RESEARCH ARTICLE

Guidelines for calculating sample size in 2x2 crossover trials : a simulation study

N.M. Siyasinghe and M.R. Sooriyarachchi
Department of Statistics, Faculty of Science, University of Colombo, Colombo 03.

Abstract: In crossover trials, patients receive two or more treatments in a random order in different periods. The sample size determination is often an important step in planning a crossover study. This paper concerns sample size calculations in 2x2 crossover trials, with random patient effects and no interaction between the treatment and the patient under two scenarios, namely the exact and the large sample size approaches. Simulation was carried out for determining the sample size for both scenarios. For varying parameter values, simulation was used for generating samples of the required size and examining whether the significance level and power of the tests are maintained. The results indicate that when the sample size was ≤ 5, neither method maintained error rates and when the sample size was >5 and < 12 only the exact approach maintained error rates. However, when the sample size is approximately > 12 both methods maintained error rates. In addition it was found that a saving in sample size can be achieved depending on the extent of the correlation between the observations on the same patient. The simulation results indicate that crossover studies should not be conducted when the anticipated sample size is ≤ 5 and when a sample size of >5 and < 12 is anticipated, the exact method of determining sample size should be used. When larger sample sizes are anticipated either method can be used but the method based on large sample size approximation is simpler.

Keywords: Crossover trial, exact method, large sample method, sample size calculation, simulation.

INTRODUCTION

Crossover trials are clinical trials in which patients are given all the medications to be studied in a random order. According to Grizzle(1965) these studies are generally conducted on patients with chronic diseases to control their symptoms. The data are analyzed according to the original intention to treat.

Ideally, clinical trials should be large enough to reliably detect the smallest possible difference in the primary outcome with treatments that are considered clinically worthwhile. According to Lee *et al.* (2005), it is not uncommon for studies to be underpowered, failing to detect even large treatment effects because of inadequate sample size. It is considered unethical to recruit patients for a study that does not have a large enough sample size for the trial to deliver meaningful information on the tested intervention. Thus, sample size should be based on scientific considerations. Several approaches are discussed (Pocock, 1983; Julious & Patterson, 2004) for calculating sample size including the power approach and the confidence interval approach. According to previous studies (Chow *et al*. 2003;Woodward, 1992), these approaches require the specification of several parameters such as between treatment and within treatment variances for the treatments under consideration, the correlation within patients and the reference improvement, which is required to be detected. Chow *et al.* (2003) explain the two different approaches for 2x2 crossover designs but do not give guidelines on when to use specific approaches. In this study simulation is extensively used to examine the problem of setting guidelines.

In this study, two situations are considered in the calculation of sample size of a crossover study as explained in Chow *et al*. (2003). These are,

(i) The exact approach
(ii) The large sample (approximate) approach.

Further, the study gives guidelines for when to use the exact approach and the large sample approach and to study how much saving in sample size can be achieved when observations on the same patient are correlated.

---

[*] Corresponding author (roshini@mail.cmb.ac.lk)

METHODS AND MATERIALS

Mixed model used: In clinical trials, it is common to assume that the patients respond consistently to treatments. However, the assumption is invalid if the patients vary randomly in their responses to the drug. For this type of situation, a random subject effects model where the subject effect is considered to be random and the treatment and period effects are considered to be fixed has to be considered (Brown & Prescott, 2006).

Chow *et al*. (2003) have explained how to calculate the sample size in a crossover design using either of the two approaches, namely the exact and the approximate. In this paper a similar 2 × 2 crossover design comparing mean responses for two groups is considered.

In the first approach the test statistic is based on the Student's t distribution, whereas in the second approach the test statistic is based on the normal distribution. In the exact approach, the sample size depends on the degrees of freedom. The calculation of sample size is therefore not straightforward. The same calculation can be done without difficulty if the approximate approach, which is based on the normal approximation, is used. Values of the inverse t distribution function need to be determined for calculating the sample sizes for the exact approach. This is done by using the approximation given in Cooke *et al*. (1982). The criteria used for determining the method to be used for calculating sample size is based on which method maintains power and significance level.

Let Y_{ijk} be the response observed from the jth (j= 1,2,..,n) subject in the ith sequence (i = 1,2) under k^{th} treatment (k=1,2). The model considered is $Y_{ijk} = \mu + t_k + p_i + s_{ijk} + e_{ijk}$, where μ is the overall mean, t_k is the kth treatment effect, p_i is the ith sequence (period) effect , s_{ijk} is the random effect of the jth subject in the ith sequence under k^{th} treatment and e_{ijk} is the error term corresponding to the jth subject in the ith sequence under k^{th} treatment.

The following mixed model is used.

$$Y_{ijk} = \mu + t_k + p_i + s_{ijk} + e_{ijk} \qquad ...(1)$$

Here, treatment effect and period effect are considered as fixed effects and subject effect as random. In this study equal allocation of patients to treatment groups are assumed and no replication is considered.

Then define the following notation.

Since μ is a constant, we can take $\mu_k = \mu + t_k$. So the new model is;

$$Y_{ijk} = \mu_k + p_i + s_{jk} + e_{jk} \qquad ...(2)$$

Here it is assumed that there is no treatment by period interaction, since a simple hypothesis test can be used only under this assumption. The subject effects S_{ij1} , S_{ij2} are assumed to be independent and identically distributed as bivariate normal random variables with mean 0 and covariance matrix

$$\Sigma = \begin{bmatrix} \sigma^2_{BT} & \rho\sigma_{BT}\sigma_{BR} \\ \rho\sigma_{BT}\sigma_{BR} & \sigma^2_{BR} \end{bmatrix}$$

where σ^2_{BT} is the variance between patients for the 'treated' group, σ^2_{BR} is the variance between patients for the 'reference' group and ρ is the correlation between subjects in the treated and reference groups.

So, S_{ij1} and S_{ij2} have a bivariate normal distribution with mean 0 and variance – covariance matrix Σ . It is assumed that the errors e_{ij1} and e_{ij2} are such that

$$e_{ij1} \sim iid \; N(0, \sigma_{WT}{}^2)$$
$$e_{ij2} \sim iid \; N(0, \sigma_{WR}{}^2)$$

where, σ^2_{WT} is the within patient variation for the treated group and σ^2_{WR} is the within patient variation for the reference group.

