

---

# Statistics

---

## Study design in clinical research: sample size estimation and power analysis

Jerrold Lerman BAsC MD FRCPC

---

*The purpose of this review is to describe the statistical methods available to determine sample size and power analysis in clinical trials. The information was obtained from standard textbooks and personal experience. Equations are provided for the calculations and suggestions are made for the use of power tables. It is concluded that sample size calculations and power analysis can be performed with the information provided and that the validity of clinical investigation would be improved by greater use of such analyses.*

*Cet article de revue décrit les méthodes statistiques utilisées au cours des épreuves cliniques pour déterminer la taille d'un échantillon et l'analyse de sa puissance. L'information provient des manuels standards et de l'expérience de l'auteur. Des équations sont fournies avec des suggestions sur l'usage des tables de puissance. En conclusion, avec cette information, il est possible d'effectuer les calculs de la taille d'un échantillon et l'analyse de sa puissance; ces analyses amélioreraient la validité d'une étude clinique si on les utilisait plus.*

Although study design is an integral component of clinical research, it appears infrequently in the anaesthesia literature.<sup>1-3</sup> This is evidenced by the absence of both sample size calculations in prospective studies and

### Key words:

STATISTICS: sample size calculation, power analysis.

---

From the Department of Anaesthesia and the Research Institute, The Hospital for Sick Children and University of Toronto, Toronto, Ontario, Canada.

Address correspondence to: Dr. J. Lerman, Department of Anaesthesia, The Hospital for Sick Children, 555 University Avenue, Toronto M5G 1X8.

Phone: (416) 813-7445. Fax: (416) 813-7543.  
email lerman@sickkids.on.ca

Accepted for publication 5th October, 1995.

power analysis in studies with negative results.<sup>1-5</sup> Two possible explanations may account for this omission from clinical anaesthesia research. First, sample size estimation and power analysis are subjects that are rarely taught to trainees and almost totally omitted from the anaesthesia literature.<sup>6</sup> Second, the mathematical expressions that are used often appear complex and overwhelming. To address similar concerns in the behavioural sciences,<sup>7</sup> Cohen developed a user-friendly approach to sample size calculation and power analysis that appears to be as accurate as the more complex mathematical expressions.<sup>8</sup> In this synopsis, the concepts of and approaches to sample size estimation and power analysis in the design and reporting of clinical research studies in anaesthesia are reviewed.

Study design is a process in which methodology and statistical analysis are organized to ensure that the null hypothesis can be accepted or rejected and that the conclusion reached reflects the truth. The null hypothesis states at the outset that the treatments under investigation have equipoise (i.e., are equal). If the study is properly designed (i.e., appropriate sample size) and the treatments differ, then the investigators are likely to conclude from their results that the treatments do indeed differ and that the null hypothesis is false and should be rejected. On the other hand, if the study is properly designed and the treatments do not differ, then the investigators are likely to conclude that the treatments do not differ and that the null hypothesis is true and should be accepted. In the second case, power analysis will clarify whether the null hypothesis was accepted correctly, on the basis of equipoise of the treatments, or incorrectly because of inadequate power.

### 1 Sample size estimation

Before undertaking a study, the investigator should first determine the minimum number of subjects (i.e., sample size estimation) that must be enrolled in each group in order that the null hypothesis can be rejected if it is

false. Sample size estimations are warranted in all clinical studies for both ethical and scientific reasons. The ethical reasons pertain to the risks of enrolling either an inadequate number of subjects or more subjects than the minimum necessary to reject the null hypothesis. In both instances, the risks include randomizing the care of subjects and/ or exposing them to unnecessary risk/harm. Consequently, the Research Ethics Boards at The Hospital for Sick Children and others require that all investigators justify the proposed sample size. The scientific reasons pertain to the enrollment of more subjects than necessary because it extends the duration of and increases the costs of clinical research studies. Thus, sample size estimation is essential to achieve excellence in clinical research.

