ISSN 2090-3359 (Print)
ISSN 2090-3367 (Online)

Advances in Decision Sciences

Volume 22(A)
22nd Anniversary Special Issue
December 2018

Michael McAleer

Editor-in-Chief
University Chair Professor
Asia University, Taiwan

Published by Asia University, Taiwan

ADS@ASIAUNIVERSITY

A Three-arm Non-inferiority Test for Heteroscedastic Data*

Jingchen Ren

School of Statistics
Beijing Normal University
Beijing, China

Xu Guo**

School of Statistics
Beijing Normal University
Beijing, China

October 2018

* The authors are grateful to a referee for helpful comments and suggestions. The research was supported by National Natural Science Foundation of China (11701034, 11601227), and the Fundamental Research Funds for Central Universities.

** **Corresponding author:** xguo12@bnu.edu.cn

Abstract

In this paper, we consider the three-arm non-inferiority trial in the statistical testing framework established by Hida and Tango (2011). As distinct from existing methods, this paper allows the data to be heteroscedastic. Several new test statistics are developed. Numerical simulations are used to illustrate the performance of the novel proposed methods, which are compared with some existing methods. It is found that a recent proposed procedure may not control the Type I Error well when the data are heteroscedastic. Among the three new methods, the Improved Score test has the best numerical performance.

Keywords: Welch t-test, Score test, Three-arm non-inferiority trial.

JEL: C12, C14, C52.

1 Introduction

Two-arm non-inferiority clinical trials with active control has been widely used in the pharmaceutical industry. However, it has several problems, such as it cannot assess the assay sensitivity because of lacking a placebo treatment, which allows us to judge the efficacy of the active control. What's more, it's sensitive to the choice of non-inferiority margin Δ, which must be determined from the effect size of the reference treatment (CPMP, 1999). The choice of Δ has long been a controversial question (Lange et al., 2005), and it is still an ongoing debate today.

In order to solve the problem mentioned above, the three-arm study including a placebo group is recommended to avoid fulfilling high quality requirements due to external validation in two-arm studies (ICH, 2000). Since then, a series of articles about the design and evaluation of non-inferiority three-arm trials have been published. See for instance D'Agostino et al. (2003) and Munk et al. (2005).

For three-arm trials, Pigeot et al. (2003) proposed to formulate non-inferiority as a fraction of the trial sensitivity. This resulted in hypotheses based on the ratio of differences in means. For a given threshold θ, the alternative hypothesis indicates that the relative efficacy of the experimental drug is more than $\theta*100$ percent of the efficacy of the reference compound compared with placebo. For this ratio hypothesis, a t-distributed test statistic was derived, assuming normal distribution and variance homogeneity (Piegot et al., 2003). Schwartz and Denne (2006) described a two-stage procedure for sample size optimization. Hasler et al. (2008) gave an extension for the case of heterogeneous group variances, and Munzell (2009) presented a non-parametric version, suggesting a rank-based test for a three-arm non-inferiority trial using relative treatment effects. These studies can be categorized to the so-called fraction methods, which all formulated the non-inferiority margin as a fraction of the trial sensitivity. Other studies including Koch and Tangen (1999), Koch and

Röhmel(2004), Röhmel(2005), Kieser and Friede(2007), and Ng (2008) are also following this line.

Hida and Tango (2011) proposed another statistical test procedure for three-arm non-inferiority trials to assess the assay sensitivity with the margin Δ defined as a pre-specified difference between treatments under the situation that the primary endpoints are normally distributed with a common, but unknown variance. To be precise, assuming that the primary clinical outcomes under experimental, reference, and placebo treatments X_i^E, X_j^R, X_k^P, are mutually independent and normally distributed with unknown variances. That is, $X_i^E \sim N(\mu_E, \sigma_E^2), i = 1, \cdots, n_E$, $X_j^R \sim N(\mu_R, \sigma_R^2), j = 1, \cdots, n_R$, and $X_k^P \sim N(\mu_P, \sigma_P^2), k = 1, \cdots, n_P$, with sample sizes n_E, n_R, and n_P not necessarily equal. Under the assumption that $\sigma_E^2 = \sigma_R^2 = \sigma_P^2$, Hida and Tango (2011) constructed two sets of hypothesis tests:

$$H_{01} : \mu_E \leq \mu_R - \Delta \text{ versus } H_1 : \mu_E > \mu_R - \Delta$$

and

$$H_{02} : \mu_R \leq \mu_P + \Delta \text{ versus } H_1 : \mu_R > \mu_P - \Delta$$

where both null hypotheses (H_{01} and H_{02}) must be simultaneously rejected by some two-tailed test at the α level, or one-tailed test at the $\alpha/2$ level (Tango, 2003).

Kwong et al. (2012) modified Hida and Tango's approach and discussed the testing procedure in cases where several new treatments are available. And Huang et al. (2015) derived appropriate testing procedures, which generalize those given in Kwong et al. (2012), for non-inferiority trials with treatments that have heterogeneous variances. Hida and Tango (2013) extended their own method (Hida and Tango, 2011) to the method for inference of the difference in the proportions of binary endpoints and derived score-based hypothesis testing and confidence intervals. Santu et al. (2016)

proposed a test procedure to improve the non-inferiority test under non-normal distribution. And many other scholars developed methods for three-arm non-inferiority testing with binary endpoints. For example, Tang and Tang (2014) proposed two asymptotic approaches for testing three-arm non-inferiority via rate difference based on Wald-type and score test statistics. Munk et al. (2010) developed likelihood ratio tests. Li and Gao (2010) used the closed testing principle to establish the hierarchical testing procedure and proposed a group sequential type design. Liu, et al. (2014) presented a three-step testing procedure and derived an optimal sample size allocation rule in an ethical and reliable manner that minimizes the total sample size. Tang et al. (2014) derived saddle-point approximations to the cumulative distribution functions of Wald-type, score and likelihood ratio test statistics and they proposed the exact unconditional, approximate unconditional and Bootstrap-resampling p-value calculation procedures for testing three-arm non-inferiority with small sample sizes.

Recently Lu et al. (2017) modified Hida and Tango (2011)'s method based on pooled estimators of the homogeneous variance. Both Hida and Tango (2011) and Lu et al. (2017) assumed the variances of X_i^E, X_j^R, X_k^P are the same. However, this may not be true in practice. In this paper, we develop several procedures for simultaneously testing H_{01} and H_{02} without common variance assumption.

The paper is organized as follows. In Section 2, we develop the test statistics. Simulation studies are reported in Section 3 where we demonstrate the superior performance of our proposed tests over existing methods. We end the paper with a discussion in Section 4.

2 Test statistics construction

In the following, we will mainly consider three kinds of test methods, namely, the Welch's t-test, the Score test, and the Improved Score test.