Consider a group, which gets treatment 1 in the first period and treatment 2 in the second period, then the model can be written as follows,

$$Y_{1j1} = \mu_1 + p_1 + s_{1j1} + e_{1j1} \qquad ...(3)$$
and
$$Y_{2j2} = \mu_2 + p_2 + s_{2j2} + e_{2j2} \qquad ...(4)$$

Estimation of the mean and variance of the treatment difference: The method of Chow *et al*. (2003) explains a procedure to measure the treatment difference of a crossover trial and this section discusses that procedure. Let ε be the measure of treatment difference, then $\varepsilon = \mu_1 - \mu_2$ (test-reference). Now take $d_{ij} = y_{ij1} - y_{ij2}$ An unbiased estimate for ε is given by,

$$\hat{\varepsilon} = \frac{1}{2n} \sum_{i=1}^{2} \sum_{j=1}^{n} d_i \text{ and } V[\hat{\varepsilon}] = \frac{\sigma_m^2}{2n},$$

where $\sigma_m^2 = \sigma^2_{BT} + \sigma^2_{BR} - 2\rho\sigma_{BT}\sigma_{BR} + \sigma^2_{WT} + \sigma^2_{WR}$

In practice, there is no prior information regarding the value of σ^2_m and it is determined by the assumed values of σ^2_{BT}, σ^2_{BR}, ρ, σ^2_{WT} and σ^2_{WR} and used for sample size

calculation at the design stage. Thus it is required to find an unbiased estimate for use in the test statistic at the analysis stage. An unbiased estimate for σ^2_m can be given by,

$$Var(\hat{\varepsilon}) = \hat{\sigma}^2_m = \frac{1}{2(n-1)} \sum_{i=1}^{2} \sum_{j=1}^{n} (d_{ij} - \overline{d}_{i.})^2 \qquad ...(5)$$

where $\quad \overline{d}_{i.} = \frac{1}{n} \sum_{j=1}^{n} d_{ij}$

Estimating sample size: The null hypothesis (H_o): $\varepsilon = 0$ and the alternative hypothesis (H_1) : $\varepsilon \neq 0$ are used to test whether the effect of two treatments is equal or not.

Under H_0, the test statistic $\dfrac{\hat{\varepsilon}}{\sqrt{\hat{\sigma}^2_m / 2n}}$ follows a t distribution

with 2n-2 degrees of freedom (Chow *et al*., 2003).

The null hypothesis is rejected at α level of significance if

$$\left| \frac{\hat{\varepsilon}}{\hat{\sigma}_m / \sqrt{2n}} \right| > t_{\alpha/2, 2n-2} \qquad ...(6)$$

The above mentioned hypothesis test will satisfy a power of $= 1 - \beta$ if

$$Pr\left[\frac{\hat{\varepsilon} - \varepsilon}{\hat{\sigma}_m / \sqrt{2n}} > t_{\alpha/2, 2n-2} \Big|_{\varepsilon = \varepsilon_R} \right] = 1 - \beta \qquad ...(7)$$

From equations (6) and (7) the corresponding sample size can be obtained by

$$n = \frac{\left[T^{-1}_{2n-2}\left(1 - \alpha/2\right) + T^{-1}_{2n-2}\left(1 - \beta\right) \right]^2 \hat{\sigma}^2_m}{2\varepsilon_R^2} \qquad ...(8)$$

[Here $T_n^{-1}(a)$ indicates the ath ordinate of the t distribution, with n degrees of freedom]

When considering the large sample approach, instead of the t distribution, the standard normal distribution is used. Then the formulae for the sample size calculation can be obtained as,

$$n = \frac{\left[Z^{-1}\left(1 - \alpha/2\right) + Z^{-1}\left(1 - \beta\right) \right]^2 \hat{\sigma}^2_m}{2\varepsilon_R^2} \qquad ...(9)$$

[Here $Z^{-1}(a)$ indicates the ath ordinate of the standard normal distribution]

Simulation studies:

(a) Description: In order to satisfy the above mentioned objectives, a simulation study was carried out. For the exact approach, the bisection method was used as the root finding technique for determining the sample size, as described in Press *et al*.(2002). The simulation study was also used for determining whether the type 1 error and the power are maintained, for both approaches. Sample sizes were determined for varying correlations for both approaches, and thereby the saving in sample size with increasing correlation was studied. Finally, based on the results of the simulation study carried out, guidelines are provided for sample size calculation in crossover trials.

A C programme was written for performing the simulation study. The C language was selected since it is efficient in doing large scale simulations. The first step of the simulation study was to set some practically plausible values for the parameters required (Sooriyarachchi & Whitehead, 1998 ; Whitehead *et al*., 2008). Usually crossover trials are associated with a small sample size, due to comparison of treatments being within patient rather than between patient, and variances within patient being usually smaller than the between variances. The between treatment standard deviation for the treated group (σ_{BT}) was examined over two values namely, 3 and 4 and the between treatment standard deviation for the reference group (σ_{BR}) was set equal to σ_{BT}, which is often assumed in crossover trials. Two values were assigned for the within patient standard deviation for the treated group (σ_{WT}) namely, 0.3 and 0.5 and again the within patient standard deviation for the reference group (σ_{WR}) set equal to σ_{WT}.

The within subject correlation coefficient was indicated by the variable ρ. The values of ρ were examined over 0, 0.3, 0.6, and 0.9. i.e. considering there is no correlation at all, some correlation, high correlation and very high correlation, respectively so that we can see and compare the outcomes for various situations. Note that although it was not considered in this study it is also possible to consider situations where σ_{BT} does not equal σ_{BR} and σ_{WT} does not equal σ_{WR}. The reference improvement is indicated by the variable named ε_R. The values of ε_R that were examined are 1.5, 2 and 3.

For each of these combinations 1000 simulations were carried out under the null and the alternative hypotheses. Under the null hypothesis, the mean difference between treatments (ε) is set to zero and under the alternative

hypothesis, the mean difference between treatments (ε) is set to ε_R.

As explained in the introduction, calculation of sample size is not straightforward for the exact case as the sample size is dependent on the degrees of freedom in this case and thus the sample size determination requires solving of a nonlinear equation in n; hence a root finding technique is needed. The method used here is the Bisection Method explained in Press *et al.* (2002).

After obtaining an estimate for the sample size, it was of interest to determine the proportion of rejections under the null and alternative hypotheses out of thousand simulations to see whether the power and the significance level are maintained. That is to simulate each sample size 1000 times and get the proportion of rejections. In order to do that, we need to simulate the model explained in the introduction. For that we need to generate s_{ijk}'s and e_{ijk}'s for each sample size.

(b) Random number generation: In simulating uncorrelated variables, Box-Muller transformation was used (Golder & Settle, 1976) and the method described in Al-Subaihi (2004) is used for simulating correlated variables.

RESULTS AND DISCUSSION

Checking whether the significance level and the power of the test are maintained was a major objective of this study. In order to do that, two probability intervals were calculated (Sooriyarachchi & Whitehead, 1998) for the true values of significance level and power. The 95% probability interval for a significance level of size *α* can be obtained by,

$$\left[\alpha \pm 2 \times \sqrt{\frac{\alpha(1-\alpha)}{1000}} \right]$$

Thus for α = 5% the probability interval is,

$$= \left[0.025 \pm 2 \times \sqrt{\frac{0.025(1-0.025)}{1000}} \right] = [0.015 , 0.035]$$

The 95% probability interval for a power of 90% can be obtained by,

$$\left[0.90 \pm 2 \times \sqrt{\frac{0.90(1-0.90)}{1000}} \right] = [0.881 , 0.919]$$

An estimate for the significance level can be obtained by the proportion of rejections of the null hypothesis when the null hypothesis is true, and an estimate for the power

of the test can be obtained by the proportion of rejections of the null hypothesis when the alternative hypothesis is true. If the corresponding proportions are within the above probability intervals, it can be concluded that the significance level/power is well maintained by the test. Table 1 gives the proportion of rejections of the null hypothesis under the null and alternative hypothesis for the exact approach and Table 2 corresponds to the similar table for the large sample approach. The proportions which are out of the confidence limits are highlighted in the tables.

