On Clinical Trials with a High Placebo Response Rate

George YH Chi*[1], Yihan Li[2], Yanning Liu[3], David Lewin[4], Pilar Lim[5]

[1]Janssen R&D, 920 US Highway 20S, Raritan, NJ 08869
[2]AbbVie, AP9A-2,1 North Waukegan Rd, North Chicago, IL 60064
[3,4,5]Janssen R&D, Trenton-Harbourton Road, Titusville, NJ 08568

Abstract

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 the apparent treatment effect since the true treatment effect has been diminished by the presence of a substantial proportion of placebo responders in the population. Analogous to an active control trial, the true treatment effect cannot be measured by the relative treatment difference. An appropriate assessment of the true 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 design.

Key Words: Adjusted treatment effect, combination test, consistency test, doubly randomized delayed start design, Joint test, Sequential parallel design with re-randomization

1. Background

1.1 The Sequential Enrichment Design

The problem of a high placebo response rate in clinical trials occurs in several therapeutic areas, but it is most often observed in trials involving subjects with psychiatric disorders. 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 extended to the entire intended study population. Another problem is that for a patient to be treated in practice, the patient has to be given placebo first to verify his/her placebo response status before the treatment can be prescribed and this would entail an ethical dilemma.

Fava et al. [2] proposed a sequential parallel design (SPD) 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 or on placebo 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

*Corresponding author: gchi@its.jnj.com Present: chionroad@gmail.com

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 if there is a differential placebo dropout rate between the two placebo arms at the end of the first period. 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 at the 2010 BASS Conference. This DRDS design involves re-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. 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, a generic DRDS design as depicted in Figure 1 may refer to a SPD ReR design or a SPD design with re-randomization if appropriate, and for convenience, some of the terminologies and notations used in Liu et al. [1] are adopted in this paper. The DRDS design has been accepted by the regulatory agencies as an innovative design. But there are some conceptual and statistical issues yet need to be resolved. To address these issues, a new methodological approach for a DRDS design that differs from the previous methods is proposed in this paper.

Table 1 displays a summary of the data from a small completed phase II study based on a DRDS design. The Period 1 data will be used later for illustrating the power and sample size of the various tests as well as generating a simulated trial using a DRDS design.

FIGURE 1

A Doubly Randomized Delayed Start (DRDS) Design for Trials with a High Placebo Response Rate

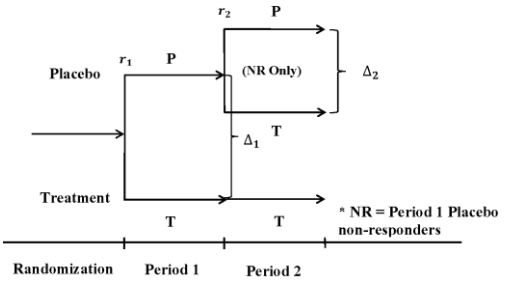


TABLE 1

Hypothetical Distributions of a $HDRS_{17}$ 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

μ_{1T}	σ_{1T}	μ_{1P}	σ_{1P}	Δ_1	σ_1
3.30	2.44	3.00	2.40	0.30	2.42

Period 2

μ_{2T}	σ_{2T}	μ_{2P}	σ_{2P}	Δ_2	σ_2
3.90	1.95	2.80	2.00	1.10	1.98

1.2 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 need to be satisfactorily resolved before a DRDS design can be applied to phase 3 trials to obtain the confirmatory evidence of effectiveness required. These issues will now be discussed.

Issue 1: The customary view considers the standard RDP design as the design of choice. But in a study population that has a substantial proportion of placebo responders, the relative treatment difference in a RDP design is only an apparent treatment effect, since it ignores the mitigating effect of the high placebo response rate on this difference. This problem is also present in the first period of a DRDS design. Therefore, the apparent treatment effect from the first period of a DRDS design would underestimate the treatment effect. Another inherent problem in this view is that even if perchance the

apparent treatment effect shows the treatment is superior to placebo (e.g. result of bias), any dosage recommendation based on an apparent dose-response relationship would likely lead to overdosing. For these two reasons, an appropriate assessment of the treatment effect adjusting for high placebo response rate is imperative.

Issue 2: A problem that is born of the view above 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 that are either unreasonable or unnecessarily stringent. The combination test that is derived from the combined statistic is used to test either the apparent treatment null hypothesis of Period 1, or a global null hypothesis which is defined as the joint apparent null of Period 1 and the enriched treatment null of Period 2. Ignoring the appropriateness of these assumptions, the rejection of these null hypotheses by their combination tests would not have solved the problem discussed in Issue 1 above.

Issue 3: A problem that is associated with the current method is that the weights used in the combined statistics depend on the placebo-to-treatment allocation ratios in Period 1 and Period 2 of a DRDS design. One can easily bias the estimate of the apparent treatment effect in favor of treatment by placing more weight on the Period 2 treatment effect estimate in the combined statistic which can be accomplished 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 is of concern. Furthermore, such inappropriate use of the Period 2 data and misleading interpretation of the purpose of the second period of a DRDS design (enrichment) are unfortunate and should be corrected.

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 true 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 assessment of the true 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.

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 the other problems discussed above.

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 were 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 follow a singly truncated bivariate normal distribution. The distributions of these cohorts in Period 2 are conditioned on 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 with some useful properties that are not available under the unconditional probability structure.

To address the above issues, a new methodology is proposed in this paper. The conditional probability structure underlying a DRDS design is described in Section 2. In Section 3, the key concept of an adjusted treatment effect is defined. In Section 4, a new combination test is derived. The concept of a consistency measure is introduced and a consistency test is defined in Section. In Section 6, a joint test composed of the combination and consistency tests is proposed for demonstrating that the treatment is effective for the intended study population. In Section 7, a simulated DRDS designed trial is presented for illustration. A summary discussion concludes the paper.

2. The DRDS Design and Its Underlying Probability Structure

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.

2.1 A DRDS Design Trial

Consider a trial with a DRDS design. Let $\Omega = \Omega_1$ denote the intended study population, and assume that there is a subpopulation of placebo responders Ω_R even though this subpopulation can't be characterized prior to the start of the trial. Let Ω_{NR} denote the placebo non-responder subpopulation. Let T denote an experimental treatment and P the placebo. In Period 1, $N_1 = n_1$ 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 denote a continuous clinical response variable of interest, $X_{1,T}$ and $X_{1,P}$ the response variables X under the treatment T and the placebo P respectively in Period 1. Let $X_{1,P} \sim N(\mu_{1,P}, \sigma_{1,P}^2)$ & $X_{1,T} \sim N(\mu_{1,T}, \sigma_{1,T}^2)$ 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, 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 $\{x_{1,P,i}, i = 1,2,\ldots,n_{1,P}\}$ and $\{x_{1,T,j}, j = 1,2,\ldots,n_{1,T}\}$ denote the observed sample responses from the placebo and treatment groups respectively. Then, $\hat{\Delta}_1 = (\hat{\mu}_{1,T} - \hat{\mu}_{1,P}) \sim N\left(\Delta_1, \frac{\sigma_1^2}{n_{1,T} R_1}\right)$, where $\hat{\mu}_{1,P} = \frac{1}{n_{1,P}} \sum_{i=1}^{n_{1,P}} x_{1,P,i}$, $\hat{\mu}_{1,T} = \frac{1}{n_{1,T}} \sum_{j=1}^{n_{1,T}} x_{1,T,j}$, and $R_1 = \frac{r_1}{1+r_1} = \frac{n_{1,P}}{n_{1,P}+n_{1,T}}$ is the fraction of placebo subjects among the entire sample of n_1 subjects. When the variances σ_1^2 and σ_2^2 of $\hat{\Delta}_1$ and $\hat{\Delta}_2$ from Period 1 and Period 2 are considered unknown, then one may estimate these unknown variances by their respective pooled sample variances given by $\hat{\sigma}_1^2 = \frac{(n_{1,T}-1)\, \hat{S}_{1,T}^2 + (n_{1,P}-1)\, \hat{S}_{1,P}^2}{(n_{1,T} + n_{1,P} - 2)}$ and $\hat{\sigma}_2^2 = \frac{(n_{2,T}-1)\, \hat{S}_{2,T}^2 + (n_{2,P}-1)\, \hat{S}_{2,P}^2}{(n_{2,T} + n_{2,P} - 2)}$ where $\hat{S}_{1,T}^2 = \frac{1}{(n_{1,T}-1)} \sum_{i=1}^{n_{1,T}} (X_{1,T,i} - \bar{X}_{1,T})^2$, $\hat{S}_{1,P}^2 = \frac{1}{(n_{1,P}-1)} \sum_{i=1}^{n_{1,P}} (X_{1,P,i} - \bar{X}_{1,P})^2$, $\hat{S}_{2,T}^2 = \frac{1}{(n_{2,T}-1)} \sum_{i=1}^{n_{2,T}} (X_{2,T,i} - \bar{X}_{2,T})^2$ and $\hat{S}_{2,P}^2 = \frac{1}{(n_{2,P}-1)} \sum_{i=1}^{n_{2,P}} (X_{2,P,i} - \bar{X}_{2,P})^2$. At the end of Period 1, a pre-specified threshold c for the response variable X will be applied to determine the response status of each placebo subject who completed the trial. At the end

