0.00	0.500	0.798	1.63	0.0091		
					0.30	0.618	0.618	1.26	0.0083		
					0.60	0.726	0.459	0.94	0.0067		
					0.90	0.834	0.300	0.63	0.0042		
				2.0	-0.60	0.274	1.215	2.48	0.0058		
					-0.30	0.382	0.998	2.04	0.0068		
					0.00	0.500	0.798	1.63	0.0072		
					0.30	0.618	0.618	1.26	0.0067		
					0.60	0.726	0.459	0.94	0.0055		
					0.90	0.834	0.300	0.63	0.0042		
			3.0	-0.60	0.274	1.215	2.48	0.0048			
				-0.30	0.382	0.998	2.04	0.0056			
				0.00	0.500	0.798	1.63	0.0059			
				0.30	0.618	0.618	1.26	0.0056			
				0.60	0.726	0.459	0.94	0.0047			
				0.90	0.834	0.300	0.63	0.0042			
				1.20	0.5	1.0	-0.60	0.274	1.215	1.24	0.0298
							-0.30	0.382	0.998	1.02	0.0341
							0.00	0.500	0.798	0.81	0.0359
							0.30	0.618	0.618	0.63	0.0321
							0.60	0.726	0.459	0.47	0.0251
							0.90	0.834	0.300	0.63	0.0219
		2.0	-0.60		0.274	1.215	1.24	0.0236			
			-0.30		0.382	0.998	1.02	0.0285			
			0.00		0.500	0.798	0.81	0.0296			
			0.30		0.618	0.618	0.63	0.0269			
			0.60		0.726	0.459	0.47	0.0210			
			0.90		0.834	0.300	0.63	0.0219			
		3.0	-0.60	0.274	1.215	1.24	0.0186				
			-0.30	0.382	0.998	1.02	0.0229				
			0.00	0.500	0.798	0.81	0.0241				
			0.30	0.618	0.618	0.63	0.0219				
			0.60	0.726	0.459	0.47	0.0170				
			0.90	0.834	0.300	0.63	0.0219				

0.025 level under most reasonable scenarios where the correlations are not too extreme and the ratio  $\kappa = \sigma_{1,T}/\sigma_{1,P}$  is expected not to deviate too much from 1, when the allocation ratios are fixed at  $r_1 = 2$  and  $r_2 = 1$ . If in a given application, it appears that it may fall into a neighborhood of some scenarios where the type I error of the joint test may be inflated, one can consider increasing the allocation ratio  $r_1$  from 2 to a higher level so that the type I error will be under control. This is an interesting and unexpected useful property which is a byproduct of the fact that the weights in the adjusted treatment effect are independent of the allocation ratios so a DRDS design has the flexibility in the choice of the allocation ratios  $r_1$  and  $r_2$  as long as they satisfy the constraint  $1 \leq r_2 \leq r_1$ . Also note that the allocation ratio of  $r_1 = 1$  is unlikely to be adopted in practice, so the increase in  $r_1$  should only be considered relative to those scenarios where the type I error appears to be inflated under a DRDS design with an allocation ratio  $r_1 = 2 \geq r_2 \geq 1$ .

In summary, Table 9a–9c show that the type I error rate of the joint test is controlled under most practical situations with the allocation ratios fixed at  $r_1 = 2$  and  $r_2 = 1$ . In a given application, under a DRDS design with allocation ratio  $r_1 = 2$  and  $r_2 = 1$ , if the situation appears to fall in one of the scenarios where type I error inflation is anticipated, then one may consider increasing the allocation ratio  $r_1$  to a level greater than 2 so that the type I error will be controlled. However, as discussed above, the type I error is expected to be under control in most practical applications.

### 6.3. Hypothetical example on HDRS<sub>17</sub> Anxiety and Somatization subscale score data

The hypothetical values presented in Table 1 are those of the distributional parameters of the HDRS<sub>17</sub> Subscale score for treatment and placebo that are derived on the basis of an exploratory early phase 2 study with a DRDS design in subjects with major depressive disorder. Although the sample size for this study is very small, they are adequate for the purpose of illustration in this paper.