Study design depends on four interdependent variables: (i) alpha ( $\alpha$ ), (ii) beta ( $\beta$ ), (iii) effect size (ES) and (iv) sample size ( $n$ ).<sup>8</sup>

(i) *Alpha* (or the *P*-value) is the probability of finding a difference between the treatments when a difference does not exist (i.e., the difference is attributable to the random selection of the subjects). This is usually expressed as a 5% chance (i.e.,  $P < 0.05$ ) that the null hypothesis is falsely rejected. The alpha value is usually two-tailed ( $\alpha_2$ ), i.e., the treatment may be greater or less than the control value. A type I statistical error, an error that occurs when groups are repeatedly compared, refers to the false claim that  $P < 0.05$  (Table). Each time another variable is compared, there is a 5% chance that the treatments will differ because of random selection of the data. After comparing several of these variables, the probability that at least one variable will differ between the treatments is approximately the product of the number of variables compared and the *P*-value (the Bonferroni inequality): i.e., if ten comparisons were performed and the *P*-value is 0.05, then the probability is  $\approx 0.5$  or 50%.\*

Three techniques have been used to minimize or account for a type I error. First, the number of comparisons should be restricted to those that are essential to address the null hypothesis. Second, the value may be corrected by decreasing it in proportion to the number of variables that are compared.<sup>9</sup> This is known as the

\*The Bonferroni inequality may overestimate the probability of making a Type I statistical error, particularly when a large number of comparisons are performed. The actual probability of obtaining at least one difference that is significant at a  $P < 0.05$  is:  $\alpha_T = 1 - (1 - \alpha)^k$ , where  $\alpha_T$  is the total probability of making a Type I error,  $\alpha$  is the significance level and  $k$  is the number of comparisons. In the case of  $\alpha = 0.05$  and  $k = 10$ , the actual  $\alpha_T$  is 0.4.<sup>9</sup>

TABLE The relationship between the null hypothesis (the premise of the study that the treatments have equipoise) and the true effects of the treatments. Two statistical errors are identified: a type I (or  $\alpha$  error) and a type II (or  $\beta$  error).

Observation	Reality	
	Treatment has no effect	Treatment has an effect
Treatment is effective Ho is rejected	Type I or $\alpha$ error (falsely reject $H_0$ ) $\alpha$	Correct conclusion (reject $H_0$ ) $1 - \beta$
Treatment has no effect Ho is accepted	Correct conclusion (accept $H_0$ ) $1 - \alpha$	Type II or $\beta$ error (falsely accept $H_0$ ) $\beta$

Bonferroni correction. However, in the presence of a large numbers of comparisons, this correction may decrease the  $\alpha$  value to the extent that it causes a type II statistical error. Third, a within-group measure of the variance of the data may be used, which reduces the probability of a type II error that is associated with the Bonferroni correction.<sup>9</sup> Thus, type I statistical errors are easy to identify and can be minimized using several techniques.

(ii) *Beta* is the probability of failing to find a difference between the treatments when a difference exists. The maximum value of  $\beta$  that is accepted in the biostatistical literature is 0.20 or a 20% chance that the null hypothesis is falsely accepted. This value is based on convention rather than any mathematical derivation. However, it is interesting that we accept a four-fold greater risk of falsely accepting the null hypothesis than we do of falsely rejecting the null hypothesis  $\alpha$ . The  $\beta$  value is usually one-tailed,  $\beta_1$ . A type II statistical error, an error that occurs when the null hypothesis is falsely accepted, occurs when the  $\beta$  error exceeds 0.2. This is usually expressed in terms of the power of the study; that is, the probability that the null hypothesis can be rejected if the treatments differ. Power is defined as  $1 - \beta$ . For a  $\beta$  of 0.2, the power is 0.8, which is the minimum power required to accept the null hypothesis. Type II statistical errors occur when the power of the study is  $< 0.8$ . Calculation of the power of a study uses the actual results of the study as described below.

(iii) *The effect size (ES)* is a measure of the smallest clinically acceptable difference between treatments normalized by the standard deviation of the data (equation 1). In the design of clinical research studies, investigators define  $\alpha$  and  $\beta$ , estimate the ES using one of the three techniques (a pilot study, published data or an educated guess based on clinical experience) and then cal-

culate the sample size. The first two techniques to estimate the ES are preferable to the third since they are based on real measurements.

