

# A Comment on Sample Size Calculation for Analysis of Covariance in Parallel Arm Studies

Guogen Shan<sup>1\*</sup> and Changxing Ma<sup>2</sup>

<sup>1</sup>Epidemiology and Biostatistics Program, Department of Environmental and Occupational Health, School of Community Health Sciences, University of Nevada Las Vegas, Las Vegas, NV 89154, USA

<sup>2</sup>Department of Biostatistics, University at Buffalo, Buffalo, NY 14214, USA

## Abstract

We compare two sample size calculation approaches for analysis of covariance with one covariate. Exact simulation studies are conducted to compare the sample size calculation based on an approach by Borm et al. (2007) (referred to as the B approach) and an exact approach (referred to as the F approach). Although the B approach and the F approach have similar performance when the correlation coefficient is small, the F approach generally has a more accurate sample size calculation as compared to the B approach. Therefore, the F approach for sample size calculation is generally recommended for use in practice.

**Keywords:** Analysis of covariance; F distribution; Sample size; Power

## Introduction

Randomized clinical trials are commonly used to confirm the efficacy of a new treatment. There are several advantages for using randomization in clinical trials, such as selection bias reduction and increased comparability among groups with potential confounding factors [1]. Balanced studies are often conducted to maximize the power of a study for a given total sample size.

Sample size calculation plays a very important role in clinical trials. It has been studied for many years and achieved significant progress [2-4]. As far as we know, sample size calculation approaches for analysis of covariance (ANCOVA) are very limited. Recently, Borm et al. [5] proposed a simple sample size calculation closed form for ANCOVA with one covariate which is considered as a baseline of the response outcome. Based on the sample size from a two sample t-test and the correlation between response outcomes and covariate values, they show that this formula has accurate sample size calculation. The other method is based on a ratio of mean squares [6,7] where the null distribution follows a F distribution and the alternative is a non-central F distribution. There is no systematic comparison between these two approaches.

We reviewed two existing sample size calculation approaches for ANCOVA with one covariate in Section 2. In Section 3, we compare the two approaches using exact simulation studies and an example from a randomized study is used to illustrate these two approaches. Section 4 is given to discussion.

## Methods

Suppose that  $Y_{ij}$  be the  $j^{\text{th}}$  response outcome for the  $i^{\text{th}}$  group,  $i=1, 2; j=1, 2, \dots, n_j$ , and  $X_{ij}$  be the associated covariate. We consider the first group as the control, and the second group as the treatment group in this article. The covariate can be viewed as the baseline for the output. The regression model for the relationship between  $Y$  and  $X$  within the  $i^{\text{th}}$  group is given as

$$Y_{ij} = \beta_{0i} + \beta_1 X_{ij} + \varepsilon_{ij}$$

where  $\beta_{0i}$  is the intercept for the  $i^{\text{th}}$  group,  $\beta_1$  is the common slope for both groups, and  $\varepsilon_{ij}$  is the measure error which follows a normal distribution [8]. The mean difference between two groups is the difference between two intercepts.

Borm et al. [5] proposed a simple sample size calculation for ANCOVA by multiplying the number of subjects for the two sample t-test by a design factor. The factor here is  $1-\rho^2$ , where  $\rho$  is the correlation coefficient between the outcome and the covariate. Sample size calculation for the two-sample t-test is based on response outcomes. Given a significance level of  $\alpha$ , a pre-specified power  $1-\beta$ , the mean difference between the treatment group and the control of  $\mu_2-\mu_1$ , and a common standard deviation of response outcome  $\sigma$ , sample size per group is calculated as

$$n = 2\sigma^2(Z_{1-\alpha/2} + Z_{1-\beta})^2 / (\mu_2 - \mu_1)^2,$$

where  $Z_d$  is the  $d$ -th percentile of a standard normal distribution. Borm et al. [5] showed that the total sample size for the ANCOVA  $N=2n(1-\rho^2)$  may not be accurate enough for small sample settings to retain the pre-specified power. They provided some power plots to show that power with this sample size formula is generally smaller than  $1-\beta$  for small sample settings. For this reason, they proposed

$$N = 2(n+1)(1-\rho^2)$$

to be used as the sample size by adding one subject for each group in the sample size calculation. They claimed that this sample size is accurate for all sample sizes.