In these tables the values taken by all the nuisance parameters (σ_{BT}^2, σ_{BR}^2, σ_{WT}^2, σ_{WR}^2, ρ) and reference improvement (ε_R) are given including whether the simulation was done under the null hypothesis (g =1) or under the alternative hypothesis (g =2). Then for each combination, the calculated sample size, the proportion of rejections of the null hypothesis, σ_m^2 and the mean value of $\hat{\sigma}_m^2$ are given. This is useful in deciding how close the mean of $\hat{\sigma}_m^2$ is to σ_m^2 and hence the unbiasedness of $\hat{\sigma}_m^2$.

When considering Table 1 corresponding to the exact method, based on the t statistic, it can be observed that usually when the sample size is very small (less than or equal to five), the estimated power is outside the probability limits and higher than the upper limit. This is because for very small sample sizes the approximation to the inverse of the t distribution, which is described in Cooke *et al.* (1982), is an overestimate resulting in a too large sample size. But when the sample size is somewhat larger (greater than 5) the power is generally well maintained. Except for a very few cases (row number 34), the significance level is usually maintained by the test.

When considering Table 2 corresponding to the large sample approximation method based on the z statistic, it can be observed that when the sample size is less than or equal to 10 in most of the cases the test is under-powered, it is worse when the sample size gets smaller. But when the sample size is equal to 11 in row numbers 6 and 27, the power is well maintained, but in row number 12 the test is under - powered. This is because the normal ordinate is an underestimate of the t ordinate for sample sizes up to about 20 (2ss-2=20 implying ss=11). It can be said that when the sample size is approximately less than 12, the test is under - powered when the large sample approximation is used. When the sample size is approximately greater than or equal to 12 the power is generally well maintained. Except for a few cases (row numbers 14, 64) for almost all the cases the significance level is well maintained.

Tables 1 and 2 show that for most of the combinations, values of the σ_m^2 and the mean of $\hat{\sigma}_m^2$ are close to each other for both situations.

In order to illustrate the results more clearly, several graphs have been plotted in addition to the two tables. Figure 1 is drawn to illustrate the variation of sample size with respect to del for different combinations of ρ, σ_{BT}, σ_{BW} and σ_{WT} for the exact approach.

Figure 1 shows how the values of nuisance parameters and the reference improvement effect the calculation of sample size for the Exact method. It can be seen that when ρ increases, the sample size required rapidly decreases, irrespective of the situation. Here ρ represents the within

Table 1: Proportion of rejections of the null hypotheses under the null and the alternative hypotheses for the exact method

No	σ_{BT}	σ_{BR}	ρ	σ_{WT}	σ_{WR}	del (ε_R)	Hypothesis	Sample size	Proportion	σ^2_m	Mean of $\hat{\sigma}^2_m$
1	3	3	0	0.3	0.3	1.5	1	44	0.019	18.18	18.27
2	3	3	0	0.3	0.3	1.5	2	44	0.893	18.18	18.27
3	3	3	0	0.3	0.3	2	1	25	0.02	18.18	18.30
4	3	3	0	0.3	0.3	2	2	25	0.903	18.18	18.30
5	3	3	0	0.3	0.3	3	1	12	0.031	18.18	18.47
6	3	3	0	0.3	0.3	3	2	12	0.907	18.18	18.47
7	3	3	0	0.5	0.5	1.5	1	45	0.022	18.5	18.62
8	3	3	0	0.5	0.5	1.5	2	45	0.904	18.5	18.62
9	3	3	0	0.5	0.5	2	1	26	0.027	18.5	18.61
10	3	3	0	0.5	0.5	2	2	26	0.907	18.5	18.61
11	3	3	0	0.5	0.5	3	1	12	0.032	18.5	18.80
12	3	3	0	0.5	0.5	3	2	12	0.903	18.5	18.80
13	3	3	0.3	0.3	0.3	1.5	1	31	0.03	12.78	12.85
14	3	3	0.3	0.3	0.3	1.5	2	31	0.891	12.78	12.85
15	3	3	0.3	0.3	0.3	2	1	18	0.027	12.78	12.85
16	3	3	0.3	0.3	0.3	2	2	18	0.907	12.78	12.85
17	3	3	0.3	0.3	0.3	3	1	9	0.022	12.78	13.06
18	3	3	0.3	0.3	0.3	3	2	9	0.905	12.78	13.06
19	3	3	0.3	0.5	0.5	1.5	1	32	0.028	13.1	13.20
20	3	3	0.3	0.5	0.5	1.5	2	32	0.884	13.1	13.20
21	3	3	0.3	0.5	0.5	2	1	19	0.029	13.1	13.17
22	3	3	0.3	0.5	0.5	2	2	19	0.909	13.1	13.17
23	3	3	0.3	0.5	0.5	3	1	9	0.021	13.1	13.39
24	3	3	0.3	0.5	0.5	3	2	9	0.893	13.1	13.39
25	3	3	0.6	0.3	0.3	1.5	1	19	0.03	7.38	7.41
26	3	3	0.6	0.3	0.3	1.5	2	19	0.913	7.38	7.41
27	3	3	0.6	0.3	0.3	2	1	11	0.025	7.38	7.50
28	3	3	0.6	0.3	0.3	2	2	11	0.9	7.38	7.50
29	3	3	0.6	0.3	0.3	3	1	6	0.02	7.38	7.46
30	3	3	0.6	0.3	0.3	3	2	6	**0.922**	7.38	7.46
31	3	3	0.6	0.5	0.5	1.5	1	19	0.027	7.7	7.73
32	3	3	0.6	0.5	0.5	1.5	2	19	0.9	7.7	7.73
33	3	3	0.6	0.5	0.5	2	1	12	0.03	7.7	7.81
34	3	3	0.6	0.5	0.5	2	2	12	**0.922**	7.7	7.81
35	3	3	0.6	0.5	0.5	3	1	6	0.022	7.7	7.78
36	3	3	0.6	0.5	0.5	3	2	6	0.91	7.7	7.78
37	3	3	0.9	0.3	0.3	1.5	1	6	0.027	1.98	1.98
38	3	3	0.9	0.3	0.3	1.5	2	6	0.901	1.98	1.98
39	3	3	0.9	0.3	0.3	2	1	4	**0.038**	1.98	1.97
40	3	3	0.9	0.3	0.3	2	2	4	0.906	1.98	1.97
41	3	3	0.9	0.3	0.3	3	1	3	0.025	1.98	1.93
42	3	3	0.9	0.3	0.3	3	2	3	**0.965**	1.98	1.93
43	3	3	0.9	0.5	0.5	1.5	1	7	0.025	2.3	2.29
44	3	3	0.9	0.5	0.5	1.5	2	7	**0.92**	2.3	2.29
45	3	3	0.9	0.5	0.5	2	1	5	0.024	2.3	2.31