Using the Period 1 data in Table 1 for the distributional parameters of the HRDS<sub>17</sub> subscale score under treatment and placebo, a major depressive disorder trial with a DRDS is simulated, where the DRDS design parameters assumed the values of  $r_1 = 2$ ,  $\pi = 0.58$ ,  $\gamma = 0.42$ ,  $r_2 = 1$ , and a Period 1 sample size of  $N_1 = 750$ . For simplicity, it is assumed that the placebo dropout rate is 0 in this simulated trial. Assuming a correlation between  $\hat{\Delta}_1$  and  $\hat{\Delta}_2$  of  $\rho_{1,2} = 0$ , this sample size was chosen to have about 69% power for the combination test, 59% power for the consistency test and 48% power for the joint test. Thus, the sample size selected is somewhat underpowered for the tests. A summary of the DRDS study design features and the simulated trial outcome statistics are given in Table 10.

From the results of the simulated trial given in Table 10, one obtains the following results for the combination test  $\hat{Z}_o$  and the consistency test  $\hat{W}_o$ :

The combined statistic is given by  $\hat{\Delta} = \alpha_1 \hat{\Delta}_1 + \alpha_2 \hat{\Delta}_2 = 0.49$  with a standard error of  $s.e.(\hat{\Delta}) = 0.16$  and a 95% CI of (0.17, 0.81). The combination Test:  $\hat{Z}_o = 3.04$  has a p-value of  $p = 0.0012$ . For the consistency test, one has  $\hat{U}_1 = 1.55$ ,  $\hat{U}_2 = 4.34$  and  $\hat{W}_o = 6.72$  with a p-value of  $p = 0.015$  with 90% CI of (4.54, 8.90).

Thus, the estimate of an adjusted treatment effect of 0.49 given by the combined statistic  $\hat{\Delta}$  is obtained as a result of adjusting for the presence of placebo responders by increasing the weight  $\alpha_{NR} = 0.42$  placed on  $\Delta_{NR}$  to the weight 0.53 by an amount  $\alpha_2 \alpha_R = 0.19(0.58) = 0.11$ .

This simulated trial shows that the apparent treatment effect  $\Delta_1$  for Period 1 is estimated to be  $\hat{\Delta}_1 = 0.29$ , and the adjusted

treatment effect  $\Delta$  is estimated to be  $\hat{\Delta} = 0.49$ . The consistency test  $\hat{W}_o = 6.72$  with a p-value of 0.015 shows that the Period 1 and Period 2 treatment effect estimates  $\hat{\Delta}_1 = 0.29$  and  $\hat{\Delta}_2 = 1.35$  are consistent. Therefore, the evidence supports the adjusted treatment effect of  $\Delta = 0.49$  as the treatment effect for the intended study population  $\Omega$ .

## 7. Summary discussion

In psychiatric trials, the presence of a relatively high proportion of placebo responders has caused many trials using a traditional randomized parallel placebo-controlled trial to fail because the treatment effect as measured by the relative treatment difference has been diluted. Various authors (Liu et al. [1], Fava et al. [3], Chen et al. [4], Huang and Tamura [5], Ivanova et al. [6], Tamura and Huang [7] and Tamura et al. [8]) have proposed a DRDS design in an attempt to resolve this problem. In their proposed methods, a combination test with certain power optimality criterion to either test the apparent treatment null hypothesis of Period 1 or global null hypothesis which is defined as the joint apparent treatment null of Period 1 and the enriched treatment null of Period 2. The weights used in the combined statistics depend on the DRDS design allocation ratios and the combined statistics may provide biased estimates of the apparent treatment effect. More importantly, it is believed that the apparent treatment effect should not be the basis for evaluating the effectiveness of the treatment since the true treatment effect has been mitigated on account of the presence of placebo responders. It can underestimate the risk/benefit ratio and it can lead to overdosing recommendation. In this paper, the concept of an adjusted treatment effect is introduced which is a weighted combination of the apparent treatment effects from Period 1 and the treatment effect from Period 2 in a DRDS design where the weights are independent of the DRDS design allocation ratios. The adjusted treatment effect is invariant in the class of DRDS design subject to the restriction that  $1 \leq r_2 \leq r_1$  which will be satisfied in practical applications. It is shown that the adjusted treatment effect can be interpreted as an adjustment of the apparent treatment effect of Period 1 by a quantity that represents an appropriately weighted amount of the treatment effect (as represented by the treatment effect from Period 2) that has been nullified by the presence of placebo responders. Therefore, the adjusted treatment effect as defined does not bias the assessment of the treatment effect in favor of the treatment. Thus, Period 2 of a DRDS design should not be viewed as providing enriched treatment effect in order to bias the adjusted treatment effect through the combined statistic, but rather as providing a measure of the treatment effect in the absence of placebo response which is exactly the information needed to make the proper adjustment. The independence of the weights from the allocation ratios in a DRDS design would allow the design to retain its flexibility in its choice of allocation ratios subject to a certain minor restriction which is needed to assure the type I error control of the joint test.