2.1 Welch's t-test

Welch's t-test, first proposed by Welch (1938), gives an approximate solution to the Behrens-Fisher problem, the problem to compare the means of two normal populations with the ratio of the populations variances unknown. Comparing with Student's t-test, Welch's t-test is more reliable when the two samples have unequal variances. The basic idea of Welch's t-test is the same as Student's t-test, but the main difference between them is that Welch's t-test uses $\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}$ instead of pooled variance in Student's t-test. Here S_1^2 and S_2^2 are the sample variance of the two samples, and n_1 and n_2 are sample sizes. Now, for testing the null hypotheses H_{01} and H_{02} in our paper, we can easily get the Welch's t-test statistic T_1 and T_2 as follows:

$$T_1 = \frac{\bar{X}_E - \bar{X}_R + \Delta}{\sqrt{\frac{S_E^2}{n_E} + \frac{S_R^2}{n_R}}};$$

$$T_2 = \frac{\bar{X}_R - \bar{X}_P - \Delta}{\sqrt{\frac{S_R^2}{n_R} + \frac{S_P^2}{n_P}}}.$$

Under null hypothesis, according to Welch-atterthwaite equation, T_1 follows a t-distribution with degree of freedom $df_1 = \frac{(\frac{S_E^2}{n_E} + \frac{S_R^2}{n_R})^2}{\frac{(\frac{S_E^2}{n_E})^2}{n_E - 1} + \frac{(\frac{S_R^2}{n_R})^2}{n_R - 1}}$, and T_2 follows another t-distribution with degree of freedom $df_2 = \frac{(\frac{S_R^2}{n_R} + \frac{S_P^2}{n_P})^2}{\frac{(\frac{S_R^2}{n_R})^2}{n_R - 1} + \frac{(\frac{S_P^2}{n_P})^2}{n_R - 1}}$, where S_E^2, S_R^2, S_P^2 denote the sample variances of the experimental, reference and placebo treatments, respectively.

When the nominal level α is given, we have sufficient reason to reject the null hypothesis H_{01} and H_{02} if and only if $T_1 > t_{\alpha/2}(df_1)$ and $T_2 > t_{\alpha/2}(df_2)$, where $t_{\alpha/2}(\upsilon)$ is the upper $100\% \times \alpha/2$ percentile of the t-distribution with υ degrees of freedom. At this time, we could say that the experimental treatment is not inferior to the reference treatment.

The type I error and the power of the Welch's t-test can be calculated as follows:

$$P_{type\,I\,error} = \sup_{H_{01} \cup H_{02}} Pr\{T_1 > t_{\alpha/2}(df_1) \cap T_2 > t_{\alpha/2}(df_2) | H_{01} \cup H_{02}\},$$

$$Power = Pr\{T_1 > t_{\alpha/2}(df_1) \cap T_2 > t_{\alpha/2}(df_2) | H_{11}, H_{12}\}.$$

Tamhane (1979) utilizes the Welch approximation of the degrees of freedom and the Bonferroni inequality based on Ury and Wiggins (1971) to make multiple comparisons of means with unequal variances. And Huang et al. (2015) discussed the type I error and power of Tamhane's method and some relative methods. They suggested that comparing with test procedure proposed by Kwong et al. (2012), Tamhane's method could control the type I error near the given critical criteria α. And we use Welch's t-test to represent Tamhane's test procedure in our following discussion.

2.2 Score test

Score test is a widely used statistical test of a simple null hypothesis that a parameter of interest θ is equal to some particular value θ_0. Normal approximation based on Score test is another solution to the Behrens-Fisher problem. The key to construct the Score statstic is the calculation of the score, which is defined as the first-order derivative of the likelihood function, and the calculation of the Fisher information matrix. Here in our problem, we use the boundary condition to construct the statistics. That's to say, we construct the Score statistic under $\mu_E = \mu_R - \Delta$ and $\mu_R = \mu_P + \Delta$. According to Jin (2009), we can obtain the Score test statistic T_1^* and T_2^* to test the null hypothesis tests H_{01} and H_{02} in the paper:

$$T_1^* = \frac{\bar{X}_E - \bar{X}_R + \Delta}{\sqrt{\frac{\tilde{\sigma}_E^2}{n_E} + \frac{\tilde{\sigma}_R^2}{n_R}}}$$

$$T_2^* = \frac{\bar{X}_R - \bar{X}_P + \Delta}{\sqrt{\frac{\tilde{\sigma}_{R1}^2}{n_R} + \frac{\tilde{\sigma}_P^2}{n_P}}}.$$

The only difference when we construct T_1^* and T_2^* is that we use data from the experimental and reference treatment in deducing T_1^*, and we use data from the reference and placebo treatment in deducing T_2^*. Thus, we only explain the meaning of the parameter in T_1^*, and readers can easily get the meaning of the parameter in T_2^* based on our following explanation. Here, $\tilde{\sigma}_E^2 = S_1^{\star 2} + t_0^2 S_{12}^{\star 2}$, $\tilde{\sigma}_R^2 = S_2^{\star 2} + (1-t_0)^2 S_{12}^{\star 2}$, $S_1^{\star 2} = \frac{n_E - 1}{n_E} S_E^2$, $S_2^{\star 2} = \frac{n_R - 1}{n_R} S_R^2$, $S_{12}^{\star 2} = (\bar{X}_E - \bar{X}_R)^2$, where S_E^2 and S_R^2 are the sample variances of the experimental and reference treatments. And t_0 is the solution in the interval (0,1) to the following cubic equation:

$$S_{12}^{\star 2}(n_E + n_R)t^3 - S_{12}^{\star 2}(2n_E + n_R)t^2 + (n_E S_2^{\star 2} + n_R S_1^{\star 2} + n_E S_{12}^{\star 2})t - n_R S_1^{\star 2} = 0.$$

Under null hypothesis, both T_1^* and T_2^* asymptotically follow the standard normal distribution. Thus, when the nominal level α is given, we have sufficient reason to reject the null hypothesis H_{01} and H_{02} if and only if $T_1^* > Z_{\alpha/2}$ and $T_2^* > Z_{\alpha/2}$, where $Z_{\alpha/2}$ is the upper $100\% \times \alpha/2$ percentile of the standard normal distribution. If we reject the null hypothesis H_{01} and H_{02} simultaneously, the non-inferiority with assay sensitivity can be claimed. Also, we can calculate the type I error and the power of the Score test as follows:

$$P_{type\,I\,error} = \sup_{H_{01} \cup H_{02}} Pr\{T_1^* > Z_{\alpha/2} \cap T_2^* > Z_{\alpha/2} | H_{01} \cup H_{02}\},$$

$$Power = Pr\{T_1^* > Z_{\alpha/2} \cap T_2^* > Z_{\alpha/2} | H_{11}, H_{12}\}.$$

2.3 Improved Score test

In Lu et al. (2017), they constructed the test statistics based on the best unbiased pooled estimators of homogeneous variance, and they took the minimum of the two statistics into consideration when calculating the power and type I error of the hypothesis tests proposed by Hida and Tango (2011). Their method overcomes some shortcomings of Hida and Tango (2011)'s method. Inspired by Lu et al. (2017), we now propose the Improved Score test which also considers the minimum of the two statistics mentioned in subsection 2.2. The statistics used in the Improved Score test is same with the Score test, T_1^* and T_2^*, but now we reject the null hypothesis H_{01} and H_{02} if and only if $min\{T_1^*, T_2^*\} > Z_\alpha$, where Z_α is the upper $100\% \times \alpha/2$ percentile of the standard normal distribution. And we may calculate the type I error and the power of the Score test with following formula:

$$P_{type\ I\ error} = \sup_{H_{01} \cup H_{02}} Pr\{min\{T_1^*, T_2^*\} > Z_\alpha | H_{01} \cup H_{02}\},$$

$$Power = Pr\{min\{T_1^*, T_2^*\} > Z_\alpha | H_{11}, H_{12}\}.$$

3 Simulation studies

In this section, we present the results of simulation studies that compare the type I errors and powers of three methods in our paper with Hida and Tango (2011)'s method and Lu et al. (2017)'s method for both homoscedastic and heteroscedastic data. In all of our simulation studies, the reported results are based on 10,000 Monte Carlo replications.