of Period 1, those placebo subjects who are identified as responders ($X_{1,P} > c$), and along with the placebo dropouts will be excluded from the second period of the study. Those placebo subjects classified as non-responders ($X_{1,P} < c$), will be re-randomized to treatment and placebo at the start of Period 2. It will 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 $\tau = \frac{c-\mu_{1,P}}{\sigma_{1,P}}$ be the standardized response threshold relative to the placebo response distribution in Period 1. Let n_2 denote the number of placebo non-responders who completed Period 1 of the study and $\gamma = \Phi(\tau) = \Phi\left(\frac{c-\mu_{1,P}}{\sigma_{1,P}}\right)$ denote the population proportion of placebo non-responders in $\Omega = \Omega_1$. Then, the ratio $\hat{\gamma} = \frac{n_2}{n_{1,P}}$ 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 in a placebo-treatment allocation ratio of $r_2 \geq 1$. For practical reason, r_2 is usually set to the value 1. Then, it follows that $n_{2,T} = n_{2,P} = \frac{n_2}{1+r_2} = \frac{\gamma n_{1,P}}{1+r_2} = \frac{\gamma n_{1,T} r_1}{1+r_2} = n_{1,T}\gamma R_{1,2}$, where $R_{1,2} = \frac{r_1}{1+r_2}$. Now without loss in generality and for obvious reason, consider relabeling the entire placebo sample in Period 1 as follows:

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

where the first $n_{2,T}$ placebo subjects $\{X_{1,P,i}, i = 1,2,\ldots,n_{2,T}\}$ are placebo non-responders that have been re-randomized in Period 2 to treatment, and the next set of $n_{2,P}$ placebo subjects $\{X_{1,P,i}, i = n_{2,T}+1, n_{2,T}+2,\ldots,n_2,\}$ are placebo non-responders that have been re-randomized in Period 2 to placebo, while the remainder of the placebo sample $\{X_{1,P,i}, i = n_2+1, n_2+2,\ldots,n_{1,P}\}$ are the placebo subjects who were placebo responders (or placebo dropouts if any) in Period 1. Note that under equal allocation in Period 2, $n_{2,P} = n_{2,T} = \frac{n_{2,P}+n_{2,T}}{2} = \frac{n_2}{2} = \frac{\gamma n_{1,P}}{2}$. Assuming that the randomization in Period 1 holds, then the placebo sample should be representative of the population $\Omega = \Omega_1$. If the *entire* untruncated placebo sample at the end of Period 1were re-randomized in Period 2 to treatment, then the pair of response variables $(X_{1,P}, X_{2,T}) \sim N(\boldsymbol{\mu_{12,T}}, \boldsymbol{\Sigma_{12,T}})$ are bivariate normal with $\boldsymbol{\mu_{12,T}} = \begin{pmatrix}\mu_{1,P}\\ \mu_{2,T}\end{pmatrix}$ and $\boldsymbol{\Sigma_{12,T}} = \begin{pmatrix}\sigma_{1,P}^2 & \rho_T\sigma_{1,P}\sigma_{2,T}\\ \rho_T\sigma_{1,P}\sigma_{2,T} & \sigma_{2,T}^2\end{pmatrix}$ 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 where ρ_T is the correlation $corr(X_{1,P}, X_{2,T})$. Similarly, if the entire placebo sample at the end of Period 1 were re-randomized in Period 2 to placebo, then the pair $(X_{1,P}, X_{2,P}) \sim N(\mu_{12,P}, \Sigma_{12,P})$, $\boldsymbol{\mu_{12,P}} = \begin{pmatrix}\mu_{1,P}\\ \mu_{2,P}\end{pmatrix}$ & $\boldsymbol{\Sigma_{12,P}} = \begin{pmatrix}\sigma_{1,P}^2 & \rho_P\sigma_{1,P}\sigma_{2,P}\\ \rho_P\sigma_{1,P}\sigma_{2,P} & \sigma_{2,P}^2\end{pmatrix}$ 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 where ρ_P is the correlation $corr(X_{1,P}, X_{2,P})$. 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., $\rho_P - \rho_T \geq 0$.

2.2 Truncated Distributions of the Two Placebo Non-Responder Cohorts in Period 2

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

normal distribution $\left(\left(X_{1,P}|X_{1,P}<c\right),\left(X_{2,T}|X_{1,P}<c\right)\right)\sim N\left(\boldsymbol{\mu}_{\boldsymbol{12,T|X_{1,P}<c}},\boldsymbol{\Sigma}_{\boldsymbol{12,T|X_{1,P}<c}}\right)$ where

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

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

where $var\left(X_{1,P}\middle|X_{1,P}<c\right)=\left[1-\tau\frac{\varphi(\tau)}{\Phi(\tau)}-\left(\frac{\varphi(\tau)}{\Phi(\tau)}\right)^2\right]\sigma_{1,P}^2$, $var(X_{2,T}|X_{1,P}<c)=\left(\rho_T^2\left[1-\tau\frac{\varphi(\tau)}{\Phi(\tau)}-\left(\frac{\varphi(\tau)}{\Phi(\tau)}\right)^2\right]\sigma_{1,P}^2+(1-\rho_T^2)\right)\sigma_{2,T}^2$, $var(X_{2,P}|X_{1,P}<c)=\left(\rho_P^2\left[1-\tau\frac{\varphi(\tau)}{\Phi(\tau)}-\left(\frac{\varphi(\tau)}{\Phi(\tau)}\right)^2\right]\sigma_{1,P}^2+(1-\rho_P^2)\right)\sigma_{2,P}^2$, $cov\left(X_{1,P},X_{2,T}\middle|X_{1,P}<c\right)=\rho_T\left[1-\tau\frac{\varphi(\tau)}{\Phi(\tau)}-\left(\frac{\varphi(\tau)}{\Phi(\tau)}\right)^2\right]\sigma_{1,P}\sigma_{2,T}$ and the correlation for the truncated $(P\rightarrow T)$ cohort given by $\rho_{T|\boldsymbol{X_{1,P}<c}}=\frac{cov(X_{1,P},X_{2,T}|X_{1,P}<c)}{\sqrt{var(X_{1,P}|X_{1,P}<c)}\sqrt{var(X_{2,T}|X_{1,P}<c)}}$. In practice, $var\left(X_{1,P}\middle|X_{1,P}<c\right)$, $var(X_{2,T}|X_{1,P<c})$ and $cov(X_{1,P},X_{2,T}|X_{1,P<c})$ may be estimated by their respective sample variances and the sample covariance given by $S^2_{X_{1,P}|X_{1,P}<c}=\frac{1}{n_{2,T}}\sum_{i=1}^{n_{2,T}}(X_{1,P,i}-\hat{X}_{1,P|X_{1,P}<c})^2$, $S^2_{X_{2,T}|X_{1,P}<c}=\frac{1}{n_{2,T}}\sum_{i=1}^{n_{2,T}}(X_{2,T,i}-\hat{X}_{2,T|X_{1,P}<c})^2$ and $S_{(X_{1,P},\ X_{2,T}|X_{1,P}<c)}=\frac{1}{n_{2,T}}\sum_{i=1}^{n_{2,T}}(X_{1,P,i}-\hat{X}_{1,P|X_{1,P}<c})(X_{2,T,i}-\hat{X}_{2,T|X_{1,P}<c})$, where $\hat{X}_{1,P|X_{1,P}<c}=\frac{1}{n_{2,T}}\sum_{i=1}^{n_{2,T}}(X_{1,P,i}|X_{1,P}<c)$ and $\hat{X}_{2,T|X_{1,P}<c}=\frac{1}{n_{2,T}}\sum_{i=1}^{n_{2,T}}(X_{2,T,i}|X_{1,P}<c)$. The sample correlation is given by $\hat{\rho}_{T|X_{1,P}<c}=S_{(X_{1,P},\ X_{2,T}|X_{1,P}<c)}/\sqrt{S^2_{X_{1,P}|X_{1,P}<c}}\sqrt{S^2_{X_{2,T}|X_{1,P}<c}}$.

