Psicológica (2013), 34, 125-143.

On using a pilot sample variance for sample size determination in the detection of differences between two means: Power consideration

Gwowen Shieh*

National Chiao Tung University, Hsinchu, Taiwan

The a priori determination of a proper sample size necessary to achieve some specified power is an important problem encountered frequently in practical studies. To establish the needed sample size for a two-sample *t* test, researchers may conduct the power analysis by specifying scientifically important values as the underlying population means while using a variance estimate obtained from related research or pilot study. In order to take account of the variability of sample variance, this article considers two approaches to sample size determinations. One provides the sample size required to guarantee with a given assurance probability that the actual power exceeds the planned power. The other gives the necessary sample size such that the expected power attains the designated power level. The suggested paradigm of adjusted sample variance combines the existing procedures into one unified framework. Numerical results are presented to illustrate the usefulness and advantages of the proposed approaches that accommodate the stochastic nature of the sample variance. More importantly, supplementary computer programs are developed to aid the usefulness and implementation of the suggested techniques. The exposition helps to clarify discrepancy in the previous demonstration and to extend the development of sample size methodology.

When designing research studies, the determination of sample size is an essential process in order to ensure there is adequate statistical power to detect scientifically credible effects. To make inferences about differences between two normal population means, the hypothesis testing procedure and corresponding sample size formula are well known and easy to apply. Specifically, Kupper and Hafner (1989) demonstrated that the particular procedure that considers statistical power performs amazingly well even for very small sample sizes. Also, recent discussions of related sample size

* Gwowen Shieh. Department of Management Science. National Chiao Tung University. 1001 Ta Hsueh Road, Hsinchu, Taiwan 30010, R.O.C. Email: gwshieh@mail.nctu.edu.tw

issues can be found in Guo and Luh (2011), Jan and Shieh (2011), and Lawson and Fisher (2011). For important general procedures, see the comprehensive treatment in Cohen (1988), Desu and Raghavarao (1990), Kraemer and Thieman (1987), Murphy and Myors (2004), and Odeh and Fox (1991). Due to the prospective nature of advance research planning, it is difficult to assess the adequacy of selected configurations for model parameters including two population means and common error variance. However, the general guideline suggests that typical sources like previously published research and successful pilot studies can offer plausible and reasonable planning values for the vital model characteristics (Thabane et al. 2010). Nonetheless, it is good practice to consider a range of design variations to provide guidance about the achieved power levels and required sample sizes for the study.

In view of the imprecision of parameter settings, Shiffler and Harwood (1985) investigated the effect on the realized α-risk of using a pilot sample variance to estimate variance on the sample size formula for testing population means. Their results suggested that researchers should recognize that the actual type I error rates may be substantially different than the nominal level. Also, Browne (1995) examined the deficiency of using a sample variance to compute the sample size needed to achieve the planned power for one- and two-sample *t*-tests. The empirical results of Browne showed that the actual power attained with the calculated sample size is quite likely to be less than the planned power. More importantly, he proposed to improve the underpowered condition by using the upper confidence limit of the sample variance rather than the sample variance estimate itself. The imputation of selected upper confidence limit for population variance in the sample size calculation can guarantee that the actual power will exceed the planned power with designated probability. In practice, the actual sample size required to guarantee the assurance probability with respect to the planned power is closely related to the underlying distributional property and magnitude of sample variance estimate from a pilot sample or related investigation.

On the other hand, Kieser and Wassmer (1996), and Julious and Owen (2006) suggested an alternative approach to accommodate the variability of sample variance for sample size determination. Specifically, Kieser and Wassmer investigated the expected power performance of the adjusted sample variance method of Browne (1995). While Julious and Owen used the unadjusted sample variance to compute the necessary sample size so that the expected power of a two-sample *t*-test will meet the planned power. Just as in the case of power assurance probability

consideration of Browne, the numerical illustration with respect to expected power showed that the sample sizes provided by the traditional formulas are too small since they neglect the imprecise nature of a variance estimate. Thus this may lead to a distorted power performance and unsatisfactory research outcome for the planned study. Notably, in his review of 30 clinical trials published along with their pilot data in top medical journals Vickers (2003) found 80% of the studies are underpowered. It should be clear that the insufficient statistical power phenomenon does not occur exclusively in only one particular scientific discipline. Also, Kraemer, Mintz, Noda, Tinklenberg and Yesavage (2006) cautioned the use of pilot studies to perform power calculations for study proposals. It seems prudent, therefore, to emphasize that the embedded properties of the sample size procedures should be well understood before they are adopted by researchers in performing sample size evaluation.

The distinct advantage of the prescribed methods is that it circumvents the uncertainty of sample variance by taking account of the underlying chi-square distribution of sample variance and permits a corrected sample size determination according to the desired assurance probability and expected power considerations. The two criteria of assurance probability and expected power are fundamentally different and provide potentially useful tool in finding sample size with good properties. However, the theoretical presentations and algebraic expressions in Browne (1995), Kieser and Wassmer (1996), and Julious and Owen (2006) appear to be diverse and incomplete. Two important aspects of these results should be pointed out.

First, Kieser and Wassmer (1996) did not specifically address the issue of how to modify the sample variance in sample size calculations so that the expected power of a two-sample *t*-test will attain the planned power. They only evaluated the expected power of the two-sample *t* test with sample sizes that are required to satisfy the selected assurance level of power. Second, although Julious and Owen (2006) studied the prescribed problem of determining the sample size under the notion of expected power, their analytical exposition for the overall or expected power function is complicated and does not conform to the adjusted sample variance approach. Consequently, the suggested sample size formula appears to be opaque and the procedure may be of less practical value in application. For pedagogical importance and practical interest, one must have a thorough understanding of the fundamental details of the sample size methodology before the technique is finally considered to be appropriate for making sound application.

In addition to the abovementioned natural formulation, a Bayesian analysis is also viable. For example, O'Hagan, Stevens and Campbell (2005) proposed to incorporate a lognormal prior distribution for the unknown variance and computed the necessary sample size to achieve a desired expected power. However, the hyper-parameters of the prior distribution still have to be imputed and are usually elicited from reliable sources such as the knowledge of experts. On the other hand, Gillett (2001) described a Bayesian approach to calculate the sample size for a designated level of expected power with respect to the posterior distribution of probable effect sizes. The uncertainty in effect size is a combination of t value and sample size from a previous study, and a normal prior distribution with specified hyper-parameter values.