Table 1 continued…

No	σ_{BT}	σ_{BR}	ρ	σ_{WT}	σ_{WR}	del (ε_R)	Hypothesis	Sample size	Proportion	σ^2_m	Mean of $\hat{\sigma}^2_m$
46	3	3	0.9	0.5	0.5	2	2	5	**0.955**	2.3	2.31
47	3	3	0.9	0.5	0.5	3	1	3	0.021	2.3	2.25
48	3	3	0.9	0.5	0.5	3	2	3	**0.937**	2.3	2.25
49	4	4	0	0.3	0.3	1.5	1	77	0.023	32.18	32.28
50	4	4	0	0.3	0.3	1.5	2	77	0.906	32.18	32.28
51	4	4	0	0.3	0.3	2	1	44	0.019	32.18	32.33
52	4	4	0	0.3	0.3	2	2	44	0.892	32.18	32.33
53	4	4	0	0.3	0.3	3	1	20	0.024	32.18	32.35
54	4	4	0	0.3	0.3	3	2	20	0.893	32.18	32.35
55	4	4	0	0.5	0.5	1.5	1	77	0.023	32.5	32.60
56	4	4	0	0.5	0.5	1.5	2	77	0.904	32.5	32.60
57	4	4	0	0.5	0.5	2	1	44	0.018	32.5	32.66
58	4	4	0	0.5	0.5	2	2	44	0.889	32.5	32.66
59	4	4	0	0.5	0.5	3	1	20	0.027	32.5	32.67
60	4	4	0	0.5	0.5	3	2	20	0.888	32.5	32.67
61	4	4	0.3	0.3	0.3	1.5	1	54	0.029	22.58	22.68
62	4	4	0.3	0.3	0.3	1.5	2	54	0.893	22.58	22.68
63	4	4	0.3	0.3	0.3	2	1	31	0.029	22.58	22.71
64	4	4	0.3	0.3	0.3	2	2	31	0.891	22.58	22.71
65	4	4	0.3	0.3	0.3	3	1	15	0.026	22.58	22.84
66	4	4	0.3	0.3	0.3	3	2	15	0.912	22.58	22.84
67	4	4	0.3	0.5	0.5	1.5	1	55	0.021	22.9	23.04
68	4	4	0.3	0.5	0.5	1.5	2	55	0.9	22.9	23.04
69	4	4	0.3	0.5	0.5	2	1	32	0.029	22.9	23.06
70	4	4	0.3	0.5	0.5	2	2	32	0.891	22.9	23.06
71	4	4	0.3	0.5	0.5	3	1	15	0.028	22.9	23.16
72	4	4	0.3	0.5	0.5	3	2	15	0.906	22.9	23.16
73	4	4	0.6	0.3	0.3	1.5	1	32	0.027	12.98	13.08
74	4	4	0.6	0.3	0.3	1.5	2	32	0.895	12.98	13.08
75	4	4	0.6	0.3	0.3	2	1	19	0.031	12.98	13.04
76	4	4	0.6	0.3	0.3	2	2	19	0.914	12.98	13.04
77	4	4	0.6	0.3	0.3	3	1	9	0.021	12.98	13.20
78	4	4	0.6	0.3	0.3	3	2	9	0.896	12.98	13.20
79	4	4	0.6	0.5	0.5	1.5	1	33	0.021	13.3	13.41
80	4	4	0.6	0.5	0.5	1.5	2	33	0.907	13.3	13.41
81	4	4	0.6	0.5	0.5	2	1	19	0.029	13.3	13.35
82	4	4	0.6	0.5	0.5	2	2	19	0.91	13.3	13.35
83	4	4	0.6	0.5	0.5	3	1	9	0.024	13.3	13.52
84	4	4	0.6	0.5	0.5	3	2	9	0.895	13.3	13.52
85	4	4	0.9	0.3	0.3	1.5	1	9	0.024	3.38	3.40
86	4	4	0.9	0.3	0.3	1.5	2	9	0.901	3.38	3.40
87	4	4	0.9	0.3	0.3	2	1	6	0.025	3.38	3.37
88	4	4	0.9	0.3	0.3	2	2	6	**0.921**	3.38	3.37
89	4	4	0.9	0.3	0.3	3	1	4	**0.039**	3.38	3.37
90	4	4	0.9	0.3	0.3	3	2	4	**0.967**	3.38	3.37
91	4	4	0.9	0.5	0.5	1.5	1	10	0.027	3.7	3.72
92	4	4	0.9	0.5	0.5	1.5	2	10	0.914	3.7	3.72
93	4	4	0.9	0.5	0.5	2	1	6	0.03	3.7	3.69
94	4	4	0.9	0.5	0.5	2	2	6	0.89	3.7	3.69
95	4	4	0.9	0.5	0.5	3	1	4	0.035	3.7	3.69
96	4	4	0.9	0.5	0.5	3	2	4	**0.954**	3.7	3.69

Table 2: Proportion of rejections of the null hypotheses under the null and the alternative hypotheses for the large sample approximation