Similarly, for the cohort $(P\rightarrow P)$ in Period 2, the sample pairs $\left\{\left(X_{1,P,n_{2,T}+i},X_{2,P,i}\right),i=1,2,\ldots,n_{2,P}\right\}$ also follows a singly truncated bivariate normal distribution with $\left(\left(X_{1,P}|X_{1,P}<c\right),\left(X_{2,P}|X_{1,P}<c\right)\right)\sim N\left(\boldsymbol{\mu}_{\boldsymbol{12,P|X_{1,P}<c}},\boldsymbol{\Sigma}_{\boldsymbol{12,P|X_{1,P}<c}}\right)$, where

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

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

The expressions for the elements of the above variance-covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{12,P|X_{1,P}<c}}$ 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

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

which may be estimated by the observed mean difference $\left(\hat{\Delta}_2|X_{1,P}<c\right)=\hat{\mu}_{2,T\,|\,X_{1,P}<c}-\hat{\mu}_{2,P\,|\,X_{1,P}<c}$ where $\hat{\mu}_{2,T\,|\,X_{1,P}<c}=\hat{X}_{2,T|X_{1,P}<c}=\frac{1}{n_{2,T}}\sum_{i=1}^{n_{2,T}=n_{2,P}}(X_{2,T,i}|X_{1,P}<c)$ and $\hat{\mu}_{2,P\,|\,X_{1,P}<c}=\hat{X}_{2,P|X_{1,P}<c}=\frac{1}{n_{2,P}}\sum_{i=1}^{n_{2,P}=n_{2,T}}(X_{2,P,i}|X_{1,P}<c)$. Thus, $E\left(\hat{\Delta}_2|X_{1,P}<c\right)=\left(\Delta_2|X_{1,P}<c\right)$.

If the duration of Period 1 is relatively short, then in the above expression for $E\left(\hat{\Delta}_2|X_{1,P}<c\right)$, the first term $\left(\mu_{2,T}-\mu_{2,P}\right)=\left(\mu_{1,T}-\mu_{1,P}\right)$, which is the apparent treatment effect from Period 1; hence the increase in the expected treatment effect in Period 2 arises primarily

from the second term $(\rho_P\sigma_{2,P} - \rho_T\sigma_{2,T})\frac{\varphi(\tau)}{\Phi(\tau)}$ 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 $(\rho_P\sigma_{2,P} - \rho_T\sigma_{2,T}) > 0$. Some of the above expressions in the conditional distributions are well-known. See for example, Johnson and Kotz [5]. Other expressions can be directly derived from them.

2.2 The Joint Distribution of $\left(\hat{\Delta}_1, \left(\hat{\Delta}_2 | \boldsymbol{X_{1,P}} < \boldsymbol{c}\right)\right)$

Now with the above derivation of the conditional probability structure underlying the two cohorts $(P \to T)$ and $(P \to P)$ in Period 2, one can establish the following lemma.

Lemma: For a DRDS design, the treatment effect estimates $\hat{\Delta}_1$ and $\hat{\Delta}_2$ from Period 1 and Period 2 follow an asymptotically normal bivariate distribution, $\left(\hat{\Delta}_1, \left(\hat{\Delta}_2 | X_{1,P} < c\right)\right) \sim \boldsymbol{\Phi}(\boldsymbol{\mu_{12}}, \boldsymbol{\Sigma_{12}})$, where $\boldsymbol{\mu_{12}} = \begin{pmatrix} \Delta_1 \\ (\Delta_2 | X_{1,P} < c) \end{pmatrix} = \begin{pmatrix} \mu_{1,T} - \mu_{1,P} \\ (\mu_{2,T} - \mu_{2,P}) + (\rho_P\sigma_{2,P} - \rho_T\sigma_{2,T})\frac{\varphi(\tau)}{\Phi(\tau)} \end{pmatrix}$,

$\boldsymbol{\Sigma_{12}} = \begin{pmatrix} var(\hat{\Delta}_1) & cov\left(\hat{\Delta}_1, \left(\hat{\Delta}_2 | X_{1,P} < c\right)\right) \\ cov\left(\hat{\Delta}_1, \left(\hat{\Delta}_2 | X_{1,P} < c\right)\right) & var\left(\hat{\Delta}_2 | X_{1,P} < c\right) \end{pmatrix}$, and

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

$$var\left(\hat{\Delta}_2 | X_{1,P} < c\right) = \frac{1}{n_{2,T}}\left(var(X_{2,T} | X_{1,P} < c) + var(X_{2,P} | X_{1,P} < c)\right) \quad (3)$$

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

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

and $cov(\hat{\Delta}_1, (\hat{\Delta}_2 | X_{1,P} < c)) = cov((\hat{\mu}_{1,T} - \hat{\mu}_{1,P}), (\hat{\mu}_{2,T \mid X_{1,P}<c} - \hat{\mu}_{2,P \mid X_{1,P}<c}))$

$$= \frac{1}{n_{1P}}\left(cov(X_{1,P}, X_{2,P} | X_{1,P} < c) - cov(X_{1,P}, X_{2,T} | X_{1,P} < c)\right) \to 0 \quad (4)$$

The proof of this lemma will be omitted since these expressions can be directly derived from the parameters of the conditional distributions of the cohorts $(P \to P)$ and $(P \to T)$.

3. The Adjusted Treatment Effect

In a trial with high placebo response rate, the first problem encountered is the inability to characterize the subpopulation of placebo responders Ω_R. Therefore, if a traditional RDP 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 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 be smaller. This reduced treatment effect termed the *apparent treatment effect* is the reason why many such trials had failed in the past. Thus, a better assessment of the treatment effect for the intended study population is needed.