The second method is an exact approach based on a ratio of mean squares,

$$T = \frac{MS_b}{MS_w},$$

where  $MS_b$  is the mean square between groups, and  $MS_w$  is the mean square within the group [7]. Under the null hypothesis with no

**\*Corresponding author:** Guogen Shan, Epidemiology and Biostatistics Program, Department of Environmental and Occupational Health, School of Community Health Sciences, University of Nevada Las Vegas, Las Vegas, NV 89154, USA, Tel: 702-895-4413; Fax: 702-895-5184; E-mail: [guogen.shan@unlv.edu](mailto:guogen.shan@unlv.edu)

**Received** December 17, 2013; **Accepted** January 09, 2014; **Published** January 16, 2014

**Citation:** Shan G, Ma C (2014) A Comment on Sample Size Calculation for Analysis of Covariance in Parallel Arm Studies. J Biomet Biostat 5: 184. doi:10.4172/2155-6180.1000184

**Copyright:** © 2014 Shan G, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

difference between the control and the treatment group, the ratio  $T$  follows a central  $F_{1,N-3}$  distribution. Given a significance level of  $\alpha$ , the threshold value is  $F_{\alpha}$ , where  $\Pr(F_{1,N-3} \geq F_{\alpha}) = \alpha$ . Under the alternative, the test statistic follows a non-central  $F_{1,N-3,\lambda}$  distribution with the non-central parameter  $\lambda = N\sigma_b^2 / \sigma_e^2$  [7], where  $\sigma_e^2 = (1 - \rho^2)\sigma^2$ ,  $\sigma_b^2 = n_1 / N(\mu_1 - \bar{\mu})^2 + n_2 / N(\mu_2 - \bar{\mu})^2$ , and  $\bar{\mu} = n_1\mu_1 / N + n_2\mu_2 / N$  is the overall response outcome mean. The power of the study is then expressed as a probability of being greater than or equal to the threshold  $F_{\alpha}$  in the non-central F distribution,  $\Pr(F_{1,N-3,\lambda} \geq F_{\alpha})$ . The required sample size is determined by increasing the sample size by one each time until the pre-specified power is reached.

### Method comparison

We referred to the approach proposed by Borm et al. [5] as the B approach, and the other based on the F distribution as the F approach. Power is calculated as the percentage of trials with significant p-values using ANCOVA based on 10000 simulations. Calculated power is presented in Table 1 for  $\alpha=0.05$ ,  $\beta=0.2$ ,  $\sigma=1$ , and  $\mu_2 - \mu_1 = 0.5$ , and Table 2 for  $\alpha=0.01$ ,  $\beta=0.2$ ,  $\sigma=1$ , and  $\mu_2 - \mu_1 = 1$ . Sample size based on the F approach is calculated using PASS 12 [9]. As can be seen from both tables, the difference between the B approach and the F approach is negligible for small  $\rho$  values. The power of the B approach is much lower than the pre-specified power for large  $\rho$  values, as shown in Table 2, the power could be as low as 52%. Although the B approach and the F approach have similar performance when  $\rho$  is small, the F approach generally has more accurate sample size calculation as compared to the B approach.

A parallel randomized clinical trial is illustrated for sample size calculation based on the B approach and the F approach. Patients with rheumatoid arthritis are randomized into one of the groups with or without leunomide [10]. The response outcome, the disease activity score, is measured before and after the treatment. The baseline

$\rho$	Borm approach		non-central F distribution	
	n per group	power	n per group	power
0	64	0.7941	64	0.7965
0.1	64	0.802	64	0.8019
0.2	62	0.7961	62	0.8057
0.3	59	0.8081	59	0.7988
0.4	54	0.803	54	0.7974
0.5	48	0.7945	49	0.8082
0.6	41	0.7957	42	0.7943
0.7	33	0.7836	34	0.7989
0.8	23	0.7758	24	0.8046
0.9	13	0.7842	14	0.8154

Table 1:  $\alpha=0.05$ , power=0.8,  $sd=1$ ,  $diff=0.5$ .