No	σ_{BT}	σ_{BR}	ρ	σ_{WT}	σ_{WR}	del (ε_R)	Hypothesis	Sample size	Proportion	σ^2_m	Mean of $\hat{\sigma}^2_m$
1	3	3	0	0.3	0.3	1.5	1	43	0.023	18.18	18.29
2	3	3	0	0.3	0.3	1.5	2	43	0.897	18.18	18.29
3	3	3	0	0.3	0.3	2	1	24	0.032	18.18	18.31
4	3	3	0	0.3	0.3	2	2	24	0.889	18.18	18.31
5	3	3	0	0.3	0.3	3	1	11	0.026	18.18	18.52
6	3	3	0	0.3	0.3	3	2	11	0.883	18.18	18.52
7	3	3	0	0.5	0.5	1.5	1	44	0.018	18.5	18.59
8	3	3	0	0.5	0.5	1.5	2	44	0.886	18.5	18.59
9	3	3	0	0.5	0.5	2	1	25	0.02	18.5	18.63
10	3	3	0	0.5	0.5	2	2	25	0.899	18.5	18.63
11	3	3	0	0.5	0.5	3	1	11	0.026	18.5	18.86
12	3	3	0	0.5	0.5	3	2	11	**0.879**	18.5	18.86
13	3	3	0.3	0.3	0.3	1.5	1	30	0.022	12.78	12.84
14	3	3	0.3	0.3	0.3	1.5	2	30	**0.876**	12.78	12.84
15	3	3	0.3	0.3	0.3	2	1	17	0.034	12.78	12.89
16	3	3	0.3	0.3	0.3	2	2	17	0.885	12.78	12.89
17	3	3	0.3	0.3	0.3	3	1	8	0.026	12.78	12.96
18	3	3	0.3	0.3	0.3	3	2	8	**0.862**	12.78	12.96
19	3	3	0.3	0.5	0.5	1.5	1	31	0.029	13.11	3.18
20	3	3	0.3	0.5	0.5	1.5	2	31	0.89	13.11	3.18
21	3	3	0.3	0.5	0.5	2	1	18	0.033	13.11	3.17
22	3	3	0.3	0.5	0.5	2	2	18	0.908	13.11	3.17
23	3	3	0.3	0.5	0.5	3	1	8	0.025	13.11	3.29
24	3	3	0.3	0.5	0.5	3	2	8	**0.848**	13.11	3.29
25	3	3	0.6	0.3	0.3	1.5	1	18	0.034	7.38	7.42
26	3	3	0.6	0.3	0.3	1.5	2	18	0.901	7.38	7.42
27	3	3	0.6	0.3	0.3	2	1	10	0.024	7.38	7.49
28	3	3	0.6	0.3	0.3	2	2	10	**0.875**	7.38	7.49
29	3	3	0.6	0.3	0.3	3	1	5	0.021	7.38	7.53
30	3	3	0.6	0.3	0.3	3	2	5	**0.866**	7.38	7.53
31	3	3	0.6	0.5	0.5	1.5	1	18	0.035	7.7	7.73
32	3	3	0.6	0.5	0.5	1.5	2	18	0.894	7.7	7.73
33	3	3	0.6	0.5	0.5	2	1	11	0.023	7.7	7.82
34	3	3	0.6	0.5	0.5	2	2	11	0.882	7.7	7.82
35	3	3	0.6	0.5	0.5	3	1	5	0.021	7.7	7.86
36	3	3	0.6	0.5	0.5	3	2	5	**0.847**	7.7	7.86
37	3	3	0.9	0.3	0.3	1.5	1	5	0.024	1.98	1.99
38	3	3	0.9	0.3	0.3	1.5	2	5	**0.845**	1.98	1.99
39	3	3	0.9	0.3	0.3	2	1	3	0.025	1.98	1.93
40	3	3	0.9	0.3	0.3	2	2	3	**0.749**	1.98	1.93
41	3	3	0.9	0.3	0.3	3	1	2	0.021	1.98	1.92
42	3	3	0.9	0.3	0.3	3	2	2	**0.599**	1.98	1.92
43	3	3	0.9	0.5	0.5	1.5	1	6	0.03	2.3	2.30
44	3	3	0.9	0.5	0.5	1.5	2	6	**0.865**	2.3	2.30
45	3	3	0.9	0.5	0.5	2	1	4	0.033	2.3	2.29
46	3	3	0.9	0.5	0.5	2	2	4	0.881	2.3	2.29
47	3	3	0.9	0.5	0.5	3	1	2	0.024	2.3	2.23
48	3	3	0.9	0.5	0.5	3	2	2	**0.557**	2.3	2.23
49	4	4	0	0.3	0.3	1.5	1	76	0.016	32.18	32.22
50	4	4	0	0.3	0.3	1.5	2	76	0.892	32.18	32.22

Table 2 continued…

No	σ_{BT}	σ_{BR}	ρ	σ_{WT}	σ_{WR}	del (ε_R)	Hypothesis	Sample size	Proportion	σ^2_m	Mean of $\hat{\sigma}^2_m$
51	4	4	0	0.3	0.3	2	1	43	0.023	32.18	32.38
52	4	4	0	0.3	0.3	2	2	43	0.895	32.18	32.38
53	4	4	0	0.3	0.3	3	1	19	0.029	32.18	32.42
54	4	4	0	0.3	0.3	3	2	19	0.883	32.18	32.42
55	4	4	0	0.5	0.5	1.5	1	76	0.016	32.53	2.55
56	4	4	0	0.5	0.5	1.5	2	76	0.887	32.5	32.55
57	4	4	0	0.5	0.5	2	1	43	0.022	32.5	32.70
58	4	4	0	0.5	0.5	2	2	43	0.896	32.5	32.70
59	4	4	0	0.5	0.5	3	1	19	0.03	32.5	32.73
60	4	4	0	0.5	0.5	3	2	19	0.882	32.5	32.73
61	4	4	0.3	0.3	0.3	1.5	1	53	0.022	22.58	22.69
62	4	4	0.3	0.3	0.3	1.5	2	53	0.904	22.58	22.69
63	4	4	0.3	0.3	0.3	2	1	30	0.023	22.58	22.69
64	4	4	0.3	0.3	0.3	2	2	30	**0.875**	22.58	22.69
65	4	4	0.3	0.3	0.3	3	1	14	0.03	22.58	22.87
66	4	4	0.3	0.3	0.3	3	2	14	0.887	22.58	22.87
67	4	4	0.3	0.5	0.5	1.5	1	54	0.025	22.9	23.00
68	4	4	0.3	0.5	0.5	1.5	2	54	0.899	22.9	23.00
69	4	4	0.3	0.5	0.5	2	1	31	0.03	22.9	23.03
70	4	4	0.3	0.5	0.5	2	2	31	0.895	22.9	23.03
71	4	4	0.3	0.5	0.5	3	1	14	0.029	22.9	23.19
72	4	4	0.3	0.5	0.5	3	2	14	0.885	22.9	23.19
73	4	4	0.6	0.3	0.3	1.5	1	31	0.028	12.98	13.05
74	4	4	0.6	0.3	0.3	1.5	2	31	0.892	12.98	13.05
75	4	4	0.6	0.3	0.3	2	1	18	0.032	12.98	13.06
76	4	4	0.6	0.3	0.3	2	2	18	0.904	12.98	13.06
77	4	4	0.6	0.3	0.3	3	1	8	0.025	12.98	13.11
78	4	4	0.6	0.3	0.3	3	2	8	**0.86**	12.98	13.11
79	4	4	0.6	0.5	0.5	1.5	1	32	0.029	13.3	13.40
80	4	4	0.6	0.5	0.5	1.5	2	32	0.887	13.3	13.40
81	4	4	0.6	0.5	0.5	2	1	18	0.033	13.3	13.37
82	4	4	0.6	0.5	0.5	2	2	18	0.898	13.3	13.37
83	4	4	0.6	0.5	0.5	3	1	8	0.025	13.3	13.43
84	4	4	0.6	0.5	0.5	3	2	8	**0.851**	13.3	13.43
85	4	4	0.9	0.3	0.3	1.5	1	8	0.029	3.38	3.38
86	4	4	0.9	0.3	0.3	1.5	2	8	**0.848**	3.38	3.38
87	4	4	0.9	0.3	0.3	2	1	5	0.024	3.38	3.40
88	4	4	0.9	0.3	0.3	2	2	5	**0.86**	3.38	3.40
89	4	4	0.9	0.3	0.3	3	1	2	0.019	3.38	3.28
90	4	4	0.9	0.3	0.3	3	2	2	**0.421**	3.38	3.28
91	4	4	0.9	0.5	0.5	1.5	1	9	0.021	3.7	3.72
92	4	4	0.9	0.5	0.5	1.5	2	9	**0.879**	3.7	3.72
93	4	4	0.9	0.5	0.5	2	1	5	0.024	3.7	3.72
94	4	4	0.9	0.5	0.5	2	2	5	**0.825**	3.7	3.72
95	4	4	0.9	0.5	0.5	3	1	3	0.025	3.7	3.61
96	4	4	0.9	0.5	0.5	3	2	3	**0.814**	3.7	3.61

patient correlation coefficient. The higher the correlation between patients, the higher the gain in sample size. Also it can be observed that, the within patient variance $(\sigma_{WT})^2$ has less effect than the between patient variance $(\sigma_{BT})^2$ on the calculation of sample size because the latter is usually much greater than the former. When the reference improvement increases the sample size becomes smaller, because the difference we want to detect is larger. Also the gain in sample size due to increasing correlation is higher for smaller reference improvement.