Essentially, the Bayesian approach is more sophisticated than the standard setup described above. It is conceivable that the accuracy of the procedures depends mostly on the appropriateness of prior distributions which often accompanies subjective hyper-parameter values. The interested reader is referred to Gillett (2001), O'Hagan et al. (2005) and the references therein for further details. Here we focus on the problem of dealing with the uncertainty inherent in a sample variance estimate under the ultimate notion of choosing a profound adjusted factor to provide sufficient sample sizes with supplied power strength for subsequent main study. Although the discussion concentrated on testing the equality of two group means, the principles and procedures are also applicable in more complicated models and it embodies all the critical issues without the distracting complications of the multiple treatment groups and more sophisticated statistical models.

In order to improve the quality of research findings, this article attempts to contribute to the derivation and evaluation of sample size methodology for two-sample t tests in two important and distinctive aspects. First, we present a simplified and constructive approach to sample size determination for the two-sample problem which computes the sample size required to guarantee that the actual power exceeds the planned power with a given assurance probability. Second, we reexamine the expected power approach to sample size determination through rigorous analytical presentations and numerical assessments.

The general formulation described in this article combines the existing procedures into one unified framework. In the process, we attempt to provide a clear and concise exposition of the fundamental theoretical arguments, and conduct exact empirical investigations to demonstrate the potential deficiency of existing findings. Thus, a well-supported and useful recommendation can be offered for empirical studies. Numerical results are

provided for a variety of situations to demonstrate the individual impact of deterministic factors and how they work as whole pertaining to the two different power considerations. Furthermore, a numerical example is presented to illustrate the usefulness and advantage of the proposed methods that account for the embedded randomness and distributional characteristic of the sample variance. For practical purposes, the computer codes are developed to facilitate the recommended procedures for computing the necessary sample size in planning research designs.

Fundamental Methodology

Consider independent random samples from two normal populations with the following formulations:

$$X_{1i} = \mu_1 + \varepsilon_{1i} \text{ and } X_{2j} = \mu_2 + \varepsilon_{2j}, \qquad (1)$$

where μ_1 and μ_2 are unknown parameters, and ε_{1i} and ε_{2j} are *iid* $N(0, \sigma^2)$ random variables, $i = 1, \ldots, N_1$ and $j = 1, ..., N_2$. For the purpose of detecting the group effect in terms of the hypothesis H_0: $\mu_1 = \mu_2$ versus H_1: $\mu_1 \neq \mu_2$, the well-known test statistic t under the model formulation in Equation 1 is of the form

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\{S_p^2(1/N_1 + 1/N_2)\}^{1/2}} \qquad (2)$$

where

$$\bar{X}_1 = \sum_{i=1}^{N_1} X_{1i}/N_1, \ \bar{X}_2 = \sum_{j=1}^{N_2} X_{2j}/N_2, \ S_p^2 = SSE/(N_1 + N_2 - 2)$$

is the usual unbiased estimator of σ^2, and

$$SSE = \sum_{i=1}^{N_1} (X_{1i} - \bar{X}_1)^2 + \sum_{j=1}^{N_2} (X_{2j} - \bar{X}_2)^2.$$

For ease of exposition, the two sample sizes N_1 and N_2 are expressed as $N_1 = N$ and $N_2 = rN_1$, respectively, where $r > 0$ is a constant. If the null hypothesis H_0: $\mu_1 = \mu_2$ is true, then the statistic t is distributed as $t(\nu)$, a central t distribution with $\nu = N_1 + N_2 - 2 = N(1 + r) - 2$ degrees of freedom; and H_0 is rejected at the significance level α if $|t| > t_{\nu,\alpha/2}$, where $t_{\nu,\alpha/2}$ is the upper $100(\alpha/2)$th percentile of the t distribution $t(\nu)$. Under the

alternative hypothesis, t has a noncentral t distribution $t(\nu, \delta)$ with ν degrees of freedom and noncentrality parameter δ

$$t \sim t(\nu, \delta), \tag{3}$$

where $\delta = d/(c\sigma^2/N)^{1/2}$, $d = \mu_1 - \mu_2$, and $c = 1 + 1/r$. With the distribution in Equation 3, the associated power function is denoted by

$$\pi(N, \sigma^2) = P\{|t(\nu, \delta)| > t_{\nu,\alpha/2}\}. \tag{4}$$

For the purpose of sample size determination, the power functions defined in Equation 4 can be employed to calculate the sample size N_t needed to attain the specified power $1 - \beta$ for the chosen significance level α and parameter values (μ_1, μ_2, σ^2). Ultimately, the resulting sample size N_t is the least integer that ensures $\pi(N_t, \sigma^2) = P\{|t(\nu_t, \delta_t)| > t_{\nu_t,\alpha/2}\} \geq 1 - \beta$, where $\nu_t = N_t(1 + r) - 2$ and $\delta_t = d/(c\sigma^2/N_t)^{1/2}$.

Although the preceding sample size determination is well documented, this article considers the practical scenario that the variance parameter is estimated with a sample value available from published research or a preliminary study. The sample size calculations under this circumstance directly follow the prescribed procedure with the obvious substitution of σ^2 with a sample variance estimate, denoted by $\hat{\sigma}_p^2$. In this case, the adapted power function is

$$\pi(N, \hat{\sigma}_p^2) = P\{|t(\nu, \hat{\delta}_p)| > t_{\nu,\alpha/2}\}, \tag{5}$$

where $\hat{\delta}_p = d/(c\hat{\sigma}_p^2/N)^{1/2}$. Accordingly, the required sample size $\widehat{N}_p$ is the least integer that guarantees $\pi(\widehat{N}_p, \hat{\sigma}_p^2)) \geq 1 - \beta$. Clearly, the calculations for sample size $\widehat{N}_p$ are straightforward to apply and do not incur any extra complexity. However, the reported sample size $\widehat{N}_p$ tends to be underestimated ($\widehat{N}_p < N_t$). Consequently, a test using $\widehat{N}_p$ as the sample size is more likely to produce an actual power smaller than the planned power, i.e., $\pi(\widehat{N}_p, \sigma^2) < 1 - \beta$. The explanation is provided next to emphasize the importance for further investigation and improvement.