Calculation of the ES is specific for the type of data under consideration. In the case of parametric data (defined as data that are continuous and whose distribution can be described by a measure of central tendency and a measure of the scatter) for two unrelated groups, ES is represented by  $d$ :<sup>8</sup>

$$d = \frac{\delta}{\sigma} = \frac{|\bar{x}_1 - \bar{x}_2|}{\sigma} \quad (\text{Eq. 1})$$

where  $\bar{x}_1$  and  $\bar{x}_2$  are the mean values of the two treatments and  $\sigma$  is the standard deviation of the treatments. These data are based on either pilot data or the literature. In this equation, the standard deviations of the two treatments are similar. (When the standard deviations differ, a mean  $\sigma$  for the two treatments is used, see Case 2 below). Because the ES does not carry a sign, absolute value brackets are included in the numerator. It is important that the values used in the calculation of the ES be clinically relevant if the interpretation of the results is to have meaning in clinical practice.

For nominal or proportional data (defined as data based on the presence or absence of a quality or attribute such as the presence or absence of vomiting), the ES is the absolute value of the difference between the incidence of these qualities, (expressed as two proportions, P1 and P2) after transformation of the proportions (see below). This difference is represented by " $h$ ":<sup>8</sup>

$$h = |\phi_1 - \phi_2| \quad (\text{Eq. 2})$$

where  $\phi_1$  is the arcsine transformation of P1 and  $\phi_2$  is the arcsine transformation of P2. Proportions which consist of only two categories (presence or absence of a quality) form a binomial distribution (rather than a normal distribution). The square root of each proportion is transformed to its arcsine value (also known as the angular transformation or inverse sine ( $\sin^{-1}$ )) by determining the angle whose sine is  $\sqrt{P}$ . The transformed values will have a distribution that approximates a normal distribution and these values can then undergo simple mathematical operations. Arcsine transformation tables for proportions are available in most statistical texts.<sup>8,10</sup>

Thus, the ES can be estimated for most of the common types of data used in clinical studies in anaesthesia, parametric and nominal data. In the case of ordinal data, which is defined as data that are discontinuous, for which sequential values have no mathematical relationship; i.e., cannot be described by a measure of central tendency and scatter, these data must be transformed to a normal distribution to apply a sample size calculation

based on parametric data or represented as categorical data. Sample size calculations can also be performed for more complex study designs including ANOVA, as described elsewhere.<sup>10</sup>

Estimation of the sample size can be based on any one of several variables being measured in a study. The most appropriate variable is the one that most closely addresses the null hypothesis of the study. However, some investigators may be tempted to use another variable to estimate the sample size, one that yields a larger ES and a correspondingly smaller sample size than would the appropriate variable! Although the smaller sample size may facilitate earlier completion of the study, it may also prevent achieving a statistically significant difference in the appropriate variable, a type II statistical error. In such a case, a power analysis should be performed (using the data generated in the study) to verify that the sample size was sufficient to reject a false null hypothesis.

*Example 1:* A study is planned to compare the anxiolytic effects of a new premedication with a placebo with the outcome variable being stress at induction of anaesthesia. The investigators planned to quantitate stress at induction using the plasma adrenaline concentration, heart rate or systolic blood pressure. Any of these three variables may be used to calculate the sample size, although the investigators based their null hypothesis on the adrenaline concentration. To obtain estimates of the mean and standard deviation of the adrenaline concentration for a sample size calculation, a literature search is performed. The published mean concentration of adrenaline at induction in the premedication group was  $\bar{x}_1 = 50$  pg, the concentration in the placebo group was  $\bar{x}_2 = 75$  pg, and the standard deviation ( $\sigma$ ) of the concentrations for both groups was 50 pg. Using equation 1,  $d$  is  $(75-50)/50$  or 0.5. This value is then used to calculate the sample size as follows.

#### (iv) Sample size calculation

##### (a) PARAMETRIC DATA – two unpaired samples

Several approaches may be used to estimate the sample size for two groups of parametric data. These are based on an iterative approach<sup>10</sup> which is detailed in Appendix A. The approach that I prefer however, is based on the ES after Cohen<sup>8</sup> because it is accurate and simple to perform. All approaches yield similar estimates of the sample size for a given set of conditions.

To estimate the sample size using the ES approach, either a simple mathematical expression is solved or tables are consulted.<sup>8-10</sup> Because the tables do not address all possible combinations of the variables, it is

preferable to solve the mathematical expression. Three common case scenarios are discussed.