Table 1 and table 2 show the results of the type I error rates of 5 different methods: Hida and Tango (2011)'s, Lu et al (2017)'s, Welch's t-test, Score test and Improved Score test (Score(I)). From the results, we can see that when the data is homoscedas-

tic, Hida and Tango (2011)'s method, Welch's t-test and Score test performs very conservatively, while the Improved Score test and Lu et al. (2017)'s method works well. That is, both methods could control the type I errors around the given significance level α=0.05. When the data is heteroscedastic, the type I error rates of Lu et al. (2017)'s method is obviously larger than not only the results of other methods but also the given significance level, 0.05, which demonstrates that Lu et al. (2017)'s method may not work well in this situation. Among other four methods, the Improved Score test works best, since other three methods are more conservative to control the type I error rate than the Improved Score test does.

Many factors may have influence on the statistical power of our hypothesis tests. We choose some of them to make a study. The first factor is the proportion of σ_E^2, σ_R^2 and σ_P^2. Here we let σ_R^2=1, and we set 6 groups of variance proportion of σ_E^2: σ_R^2: σ_P^2: 0.2:1:5, 0.25:1:4, 0.4:1:2.5, 0.5:1:2, 0.8:1:1.25, and 1:1:1. The second factor is sample size and the proportion of three group samples. Statistical power may have much difference under small sample size and large sample size, and it may change greatly when the proportion of the three group sample sizes varies. Thus, we set 6 groups to learn the effect of this factor, that is, n_E:n_R:n_P equals to 20:20:20, 30:30:30, 50:50:50, 100:100:100, 120:120:120 and 150:100:50. Another factor is the distance between μ_E and $\mu_P - \Delta$ and the distance between $\mu_P - \Delta$ and μ_R. Recall that our statistic is constructed under the boundary condition, so it's natural to think that with the distance of μ_E and $\mu_P - \Delta$ or the distance of $\mu_P - \Delta$ and μ_R becoming larger, the statistical power may change accordingly. In order to make our simulation more conveniently, we fix Δ=0.5, μ_E=3, and μ_R=2.6, then we change μ_P to discuss the influence of the distance between μ_E and $\mu_P - \Delta$ and the distance between $\mu_P - \Delta$ and μ_R. We set 5 different value of μ_P: 2, 1.6, 1.2, 0.8, and 0.4.

With the simulation design mentioned above, the results are as following. Tables 3-12 show the results of the statistical powers of the five approaches. When the data is

homoscedastic, for all the simulation results, the powers of Lu et al. (2017)'s method and the Improved Score test are consistently larger than other three methods. Lu et al. (2017)'s method and the Improved Score test are almost equally efficient. When the data is heteroscedastic, as mentioned above, Lu et al. (2017)'s method could not control the type I error rate, thus, we only compare the other four methods' statistical powers. Under most circumstances, especially when the sample sizes are relatively small, the Improved Score test has the largest power among the four methods. Hida and Tango (2011)'s method has the second large powers, while Welch's t-test and Score test has much smaller power than the other two. While when the sample sizes are large (more than 100), no matter the sample sizes in each group were equal or not, Hida and Tango (2011)'s method has the largest power, the Improved Score test has the second large power. But at this time, the discrepancy of statistical power of the four methods is trivial. Thus, from the statistical power point, the Improved Score test performs well, especially when the sample sizes are relatively small.

According to the demonstration above, the Improved Score test performs well no matter the data is homoscedastic or heteroscedastic, and thus, when the variances of the experimental, reference and placebo treatments are not the same, we may choose the Improved Score test to test whether the non-inferiority with assay sensitivity can be claimed.

4 Conclusions and discussions

In this paper, we consider the three-arm non-inferiority trial in the statistical testing framework established by Hida and Tango (2011). Different from existing methods, in this paper, we allow the data to be heteroscedastic. We propose several test statistics. Numerical studies are used to illustrate the performance of our proposed method and compare with some existing methods. When the data is heteroscedastic, Lu et al.

(2017)'s method could not control the type I error rates, and Hida and Tango (2011)'s method, Welch's t-test as well as Score test are conservative in controlling the type I error rates. While the improved Score test overcomes this problem, performs well in controlling the type I error rates. Further, the Improved Score test has the largest power among all the methods we discussed in the paper in almost all the conditions, thus, we could say that the Improved Score test is a valid test for the three-arm non-inferiority problem.

In this paper, we only discussed the endpoints of normally distributed data sets, and other endpoints such as binary data are not considered. In the future, we may assess the validity of our method in a three-arm inferiority trial with the placebo treatment having other endpoints.

References

[1] D'Agostino, R. B., Massaro, J. M. and Sullivan, L. M. (2003). Non-inferiority trials: design concepts and issuesthe encounters of academic consultants in statistics. *Statistics in Medicine*, **22**, 169-186.

[2] Ghosh, S., Chatterjee, A. and Ghosh, S.(2016). Non-inferiority test based on transformations for non-normal distributions. *Computational Statistics and Data Analysis*, **113**.

[3] Hasler, M., Vonk, R. and Hothorn, L. A. (2008). Assessing non-inferiority of a new treatment in a three-arm trial in the presence of heteroscedasticity. *Statistics in Medicine*, **27**, 490-503.

[4] Hida, E. and Tango, T. (2011). On the three-arm non-inferiority trial including a placebo with a prespecified margin. *Statistics in Medicine*, **30**, 224-231.

[5] Hida, E. and Tango, T. (2013). Three-arm non-inferiority trials with a prespecified margin for inference of the difference in the proportions of binary endpoints. *Journal of Biopharmaceutical Statistics*, **23(4)**, 774-789.

[6] Higuchi, T., Murasaki, M. and Kamijima, K. (2009). Clinical evaluation of duloxetine in the treatment of major depressive disorder-placebo- and paroxetine-controlled double-blind comparative study. *Japanese Journal of Clinical Psychopharmacology*, **12**, 1613-1634.

[7] Huang, L. C., Wen, M. J. and Cheung, S. H. (2015). Noninferiority studies with multiple new treatments and heterogeneous variances. *Journal of Biopharmaceutical Statistics*,**25(5)**, 958-971.

[8] Jin, H., Zheng, Sh. T. and Chen, W. Q. (2009). Normal approximation to the Behrens-Fisher problem. *Statistical Reasearch*, **26(11)**,106-108.

[9] Kieser, M. and Friede, T. (2007). Planning and analysis of three-arm non-inferiority trials with binary endpoints. *Statistics in Medicine*, **26**, 253-273.

[10] Koch, G. G. and Tangen, C. M. (1999). Nonparametric analysis of covariance and its role in noninferiority clinical trials. *Drug Information Journal*, **33**, 1145-1159.

[11] Koch, A. and Röhmel, J. (2004). Hypothesis testing in the 'gold standard' design for proving the efficacy of an experimental treatment relative to placebo and a reference. *Journal of Biopharmaceutical Statistics*, **14**, 315-325.