3.1 An Adjusted Treatment Effect

Under the assumption $\sigma_{1,P}^2 = \sigma_{1,T}^2$, denote their common variance by σ_1^2, and hence $\sigma_{\hat{\Delta}_1}^2 = \frac{\sigma_1^2}{n_{1,T}R_1}$. Similarly, one may assume without loss in generality that in Period 2, the conditional variances are equal, i.e., $\sigma_{2,T|X_{1,P}<c}^2 = var\left(X_{2,T} | X_{1,P} < c\right) = \sigma_{2,P|X_{1,P}<c}^2 =$

$var(X_{2,P}|X_{1,P} < c) = \sigma_2^2$, 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 $\sigma^2_{(\hat{\Delta}_2|X_{1,P}<c)}$, and hence here one has $\sigma^2_{(\hat{\Delta}_2|X_{1,P}<c)} = \frac{\sigma_2^2}{n_{2,T}R_2}$. If one were to combine the treatment effect estimate $\hat{\Delta}_1$ from Period 1 and $(\hat{\Delta}_2|X_{1,P} < c)$ from Period 2 using weights defined through their inverse variances following the method of weighted least square [6], then the least square estimator of the treatment effect is given by

$$\hat{\Delta} = \alpha_1\hat{\Delta}_1 + \alpha_2(\hat{\Delta}_2|X_{1,P} < c) \tag{5}$$

where the coefficients α_1 and α_2 are given in general by

$$\alpha_1 = 1 - \alpha_2 \text{ where } \alpha_2 = \frac{\frac{\sigma_1^2}{n_{1,T}R_1} - cov(\hat{\Delta}_1,(\hat{\Delta}_2|\, X_{1,P}<c))}{\frac{\sigma_1^2}{n_{1,T}R_1}+\frac{\sigma_2^2}{n_{2,T}R_2} - 2cov(\hat{\Delta}_1,(\hat{\Delta}_2|\, X_{1,P}<c))} \tag{6}$$

Now, since $cov(\hat{\Delta}_1, (\hat{\Delta}_2|\, X_{1,P} < c)) \to 0$ asymptotically as noted earlier, hence under large sample, α_2 in Eqn. (6) is 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 $n_{2,T} = n_{1,T}\gamma R_{12}$, and $\gamma = \Phi(\tau)$ is the population proportion of placebo non-responders which can be consistently estimated by the fraction of placebo non-responders at the end of Period 1under the previous assumptions.

Now, in a DRDS design, for practical reasons, the following constraints on the allocation ratios are expected $1 \le r_2 \le r_1$. Based on this restriction, the ratio $\left(\frac{R_1}{R_{12}R_2}\right)$ in the above expression for α_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

$$\alpha_2 = \frac{1}{1+\left(\frac{\sigma_2}{\sigma_1}\right)^2\frac{2}{\gamma}} \text{ and } \alpha_1 = 1 - \alpha_2 \tag{7}$$

which minimizes the weight α_2 placed on $(\Delta_2|X_{1,P} < c)$, the Period 2 treatment effect.

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

$$\Delta = \alpha_1\Delta_1 + \alpha_2(\Delta_2|X_{1,P} < c) \tag{8}$$

where the coefficients α_1 and α_2 are as defined in Eqn. (7).

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 constraint $1 \le r_2 \le r_1$. More importantly, α_2 represents the smallest possible weight assigned to Δ_2 under a DRDS design subject to the above restriction and α_2 is actually attained under a DRDS design with equal allocation. Also, with α_2 so defined, the actual DRDS design can still assume allocation ratios other than equal allocation provided the allocation ratios satisfy the above constraints. Thus, if a given DRDS design adopts an allocation ratio $r_1 > 1$, it will improve the precision of estimates, but the estimate of the adjusted treatment effect as defined in Eqn. (8) remains *invariant under the constraints*.

Remark 2: The fact that the weights defined in Eqn. (7) for the adjusted treatment effect as defined in Eqn. (8) are independent of the allocation ratios r_1 and r_2 as long as they satisfy the constraints $1 \le r_2 \le r_1$ allows one to freely choose a DRDS design with any allocation ratios r_1 and r_2 as long as they satisfy the constraints. This flexibility will be needed to assure the type I error control of the joint test to be discussed in Section 5.1.

Note: The combined statistic as given in Eqn. (5) will not necessarily retain the efficiency property of a least square estimator in light of the weights as defined in Eqn. (7) 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 allocation ratios to avoid bias favoring the treatment. Avoiding bias seems to be more important than optimal efficiency, because an appropriate definition of adjusted treatment effect is critical in a proper assessment of the treatment effect for the intended population.

3.2 Interpretation of the Adjusted Treatment Effect

If one were able to characterize the subpopulation Ω_R of placebo responders and the subpopulation Ω_{NR} of placebo non-responders, then for the overall study population $\Omega = \Omega_1$ in Period 1 of a DRDS design, the apparent treatment effect Δ_1 can be expressed as

$$\Delta_1 = \alpha_R\Delta_R + \alpha_{NR}\Delta_{NR} \quad (9)$$

Then, the adjusted treatment effect given by Eqn. (8) becomes

$$\Delta = \alpha_1\Delta_1 + \alpha_2\left(\Delta_2|X_{1,P} < c\right) = \alpha_1[\alpha_R\Delta_R + \alpha_{NR}\Delta_{NR}] + \alpha_2\left(\Delta_2|X_{1,P} < c\right)$$

where $\left(\Delta_2|X_{1,P} < c\right) \cong \Delta_{NR}$ assuming that the distribution of the placebo responders/non-responders among the placebo dropouts, if any, is the same as the population distribution. Hence, from the fact that $\alpha_1 = (1 - \alpha_2)$ and $\alpha_R = 1 - \alpha_{NR}$, one has

$$\Delta \cong (\alpha_R - \alpha_2\alpha_R)\Delta_R + (\alpha_{NR} + \alpha_2\alpha_R)\Delta_{NR} \quad (10)$$

Upon comparing Eqn. (9) and Eqn. (10), one notes that the adjusted treatment effect Δ as defined in Eqn. (8) can be viewed as a weighted average of Δ_R and Δ_{NR} as in Eqn. (9) for Δ_1 except now the weight for Δ_R has been decreased by the fractional amount $\alpha_2\alpha_R$ while the weight for Δ_{NR} has been increased by the same fractional amount $\alpha_2\alpha_R$. Thus, Eqn. (10) shows that the adjusted treatment effect Δ can be viewed as a weighted average of the treatment effect Δ_R and Δ_{NR} and hence represents a treatment effect for the intended 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 Ω_R in $\Omega = \Omega_1$ which can be seen as follows. Now, Eqn. (10) can also be rearranged as follows:

$$\Delta \cong \Delta_1 + \alpha_2[\alpha_R(\Delta_{NR} - \Delta_R)] \quad (11)$$

Now the quantity $[\alpha_R(\Delta_{NR} - \Delta_R)]$ represents the total amount of expected treatment effect Δ_{NR} that is not observed in Ω_R due to the placebo response in Ω_R. Furthermore, 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 Ω_R. Hence, in light of the weights used in the definition of adjusted treatment effect in Eqn. (8), it follows that $\alpha_2[\alpha_R(\Delta_2 - \Delta_R)]$ represents an appropriately weighted amount of $[\alpha_R(\Delta_2 - \Delta_R)]$ from Period 2 that needs to be added to the apparent treatment effect Δ_1 from Period 1 to account for the presence of placebo responders Ω_R.

4. The Combination Test

The adjusted treatment null hypothesis and its alternative are now defined as follows:

$$H_{o,Adj}: \Delta = \alpha_1\Delta_1 + \alpha_2\left(\Delta_2|X_{1,P} < c\right) \leq 0 \;\; vs. H_{a,Adj}: \Delta = \alpha_1\Delta_1 + \alpha_2\left(\Delta_2|X_{1,P} < c\right) > 0 \quad (12)$$

The adjusted null hypothesis is a stronger null hypothesis than the global null hypothesis because the parameter space defined by $\left\{\left(\Delta_1, \left(\Delta_2|X_{1,P} < c\right)\right) \mid \Delta_1 \leq 0 \;\&\; \left(\Delta_2|X_{1,P} < c\right) \leq 0\right\}$ includes the third quadrant as illustrated in Figure 2. Let the estimate of the adjusted treatment effect Δ be given by the least square estimator as defined by Eqn. (5) with weights defined by Eqn. (7), that is, $\hat{\Delta} = \alpha_1\hat{\Delta}_1 + \alpha_2\left(\hat{\Delta}_2|X_{1,P} < c\right)$

FIGURE 2

Region of the Parameter Space for the Adjusted Treatment Null

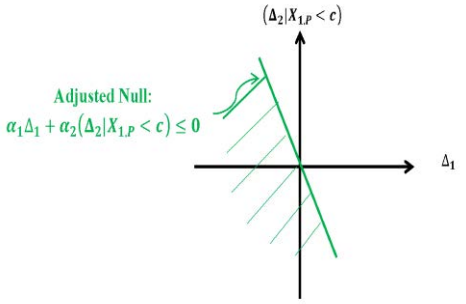
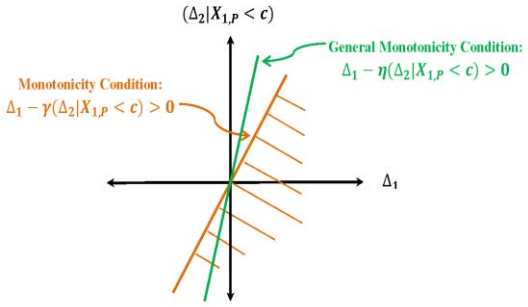