A new combined statistic is derived to test the adjusted treatment null hypothesis. In order for the adjusted treatment effectiveness claim to be extendable to the intended study population, a consistency measure is introduced to assess the consistency between the treatment effects from the two periods. The general monotonicity condition which has been suggested by some as a criterion for extendibility of the treatment effectiveness claim to the intended study population appears to be too stringent because it is analogous to requiring the treatment to be at least as effective as the control in an active control trial. It is shown that the consistency condition is a natural generalization of the monotonicity condition and it is less stringent and does not

**Table 10**

Summary Statistics from a Simulated MDD Trial with the Specified DRDS Design Parameter Values and the Hypothetical Distributions of a HDRS<sub>17</sub> Subscale Score under Treatment and Placebo as given in the First Row of Tables 1 and 3 ( $r_1 = 2, R_1 = 2/3, \pi = 0.58, \gamma = 0.42, r_2 = 1, R_2 = 1/2$ ) ( $c_\alpha = 1.96, c_{\alpha,W} = 1.60, N_1 = 750$  with 70%, 59%, 48% Power for  $\hat{Z}_o, \hat{W}_o, (\hat{Z}_o, \hat{W}_o)$ ) ( $\mu_{1T} = 3.30, \sigma_{1T} = 2.44, \mu_{1P} = 3.00, \sigma_{1P} = 2.40$ ).

Period 1			Period 2			Test Statistics	
$n_{1T}$	$\hat{\mu}_{1,T}$	$\hat{\sigma}_{1,T}$	$n_{2T}$	$\hat{\mu}_{2,T X_{1,P}<C}$	$\hat{\sigma}_{2,T X_{1,P}<C}$	Combined Statistic $\hat{\Delta}$	0.49
250	3.28	2.48	105	2.89	2.42	var( $\hat{\Delta}$ )	0.026
						95% CI	(0.17, 0.81)
$n_{1P}$	$\hat{\mu}_{1,P}$	$\hat{\sigma}_{1,P}$	$n_{2P}$	$\hat{\mu}_{2,P X_{1,P}<C}$	$\hat{\sigma}_{2,P X_{1,P}<C}$	Combination Test $\hat{Z}_o$	3.04
500	2.99	2.35	105	1.54	2.07	p-value	0.0012
						Consistency Measure	
$\hat{\Delta}_1$	s.e.( $\hat{\Delta}_1$ )	var( $\hat{\Delta}_1$ )	$\hat{\Delta}_{2 X_{1,P}<C}$	s.e.( $\hat{\Delta}_{2 X_{1,P}<C}$ )	var( $\hat{\Delta}_{2 X_{1,P}<C}$ )	$\hat{\Delta}_1 \hat{\Delta}_{2 X_{1,P}<C}$	0.39
0.29	0.18	0.034	1.35	0.44	0.19	Consistency Test $\hat{W}_o$	6.72
$\alpha_1$	$\alpha_2$	$\hat{\Delta}$	$\hat{U}_1$	$\hat{U}_2$	$\hat{\Delta}_1 \hat{\Delta}_{2 X_{1,P}<C}$	p-value	0.015
0.81	0.19	0.49	1.55	4.34	0.39	90% CI	(4.54, 8.90)

require the specification of a non-inferiority margin. It is suggested that the rejection of the consistency null by the consistency test should provide the additional evidence needed to be able to extend the adjusted treatment effectiveness claim to the intended study population.