[12] Kwong, K. S., Cheung, S. H., Hayter, A. J. and Wen, M. J. (2012). Extension of three-arm non-inferiority studies to trials with multiple new treatments. *Statistics in Medicine*, **31**, 2833-2843.

[13] Lange, S. and Freitag, G. (2005). Choice of delta: requirements and reality-results of a systematic review. *Biometrical Journal*, **47**, 12-27.

[14] Li, G. and Gao, S. (2010). A group sequential type design for three-arm non-inferiority trials with binary endpoints. *Biometrical Journal*, **52**, 504-518.

[15] Liu, J. T., Tzeng, C. S. and Tsou, H. H. (2014). Establishing non-inferiority of a new treatment in a three-arm trial: apply a step-down hierarchical model in a papulopustular acne study and an oral prophylactic antibiotics study. *International Journal of Statistics in Medical Research*, **3**, 11-20.

[16] Lu, H. Z., Jin, H. and Zeng, W. X. (2017). A more efficient three-arm non-inferiority test based on pooled estimators of the homogeneous variance. *Statistical Methods in Medical Research*, in press.

[17] Munk, A., Mielke, M., Skipka, G. and Freitag, G. (2010). Testing non-inferiority in three-armed clinical trials based on likelihood ratio statistics. *Canadian Journal of Statistics*, **35(3)**, 413-431.

[18] Munk, A. and Trampisch, H.J. (2005). Therapeutic equivalence-clinical issues and statistical methodology in non-inferiority trials. *Biometrical Journal*, **47**, 7-9.

[19] Munzell, U. (2009). Nonparametric non-inferiority analyses in the three-arm design with active control and placebo.*Statistics in Medicine*, **28**, 3643-3656.

[20] Ng, T. H. (2008). Non-inferiority hypotheses and choice of noninferiority margin. *Statistics in Medicine*, **27**, 5392-5406.

[21] Pigeot, I., Schäfer, J., Röhmel, J. and Hauschke, D. (2003). Assessing non-inferiority of a new treatment in a three-arm clinical trial including a placebo. *Statistics in Medicine*, **22**, 883-899.

[22] Röhmel, J. (2005). On confidence bounds for the ratio of net differences in the 'gold standard' design with reference, experimental, and placebo treatment. *Biometrical Journal*, **47**, 799-806.

[23] Schwartz, T. A. and Denne, J. S. (2006). A two-stage sample size recalculation procedure for placebo-and active-controlled non-inferiority trials. *Statistics in Medicine*, **25**, 3396-3406.

[24] Tamhane, A. C. (1979). A comparison of procedures for multiple comparisons of means with unequal variances. *Journal of the American Statistical Association*, **74**, 471-480.

[25] Tang, M. L. and Tang, N. S. (2004). Tests of non-inferiority via rate difference for three-arm clinical trials with placebo. *Journal of Biopharmaceutical Statistics*, **14(2)**, 337-347.

[26] Tang, N. S., Yu, B. and Tang, M. L. (2014). Testing non-inferiority of a new treatment in three-arm clinical trials with binary endpoints. *Bmc Medical Research Methodology*, **14(1)**, 134.

[27] Tango, T. (2003). *Randomized controlled trial (in Japanese)*. Tokyo: Asakura Publishing.

[28] Ury, H. K. and Wiggins, A. D. (1971). Large sample and other multiple comparisons among means. *British Journal of Mathematical and Statistical Psychology*, **24**, 174-194..

[29] Welch,B.L. (1938). The significance of the difference between two means when the population variance are unequal. *Biometrika*, **29(3/4)**, 350-362.

[30] CPMP Concept Paper on the development of a Committee for Proprietary Medicinal Products(CPMP). Points to Consider on biostatistical=methodological issues arising from recent CPMP discussions on licensing applications: Choice of delta, www.emea.eu.int/pdfs/human/ewp/215899en.pdf (1999).

[31] ICH Harmonised Tripartite Guideline. Choice of control group and related issues in clinical trials, www.ich.org/pdfICH/e10step4.pdf(2000).

Table 1: Comparing the type I error rates of five approaches

$\mu_E = 3$,$\mu_R = 2.6$,$\alpha = 0.05$,$\sigma_R^2 = 1$,$n_R = n_p = 30$

Δ	n_E	σ_E^2	σ_P^2	Welch's t	Hida	Score	Score(I)	Lu
0.4	20	0.2	5	2.35%	2.55%	1.79%	4.21%	7.91%
$\mu_P = 2.6$		0.25	4	2.49%	2.59%	2.01%	4.29%	7.55%
		0.4	2.5	2.20%	2.40%	1.86%	3.86%	6.96%
		0.5	2	2.57%	2.76%	1.66%	3.98%	6.78%
		0.8	1.25	2.53%	2.26%	1.60%	4.23%	5.91%
		1	1	2.43%	2.39%	1.99%	4.07%	4.82%
	30	0.2	5	2.30%	2.94%	1.88%	4.77%	8.75%
		0.25	4	2.33%	2.64%	2.04%	4.11%	8.68%
		0.4	2.5	2.36%	2.77%	1.84%	4.28%	8.05%
		0.5	2	2.58%	2.69%	1.92%	4.48%	7.21%
		0.8	1.25	2.44%	2.11%	1.90%	4.11%	5.91%
		1	1	2.75%	2.48%	1.82%	3.77%	5.37%
0.5	20	0.2	5	2.55%	2.63%	1.85%	4.15%	7.67%
$\mu_P = 2.5$		0.25	4	2.35%	2.52%	1.64%	4.23%	7.48%
		0.4	2.5	2.23%	2.45%	1.49%	4.03%	7.14%
		0.5	2	2.55%	2.41%	1.40%	4.07%	7.00%
		0.8	1.25	2.59%	2.53%	1.47%	3.80%	5.59%
		1	1	2.58%	2.49%	1.58%	3.76%	5.08%
	30	0.2	5	2.49%	2.62%	1.73%	4.00%	8.79%
		0.25	4	2.14%	2.62%	1.68%	3.83%	8.54%
		0.4	2.5	2.58%	2.64%	1.71%	3.88%	8.43%
		0.5	2	2.50%	2.62%	1.60%	3.86%	7.55%
		0.8	1.25	2.46%	2.49%	1.39%	3.88%	6.53%
		1	1	2.61%	2.44%	1.55%	3.88%	5.17%
0.6	20	0.2	5	2.57%	2.73%	1.48%	3.77%	7.74%
$\mu_P = 2.4$		0.25	4	2.41%	2.42%	1.53%	3.87%	7.35%
		0.4	2.5	2.61%	2.34%	1.48%	3.48%	6.85%
		0.5	2	2.46%	2.81%	1.30%	3.72%	6.69%
		0.8	1.25	2.41%	2.61%	1.26%	3.48%	6.08%
		1	1	2.43%	2.61%	1.39%	3.55%	4.98%
	30	0.2	5	2.26%	2.48%	1.40%	3.80%	8.85%
		0.25	4	2.52%	2.59%	1.51%	3.72%	8.42%
		0.4	2.5	2.40%	2.75%	1.58%	3.66%	8.02%
		0.5	2	2.51%	2.65%	1.46%	3.63%	6.91%
		0.8	1.25	2.60%	2.52%	1.28%	3.54%	5.84%
		1	1	2.53%	2.40%	1.35%	3.23%	5.07%