FIGURE 3

Region Defined by the General Monotonicity Condition



Then, it follows that $E(\hat{\Delta}) = \Delta$ and $var(\hat{\Delta}) = \Sigma_{\hat{\Delta}}^2 = \alpha_1^2 var(\hat{\Delta}_1) + \alpha_2^2 var(\hat{\Delta}_2|X_{1,P} < c) + 2\alpha_1\alpha_2 cov\left(\hat{\Delta}_1, (\hat{\Delta}_2|X_{1,P} < c)\right)$, where $var(\hat{\Delta}_1)$, $var(\hat{\Delta}_2|X_{1,P} < c)$ and $cov\left(\hat{\Delta}_1, (\hat{\Delta}_2|X_{1,P} < c)\right)$ are given in Eqn. (2), Eqn. (3) and Eqn. (4). The combination test for testing the adjusted null hypothesis is then given by $\hat{Z} = \frac{(\hat{\Delta}-\Delta)}{\sqrt{var(\hat{\Delta})}}$

$$= \frac{\left[\alpha_1\hat{\Delta}_1 + \alpha_2(\hat{\Delta}_2|X_{1,P} < c)\right] - \left[\alpha_1\Delta_1 + \alpha_2(\Delta_2|X_{1,P} < c)\right]}{\sqrt{\alpha_1^2(\gamma,\sigma_1,\sigma_2)var(\hat{\Delta}_1) + 2\alpha_1\alpha_2 cov\left(\hat{\Delta}_1, (\hat{\Delta}_2|X_{1,P} < c)\right) + \alpha_2^2(\gamma,\sigma_1,\sigma_2)var\left((\hat{\Delta}_2|X_{1,P} < c)\right)}}$$

4.1 The Type I Error, Power and Sample Size for the Combination Test

The type I error for the combination test is given by $\alpha = P(\hat{Z} > c_\alpha | H_{o,Adj}) = P(\hat{Z}_o > c_\alpha)$ where $\hat{Z}_o = \left(\frac{((\alpha_1\hat{\Delta}_1+\alpha_2(\hat{\Delta}_2|X_{1,P}<c)))}{\sqrt{var(\alpha_1\hat{\Delta}_1+\alpha_2(\hat{\Delta}_2|X_{1,P}<c))}}\right)$, $var\left(\alpha_1\hat{\Delta}_1 + \alpha_2(\hat{\Delta}_2|X_{1,P} < c)\right) = \alpha_1^2(\gamma,\sigma_1,\sigma_2)var(\hat{\Delta}_1)$ $+ 2\alpha_1\alpha_2 cov\left(\hat{\Delta}_1, (\hat{\Delta}_2|X_{1,P} < c)\right) + \alpha_2^2(\gamma,\sigma_1,\sigma_2)var\left((\hat{\Delta}_2|X_{1,P} < c)\right)$. The power of the combination test at a specified alternative (Δ_1, Δ_2) in the first quadrant is given by $1 - \beta = P(\hat{Z}_o > c_\alpha \,|\, H_{a,Adj}: (\Delta_1, \Delta_2)\ in\ 1st\ Quadrant, \rho_P > \rho_T) =$

$P\left(\hat{Z}_a > c_\alpha - \frac{\alpha_1\Delta_1+\alpha_2(\Delta_2|X_{1,P}<c)}{\Sigma_{\hat{\Delta},a}} | H_{a,Adj}: (\Delta_1, \Delta_2)\ in\ 1st\ Quadrant, \rho_P > \rho_T\right)$, where $\Sigma_{\hat{\Delta},a} = \Sigma_{\hat{\Delta},o}$ and $\hat{Z}_a = \frac{\left(\alpha_1\hat{\Delta}_1+\alpha_2(\hat{\Delta}_2|X_{1,P}<c)\right)-\left(\alpha_1\Delta_1+\alpha_2(\Delta_2|X_{1,P}<c)\right)}{\Sigma_{\hat{\Delta},a}} \sim N(0,1)$. From the above power function, the sample size formula is given by:

$$n_{1T} = \left(\frac{c_\alpha + c_{1-\beta}}{\alpha_1\Delta_1+\alpha_2(\Delta_2|X_{1,P}<c)}\right)^2 \left(\alpha_1^2\frac{\sigma_1^2}{R_1} + \alpha_2^2\frac{\sigma^2_{\hat{\Delta}_2|X_{1P}<c}}{\gamma R_{12}} + 2\alpha_1\alpha_2\rho_{1,2}\frac{\sigma_1}{\sqrt{R_1}}\frac{\sigma_{\hat{\Delta}_2|X_{1P}<c}}{\sqrt{\gamma R_{12}}}\right)$$

Table 2provides the power and sample size for selected scenarios based on Table 1 data.

5. Consistency

The null space as depicted in Figure 2 shows that there is an area in the alternative space that is situated inside the second quadrant. This suggests that even though the probability may be small, the adjusted treatment null may be rejected by the combination test, but the Period 1 treatment effect Δ_1 may be negative. From Eqn. (9), such a negative Δ_1 implies that the treatment may perform worse than placebo in the subpopulation Ω_R. Now in the subpopulation Ω_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

than placebo by a given non-inferiority margin $\delta > 0$. So, what should be an equivalent non-inferiority margin for assessing the effectiveness of a treatment in the subpopulation Ω_R? As a condition required for an apparent treatment effectiveness claim to be extendable to the intended population, Tamura et al. [7] introduced a monotonicity condition for the case under binary outcome. The condition requires that each placebo responder is also a treatment responder. Under binary outcome, this monotonicity condition is equivalent to requiring that the treatment be at least as effective as placebo. Now for continuous outcome, one can see from Eqn. (9) that this condition is equivalent to the following condition since under the earlier assumptions on the placebo dropouts if any, $\alpha_{NR} = \gamma = \Phi(\tau)$ and $\Delta_{NR} = \Delta_2$:

$$\Delta_1 = \alpha_R\Delta_R + \alpha_{NR}\Delta_{NR} > \gamma\left(\Delta_2|X_{1,P} < c\right) \text{ or } \left(\Delta_2|X_{1,P} < c\right) < \frac{1}{\gamma}\Delta_1 \quad (13)$$

This condition is depicted in Figure 3. It is clear that this condition is quite stringent and in addition, superiority is also not required for a non-inferiority trial. The reason Tamura et al [7] need a stringent condition is because they want to estimate the apparent treatment effect. 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 (see Figure 3) is to require that

$$\left(\Delta_2|X_{1,P} < c\right) < \eta\Delta_1, \text{ for some } \eta > \frac{1}{\gamma} \quad (14)$$

The slope η can be viewed here as the equivalent of a non-inferiority margin δ. But how should η be determined? This would be a challenging problem. But even the general monotonicity condition as defined by Eqn. (14) is unnecessarily stringent because the general monotonicity conditions as defined by Eqn. (14) places a constraint on the expected Period 2 treatment effect $\left(\Delta_2|X_{1,P} < c\right)$. This constraint is really not necessary because from Eqn. (1), one has

$$\left(\Delta_2|X_{1,P} < c\right) = \left(\mu_{2,T} - \mu_{2,P}\right) + \left(\rho_P\sigma_{2,P} - \rho_T\sigma_{2,T}\right)\left(\frac{\varphi(\tau)}{\Phi(\tau)}\right) \cong \Delta_1 + \left(\rho_P\sigma_{2,P} - \rho_T\sigma_{2,T}\right)\left(\frac{\varphi(\tau)}{\Phi(\tau)}\right) \quad (15)$$

From Eqn. (15), one can see that the magnitude of the expected Period 2 treatment effect $\left(\Delta_2|X_{1,P} < c\right)$ is determined by the magnitude of the Period 1 treatment effect Δ_1 and the term $\left(\rho_P\sigma_{2,P} - \rho_T\sigma_{2,T}\right)\left(\frac{\varphi(\tau)}{\Phi(\tau)}\right)$, whose magnitude cannot be arbitrarily large. Therefore, one should consider relaxing the condition by letting $\eta \to \infty$. Now if one let $\eta \to \infty$, the line $\left(\Delta_2|X_{1,P} < c\right) = \eta\Delta_1 \to$ the $\left(\Delta_2|X_{1,P} < c\right)$ – axis. This then naturally leads to the consistency condition to be introduced in the next section.