Regarding the underlying randomness of the sample variance $\hat{\sigma}_p^2$, it is commonly assumed that it has the following distribution

$$\upsilon \cdot \hat{\sigma}_p^2 \sim \sigma^2 \chi^2(\upsilon), \tag{6}$$

where $\chi^2(\upsilon)$ is a chi-square distribution with υ degrees of freedom. It is well known that $\hat{\sigma}_p^2$ is an unbiased estimator of σ^2, however, the right-skewness of the chi-square distribution gives $P\{\hat{\sigma}_p^2 < \sigma^2\} = P\{\chi^2(\upsilon) < \upsilon\} > 0.5$. Hence, an observed sample variance $\hat{\sigma}_p^2 < \sigma^2$ occurs more often than $\hat{\sigma}_p^2 >$

σ^2. With the designated mean difference d, significance level α, power level $1 - \beta$, and selected sample sizes $\widehat{N}_p$ and N_t, both power functions basically give the identical value $\pi(\widehat{N}_p, \hat{\sigma}_p^2) \doteq \pi(N_t, \sigma^2) \doteq 1 - \beta$. Note that the generic power function in Equation 4 is a monotone function of noncentrality δ for fixed degrees of freedom ν, and that the discrepancy between the power values computed with different degrees of freedom is comparatively negligible. Thus, the identity $\pi(\widehat{N}_p, \hat{\sigma}_p^2) \doteq \pi(N_t, \sigma^2)$ implies the approximate equivalence between the two noncentrality parameters or $\hat{\sigma}_p^2/\widehat{N}_p \doteq \sigma^2/N_t$. Moreover, it leads to $\widehat{N}_p < N_t$ and $\pi(\widehat{N}_p, \sigma^2) < \pi(N_t, \sigma^2) \doteq 1 - \beta$ if $\hat{\sigma}_p^2 < \sigma^2$. Hence, the calculated sample size $\widehat{N}_p$ based on observed sample variance $\hat{\sigma}_p^2$ has a high probability of giving insufficient power because $P\{\hat{\sigma}_p^2 < \sigma^2\} > 0.5$. This phenomenon has been demonstrated in the numerical presentations of Browne (1995).

In view of the deficiency in the sample size calculation associated with power formula in Equation 5, it is of both theoretical interest and practical importance to consider alternative procedures. We apply the general idea in Browne (1995) in the following two different approaches to sample size determination with the use of an adjusted form of pilot sample variance $\hat{\sigma}_a^2 = a \cdot \hat{\sigma}_p^2$, where a is a constant to be chosen. With the use of a multiple of sample variance $\hat{\sigma}_a^2$ in place of $\hat{\sigma}_p^2$ in Equation 5, the power function for sample size calculation is modified as

$$\pi(N, \hat{\sigma}_a^2) = P\{|t(\nu, \hat{\delta}_a)| > t_{\nu, \alpha/2}\}, \qquad (7)$$

where $\hat{\delta}_a = d/(c\,\hat{\sigma}_a^2/N)^{1/2}$. Accordingly, the required sample size $\widehat{N}_a$ is the minimum sample size so that $\pi(\widehat{N}_a, \hat{\sigma}_a^2) \geq 1 - \beta$ and the actual power is the value of $\pi(\widehat{N}_a, \sigma^2)$.

To account for the stochastic nature of pilot sample variance in sample size determination, the overall or unconditional considerations of assurance probability and expected power are presented next. It is noteworthy that the two principles of assurance probability and expected width are closely related to the two standard criteria of consistency and unbiasedness in statistical point estimation, respectively. In other words, these two measures impose unique and distinct aspects of desirable characteristics on the resulting power behavior, and each principle has conceptual and empirical implications in its own right.

Assurance Probability Approach

Since $\hat{\sigma}_p^2$ is a random variable with the scaled chi-square distribution given in Equation 6, the power function defined in Equation 7 is also a random variable. Accordingly, the power function of Equation 7 based on an observed sample variance $\hat{\sigma}_p^2$ in $\hat{\sigma}_a^2 = a \cdot \hat{\sigma}_p^2$ can be viewed as an estimate because the sample variance estimate $\hat{\sigma}_p^2$ represents a realization of all possible outcomes. It is constructive to ensure the actual power is at least as large as the planned power with a designated assurance probability.

Specifically, given mean difference d, significance level α, planned power $1 - \beta$ and assurance probability $1 - \gamma$, this procedure purports to find the proper correction factor a so that it satisfies the equality

$$\Gamma(\upsilon, \hat{\sigma}_a^2) = P\{\pi(\widehat{N}_a, \sigma^2) \geq 1 - \beta\} = 1 - \gamma. \quad (8)$$

It is important to note that the assurance probability $\Gamma(\upsilon, \hat{\sigma}_a^2)$ depends on the unknown parameter σ^2 and cannot be evaluated in practice. However, a useful and accurate approximation can be obtained. Following an argument similar to that employed above for the direct use of uncorrected sample variance, it is clear that $\pi(\widehat{N}_a, \hat{\sigma}_a^2) \doteq \pi(N_t, \sigma^2) \doteq 1 - \beta$ and $\hat{\sigma}_a^2/\widehat{N}_a \doteq \sigma^2/N_t$. Thus,

$$P\{\pi(\widehat{N}_a, \sigma^2) \geq 1 - \beta\} \doteq P\{\pi(\widehat{N}_a, \sigma^2) \geq \pi(N_t, \sigma^2)\} =$$
$$= P(\widehat{N}_a \geq N_t) \doteq P(\widehat{N}_a \geq \sigma^2). \quad (9)$$

With the specified distribution of sample variance given in Equation 6, the assurance probability $\Gamma(\upsilon, \hat{\sigma}_a^2)$ can be well approximated by

$$\Gamma(\upsilon, \hat{\sigma}_a^2) \doteq P(K \geq \upsilon/a), \quad (10)$$

where $K = \upsilon \cdot \hat{\sigma}_p^2/\sigma^2 \sim \chi^2(\upsilon)$. It is important to note that the approximation is independent of the parameter values of mean difference d and error variance σ^2. For the selected assurance level of $1 - \gamma$, the assurance probability $\Gamma(\upsilon, \hat{\sigma}_a^2) \doteq 1 - \gamma$ if $a = g$ where

$$g = \upsilon/\chi_{\upsilon,\gamma}^2 \quad (11)$$

and $\chi_{\upsilon,\gamma}^2$ is the $100 \cdot \gamma$th percentile of a chi-square distribution with υ degrees of freedom.