Table 2: Comparing the type I error rates of five approaches

$\mu_E = 3$,$\mu_R = 2.6$,$\alpha = 0.05$,$\sigma_R^2 = 1$,$n_R = n_p = 30$

Δ	n_E	σ_E^2	σ_P^2	Welch's t	Hida	Score	Score(I)	Lu
0.8	20	0.2	5	2.51%	2.68%	1.33%	3.29%	7.92%
$\mu_P = 2.2$		0.25	4	2.62%	2.77%	1.14%	3.35%	7.72%
		0.4	2.5	2.53%	2.64%	1.13%	3.00%	7.38%
		0.5	2	2.50%	2.41%	0.99%	3.44%	6.91%
		0.8	1.25	2.40%	2.75%	0.98%	2.73%	5.70%
		1	1	2.35%	2.45%	0.82%	2.92%	5.25%
	30	0.2	5	2.58%	2.69%	1.14%	3.28%	8.76%
		0.25	4	2.44%	2.58%	0.94%	2.86%	8.48%
		0.4	2.5	2.52%	2.46%	1.31%	2.84%	7.48%
		0.5	2	2.43%	2.44%	1.16%	2.92%	7.39%
		0.8	1.25	2.33%	2.52%	1.15%	2.57%	5.55%
		1	1	2.56%	2.35%	0.93%	2.67%	5.57%
1	20	0.2	5	2.45%	2.78%	0.91%	2.77%	7.81%
$\mu_P = 2$		0.25	4	2.34%	2.78%	0.80%	2.60%	7.82%
		0.4	2.5	2.50%	2.58%	0.84%	2.39%	7.16%
		0.5	2	2.38%	2.58%	0.55%	2.13%	6.85%
		0.8	1.25	2.57%	2.53%	0.63%	1.91%	5.61%
		1	1	2.42%	2.39%	0.59%	2.01%	5.36%
	30	0.2	5	2.53%	2.59%	0.83%	3.22%	8.77%
		0.25	4	2.63%	2.41%	0.79%	2.48%	8.78%
		0.4	2.5	2.51%	2.62%	0.71%	2.41%	7.86%
		0.5	2	2.70%	2.43%	0.61%	2.56%	7.23%
		0.8	1.25	2.67%	2.45%	0.60%	1.84%	6.34%
		1	1	2.55%	2.33%	0.48%	2.07%	4.75%

Table 3: Comparing the statistical powers of four approaches

$\mu_E = 3$,$\mu_R = 2.6$,$\Delta = 0.5$,$\alpha = 0.05$,$\sigma_R^2 = 1$

μ_P	$n_E : n_R : n_P$	σ_E^2	σ_P^2	Welch's t	Hida	Score	Score(I)
2	20:20:20	0.2	5	0.0322	0.0355	0.0199	0.0512
		0.25	4	0.0302	0.0319	0.0189	0.0489
		0.4	2.5	0.031	0.0313	0.0221	0.0525
		0.5	2	0.0311	0.0311	0.0163	0.0502
		0.8	1.25	0.0247	0.0243	0.0161	0.0457
		1	1	0.0227	0.0219	0.0116	0.0417
	30:30:30	0.2	5	0.0417	0.0414	0.027	0.0643
		0.25	4	0.0357	0.0448	0.0292	0.0648
		0.4	2.5	0.0467	0.0452	0.0317	0.068
		0.5	2	0.0432	0.0455	0.0301	0.0716
		0.8	1.25	0.0398	0.041	0.0279	0.0712
		1	1	0.0443	0.0419	0.0281	0.066
	50:50:50	0.2	5	0.0467	0.0509	0.0396	0.0744
		0.25	4	0.0514	0.0506	0.0359	0.0776
		0.4	2.5	0.058	0.053	0.0382	0.09
		0.5	2	0.0622	0.0638	0.0445	0.087
		0.8	1.25	0.0697	0.0656	0.0476	0.0985
		1	1	0.0655	0.067	0.051	0.1031
	100:100:100	0.2	5	0.0602	0.0615	0.0506	0.0941
		0.25	4	0.0607	0.0623	0.0556	0.105
		0.4	2.5	0.08	0.0754	0.0632	0.1178
		0.5	2	0.0853	0.0855	0.0624	0.1184
		0.8	1.25	0.0966	0.1019	0.0777	0.1452
		1	1	0.1002	0.1037	0.0858	0.146
	120:120:120	0.2	5	0.0497	0.1053	0.0413	0.0781
		0.25	4	0.0525	0.1118	0.0423	0.0798
		0.4	2.5	0.0612	0.1097	0.0439	0.0924
		0.5	2	0.0735	0.0976	0.046	0.0982
		0.8	1.25	0.0834	0.0945	0.052	0.1112
		1	1	0.0904	0.0877	0.0618	0.1219
	150:100:50	0.2	5	0.0495	0.1018	0.035	0.0658
		0.25	4	0.0469	0.1023	0.0349	0.077
		0.4	2.5	0.056	0.0928	0.0411	0.0841
		0.5	2	0.0657	0.0946	0.044	0.0866
		0.8	1.25	0.0741	0.0912	0.0476	0.1013
		1	1	0.083	0.0844	0.0567	0.1049
1.6	20:20:20	0.2	5	0.1169	0.1299	0.0849	0.1761
		0.25	4	0.1324	0.1427	0.104	0.197
		0.4	2.5	0.17	0.1702	0.124	0.2435
		0.5	2	0.1777	0.1835	0.1322	0.2693
		0.8	1.25	0.2039	0.2071	0.1543	0.3085
		1	1	0.211	0.217	0.1615	0.3149

Table 4: Comparing the statistical powers of four approaches

$\mu_E = 3$,$\mu_R = 2.6$,$\Delta = 0.5$,$\alpha = 0.05$,$\sigma_R^2 = 1$

μ_P	$n_E : n_R : n_P$	σ_E^2	σ_P^2	Welch's t	Hida	Score	Score(I)
1.6	30:30:30	0.2	5	0.1873	0.1913	0.1512	0.2523
		0.25	4	0.2144	0.222	0.1783	0.2948
		0.4	2.5	0.2858	0.29	0.2388	0.3732
		0.5	2	0.3335	0.3244	0.2602	0.4166
		0.8	1.25	0.3883	0.3879	0.3234	0.4902
		1	1	0.4198	0.4191	0.3483	0.5217
	50:50:50	0.2	5	0.2906	0.3	0.2635	0.3808
		0.25	4	0.3471	0.34	0.2954	0.4326
		0.4	2.5	0.462	0.4709	0.4076	0.5596
		0.5	2	0.5268	0.5276	0.4682	0.6095
		0.8	1.25	0.6382	0.6502	0.5842	0.7244
		1	1	0.6922	0.6966	0.6345	0.7682
	100:100:100	0.2	5	0.527	0.5351	0.4951	0.6242
		0.25	4	0.6036	0.6102	0.5681	0.6902
		0.4	2.5	0.7625	0.7582	0.7168	0.828
		0.5	2	0.8162	0.8283	0.796	0.8737
		0.8	1.25	0.9099	0.9106	0.8897	0.9409
		1	1	0.9398	0.9405	0.9234	0.9624
	120:120:120	0.2	5	0.3693	0.528	0.3197	0.4614
		0.25	4	0.4268	0.5935	0.3767	0.5165
		0.4	2.5	0.5978	0.7028	0.5221	0.6719
		0.5	2	0.6757	0.7561	0.6093	0.74
		0.8	1.25	0.8226	0.848	0.7729	0.8653
		1	1	0.8789	0.8788	0.8348	0.9144
	150:100:50	0.2	5	0.3197	0.4748	0.2597	0.4016
		0.25	4	0.3771	0.5235	0.3183	0.4568
		0.4	2.5	0.5284	0.6309	0.446	0.5991
		0.5	2	0.596	0.6783	0.5266	0.6637
		0.8	1.25	0.7557	0.7787	0.6757	0.8073
		1	1	0.8129	0.828	0.7514	0.8527
1.2	20:20:20	0.2	5	0.3037	0.3267	0.2495	0.4201
		0.25	4	0.356	0.377	0.3028	0.4669
		0.4	2.5	0.4823	0.4689	0.3938	0.5918
		0.5	2	0.5163	0.5171	0.4572	0.6425
		0.8	1.25	0.5854	0.5903	0.5357	0.7233
		1	1	0.591	0.6015	0.5386	0.7314
	30:30:30	0.2	5	0.4953	0.5071	0.4287	0.5903
		0.25	4	0.5695	0.5729	0.5046	0.6663
		0.4	2.5	0.7249	0.7109	0.6627	0.798
		0.5	2	0.7732	0.7725	0.722	0.8448
		0.8	1.25	0.8562	0.8513	0.8167	0.9069
		1	1	0.8546	0.8609	0.8393	0.9208