5.1 A Measure of Consistency and A Test for Consistency

Let the consistency measure Γ between Δ_1 and $\left(\Delta_2|X_{1,P} < c\right)$ be defined as $\Gamma = \Delta_1\left(\Delta_2|X_{1,P} < c\right)$. Then the consistency null and alternative hypotheses are defined as:

$$H_{o,C}: \Lambda = \Delta_1\left(\Delta_2|X_{1,P} < c\right) \le 0 \quad vs. \quad H_{o,C}: \Lambda = \Delta_1\left(\Delta_2|X_{1,P} < c\right) > 0 \quad (16)$$

The consistency null hypothesis is depicted by the shaded region in Figure 4. Now consider the following statistic: $\hat{\Gamma} = \hat{\Delta}_1\left(\hat{\Delta}_2|X_{1,P} < c\right) - \widehat{cov}\left(\hat{\Delta}_1, \left(\hat{\Delta}_2|X_{1,P} < c\right)\right)$. Then, one has $E\left(\hat{\Gamma}\right) = E\left(\hat{\Delta}_1\left(\hat{\Delta}_2|X_{1,P} < c\right) - cov\left(\hat{\Delta}_1, \left(\hat{\Delta}_2|X_{1,P} < c\right)\right)\right) = \Gamma$. The variance of $\hat{\Gamma}$ is given asymptotically by $var\left(\hat{\Gamma}\right) = \left[var\left(\hat{\Delta}_1\right)var\left(\hat{\Delta}_2|X_{1,P} < c\right)\right] + cov^2\left(\hat{\Delta}_1, \left(\hat{\Delta}_2|X_{1,P} < c\right)\right) + \left[\left(\Delta_2|X_{1,P} < c\right)^2 var\left(\hat{\Delta}_1\right)\right] + \left[\Delta_1^2 var\left(\hat{\Delta}_2|X_{1,P} < c\right)\right] + \left[4\Delta_1\left(\Delta_2|X_{1,P} < c\right)cov\left(\hat{\Delta}_1, \left(\hat{\Delta}_2|X_{1,P} < c\right)\right)\right] + \Delta_1^2\left(\Delta_2|X_{1,P} < c\right)^2$. The consistency test is then defined by: $\widehat{W} = \frac{\hat{\Gamma} - E\left(\hat{\Gamma}\right)}{\sqrt{var\left(\hat{\Gamma}\right)}} = \frac{\left[\hat{\Delta}_1\left(\hat{\Delta}_2|X_{1,P} < c\right) - \widehat{cov}\left(\hat{\Delta}_1, \left(\hat{\Delta}_2|X_{1,P} < c\right)\right)\right] - \Delta_1\left(\Delta_2|X_{1,P} < c\right)}{\sqrt{\left[var\left(\hat{\Delta}_1\right)var\left(\hat{\Delta}_2|X_{1,P} < c\right)\right] + cov^2\left(\hat{\Delta}_1, \left(\hat{\Delta}_2|X_{1,P} < c\right)\right) + \left[\left(\Delta_2|X_{1,P} < c\right)^2 var\left(\hat{\Delta}_1\right)\right] + \left[\Delta_1^2 var\left(\hat{\Delta}_2|X_{1,P} < c\right)\right] + \left[4\Delta_1\Delta_2 cov\left(\hat{\Delta}_1, \left(\hat{\Delta}_2|X_{1,P} < c\right)\right)\right] + \Delta_1^2\Delta_2^2}}$

where $\widehat{cov}\left(\hat{\Delta}_1, (\hat{\Delta}_2|X_{1,P}<c)\right) = \frac{1}{n_{1P}}\left(S_{X_{1,P}<c,X_{2,P}|X_{1,P}} - S_{X_{1,P}<c,X_{2,T}|X_{1,P}}\right)$, $S_{X_{1,P}<c,X_{2,P}|X_{1,P}}$ 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 $(P \to P)$ and $(P \to T)$.

FIGURE 4

Region of the Parameter Space for the Consistency Null

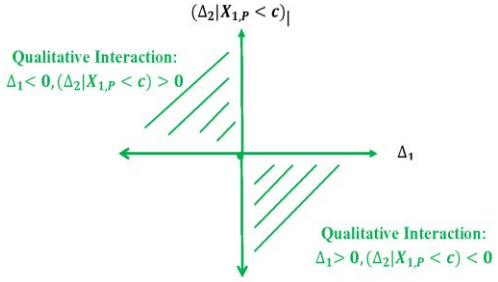


TABLE 2

Powers for the Combination, the Consistency and the Joint Tests at $c_\alpha = 0.025$, $c_{\alpha,\alpha\_W} = 1.60$ for the Specified DRDS Design Parameter Values and the Hypothetical Distributions of a DRDS$_{17}$ Subscale Score (DRDS Design Parameter Values: $r_1 = 2$, $r2 = 1$, $\gamma = 0.44$, for First Scenario, $\gamma = 0.42$ for Second Scenario))

μ_{11}	μ_{1F}	Δ_1	σ_1	ρ_P	ρ_Γ	Δ_{21C}	σ_{21C}	N_1	$P(\hat{Z}_o > c_{\alpha,025}\|H_a)$	$P(\hat{W}_O > c_{\alpha,025,W}\|H_a)$	$P(\hat{Z}_o > c_{\alpha,025}, \hat{W}_O > c_{\alpha,025,W}\|H_a)$
3.50	3.50	0.00	2.42	0.80	0.80	1.48	3.23	750	0.026	0.050	0.001
3.50	3.10	0.40	2.42	0.80	0.20	1.48	3.23	750	0.87	0.75	0.74
								840	0.90	0.79	0.78
								990	0.94	0.84	0.84
				0.80	0.50	0.96	3.32	750	0.77	0.60	0.59
								840	0.81	0.65	0.64
								990	0.87	0.72	0.71
3.30	3.00	0.30	2.42	0.80	0.20	1.43	3.18	750	0.70	0.59	0.48
								840	0.75	0.63	0.55
								990	0.81	0.68	0.65
				0.80	0.50	0.88	3.28	750	0.55	0.46	0.31
								840	0.60	0.49	0.36
								990	0.67	0.55	0.45

5.2 The Type I Error, Power and Sample Size for the Consistency Test

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

$$\alpha = P\left(\frac{\hat{\Delta}_1(\hat{\Delta}_2|X_{1,P}<c) - \widehat{cov}\left(\hat{\Delta}_1, (\hat{\Delta}_2|X_{1,P}<c)\right)}{\sqrt{\left[var(\hat{\Delta}_1)var(\hat{\Delta}_2|X_{1,P}<c)\right] + \left[(\Delta_2|X_{1,P}<c)^2 var(\hat{\Delta}_1)\right] + \left[\Delta_1^2 var(\hat{\Delta}_2|X_{1,P}<c)\right]}} > c_{\alpha,W}\right)$$

Since $\widehat{W}_o = \widehat{U}_1\widehat{U}_2 = \frac{\hat{\Delta}_1}{\sqrt{var(\hat{\Delta}_1)}} \frac{(\hat{\Delta}_2|X_{1,P}<c)}{\sqrt{var(\hat{\Delta}_2|X_{1,P}<c)}}$ is not normally distributed and has a distribution with heavy tail, its critical values are somewhat larger for the same significance level α as compared to the critical values from a normal distribution. In light of the joint test to be proposed later, it is recommended that the significance level for the consistency test be chosen at the one-sided 0.05 level with a critical value of 1.60 instead of the one-sided 0.025 level with a critical value of 2.18.