For the purpose of sample size determination, the modified power function given in Equation 7 with $a = g$ can be employed to calculate the sample size $\widehat{N}_g$ needed to test hypothesis H_0: $\mu_1 = \mu_2$ versus H_1: $\mu_1 \neq \mu_2$ in order to attain the specified assurance probability $(1 - \gamma)$ for planned power $(1 - \beta)$ with significance level α and mean difference d. Accordingly, the computed sample size $\widehat{N}_g$ varies with the actual value of the pilot sample

variance from one application to another. Let $z_{\alpha/2}$ and z_β denote the upper 100(α/2)th and 100·βth percentiles of the standard normal distribution, respectively. Since the sample size $\widehat{N}_g$ conditional on $\hat{\sigma}_g^2 = g\cdot\hat{\sigma}_p^2$ is approximately equivalent to $\widehat{N}_g \doteq c\hat{\sigma}_g^2\ (z_{\alpha/2} + z_\beta)^2/d^2$, the unconditional or expected sample size $N(\upsilon, \hat{\sigma}_g^2)$ is $N(\upsilon, \hat{\sigma}_g^2) = E\{\widehat{N}_g\} \doteq E\{c\hat{\sigma}_g^2\ (z_{\alpha/2} + z_\beta)^2/d^2\} = cg\cdot\sigma^2(z_{\alpha/2} + z_\beta)^2/d^2$.

Similarly, the expected sample size corresponding to the usual procedure without sample variance correction is $N(\upsilon, \widehat{\sigma}_p^2) = E\{\widehat{N}_p\} \doteq E\{c\widehat{\sigma}_p^2\ (z_{\alpha/2} + z_\beta)^2/d^2\} = c\cdot\sigma^2(z_{\alpha/2} + z_\beta)^2/d^2$. Obviously, $\widehat{N}_p < \widehat{N}_g$ and $N(\upsilon,\hat{\sigma}_p^2) < N(\upsilon, \widehat{\sigma}_g^2)$ if $g > 1$. Note that the approximation in Equation 10 was also presented in Kieser and Wassmer (1996) to justify the simulation results in Browne (1995). However, a close examination shows the theoretical justification in Kieser and Wassmer (1996) involves some minor and unstated simplifications about the interchange of degrees of freedom $\hat{\nu}_g = \widehat{N}_g(1 + r) - 2$ and $\nu_t = N_t(1 + r) - 2$ in our notation. This is the key argument that prevents a direct proof of $N_t < \widehat{N}_g$ in the foregoing exposition and should be explicitly stated.

In order to examine the performance of the prescribed approximate method for sample size determinations, numerical investigations are performed for the detection of mean difference. The actual computation of the exact assurance probability $\Gamma(\upsilon,\ \widehat{\sigma}_g^2)$ in Equation 8 requires the evaluation of noncentral *t* cumulative distribution function and the one-dimensional integration with respect to a chi-square distribution. Specifically, for each possible value of the sample variance $\widehat{\sigma}_p^2$ with a scaled chi-square distribution, the least sample size $\widehat{N}_g$ is computed so that the corresponding power $\pi(\widehat{N}_g,\ \widehat{\sigma}_g^2) \geq 1 - \beta$, where $\widehat{\sigma}_g^2 = g\cdot\widehat{\sigma}_p^2$ and g is calculated as in Equation 11. Also, the actual power $\pi(\widehat{N}_g, \sigma^2)$ obtained with sample size $\widehat{N}_g$ is evaluated. Then the exact assurance probability is the expected value $\Gamma(\upsilon,\ \widehat{\sigma}_g^2) = E_K[I(\widehat{N}_g,\ \sigma^2)]$, where $E_K[\cdot]$ denotes the expectation taken with respect to the chi-square distribution of $K \sim \chi^2(\upsilon)$ given in Equation 10, and $I(\upsilon,\ \widehat{\sigma}_g^2) = 1$ if $\pi(\widehat{N}_g, \sigma^2) \geq 1 - \beta$ and $I(\upsilon,\ \widehat{\sigma}_g^2) = 0$ if $\pi(\widehat{N}_g,\ \sigma^2) < 1 - \beta$. Likewise, the exact expected sample size is the expected value $N(\upsilon,\ \widehat{\sigma}_g^2) = E_K[\widehat{N}_g]$. Moreover, the exact values of $\Gamma(\upsilon,\ \widehat{\sigma}_p^2)$ and $N(\upsilon, \widehat{\sigma}_p^2)$ can be obtained with the same computing procedures. It should be noted that neither Browne (1995) nor Kieser and Wassmer (1996) conducted exact computations of assurance probability. Only simulation results were reported in Browne (1995).

To facilitate the application of sample size planning under assurance probability consideration, selected comparisons of exact and approximate results associated with and without sample variance modification were performed. The adequacy of the approximate procedure is determined by the error = exact assurance probability – approximate assurance probability. For the chosen model configurations of $N_1 = N_2$, $\sigma^2 = 1$, planned power $(1 - \beta) = 0.90$, assurance probability $(1 - \gamma) = 0.80$, and $\alpha = 0.05$, the assurance probabilities $\Gamma(\upsilon, \widehat{\sigma}_p^2)$ and $\Gamma(\upsilon, \widehat{\sigma}_g^2)$ for sample variance degrees of freedom $\upsilon = 10, 50, 100$, and 500 are listed in Tables 1-3 for $d = 0.25, 0.50$ and 1, respectively. In addition, the corresponding exact and approximate magnitudes of expected sample sizes $N(\upsilon, \widehat{\sigma}_p^2)$ and $N(\upsilon, \widehat{\sigma}_g^2)$ are also presented. The correction multiplier g in $\widehat{\sigma}_g^2$ is $g = \upsilon/\chi^2_{\upsilon,0.20} = 1.6184$, 1.2063, 1.1371 and 1.0566 for $\upsilon = 10, 50, 100$, and 500, respectively.

Examination of the exact values of $\Gamma(\upsilon, \widehat{\sigma}_p^2)$ and $\Gamma(\upsilon, \widehat{\sigma}_g^2)$ in Tables 1-3 confirms the undesired behavior of the usual procedure by directly substituting sample variance $\widehat{\sigma}_p^2$ for σ^2 in sample size calculation. Specifically, the values of $\Gamma(\upsilon, \widehat{\sigma}_p^2)$ in Tables 1-3 fall within the interval of 0.4408 and 0.4971, which appears to be too low to be satisfactory in power assurance. On the contrary, the results correspond to the case of corrected counterpart $\widehat{\sigma}_g^2$ are nearly equivalent to the designated level of 0.80. Certainly, the discrepancy in assurance probability $\Gamma(\upsilon, \widehat{\sigma}_p^2)$ and $\Gamma(\upsilon, \widehat{\sigma}_g^2)$ simply reflects the substantial difference between the expected sample sizes of $N(\upsilon, \widehat{\sigma}_p^2)$ and $N(\upsilon, \widehat{\sigma}_g^2)$. Regarding the performance of suggested approximate formulas, there is a close agreement between the exact and approximate values of expected sample size and assurance probability. The largest absolute error of expected sample size is 1.52 while the maximum absolute error of assurance probability is 0.0055. According to these findings, the performance of the approximate method given in Equation 10 with $a = g$ seems to be reasonably good for the range of model specifications considered here. Therefore, the required sample size to ensure sufficient assurance probability can be computed with the adapted procedure with adjusted sample variance $\widehat{\sigma}_g^2$ so that the possible low statistical power problem for detecting mean difference can be recognized in advance.