Table 5: Comparing the statistical powers of four approaches

$\mu_E = 3, \mu_R = 2.6, \Delta = 0.5, \alpha = 0.05, \sigma_R^2 = 1$

μ_P	$n_E : n_R : n_P$	σ_E^2	σ_P^2	Welch's t	Hida	Score	Score(I)
1.2	50:50:50	0.2	5	0.7273	0.7331	0.6853	0.8015
		0.25	4	0.8008	0.7992	0.7666	0.8586
		0.4	2.5	0.9176	0.924	0.8981	0.9489
		0.5	2	0.952	0.9557	0.9364	0.9687
		0.8	1.25	0.9808	0.9833	0.9801	0.9918
		1	1	0.9843	0.9868	0.9845	0.9938
	100:100:100	0.2	5	0.9561	0.9542	0.9427	0.9733
		0.25	4	0.9796	0.9828	0.9722	0.9906
		0.4	2.5	0.9972	0.9979	0.9971	0.9986
		0.5	2	0.9993	0.9994	0.9991	0.9996
		0.8	1.25	1	0.9999	1	1
		1	1	1	1	1	1
	120:120:120	0.2	5	0.8337	0.913	0.8033	0.8802
		0.25	4	0.8997	0.9578	0.8737	0.9348
		0.4	2.5	0.978	0.9901	0.9669	0.9852
		0.5	2	0.9916	0.995	0.9883	0.9949
		0.8	1.25	0.9993	0.9995	0.9988	0.9995
		1	1	0.9998	1	0.9998	0.9998
	150:100:50	0.2	5	0.7642	0.8679	0.7135	0.8241
		0.25	4	0.8403	0.9144	0.7908	0.8853
		0.4	2.5	0.9519	0.9705	0.9295	0.9713
		0.5	2	0.979	0.9887	0.965	0.9879
		0.8	1.25	0.9971	0.9977	0.9939	0.9984
		1	1	0.9988	0.9991	0.9985	0.9997
0.8	20:20:20	0.2	5	0.5772	0.5958	0.5059	0.6935
		0.25	4	0.6449	0.6641	0.5795	0.7506
		0.4	2.5	0.7681	0.771	0.7267	0.8606
		0.5	2	0.7978	0.802	0.775	0.8899
		0.8	1.25	0.7939	0.798	0.796	0.9003
		1	1	0.7752	0.7711	0.7859	0.8778
	30:30:30	0.2	5	0.8071	0.8098	0.7531	0.8635
		0.25	4	0.8622	0.8708	0.8277	0.9136
		0.4	2.5	0.9444	0.9476	0.9267	0.9705
		0.5	2	0.955	0.9596	0.9513	0.9767
		0.8	1.25	0.9471	0.9496	0.9469	0.9776
		1	1	0.9225	0.9316	0.9373	0.9661
	50:50:50	0.2	5	0.9612	0.962	0.9473	0.9747
		0.25	4	0.9826	0.9799	0.9734	0.9916
		0.4	2.5	0.9973	0.9983	0.9964	0.9993
		0.5	2	0.999	0.9996	0.9988	0.9996
		0.8	1.25	0.9962	0.9959	0.9971	0.9988
		1	1	0.9934	0.9944	0.9951	0.9978

Table 6: Comparing the statistical powers of four approaches

$\mu_E = 3$,$\mu_R = 2.6$,$\Delta = 0.5$,$\alpha = 0.05$,$\sigma_R^2 = 1$

μ_P	$n_E : n_R : n_P$	σ_E^2	σ_P^2	Welch's t	Hida	Score	Score(I)
0.8	100:100:100	0.2	5	0.9997	0.9996	0.9995	0.9996
		0.25	4	0.9998	1	0.9999	1
		0.4	2.5	1	1	1	1
		0.5	2	1	1	1	1
		0.8	1.25	1	1	1	1
		1	1	1	1	0.9999	1
	120:120:120	0.2	5	0.9885	0.9961	0.9826	0.9934
		0.25	4	0.9959	0.9993	0.9929	0.9992
		0.4	2.5	0.9999	1	1	1
		0.5	2	1	1	1	1
		0.8	1.25	1	1	1	1
		1	1	1	1	1	1
	150:100:50	0.2	5	0.9704	0.9911	0.9586	0.9838
		0.25	4	0.9891	0.9966	0.9834	0.9944
		0.4	2.5	0.9992	0.9998	0.9987	0.9998
		0.5	2	1	1	0.9998	1
		0.8	1.25	1	1	0.9999	1
		1	1	1	1	1	1
0.4	20:20:20	0.2	5	0.7949	0.8043	0.7544	0.8798
		0.25	4	0.8441	0.8495	0.8139	0.911
		0.4	2.5	0.8793	0.884	0.8904	0.9485
		0.5	2	0.8809	0.8794	0.891	0.9506
		0.8	1.25	0.8306	0.8329	0.8558	0.9129
		1	1	0.7924	0.7899	0.8176	0.8907
	30:30:30	0.2	5	0.9548	0.953	0.9338	0.9756
		0.25	4	0.9712	0.9742	0.9616	0.9858
		0.4	2.5	0.9797	0.9826	0.9822	0.9929
		0.5	2	0.9775	0.9747	0.9795	0.9909
		0.8	1.25	0.9529	0.9501	0.9596	0.9782
		1	1	0.9322	0.9263	0.9355	0.968
	50:50:50	0.2	5	0.9985	0.9983	0.998	0.9988
		0.25	4	0.9998	0.999	0.9993	0.9995
		0.4	2.5	0.9995	0.9997	0.9998	1
		0.5	2	0.9993	0.9995	0.9992	0.9996
		0.8	1.25	0.9962	0.9966	0.9967	0.9987
		1	1	0.9932	0.9942	0.9949	0.998
	100:100:100	0.2	5	1	0.9999	1	1
		0.25	4	22 1	1	1	1
		0.4	2.5	1	1	1	1
		0.5	2	1	1	1	1
		0.8	1.25	1	1	1	1
		1	1	1	0.9999	1	1