$\rho$	Borm approach		non-central F distribution	
	n per group	power	n per group	power
0	25	0.7918	26	0.806
0.1	25	0.7891	25	0.7917
0.2	24	0.7874	25	0.8059
0.3	23	0.7966	24	0.8096
0.4	21	0.782	22	0.8037
0.5	19	0.7839	20	0.7996
0.6	16	0.7629	17	0.795
0.7	13	0.7452	14	0.7975
0.8	9	0.6765	11	0.8117
0.9	5	0.5197	7	0.8126

Table 2:  $\alpha=0.01$ , power=0.8,  $sd=1$ ,  $diff=1$ .

measurement is considered as the covariate in the ANCOVA model. This example is also used by Borm et al. [5]. The standard deviation is estimated as  $\sigma=1.2$ . At a significance level of  $\alpha=0.01$  and 90% power, the sample size calculations based on the B approach to detect a mean difference of  $\mu_2 - \mu_1 = 0.6$  are 122, 86, and 46 as total sample sizes for  $\rho=0.7, 0.8$ , and  $0.9$ , respectively. It needs total sample sizes of 126, 90, and 50 using the F approach. The sample size from the B approach is less than that from the F approach. The sample size from the B approach may not attain the pre-specified power of the study.

### Conclusions

The sample size calculation formula proposed by Borm et al. [5] has a closed form, and it is computationally easy. Power of the study may be lower than the pre-specified power when  $\rho$  is large. Power of the F approach is closer to the pre-specified power for all  $\rho$  values. The code written in R for the sample size calculation for both methods is available from the first author. In a study with multiple covariates, Borm et al. [5] recommended using  $1 - R^2$  as the design factor. We consider the comparison between this approach and the approach based on the ratio of the mean squares as future work. Another possible future work would be sample size calculation based on exact approaches [11-14].

### Acknowledgment

Dr. Shan's research is partially supported by a Faculty Opportunity Awards from UNLV.

### References

- Wilding GE, Shan G, Hutson AD (2012) Exact two-stage designs for phase II activity trials with rank-based endpoints. *Contemp Clin Trials* 33: 332-341.
- Simon R (1989) Optimal two-stage designs for phase II clinical trials. *Control Clin Trials* 10: 1-10.
- Shan G, Hutson AD, Wilding GE (2012) Two-stage k-sample designs for the ordered alternative problem. *Pharm Stat* 11: 287-294.
- Shein-Chung C, Wang H, Shao J (2003) *Sample Size Calculations in Clinical Research* (Chapman & Hall/CRC Biostatistics Series). (2ndedn), CRC Press, USA.
- Borm GF, Fransen J, Lemmens WA (2007) A simple sample size formula for analysis of covariance in randomized clinical trials. *J Clin Epidemiol* 60: 1234-1238.
- Erdfelder E, Faul F, Buchner A (1996) GPOWER: A general power analysis program. *Behavior Research Methods, Instruments, & Computers* 28: 1-11.
- Keppel G, Wickens TD (2004) *Design and Analysis: A Researcher's Handbook*. (4thedn), Pearson, Prentice Hall, USA.
- Shan G, Vexler A, Wilding G, Hutson A (2011) Simple and Exact Empirical Likelihood Ratio Tests for Normality Based on Moment Relations. *Communications in Statistics: Simulation and Computation* 40:129-146.
- Hintze J (2013) PASS 12. NCSS, LLC. Kaysville, Utah, USA.
- Dougados M, Emery P, Lemmel EM, Zerbini CA, Brin S, et al. (2005) When a DMARD fails, should patients switch to sulfasalazine or add sulfasalazine to continuing leflunomide? *Ann Rheum Dis* 64: 44-51.
- Shan G, Ma C, Hutson AD, Wilding GE (2012) An efficient and exact approach for detecting trends with binary endpoints. *Stat Med* 31: 155-164.
- Shan G, Ma C, Hutson AD, Wilding GE (2013) Some tests for detecting trends based on the modified Baumgartner-Weiß-Schindler statistics. *Computational Statistics & Data Analysis* 57: 246-261.
- Shan G, Wang W (2013) ExactCliff: An R Package for Computing Exact Confidence Intervals for the Difference of Two Proportions. *The R Journal* 5(2): 62-71.
- Shan G (2013) A note on exact conditional and unconditional tests for Hardy-Weinberg equilibrium. *Hum Hered* 76: 10-17.