Table 1. Assurance probability and expected power using unadjusted and adjusted pilot sample variance when $N_1 = N_2 = N$, $\sigma^2 = 1$, and $d = 0.25$

		$\hat{\sigma}_p^2$			$\hat{\sigma}_g^2 = g\hat{\sigma}_p^2$			$\hat{\sigma}_h^2 = h\hat{\sigma}_p^2$		
υ		$N(\upsilon, \hat{\sigma}_p^2)$	$\Gamma(\upsilon, \hat{\sigma}_p^2)$	$\Pi(\upsilon, \hat{\sigma}_p^2)$	$N(\upsilon, \hat{\sigma}_g^2)$	$\Gamma(\upsilon, \hat{\sigma}_g^2)$	$\Pi(\upsilon, \hat{\sigma}_g^2)$	$N(\upsilon, \hat{\sigma}_h^2)$	$\Gamma(\upsilon, \hat{\sigma}_h^2)$	$\Pi(\upsilon, \hat{\sigma}_h^2)$
10	Approximation	336.24	0.4405	0.8357	544.15	0.8000	0.9385	437.28	0.6592	0.9000
	Exact method	337.69	0.4408	0.8363	545.59	0.8003	0.9387	438.73	0.6597	0.9004
	Error	−1.45	−0.0003	−0.0006	−1.44	−0.0003	−0.0002	−1.44	−0.0005	−0.0004
50	Approximation	336.24	0.4734	0.8858	405.60	0.8000	0.9322	354.09	0.5751	0.9000
	Exact method	337.69	0.4748	0.8862	407.06	0.8005	0.9324	355.54	0.5761	0.9004
	Error	−1.46	−0.0014	−0.0005	−1.45	−0.0005	−0.0003	−1.46	−0.0010	−0.0004
100	Approximation	336.24	0.4812	0.8928	382.33	0.8000	0.9259	345.03	0.5535	0.9000
	Exact method	337.69	0.4828	0.8932	383.78	0.8010	0.9262	346.49	0.5550	0.9004
	Error	−1.46	−0.0016	−0.0004	−1.45	−0.0010	−0.0003	−1.46	−0.0015	−0.0004
500	Approximation	336.24	0.4916	0.8985	355.27	0.8000	0.9135	337.98	0.5241	0.9000
	Exact method	337.70	0.4960	0.8990	356.73	0.8026	0.9139	339.43	0.5287	0.9004
	Error	−1.46	−0.0045	−0.0004	−1.46	−0.0026	−0.0004	−1.46	−0.0046	−0.0004

Note: $\hat{\sigma}_p^2$ is the pilot sample variance, and $\hat{\sigma}_g^2 = g\hat{\sigma}_p^2$ and $\hat{\sigma}_h^2 = h\hat{\sigma}_p^2$ are adjusted variances with g and h determined by Equations 11 and 15, respectively.

Table 2. Assurance probability and expected power using unadjusted and adjusted pilot sample variance when $N_1 = N_2 = N$, $\sigma^2 = 1$, and $d = 0.50$

		$\hat{\sigma}_p^2$			$\hat{\sigma}_g^2 = g\hat{\sigma}_p^2$			$\hat{\sigma}_h^2 = h\hat{\sigma}_p^2$		
υ		$N(\upsilon, \hat{\sigma}_p^2)$	$\Gamma(\upsilon, \hat{\sigma}_p^2)$	$\Pi(\upsilon, \hat{\sigma}_p^2)$	$N(\upsilon, \hat{\sigma}_g^2)$	$\Gamma(\upsilon, \hat{\sigma}_g^2)$	$\Pi(\upsilon, \hat{\sigma}_g^2)$	$N(\upsilon, \hat{\sigma}_h^2)$	$\Gamma(\upsilon, \hat{\sigma}_h^2)$	$\Pi(\upsilon, \hat{\sigma}_h^2)$
10	Approximation	84.06	0.4405	0.8357	136.04	0.8000	0.9385	109.32	0.6592	0.9000
	Exact method	85.53	0.4408	0.8381	137.50	0.8003	0.9394	110.79	0.6597	0.9015
	Error	−1.47	−0.0003	−0.0024	−1.46	−0.0003	−0.0010	−1.47	−0.0005	−0.0015
50	Approximation	84.06	0.4734	0.8858	101.40	0.8000	0.9322	88.52	0.5751	0.9000
	Exact method	85.53	0.4743	0.8876	102.87	0.8005	0.9333	89.99	0.5761	0.9016
	Error	−1.47	−0.0009	−0.0018	−1.47	−0.0005	−0.0011	−1.47	−0.0010	−0.0016
100	Approximation	84.06	0.4812	0.8928	95.58	0.8000	0.9259	86.26	0.5535	0.9000
	Exact method	85.53	0.4821	0.8945	97.05	0.8005	0.9271	87.73	0.5544	0.9016
	Error	−1.47	−0.0009	−0.0018	−1.47	−0.0005	−0.0013	−1.47	−0.0009	−0.0016
500	Approximation	84.06	0.4916	0.8985	88.82	0.8000	0.9135	84.49	0.5241	0.9000
	Exact method	85.53	0.4939	0.9002	90.29	0.8018	0.9150	85.96	0.5265	0.9017
	Error	−1.47	−0.0023	−0.0017	−1.47	−0.0018	−0.0015	−1.47	−0.0025	−0.0017

Note: $\hat{\sigma}_p^2$ is the pilot sample variance, and $\hat{\sigma}_g^2 = g\hat{\sigma}_p^2$ and $\hat{\sigma}_h^2 = h\hat{\sigma}_p^2$ are adjusted variances with g and h determined by Equations 11 and 15, respectively.