Therefore, a joint test consisting of the combination test and the consistency test is proposed for testing the adjusted null and the consistency null. In most practical applications, the type I error of the joint test should be under control. Indeed the conditional probability structure underlying a DRDS design shows that the Period 2 treatment effect cannot be arbitrarily large. However, in a given application, if specific scenario suggests that the type I error may be inflated, then an appropriate choice of the allocation ratios can be selected for the DRDS design to assure the type I error control. The independence of the weights in the adjusted treatment effect from the allocation ratios in a DRDS design subject to a certain minor restriction would allow a DRDS design to retain this needed flexibility in its choice of the allocation ratios. The power of the joint test is not expected to be high and therefore the proposed methodology is not expected to increase efficiency compared to a standard randomized parallel design. But the proposed method would allow an unbiased estimate of the adjusted treatment effect which represents an appropriate assessment of the true treatment effect in the intended study population which is something that a standard randomized parallel design can never provide.

A successful outcome based on the proposed methodology should provide the confidence required of the evidence provided by a DRDS design to support the treatment effectiveness claim for the intended study population. The estimated adjusted treatment effect should also provide crucial information needed for making appropriate benefit/risk analysis and dosage recommendation.

**Acknowledgment**

The authors wish to thank Kim DeWoody and Shif Mariam for their consistent interest and support of research. In addition,

appreciation is extended to Hung Kung Liu for his insights and suggestions all of which helped to improve the content and guide the direction of this paper. The first author wishes to thank Ed Davis for the opportunity to be initiated into clinical trials at UNC in the early years and to Satya Dubey, Bob O'Neill, Ray Lipicky and Bob Temple for having imbued the necessary regulatory perspective during the years at FDA which is inherent in the proposed formulation and approach to this problem as presented in this paper.

**References**

- [1] Q. Liu, P. Lim, J. Singh, D. Lewin, B. Schwab, Kent, Doubly randomized delayed start design for enrichment studies with responders or non-responders, *J. Biopharm. Stat.* 22 (2012) 737–757.
- [2] M. Fava, A.E. Evins, D.J. Dorer, D. Schoenfeld, The problem of the placebo response in clinical trials for psychiatric disorders: culprits, possible remedies, and a novel study design approach, *Psychother. Psychosom.* 72 (2003) 115–227.
- [3] R.J. Temple, Special study designs: early escape, enrichment, studies in non-responders, *Commun. Stat. – Theory Methods* 23 (1994) 499–531.
- [4] Y.F. Chen, Y. Yang, H.M.J. Hung, S.J. Wang, Evaluation of performance of some enrichment designs dealing with high placebo response in psychiatric clinical trials, *Contemp. Clin. Trials* 32 (4) (2011) 592–604.
- [5] X. Huang, R.N. Tamura, Comparison of test statistics for the sequential parallel design, *Stat. Biopharm. Res.* 2 (1) (2010) 42–50.
- [6] A. Ivanova, B. Qaqish, D. Schoenfeld, Optimality, sample size and power calculations for the sequential parallel comparison design, *Stat. Med.* 30 (23) (2011) 2793–2803.
- [7] R. Tamura, X. Huang, An examination of the efficiency of the sequential parallel design in psychiatric clinical trials, *Clin. Trials* 4 (2007) 309–317.
- [8] R. Tamura, X. Huang, D. Boos, Estimation of treatment effect for the sequential parallel design, *Stat. Med.* 30 (30) (2011) 3496–3506.
- [9] N.L. Johnson, S. Kotz, *Distributions in Statistics: Continuous Multivariate Distributions*, John Wiley & Sons, Inc., New York, 1972.
- [10] A.V. Gajjar, K. Subrahmaniam, On the sample correlation coefficient in the truncated bivariate normal distribution, *Commun. Stat. Ser. B* 7 (5) (1978) 455–477.
- [11] S. Rosenbaum, Moments of a truncated bivariate normal distribution, *J. R. Stat. Soc. Ser. B* 23 (2) (1961) 405–408.
- [12] S.M. Shah, N.T. Parikh, Moments of single and doubly truncated standard bivariate normal distribution, *Vidya (Gujarat Univ.)* 7 (1964) 82–91.
- [13] G.M. Tallis, The moment generating function of the truncated multi-normal distribution, *J. R. Stat. Soc. Ser. B* 23 (1961) 223–229.
- [14] R.J. Serfling, *Approximation Theorems of Mathematical Statistics*, Wiley, New York, 1980.