Table 7: Comparing the statistical powers of four approaches

$\mu_E = 3$,$\mu_R = 2.6$,$\Delta = 0.5$,$\alpha = 0.05$,$\sigma_R^2 = 1$

μ_P	$n_E : n_R : n_P$	σ_E^2	σ_P^2	Welch's t	Hida	Score	Score(I)
0.4	120:120:120	0.2	5	0.9999	1	0.9997	0.9998
		0.25	4	1	1	1	1
		0.4	2.5	1	1	1	1
		0.5	2	1	1	1	1
		0.8	1.25	1	1	1	1
		1	1	1	1	1	1
	150:100:50	0.2	5	0.9989	0.9998	0.998	0.9994
		0.25	4	0.9998	1	0.9999	0.9998
		0.4	2.5	1	1	1	1
		0.5	2	1	1	1	1
		0.8	1.25	1	1	1	1
		1	1	1	1	1	1

Table 8: Comparing the statistical powers of four approaches

$\mu_E = 3, \mu_R = 2.6, \Delta = 0.5, \alpha = 0.05, \sigma_R^2 = 1$

μ_P	$n_E : n_R : n_P$	σ_E^2	σ_P^2	Welch's t	Hida	Score	Score(I)
2	20:20:20	0.2	5	0.0406	0.046	0.0381	0.0969
		0.25	4	0.0427	0.0494	0.0375	0.0947
		0.4	2.5	0.0476	0.0482	0.0379	0.1034
		0.5	2	0.0497	0.0477	0.0405	0.1009
		0.8	1.25	0.0438	0.0427	0.0371	0.1136
		1	1	0.0394	0.0439	0.036	0.1079
	30:30:30	0.2	5	0.0757	0.0832	0.0653	0.134
		0.25	4	0.0868	0.0889	0.0763	0.1514
		0.4	2.5	0.0957	0.0997	0.0823	0.1717
		0.5	2	0.0979	0.1094	0.0922	0.1906
		0.8	1.25	0.1137	0.1152	0.0989	0.1994
		1	1	0.1085	0.1073	0.0966	0.208
	50:50:50	0.2	5	0.1272	0.1294	0.1204	0.2081
		0.25	4	0.1518	0.1491	0.1354	0.2287
		0.4	2.5	0.1847	0.2034	0.1755	0.281
		0.5	2	0.2157	0.2176	0.1959	0.311
		0.8	1.25	0.2554	0.2692	0.2324	0.3671
		1	1	0.2758	0.271	0.2453	0.4011
	100:100:100	0.2	5	0.2173	0.2256	0.2159	0.3304
		0.25	4	0.2673	0.269	0.2469	0.3592
		0.4	2.5	0.3541	0.3506	0.3374	0.4708
		0.5	2	0.4108	0.3975	0.3958	0.5134
		0.8	1.25	0.5167	0.5177	0.4918	0.6264
		1	1	0.5583	0.5597	0.5317	0.6672
	120:120:120	0.2	5	0.1617	0.2775	0.1442	0.2371
		0.25	4	0.1904	0.3085	0.1692	0.2747
		0.4	2.5	0.2602	0.3644	0.2384	0.3594
		0.5	2	0.3152	0.3792	0.2661	0.3946
		0.8	1.25	0.4081	0.438	0.3728	0.5056
		1	1	0.4744	0.47	0.4259	0.5672
	150:100:50	0.2	5	0.141	0.2636	0.1217	0.2097
		0.25	4	0.1651	0.2715	0.1442	0.2396
		0.4	2.5	0.2228	0.3238	0.196	0.3093
		0.5	2	0.2664	0.3442	0.2332	0.3535
		0.8	1.25	0.3561	0.3778	0.3144	0.4459
		1	1	0.4014	0.408	0.3611	0.4931
1.6	20:20:20	0.2	5	0.1511	0.1578	0.1271	0.2652
		0.25	4	0.1703	0.1815	0.1502	0.2975
		0.4	2.5	0.2113	0.2218	0.2004	0.3545
		0.5	2	0.231	0.2413	0.2217	0.3967
		0.8	1.25	0.2658	0.2675	0.2417	0.4215
		1	1	0.2608	0.2717	0.2533	0.4383

Table 9: Comparing the statistical powers of four approaches

$\mu_E = 3$,$\mu_R = 2.6$,$\Delta = 0.5$,$\alpha = 0.05$,$\sigma_R^2 = 1$

μ_P	$n_E : n_R : n_P$	σ_E^2	σ_P^2	Welch's t	Hida	Score	Score(I)
1.6	30:30:30	0.2	5	0.2812	0.3013	0.27	0.4239
		0.25	4	0.3315	0.3335	0.315	0.4692
		0.4	2.5	0.4286	0.4407	0.4039	0.5795
		0.5	2	0.4835	0.4805	0.4483	0.6282
		0.8	1.25	0.528	0.5364	0.5214	0.6992
		1	1	0.5404	0.5366	0.5239	0.6958
	50:50:50	0.2	5	0.5022	0.5092	0.4876	0.617
		0.25	4	0.5789	0.5777	0.5586	0.6808
		0.4	2.5	0.7315	0.7305	0.7009	0.8191
		0.5	2	0.788	0.7862	0.7687	0.8705
		0.8	1.25	0.8536	0.8582	0.8524	0.9264
		1	1	0.8683	0.873	0.8671	0.9326
	100:100:100	0.2	5	0.8163	0.8068	0.8016	0.8799
		0.25	4	0.8751	0.8744	0.8685	0.9228
		0.4	2.5	0.9597	0.9619	0.9544	0.9805
		0.5	2	0.9807	0.9793	0.9771	0.989
		0.8	1.25	0.995	0.9962	0.9955	0.9981
		1	1	0.9974	0.9975	0.9979	0.9987
	120:120:120	0.2	5	0.6349	0.7743	0.6023	0.7225
		0.25	4	0.7052	0.83	0.6791	0.7998
		0.4	2.5	0.8674	0.9247	0.8487	0.9163
		0.5	2	0.9135	0.9507	0.9071	0.9553
		0.8	1.25	0.9846	0.9865	0.9767	0.9913
		1	1	0.9907	0.9922	0.9897	0.996
	150:100:50	0.2	5	0.5454	0.7032	0.5135	0.6537
		0.25	4	0.6276	0.7713	0.5932	0.7312
		0.4	2.5	0.7992	0.8701	0.7791	0.8692
		0.5	2	0.8794	0.9166	0.8442	0.9133
		0.8	1.25	0.9594	0.9675	0.9504	0.9744
		1	1	0.9801	0.9806	0.9735	0.9898
1.2	20:20:20	0.2	5	0.3446	0.3594	0.3265	0.5094
		0.25	4	0.393	0.4012	0.372	0.5655
		0.4	2.5	0.4769	0.483	0.4702	0.6668
		0.5	2	0.5153	0.5156	0.5077	0.6846
		0.8	1.25	0.5295	0.5174	0.5421	0.7042
		1	1	0.5047	0.503	0.5235	0.6756
	30:30:30	0.2	5	0.5983	0.612	0.5844	0.7412
		0.25	4	0.56761	0.6844	0.6505	0.7957
		0.4	2.5	0.7733	0.7829	0.7725	0.8763
		0.5	2	0.7941	0.7969	0.7992	0.8899
		0.8	1.25	0.7784	0.7753	0.7939	0.8823
		1	1	0.745	0.7474	0.7615	0.8549