The Power of the Consistency Test is given by: $1-\beta = P(\widehat{W}_o > c_\alpha|H_{a,C}) = P\left(\widehat{W}_a > \frac{c_{\alpha,W}\sqrt{var(\widehat{W}_o)} - \Delta_1(\Delta_2|X_{1,P}<c)}{\sqrt{var(\widehat{W}_a)}}\,|H_{a,C}\right)$, where $\widehat{W}_a = \frac{\hat{\Delta}_1(\hat{\Delta}_2|X_{1,P}<c) - \widehat{cov}\left(\hat{\Delta}_1, (\hat{\Delta}_2|X_{1,P}<c)\right) - \Delta_1(\Delta_2|X_{1,P}<c)}{\sqrt{var(\widehat{W}_a)}}$,
$var(\widehat{W}_o) = var(\hat{\Delta}_1)var(\hat{\Delta}_2|X_{1,P}<c)$, $var(\widehat{W}_a) = var(\widehat{W}_o) + cov^2\left(\hat{\Delta}_1, (\hat{\Delta}_2|X_{1,P}<c)\right) + (\Delta_2|X_{1,P}<c)^2 var(\hat{\Delta}_1) + \Delta_1^2 var(\hat{\Delta}_2|X_{1,P}<c) + 4\Delta_1(\Delta_2|X_{1,P}<c)cov\left(\hat{\Delta}_1, (\hat{\Delta}_2|X_{1,P}<c)\right) + \Delta_1^2(\Delta_2|X_{1,P}<c)^2$. Note that the power can alsobe evaluated by viewing $\hat{\Delta}_1$ and $(\hat{\Delta}_2|X_{1,P}<c)$ as having an asymptotic bivariate normal distribution given by

$$1-\beta = P\left(\widehat{V}_2 > \frac{\left(c_{\alpha,W}\sqrt{1+\rho_{1,2}^2+U_1^2+U_2^2} + \rho_{1,2}\right) - U_1U_2 - U_2\widehat{V}_1}{U_1+\widehat{V}_1}\,|\,H_{a,QIA}\right)$$
$$= \int_{-\infty}^{\infty} \varphi_{\widehat{V}_1}(x) \int_{\frac{\left(c_{\alpha,W}\sqrt{1+\rho_{1,2}^2+U_1^2+U_2^2} + \rho_{1,2}\right) - U_1U_2 - U_2x - \rho_{1,2}x(x+U_1)}{(x+U_1)\sqrt{1-\rho_{1,2}^2}}}^{\infty} \varphi(y)\,dydx$$

where $U_i = \frac{\Delta_i}{\frac{\sigma_i}{\sqrt{n_{i,T}R_i}}}$, i=1,2 and $(\hat{V}_1, \hat{V}_2) \sim \boldsymbol{N}(\boldsymbol{\mu_{1,2}}, \boldsymbol{\Sigma_{1,2}^2})$ where $\boldsymbol{\mu_{1,2}} = \begin{pmatrix}0\\0\end{pmatrix}$ and $\boldsymbol{\Sigma_{1,2}^2} = \begin{pmatrix}1 & \rho_{1,2}\\ \rho_{1,2} & 1\end{pmatrix}$ and $\rho_{1,2} = corr\left(\hat{\Delta}_1, \left(\hat{\Delta}_2 | X_{1,P} < c\right)\right) = \rho_{\hat{\Delta}_1, (\hat{\Delta}_2 | X_{1,P} < c)}$ 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 $n_{1,T}$. Conversely, to calculate the sample size, one can simply 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 2 based on the example given 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 ($\hat{Z}_o > c_{0.025},\ \widehat{W}_o > c_{0.05,W}$) is proposed here for simultaneously testing the adjusted treatment null and the consistency null.

6.1 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 $\left(\Delta_2 | X_{1,P} < c\right)$ – axis. Let $\left(0, \left(\Delta_2 | X_{1,P} < c\right)\right)$ be a point on the positive $\left(\Delta_2 | X_{1,P} < c\right)$ – axis on the boundary of the joint null. The point $\left(0, \left(\Delta_2 | X_{1,P} < c\right)\right)$ is on the boundary of the consistency null, but is in the alternative space of the adjusted null. Hence, the type I error for the joint test is given by

$$\alpha = P\left(\hat{Z}_a > c_\alpha - \frac{\alpha_2\left(\Delta_2 | X_{1,P} < c\right)}{\sqrt{\alpha_1^2 var\left(\hat{\Delta}_1\right) + \alpha_2^2 var\left(\hat{\Delta}_2 | X_{1,P} < c\right) + 2\alpha_1\alpha_2 cov\left(\hat{\Delta}_1, \left(\hat{\Delta}_2 | X_{1,P} < c\right)\right)}}, \widehat{W}_o > c_{\alpha,W}\right)$$

In light of the Eqn. (15), one has $\left(\Delta_2 | X_{1,P} < c\right) \cong \Delta_1 + \left(\rho_P\sigma_{2,P} - \rho_T\sigma_{2,T}\right)\left(\frac{\varphi(\tau)}{\Phi(\tau)}\right)$. Therefore, at the boundary point $\left(0, \left(\Delta_2 | X_{1,P} < c\right)\right)$, since $\Delta_1 = 0$, one has $\left(\Delta_2 | X_{1,P} < c\right) \cong \left(\rho_P\sigma_{2,P} - \rho_T\sigma_{2,T}\right)\left(\frac{\varphi(\tau)}{\Phi(\tau)}\right)$. Since $\hat{Z}_a$ and $\widehat{W}_o$ are asymptotically independent, one has

$$\alpha \le P\left(\hat{Z}_a > c_\alpha - \frac{\sqrt{n_{1,T}}(\rho_P - \rho_T)\left(\frac{\varphi(\tau)}{\Phi(\tau)}\right)}{\sqrt{\left(\frac{\alpha_1}{\alpha_2}\right)^2\frac{1}{R_1} + \frac{1}{\gamma R_{12}}\left(2 + \left(\left(\rho_T^2 + \rho_T^2\right)\left(\left[1 - \tau\frac{\varphi(\tau)}{\Phi(\tau)} - \left(\frac{\varphi(\tau)}{\Phi(\tau)}\right)^2\right]\sigma_1^2 - 1\right)\right)\right) + 2\left(\frac{\alpha_1}{\alpha_2}\right)\frac{1}{r_1}(\rho_P - \rho_T)\left[1 - \tau\left(\frac{\varphi(\tau)}{\Phi(\tau)}\right) - \left(\frac{\varphi(\tau)}{\Phi(\tau)}\right)^2\right]}}\right) \times 0.05$$

Table 3a and Table 3b provide the type I error rates for the joint test at the boundary points on the positive $\left(\Delta_2 | X_{1,P} < c\right)$ – axis derived from selected values of the parameters ρ_P, ρ_T, $\sigma_1 = \sigma_{1,P}$, $\kappa = \frac{\sigma_{1,T}}{\sigma_{1,P}} = \frac{\sigma_{2,T}}{\sigma_{2,P}}$ and τ with the allocation ratios $r_1 = 1,2,3$ and $r_2 = 1$. As Table 3a & Table 3b illustrate, under a given scenario, 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 inside the power expression of the combination test on the right side of the above inequality and a corresponding reduction in the power of the combination test. This fact holds true across all scenarios. Therefore, from Table 3a & Table 3b, it suggests that the type I error rate of the joint test is controlled at the one-sided 0.025 level under most reasonable scenarios where the correlations are not too

extreme and the ratio $\kappa = \sigma_{1,T}/\sigma_{1,P}$ does not deviate too far from 1, while holding the allocation ratios 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 then 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 constraints $1 \le r_2 \le r_1$. 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 evaluated relative to those scenarios where the type I error appears to be inflated under a DRDS design with an allocation ratio $r_1 = 2 \ge r_2 \ge 1$.