Table 3. Assurance probability and expected power using unadjusted and adjusted pilot sample variance when $N_1 = N_2 = N$, $\sigma^2 = 1$, and $d = 1.00$

		$\hat{\sigma}_p^2$			$\hat{\sigma}_g^2 = g\hat{\sigma}_p^2$			$\hat{\sigma}_h^2 = h\hat{\sigma}_p^2$		
υ		$N(\upsilon, \hat{\sigma}_p^2)$	$\Gamma(\upsilon, \hat{\sigma}_p^2)$	$\Pi(\upsilon, \hat{\sigma}_p^2)$	$N(\upsilon, \hat{\sigma}_g^2)$	$\Gamma(\upsilon, \hat{\sigma}_g^2)$	$\Pi(\upsilon, \hat{\sigma}_g^2)$	$N(\upsilon, \hat{\sigma}_h^2)$	$\Gamma(\upsilon, \hat{\sigma}_h^2)$	$\Pi(\upsilon, \hat{\sigma}_h^2)$
10	Approximation	21.01	0.4405	0.8357	34.01	0.8000	0.9385	27.33	0.6592	0.9000
	Exact method	22.53	0.4413	0.8455	35.50	0.8003	0.9423	28.83	0.6597	0.9061
	Error	−1.52	−0.0008	−0.0098	−1.49	−0.0003	−0.0038	−1.50	−0.0005	−0.0061
50	Approximation	21.01	0.4734	0.8858	25.35	0.8000	0.9322	22.13	0.5751	0.9000
	Exact method	22.52	0.4759	0.8930	26.85	0.8014	0.9366	23.64	0.5772	0.9064
	Error	−1.51	−0.0025	−0.0072	−1.50	−0.0014	−0.0045	−1.51	−0.0021	−0.0064
100	Approximation	21.01	0.4812	0.8928	23.90	0.8000	0.9259	21.56	0.5535	0.9000
	Exact method	22.52	0.4841	0.8996	25.40	0.8020	0.9308	23.07	0.5564	0.9064
	Error	−1.51	−0.0029	−0.0069	−1.50	−0.0020	−0.0049	−1.50	−0.0028	−0.0064
500	Approximation	21.01	0.4916	0.8985	22.20	0.8000	0.9135	21.12	0.5241	0.9000
	Exact method	22.52	0.4971	0.9051	23.71	0.8042	0.9192	22.63	0.5298	0.9065
	Error	−1.51	−0.0055	−0.0066	−1.50	−0.0042	−0.0057	−1.51	−0.0057	−0.0065

Note: $\hat{\sigma}_p^2$ is the pilot sample variance, and $\hat{\sigma}_g^2 = g\hat{\sigma}_p^2$ and $\hat{\sigma}_h^2 = h\hat{\sigma}_p^2$ are adjusted variances with g and h determined by Equations 11 and 15, respectively.

Expected Power Approach

Instead of assurance probability appraisal, an alternative criterion for sample size determination is agreement of expected actual power with planned power. Following the underlying assumption of sample variance defined in Equation 6, we continue the adjusted sample variance strategy in order to provide a unified framework for both the assurance probability and expected power considerations.

Accordingly, given mean difference d, significance level α, and planned power $1 - \beta$, it is desired to find the proper adjusted factor a so that the equality is fulfilled

$$\Pi(\upsilon, \hat{\sigma}_a^2) = E_K[\pi(\widehat{N}_a, \sigma^2)] = 1 - \beta. \quad (12)$$

where $E_K[\cdot]$ denotes the expectation taken with respect to the chi-square distribution of $K \sim \chi^2(\upsilon)$ given in Equation 10. It is obvious that the evaluation of expected power $\Pi(\upsilon, \hat{\sigma}_a^2)$ in Equation 12 requires the specification of presumably unknown parameter σ^2. A feasible approximation is presented in the following.

Note that $\pi(\widehat{N}_a, \sigma^2) \doteq P\{|Z + \bar{\delta}_a| > z_{\alpha/2}\}$ for moderate sample size $\widehat{N}_a$ where Z has a standard normal distribution and $\bar{\delta}_a = d/(c\sigma^2/\widehat{N}_a)^{1/2}$. Since $\widehat{N}_a$ is the minimum sample size so that $\pi(\widehat{N}_a, \hat{\sigma}_a^2) \geq 1 - \beta$, it follows that $\widehat{N}_a \doteq c\hat{\sigma}_a^2(z_{\alpha/2} + z_\beta)^2/d^2$ and $\bar{\delta}_a \doteq a^{1/2}(z_{\alpha/2} + z_\beta)(K/\upsilon)^{1/2}$. Hence the actual power conditional on K is

$$\pi(\widehat{N}_a, \sigma^2) \doteq P\{(Z + z_{\alpha/2})/(K/\upsilon)^{1/2} < -a^{1/2}(z_{\alpha/2} + z_\beta)\} + P\{(Z + z_{\alpha/2})/(K/\upsilon)^{1/2} < a^{1/2}(z_{\alpha/2} + z_\beta)\}. \quad (13)$$

It follows from the definition of a noncentral t distribution (Rencher, 2000, p. 102) that the suggest approximation to the expected power is of the form

$$\Pi(\upsilon, \hat{\sigma}_a^2) \doteq P\{T < -a^{1/2}(z_{\alpha/2} + z_\beta)\} + P\{T < a^{1/2}(z_{\alpha/2} + z_\beta)\}, \quad (14)$$

where $T \sim t(\upsilon, z_{\alpha/2})$. An important aspect of the simplified expression is that it does not depend on the unknown variance σ^2. Since the cumulative distribution function of a noncentral t distribution is readily embedded in modern statistical packages such as the SAS system, a standard iterative search can be conducted to find the correction factor $a = h$ so that

$$P\{T < -h^{1/2}(z_{\alpha/2} + z_\beta)\} + P\{T < h^{1/2}(z_{\alpha/2} + z_\beta)\} = 1 - \beta. \quad (15)$$

The search can be simplified with $P\{T < h^{1/2}(z_{\alpha/2} + z_\beta)\} = 1 - \beta$ because $P\{T < -h^{1/2}(z_{\alpha/2} + z_\beta)\}$ is generally negligible for large power $(1 - \beta)$. Although similar formula of expected power was described in

Equation 5 of Kieser and Wassmer (1996), they did not specifically address the issue of how to modify the sample variance in sample size calculations so that the expected power of a two-sample *t*-test will attain the planned power. In fact, their presentation focused on the expected power calculations of the two-sample *t* test with sample sizes that are required to satisfy the selected assurance level of power.