Figure 2 is drawn in order to see whether the significance level is maintained by the test for the exact approach. It shows a graph of the proportion of rejections of the null hypothesis, when the null hypothesis is true versus del for different combinations of ρ, σ_{BT} and σ_{WT} for the exact approach.

Two coloured lines represent the band within which the proportion should lie, in order to maintain the significance level. The corresponding sample size is shown near the points, which are out of the bands. Similar results as shown by Table 1 are illustrated here.

Figure 3 is drawn in order to see how well the power is maintained by the test for the exact approach. It gives a graph of the proportion of rejections of the null hypothesis, when the alternative hypothesis is true versus del for different combinations of ρ, σ_{BT} and σ_{WT} for the exact approach.

When considering Figure 3 it can be seen that most of the sample sizes which lie outside the bands are very small numbers except for 12, which is very close to the upper limit. When the sample size is five or less, many points lie outside the band. The reason is the imprecision of the approximation used in calculating the inverse t distribution values when the sample size is less than or equal to five. When ρ is very high (0.9) and the reference improvement is large (3), there is a higher tendency in obtaining a small sample size, hence a higher number of points can be observed outside the bands in those combinations.

From Figures 2 and 3, it can be seen that similar results are obtained as per the table for the exact approach.

Figures 4 and 5 are the corresponding graphs to Figures 1 and 2, for the large sample approach respectively. The conclusions drawn from Figure 4 are same as those drawn from Figure 1. Figure 5 illustrates similar results as given in Table 2.

Figure 6 is the corresponding graph to Figure 3, for the large sample approach and illustrates similar results as in Table 2. When ρ is very high (0.9) and the reference improvement is large (3), there is a higher tendency in obtaining a small sample size.

CONCLUSION

This study deals with the sample size calculation of crossover trials under two situations in the power approach, namely the exact and large sample methods (Chow *et al*. 2003).

From the results of the simulation study the following guidelines can be given. It was seen that when the sample size is very small (less than or equal to five), neither method maintains error rates. i.e. even the exact method which was based on the t distribution failed for very small sample sizes. This is because the approximation used to calculate the inverse of the t distribution which is described in Cooke *et al*. (1982), in determining the sample size, is not accurate for very small sample sizes. Also it is evident that when the sample size is fairly large in terms of crossover studies (5 < sample size < 12), only the exact approach has maintained error rates. This is because the normal ordinate is an underestimate of the t ordinate for sample sizes within that range. That means within the specified sample limits the exact approach should be used.

When the sample size is large in terms of crossover studies ($\geq$ 12) both methods have maintained error rates. i.e. for the sample sizes greater than about eleven, the large sample approach, which is much simpler than the exact approach, can be used for sample size determination instead of the exact approach, which needs numerical methods. A higher reduction in sample size can be achieved when the within patient correlation is high.

A better approximation for the inverse distribution of the t distribution should be found when calculating sample sizes, which are believed to be very small. The study is done for 2 x 2 crossover trials, which consider only two treatments and two periods. As an extension one can consider more treatments and periods. Also in this study there are no replications of treatments, i.e. a treatment is given to a set of patients (subjects) only once. One can extend this study to have replications.