TABLE 3a

Type I Error Rate for Joint Test at a Boundary Point on the Positive ($\Delta_2|X_{1,P}<c$)-Axis for Selected Parameter Values of ρ_T, ρ_P, σ_1 and τ (DRDS Design Parameter Values: $r_1 = 1, 2, 3$, $N_1 = 990$)

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

$P(\hat{Z}_o > c_{0.025} - \alpha_2(\Delta_2|X_{1,P} < c)/\text{std}(\hat{\Delta})\ \&\ \widehat{W}_o > c_{0.025,W}\,|(0, \Delta_2|X_{1,P} < c))$

ρ_P	ρ_T	σ_1	κ	r_1	τ	$\gamma = \Phi(\tau)$	$\phi(\tau)/\Phi(\tau)$	$\Delta_2\|X_{1,P} < c$	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
				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
				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

TABLE 3b

Type I Error Rate for Joint Test at a Boundary Point on the Positive ($\Delta_2|X_{1,P}<c$)-Axis for Selected Parameter Values of ρ_T, ρ_P, σ_1 and τ (DRDS Design Parameter Values: $r_1 = 1, 2, 3$, $N_1 = 990$)

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

$P(\hat{Z}_o > c_{0.025} - \alpha_2(\Delta_2|X_{1,P} < c)/\text{std}(\hat{\Delta})\ \&\ \widehat{W}_o > c_{0.025,W}\,|(0, \Delta_2|X_{1,P} < c))$

ρ_P	ρ_T	σ_1	κ	r_1	τ	$\gamma = \Phi(\tau)$	$\phi(\tau)/\Phi(\tau)$	$\Delta_2\|X_{1,P} < c$	Type I Error Rate
0.90	0.10	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
				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
				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

Figure 5 illustrates the rejection region for the joint test. The power of the joint test ($\hat{Z}_o > c_{0.025}$, $\widehat{W}_o > c_{0.05,W}$) is displayed in the last column of TABLE 2. As expected, the power will be relatively low.

7. A Simulated Trial with a DRDS Design

A major depressive disorder trial using the $HRDS_{17}$ subscale score data from Period 1 of Table 1is simulated using a DRDS design with $r_1 = 2$, $\pi = 0.58$, $\gamma = 0.42$, $r_2 = 1$, and $N_1 = 750$. For simplicity, it is assumed that the placebo dropout rate is 0 and a correlation between $\hat{\Delta}_1$ and $\hat{\Delta}_2$ of $\rho_{1,2} = 0$. This sample size provides 69% power for the combination test, 59% for the consistency test and 48% for the joint test. Thus, the trial is somewhat underpowered for the tests. A summary of the DRDS design and the simulated trial outcome statistics are given in Table 4.

The estimate of an adjusted treatment effect of 0.49 given by the combined statistic $\hat{\Delta}$ is obtained as a result of adjusting the apparent treatment effect $\hat{\Delta}_1 = 0.29$ of Period 1for the presence of placebo responders by increasing the weight α_{NR} placed on Δ_{NR} from 0.42 to the weight 0.53 by an amount $\alpha_2\alpha_R = 0.11$. This simulated trial shows that the apparent treatment effect Δ_1 for Period 1 is estimated to be $\hat{\Delta}_1 = 0.29$, and the adjusted treatment effect Δ is estimated to be $\hat{\Delta} = 0.49$ with a 95% CI of (0.17, 0.81). The consistency test $\widehat{W}_o = \widehat{U}_1 \times \widehat{U}_2 = 1.55 \times 4.34 = 6.72$ with a p-value of 0.015 and 95% CI of (4.54, 8.90) shows that the 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 an appropriate assessment of the true treatment effect for the intended study population $\Omega = \Omega_1$.

FIGURE 5

Rejection Region in the Alternative Space

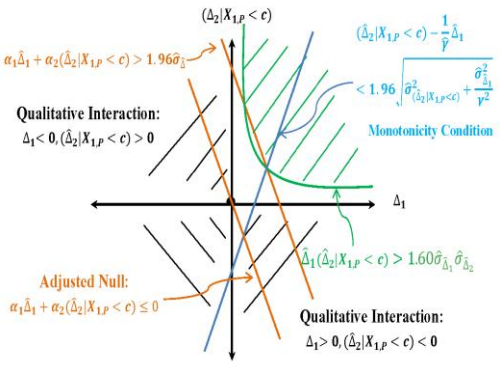


TABLE 4

Summary Statistics from a Simulated Trial with the Specified DRDS Design Parameter Values and the Hypothetical Distributions of a $HDRS_{17}$ Subscale Score given in Table 1

($r_1 = 2$, $R_1 = 2/3$, $\pi = 0.58$, $\gamma = 0.42$, $r_2 = 1$, $R_2 = 1/2$)

($c_\alpha = 1.96$, $c_{\alpha,W} = 1.60$, N1 = 750 with 70%, 59%, 48% Power for $\tilde{Z}_o$, $\tilde{W}_o$, $(\tilde{Z}_o, \tilde{W}_o)$)

($\mu_{1T} = 3.30$, $\sigma_{1T} = 2.44$, $\mu_{1P} = 3.00$, $\sigma_{1P} = 2.40$)

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}	$\hat{\mu}_{2,P\|X_{1,P}<C}$	$\hat{\sigma}_{2,P\|X_{1,P}<C}$		
500	2.99	2.35	105	1.54	2.07	Combination Test $\tilde{Z}_o$	3.04
						p-value	0.0012
$\hat{\Delta}_1$	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($\hat{\Delta}_{2\|X_{1,P}<C}$)	Consistency Measure	
0.29	0.18	0.034	1.35	0.44	0.19	$\hat{\Delta}_1\hat{\Delta}_{2\|X_{1,P}<C}$	0.39
						Consistency Test $\tilde{W}_o$	6.72
α_1	α_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)

8. Summary Discussion

A new methodology is introduced in this paper for the design and analysis of a DRDS design for investigating the treatment effect in a study population that is expected to have a high placebo response rate. The proposed approach includes a new concept of adjusted treatment effect, a new combination test, a new consistency generalizing the general monotonicity condition, and their joint test all of which are based on a conditional probability structure underlying a DRDS design. In order to maintain type I error control of the joint test, it is recommended that a phase 2 study be first conducted to assess the likely scenario for the specific application at hand. Then, an appropriate choice of the placebo-treatment allocation ratios for Period 1 and Period 2 can be made in the DRDS design for the confirmatory trial to assure the control of the type I error of the joint test.

Acknowledgment: The author wishes to thank Kim DeWoody for her consistent interest and support of this research, and also to express his appreciation to Hung Kung Liu for his critical comments and discussions which helped guide the direction of this paper.

References

[1] Liu Q, Lim P, Singh J, Lewin D, Schwab B, Kent (2012). Doubly randomized delayed start design for enrichment studies with responders or non-responders, *J. of Biopharmaceutical Statistics* 22: 737-757.

[2] Fava M, Evins AE, Dorer DJ, and Schoenfeld D (2003). The problem of the placebo response in clinical trials for psychiatric disorders: culprits, possible remedies, and a novel study design approach, *Psychotherapy and Psychosomatics* 72: 115-227.

[3] Temple RJ (1994). Special study designs: early escape, enrichment, studies in non-responders, *Communications in Statistics – Theory and Methods* 23: 499-531.

[4] Chen YF, Yang Y, Hung HMJ, Wang SJ (2011). Evaluation of performance of some enrichment designs dealing with high placebo response in psychiatric clinical trials, *Contemporary Clinical Trials* 1; 32(4): 592-604.

[5] Johnson NL and Kotz S (1972). Distributions in Statistics: Continuous Multivariate Distributions, John Wiley & Sons, Inc., New York.

[6] Serfling RJ. Approximation Theorems of Mathematical Statistics. Wiley, New York, 1980.

[7] Tamura R, Huang X, Boos D (2011). Estimation of treatment effect for the sequential parallel design, *Statistics in Medicine* 30(30): 3496-3506.