Table 10: Comparing the statistical powers of four approaches

$\mu_E = 3, \mu_R = 2.6, \Delta = 0.5, \alpha = 0.05, \sigma_R^2 = 1$

μ_P	$n_E : n_R : n_P$	σ_E^2	σ_P^2	Welch's t	Hida	Score	Score(I)
1.2	50:50:50	0.2	5	0.8734	0.8699	0.8576	0.9265
		0.25	4	0.9194	0.9215	0.9083	0.9579
		0.4	2.5	0.9696	0.9692	0.969	0.9887
		0.5	2	0.9722	0.9719	0.9738	0.9889
		0.8	1.25	0.9543	0.9565	0.9568	0.9806
		1	1	0.932	0.9331	0.937	0.9687
	100:100:100	0.2	5	0.9933	0.9937	0.9912	0.9974
		0.25	4	0.9987	0.9983	0.9976	0.9997
		0.4	2.5	1	0.9999	1	1
		0.5	2	0.9999	0.9999	0.9999	0.9999
		0.8	1.25	0.9996	0.9994	0.9996	0.9997
		1	1	0.9984	0.9984	0.9983	0.9993
	120:120:120	0.2	5	0.9453	0.9781	0.9441	0.9721
		0.25	4	0.9792	0.9919	0.971	0.9898
		0.4	2.5	0.9991	0.9991	0.997	0.999
		0.5	2	0.9997	1	0.9991	1
		0.8	1.25	0.9997	0.9999	0.9999	1
		1	1	1	0.9998	0.9999	0.9998
	150:100:50	0.2	5	0.9014	0.9595	0.8893	0.9431
		0.25	4	0.9539	0.9814	0.9421	0.973
		0.4	2.5	0.9938	0.9965	0.9904	0.9975
		0.5	2	0.9977	0.999	0.9979	0.9995
		0.8	1.25	0.9995	0.9998	0.9999	1
		1	1	0.9999	0.9996	0.9999	0.9998
0.8	20:20:20	0.2	5	0.5679	0.5779	0.5546	0.735
		0.25	4	0.6172	0.6318	0.6085	0.7784
		0.4	2.5	0.6671	0.6713	0.6775	0.8243
		0.5	2	0.6637	0.6683	0.6875	0.8117
		0.8	1.25	0.6091	0.6248	0.6515	0.7572
		1	1	0.5691	0.5774	0.6038	0.7356
	30:30:30	0.2	5	0.8318	0.8455	0.8326	0.9105
		0.25	4	0.8637	0.8729	0.8727	0.9328
		0.4	2.5	0.8807	0.8814	0.8903	0.9414
		0.5	2	0.8612	0.8653	0.8812	0.9331
		0.8	1.25	0.7916	0.7994	0.8215	0.8845
		1	1	0.7608	0.7567	0.7738	0.8599
	50:50:50	0.2	5	0.9829	0.9832	0.981	0.993
		0.25	4	0.9689	0.9915	0.9888	0.9964
		0.4	2.5	0.9848	0.9842	0.9874	0.9961
		0.5	2	0.9826	0.9795	0.9816	0.9918
		0.8	1.25	0.9552	0.9568	0.9589	0.9814
		1	1	0.9331	0.9406	0.9399	0.9689

Table 11: Comparing the statistical powers of four approaches

$\mu_E = 3$,$\mu_R = 2.6$,$\Delta = 0.5$,$\alpha = 0.05$,$\sigma_R^2 = 1$

μ_P	$n_E : n_R : n_P$	σ_E^2	σ_P^2	Welch's t	Hida	Score	Score(I)
0.8	100:100:100	0.2	5	0.9999	1	0.9999	1
		0.25	4	1	1	1	1
		0.4	2.5	1	1	1	1
		0.5	2	0.9999	1	0.9999	1
		0.8	1.25	0.9995	0.9992	0.9996	0.9997
		1	1	0.9984	0.9984	0.9993	0.9994
	120:120:120	0.2	5	0.9985	0.9997	0.9982	0.9991
		0.25	4	0.9998	1	0.9999	0.9999
		0.4	2.5	1	1	1	1
		0.5	2	1	1	1	1
		0.8	1.25	0.9999	0.9997	0.9999	1
		1	1	0.9994	0.9999	0.9999	1
	150:100:50	0.2	5	0.9933	0.9988	0.9906	0.9967
		0.25	4	0.9988	0.9994	0.9983	0.9998
		0.4	2.5	1	1	1	1
		0.5	2	1	0.9999	1	1
		0.8	1.25	1	0.9997	0.9998	1
		1	1	0.9996	0.9996	0.9995	1
0.4	20:20:20	0.2	5	0.7052	0.7262	0.7165	0.8475
		0.25	4	0.7247	0.7343	0.7461	0.86
		0.4	2.5	0.7188	0.7253	0.7523	0.8459
		0.5	2	0.6971	0.6962	0.7349	0.8201
		0.8	1.25	0.6267	0.622	0.664	0.7622
		1	1	0.579	0.5783	0.6077	0.7268
	30:30:30	0.2	5	0.9088	0.9143	0.9224	0.9633
		0.25	4	0.9143	0.9177	0.9258	0.9613
		0.4	2.5	0.8873	0.8878	0.9081	0.942
		0.5	2	0.8626	0.858	0.8851	0.9322
		0.8	1.25	0.803	0.8059	0.8213	0.8916
		1	1	0.7559	0.7624	0.7737	0.8656
	50:50:50	0.2	5	0.9917	0.9939	0.9927	0.9989
		0.25	4	0.9926	0.992	0.9926	0.9963
		0.4	2.5	0.985	0.9853	0.987	0.9946
		0.5	2	0.9795	0.9787	0.9808	0.9921
		0.8	1.25	0.9556	0.9537	0.9603	0.982
		1	1	0.9314	0.9336	0.9424	0.9668
	100:100:100	0.2	5	1	1	0.9999	1
		0.25	4	27 1	1	1	1
		0.4	2.5	1	1	1	1
		0.5	2	0.9999	0.9999	1	1
		0.8	1.25	0.9995	0.9995	0.9997	0.9997
		1	1	0.9984	0.9982	0.9987	0.9998

Table 12: Comparing the statistical powers of four approaches

$\mu_E = 3$,$\mu_R = 2.6$,$\Delta = 0.5$,$\alpha = 0.05$,$\sigma_R^2 = 1$

μ_P	$n_E : n_R : n_P$	σ_E^2	σ_P^2	Welch's t	Hida	Score	Score(I)
0.4	120:120:120	0.2	5	1	1	0.9999	1
		0.25	4	1	1	1	1
		0.4	2.5	1	1	1	1
		0.5	2	1	1	1	1
		0.8	1.25	0.9999	1	0.9999	0.9999
		1	1	0.9998	0.9999	0.9998	1
	150:100:50	0.2	5	0.9999	1	0.9997	0.9999
		0.25	4	1	1	1	1
		0.4	2.5	1	1	1	1
		0.5	2	1	1	0.9999	1
		0.8	1.25	1	0.9998	0.9997	1
		1	1	0.9998	0.9997	0.9997	0.9998