An assumption used in this study is that of equality of within and between patient variances for both treated and reference groups (i.e. $\sigma_{BT}{}^2 = \sigma_{BR}{}^2$ and $\sigma_{WT}{}^2 = \sigma_{WR}{}^2$) and equal allocation of patients to both groups. Thus further investigation can be carried out taking different values for these parameters.
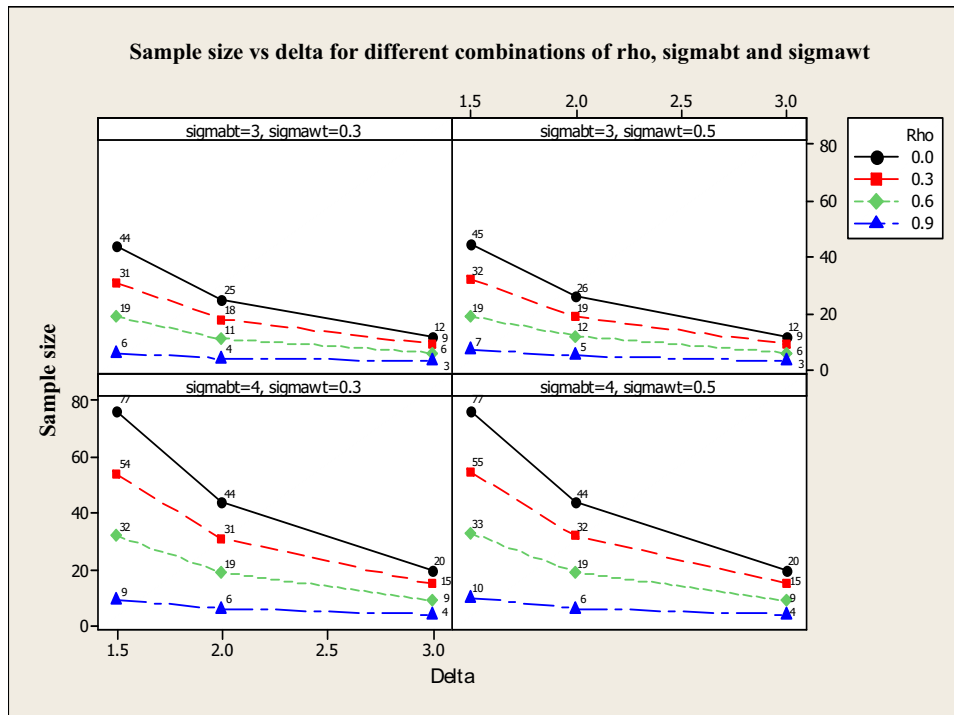
Figure 1: Graph of sample size versus delta for different combinations of ρ, σ_{BT} and σ_{WT} for the exact approach
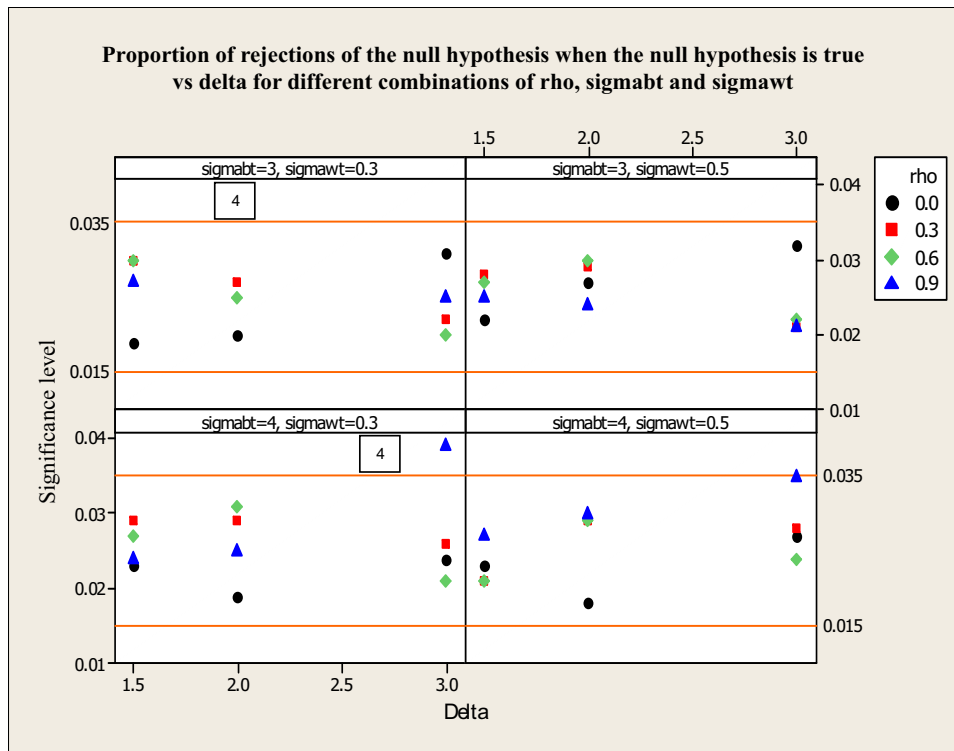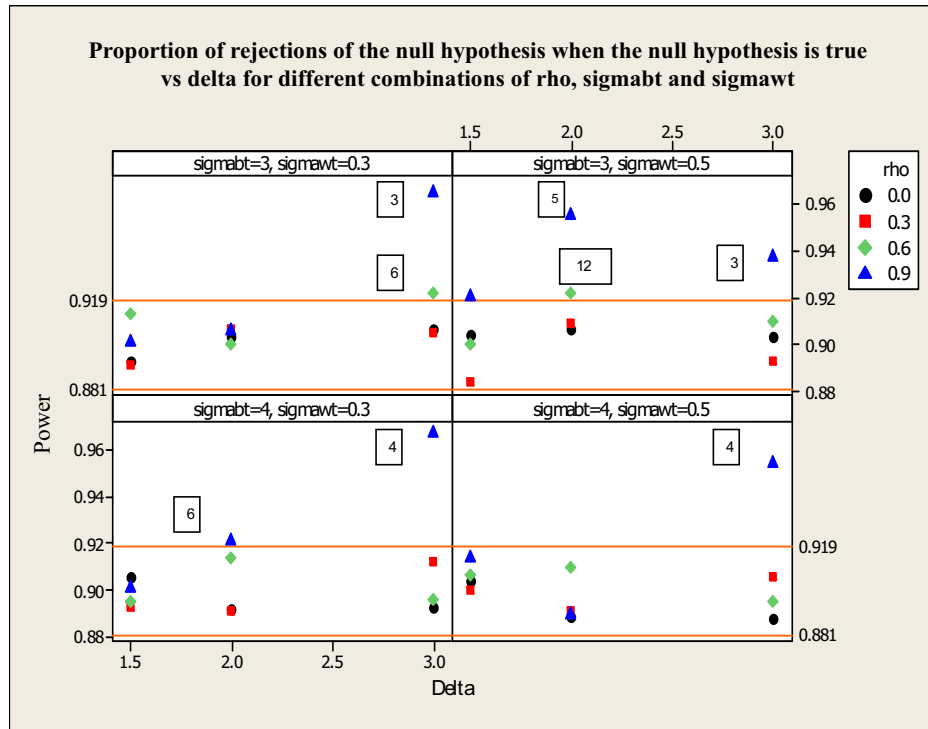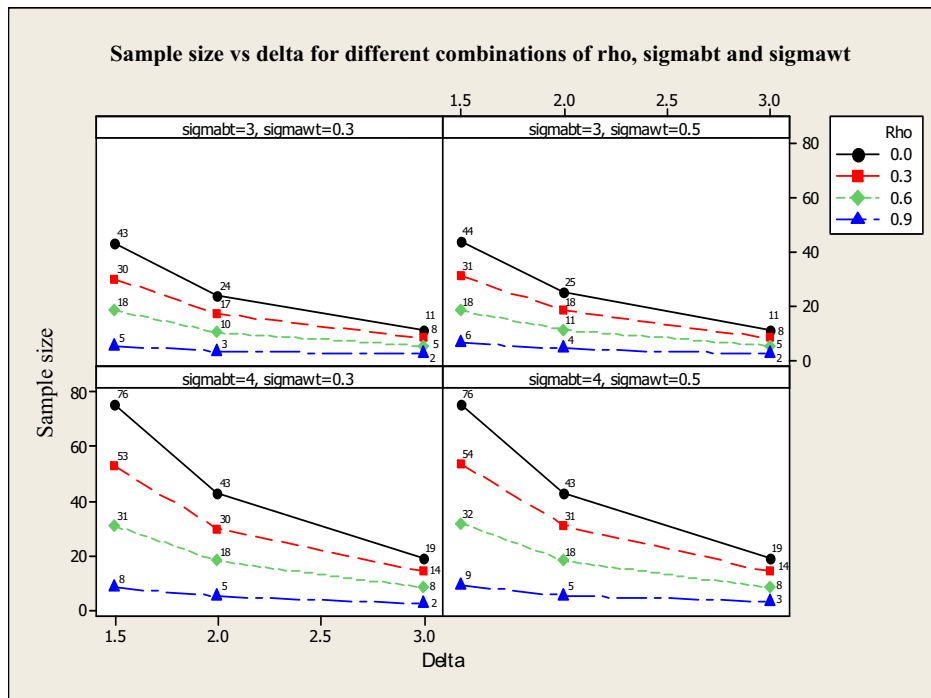

Figure 2: Graph of proportion of rejections of the null hypothesis, when the null hypothesis is true versus delta for different combinations of ρ, σ_{BT} and σ_{WT} for the exact approach

Figure 3: Graph of proportion of rejections of the null hypothesis, when the alternative hypothesis is true versus delta for different combinations of ρ, σ_{BT} and σ_{WT} for the exact approach