For the practical purpose of sample size determination, the modified power function given in Equation 7 with $a = h$ can be employed to calculate the sample size $\widehat{N}_h$ needed to test hypothesis H_0: $\mu_1 = \mu_2$ versus H_1: $\mu_1 \neq \mu_2$ in order to attain the expected power $(1 - \beta)$ with the chosen significance level α and mean difference d. In this case, since the sample size $\widehat{N}_h$ conditional on $\widehat{\sigma}_h^2 = h \cdot \widehat{\sigma}_p^2$ is approximately equivalent to $\widehat{N}_h \doteq c\,\widehat{\sigma}_h^2 (z_{\alpha/2} + z_\beta)^2/d^2$, the unconditional or expected sample size $N(\upsilon, \widehat{\sigma}_h^2)$ is $N(\upsilon, \widehat{\sigma}_h^2) = E\{\widehat{N}_h\} \doteq E\{c\,\widehat{\sigma}_h^2 (z_{\alpha/2} + z_\beta)^2/d^2\} = ch \cdot \sigma^2 (z_{\alpha/2} + z_\beta)^2/d^2$. As described earlier, the expected sample size corresponding to the usual procedure without sample variance correction is $N(\upsilon, \widehat{\sigma}_p^2) \doteq c \cdot \sigma^2 (z_{\alpha/2} + z_\beta)^2/d^2$. Hence, $\widehat{N}_p < \widehat{N}_h$ and $N(\upsilon, \widehat{\sigma}_p^2) < N(\upsilon, \widehat{\sigma}_h^2)$ if $h > 1$.

To enhance the applicability of the proposed approximate procedure for sample size calculations, the accuracy of correction factor h is demonstrated with the differences between the exact and approximate expected power. Due to the complexity in the computation of the exact expected power, a computing algorithm similar to the one in the assurance probability approach is developed. As in the empirical investigation of assurance probability, selected comparisons of exact and approximate results associated with and without sample variance modification were performed. Likewise, the adequacy of the approximate procedure is determined by the error = exact expected power − approximate expected power. For the chosen model configurations of $N_1 = N_2$, $\sigma^2 = 1$, planned power $(1 - \beta) = 0.90$, and $\alpha = 0.05$, the expected power $\Pi(\upsilon, \widehat{\sigma}_p^2)$ and $\Pi(\upsilon, \widehat{\sigma}_h^2)$ for sample variance degrees of freedom $\upsilon = 10$, 50, 100, and 500 are listed in Tables 1-3 for $d = 0.25$, 0.50 and 1, respectively. For ease of reference, we also summarize the corresponding exact and approximate results of expected sample sizes $N(\upsilon, \widehat{\sigma}_p^2)$ and $N(\upsilon, \widehat{\sigma}_h^2)$. The computed correction value h in $\widehat{\sigma}_h^2$ is $h = 1.3005$, 1.0531, 1.0262 and 1.0052 for $\upsilon =$ 10, 50, 100, and 500, respectively.

As can be seen from the results in Tables 1-3, it should not be surprising in view of the exact values that the expected sample size $N(\upsilon, \widehat{\sigma}_p^2) < N(\upsilon, \widehat{\sigma}_h^2)$ and expected power $\Pi(\upsilon, \widehat{\sigma}_p^2) < \Pi(\upsilon, \widehat{\sigma}_h^2)$ because the reported

correction $h > 1$. Hence an adjusted sample variance is required to perform sample size calculation for the designated expected power, especially the degrees of freedom υ of the pilot sample variance is small. For the power level $1 - \beta = 0.90$, the exact and approximated expected power $\Pi(\upsilon, \widehat{\sigma}_h^2)$ are almost identical with the largest absolute error 0.0065 for the case of $\upsilon = 500$ in Table 3. The errors associated with expected power $\Pi(\upsilon, \widehat{\sigma}_p^2)$ are slightly larger than those of $\Pi(\upsilon, \widehat{\sigma}_h^2)$, but they are less than 0.01 and are generally acceptable. Therefore the accurate approximate procedure given in Equation 14 with $a = h$ or Equation 15 can be employed to calculate the sample size needed for the chosen level of expected power. Without the modification of a pilot sample variance, the resulting sample size and expected power may be too small to be satisfactory. For completeness, the expected power $\Pi(\upsilon, \widehat{\sigma}_g^2)$ and assurance probability $\Gamma(\upsilon, \widehat{\sigma}_h^2)$ are also presented in Tables 1-3. The contrasting values of $\Gamma(\upsilon, \widehat{\sigma}_g^2)$, $\Pi(\upsilon, \widehat{\sigma}_g^2)$, $\Gamma(\upsilon, \widehat{\sigma}_h^2)$, $\Pi(\upsilon, \widehat{\sigma}_h^2)$ reveal that the difference between $\Gamma(\upsilon, \widehat{\sigma}_g^2)$ and $\Gamma(\upsilon, \widehat{\sigma}_h^2)$ is greater than that between $\Pi(\upsilon, \widehat{\sigma}_g^2)$ and $\Pi(\upsilon, \widehat{\sigma}_h^2)$. In other words, the change of adjusted factor from g to h ($g > h$) incurs more substantial decline in assurance probability than expected power. Hence, the assurance probability criterion is more sensitive to the adjustment of sample variance. Although the variance parameter of pilot sample variance is fixed as $\sigma^2 = 1$ throughout the numerical comparison, the results in Tables 1-3 are also applicable for any magnitude of variance as long as the ratio of mean difference and standard deviation $d/\sigma = 0.25, 0.50$ and 1.

Although the expected power approach has been considered in Kieser and Wassmer (1996) and Julious and Owen (2006), their results are insufficient and cumbersome. First, Kieser and Wassmer (1996) presented an approximation for the computation of $\Pi(\upsilon, \widehat{\sigma}_g^2)$ and did not address the notion about the search for a correction factor h so that the planned expected power is satisfied. On the other hand, Julious and Owen (2006) derived a formula for $\Pi(\upsilon, \widehat{\sigma}_p^2)$ and used the equation to compute the sample size needed to achieve the required expected power. Although the presented formula in Equation 9 of Julious and Owen (2006) appears to give similar results, their analytical derivation is complicated and is very different from the exposition presented in this article. In addition, no exact formulation and examination about the expected power of their approximation were provided. They suggested the derived formulas for accurate sample size calculations when the number of degrees of freedom of preliminary sample variance is less than 200. However, the values of expected power $\Pi(\upsilon, \widehat{\sigma}_p^2)$ and $\Pi(\upsilon, \widehat{\sigma}_h^2)$ are nearly equivalent for $\upsilon \geq 100$

in Tables 1-3. Hence, it suggested that the correction of sample size is fruitful when degrees of freedom $\upsilon < 100$. On the other hand, the use of adjustment of sample variance in sample size determination is always recommended in view of the fact that the assurance probability $\Gamma(\upsilon, \widehat{\sigma}_p^2)$ is far less than $\Gamma(\upsilon, \widehat{\sigma}_g^2)$ even for values of υ as large as 500 in Tables 1-3.

Numerical Example

For illustration, the aforementioned two methods for sample size calculations for the two-sided two-sample *t*-test defined in Equation 2 are exemplified with balanced group sizes. Assume an estimate of sample variance $\widehat{\sigma}_p^2 = 100$ with $\upsilon = 50$ degrees of freedom is available from a similar study. With the mean difference is $d = 5$ units, it follows from the suggested procedure that a sample size of $N_1 = N_2 = 103$ is required to guarantee a chance of $1 - \gamma = 0.80$ that the actual power exceeds 0.90 with $\alpha = 0.05$. On the other hand, the minimum sample size of $N_1 = N_2 = 90$ is necessary to ensure the expected power is at least $1 - \beta = 0.90$ with $\alpha = 0.05$. It is interesting to note that the approximate assurance probability with the smaller sample size of $N_1 = N_2 = 90$ rather than 103 turns out to be about 0.5751. In contrast, the approximate expected power attained with the sample size $N_1 = N_2 = 103$ instead of 90 is as high as 0.9322. Accordingly, the two considerations of assurance probability and expected power are fundamentally distinct and yield markedly different sample sizes. The differential phenomenon should continue to exist for other settings. Furthermore, the sample size of $N_1 = N_2 = N_t = 86$ given by the conventional sample size formula using sample variance as the population variance provides an assurance probability of only $P\{\hat{\sigma}^2 > \sigma^2\} = P\{\chi^2(50) > 50\} = 0.4734$ that the actual power will attain a planned power. Moreover, the corresponding approximate expected power is 0.8858. The empirical assurance results help to exemplify the entrenched difference in sample size determination between a stochastic sample variance and the stationary population variance. The SAS/IML (SAS Institute, 2011) programs employed to perform the sample size calculations are available upon request.

Concluding Remarks

Sample size determination is an important step in study planning. This paper describes the methodology for determining the necessary sample size of two-sample *t*-tests with adjusted sample variance under assurance probability and expected power considerations. The presented methods

permit imprecision in the variance estimate and exploit the stochastic distribution of sample variance in the calculations. In contrast, a direct use of a sample variance from pilot study or previous research can lead to serious underestimation of the necessary sample size and distortion of the desired power in detecting treatment differences. Although the proposed techniques are described in the context of a two-sample problem regarding the equality of two group means, the principles and procedures apply to more complex settings such as ANOVA and linear regression models. These techniques should prove efficacious in conducting a priori power analysis and sample size calculation under circumstances of imprecise information based on the findings of published research and preliminary study.

REFERENCES

Browne, R. H. (1995). On the use of a pilot sample for sample size determination. *Statistics in Medicine*, 14, 1933-1940.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Desu, M. M., & Raghavarao, D. (1990). *Sample Size Methodology*. Boston: Academic Press.

Guo, J. H., Chen, H. J., & Luh, W. M. (2011). Sample size planning with the cost constraint for testing superiority and equivalence of two independent groups. *British Journal of Mathematical and Statistical Psychology*, 64, 439-461.

Gillett, R. (2001). Sample size determination for a *t* test given a *t* value from a previous study: A FORTRAN 77 program. *Behavior Research Methods, Instruments & Computers*, 33, 544-548.

Jan, S. L., & Shieh, G. (2011). Optimal sample sizes for Welch's test under various allocation and cost considerations. *Behavior Research Methods*, 43, 1014-1022.

Julious, S. A., & Owen, R. J. (2006). Sample size calculations for clinical studies allowing for uncertainty about the variance. *Pharmaceutical Statistics*, 5, 29-37.

Kieser, M., & Wassmer, G. (1996). On the use of the upper confidence limit for the variance from a pilot sample for sample size determination. *Biometrical Journal*, 38, 941-949.

Kraemer, H. C., Mintz, J., Noda, A., Tinklenberg, J., & Yesavage, J. A. (2006). Caution regarding the use of pilot studies to guide power calculations for study proposals. *Archives of General Psychiatry*, 63, 484-489.

Kraemer, H. C., & Thiemann, S. (1987). *How Many Subjects? Statistical Power Analysis in Research*. Newbury Park, CA: Sage.

Kupper, L. L., & Hafner, K. B. (1989). How appropriate are popular sample size formulas? *The American Statistician*, 43, 101-105.

Lawson, C. A., & Fisher, A. V. (2011). It's in the sample: The effects of sample size and sample diversity on the breadth of inductive generalization. *Journal of Experimental Child Psychology*, 110, 499-519.

Murphy, K. R., & Myors, B. (2004). *Statistical Power Analysis: A Simple and General Model for Traditional and Modern Hypothesis Tests* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.

Odeh, R. E., & Fox, M. (1991). *Sample Size Choice: Charts for Experiments with Linear Models* (2nd ed.). New York: Marcel Dekker.

O'Hagan, A., Stevens, J. W., & Campbell, M. J. (2005). Assurance in clinical design. *Pharmaceutical Statistics*, 4, 187-201.

Rencher, A. C. (2000). *Linear models in statistics*. New York: Wiley.

SAS Institute (2011). *SAS/IML user's guide*, *Version 9.3*. Cary, NC: author.

Shiffler, R. E., & Harwood, G. B. (1985). An empirical assessment of realized α-risk when testing hypotheses. *Educational and Psychological Measurement*, 45, 811-823.

Thabane, L., Ma, J., Chu, R., Cheng, J., Ismalia, A., Rios, L. P., Robson, R., Thabane, M., Giangregorio, L., & Goldsmith, C. H. (2010). A tutorial on pilot studies: The what, why and how. *BMC Medical Research Methodology*, 10:1.

Vickers, A. J. (2003). Underpowering in randomized trials reporting a sample size calculation. *Journal of Clinical Epidemiology*, 56, 717-720.

(Manuscript received: 17 April 2012; accepted: 16 May 2012)