Figure 4: Graph of sample size versus delta for different combinations of ρ, σ_{BT} and σ_{WT} for the large sample approach
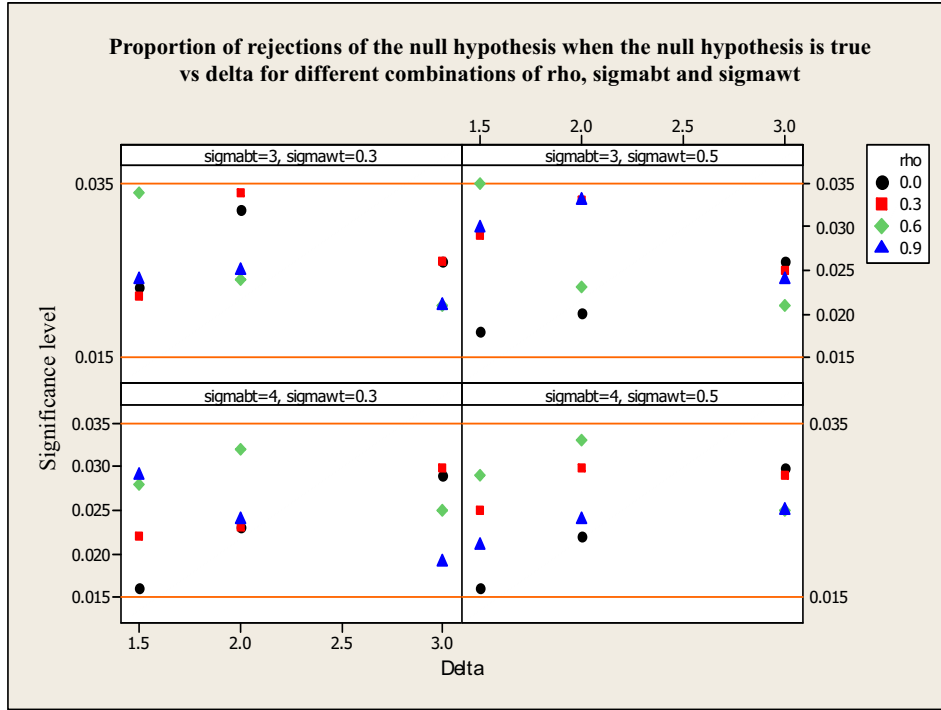
Figure 5 : Graph of proportion of rejections of the null hypothesis, when the null hypothesis is true versus delta for different combinations of ρ, σ_{BT} and σ_{WT} for the large sample approach
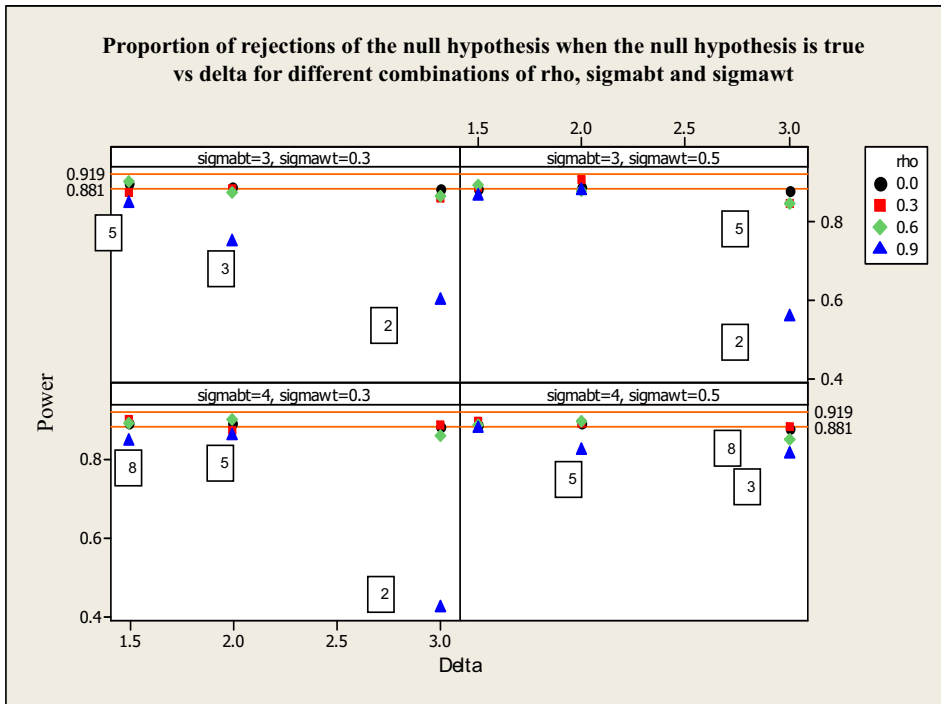

Figure 6: Graph of proportion of rejections of the null hypothesis, when the alternative hypothesis is true versus delta for different combinations of ρ, σ_{BT} and σ_{WT} for the large sample approach

In the model specified, it was assumed that there is no treatment by period interaction (that is, the effect of the treatment remains consistent over the two periods). If such an interaction is present, a simple t test cannot be used in testing the hypothesis and a modelling approach will have to be used (Jones & Kenward, 2003). This requires a new investigation to be carried out.

References

1. Al-Subaihi A.A. (2004). Simulating correlated multivariate pseudorandom numbers. *Journal of Statistical Software* **9**(i04) *http://ideas.repec.org/a/jss/jstsof/09i04.html.*

2. Brown H. & Prescott R. (2006). *Applied Mixed Models in Medicine.* John Wiley & Sons, Inc., Chichester, UK.

3. Chow S.C., Shao J. & Wang H. (2003). *Sample Size Calculations in Clinical Research*. Chapman and Hall/ CRC Press, New York, USA.

4. Cooke D., Craven A.H. & Clarke G.M. (1982). *Basic Statistical Computing*. Edward Arnold Press, London, UK.

5. Golder E.R. & Settle J.G. (1976). The Box-Muller method for generating pseudo-random normal deviates. *Applied Statistics* **25**(1):12-20.

6. Grizzle J.E. (1965). The two-period change-over design and its use in clinical trials. *Biometrics* **21**(2): 467-480.

7. Jones B. & Kenward M. (2003). *The Analysis of Crossover Trials*. Chapman and Hall / CRC Press, New York, USA.

8. Julious S.A. & Patterson S.D. (2004). Sample sizes for estimation in clinical trials. *Pharmaceutical Statistics* **3**(3):213-215

9. Lee C., Lee L.H., Christoper L.W., Chen M. & Benjamin R. (2005). *Clinical Trials of Drugs and Biopharmaceuticals*. Chapman and Hall /CRC Press, New York, USA.

10. Pocock S.J. (1983). *Clinical Trials: A Practical Approach.* John Wiley & Sons, Inc., Chichester, UK.

11. Press W.H., Teukolsky S.A., Vetterling W.T. & Flannery B.P. (2002). *Numerical Recipes in C++*, pp. 357-358. Cambridge University Press, Cambridge, UK.

12. Sooriyarachchi M.R. & Whitehead J. (1998). A method for sequential analysis of survival data with non-proportional hazards. *Biometrics* **54**(3): 1072-1084.

13. Whitehead A., Sooriyarachchi M.R., Whitehead J. & Bolland K. (2008). Incorporating intermediate binary responses into interim analyses of clinical trials: a comparison of four methods. *Statistics in Medicine* **27**(10): 1646 – 1666.

14. Woodward M. (1992). Formulae for sample size, power and minimum detectable risk in medical studies. *The Statistician* **41**(2):185-196.