CASE 1:  $n_1 = n_2, \sigma_1 = \sigma_2$

In this case, both the sample size ( $n$ ) and the standard deviation ( $\sigma$ ) of the data in each of the two groups are equal. Here, sample size estimation is related to the ES by the expression:<sup>8</sup>

$$n = \frac{n_{0.1}}{100 ES^2} + 1 \tag{Eq. 3}$$

where  $n_{0.1}$  is the sample size for an ES of 0.1,  $\alpha_2$  of 0.05 and  $\beta_1$  is 0.2. Equation 3 accurately predicts the sample size for the conditions;  $\alpha_2 = 0.05$ ,  $\beta_1 \leq 0.3$  and ES between 0.2 and 1.0. For ES values outside this range, equation 3 must be modified.<sup>8</sup> Equation 3 may be simplified for the common set of conditions,  $\alpha_2 = 0.05$  and  $\beta_1 = 0.2$  by substituting  $n_{0.1} = 1570$ :<sup>8</sup>

$$n = \frac{15.7}{ES^2} + 1 \tag{Eq. 4}$$

The sample size is calculated by substituting the value for ES into equation 4. For  $\beta$  values between 0.2 and 0.05, the relationships between  $n$  and ES are illustrated in Figure 1. For each  $\beta$  value, there is a corresponding value for  $n_{0.1}$  based on Table 2.4.1 in reference 8: for  $\beta$  of 0.1, the numerator of equation 4 is 21.0 and for  $\beta$  of 0.05, it is 26.0. By substituting these values into equation 4, the sample size can be calculated for a range of values of ES. In addition, sample size may be calculated using either of two small but user-friendly statistics packages (Primer of Biostatistics (McGraw-Hill, New York, 1992: version 3.0) and InStat™ (GraphPad. Software, San Diego, CA, 1993: version 2.03)).

*Example 2:* Continuing the example of the new pre-medicant, the sample size required for an ES of 0.5,  $\alpha_2$  of 0.05 and a  $\beta_1$  of 0.2 is 64 subjects in each group (equation 4). However, suppose the investigators found a more recent study in which the adrenaline concentration after the placebo was only 25 pg, one-half that previously published. Using  $\bar{x}_2 = 25$  pg, then ES increases 100% (to a value of 1.0) and the sample size decreases from 64 to 17 subjects in each group (Figure 1).

When several values for the ES exist (based on the results from different studies), the ES that best approximates the clinical study assumptions should be used to estimate the sample size. If the best approximate is not obvious, then the authors are best advised to follow a conservative approach and choose the smaller ES (i.e., the one that yields a larger sample size). In the final analysis, if the investigators overestimate the ES and use

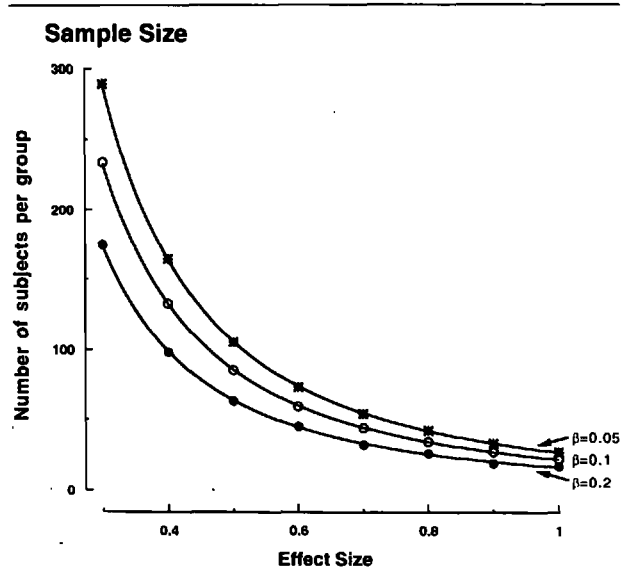


FIGURE 1 The relationship between sample size and ES for the unpaired t test for  $\beta$  values between 0.05 (power is 0.95) and 0.20 (power is 0.80) based on equation 3.<sup>8</sup> As the ES decreases, the sample size, as well as the difference in the sample size among the three  $\beta$  values, increase (Figure 1). By doubling the ES, the sample size decreases by four-fold as predicted by equation 3.

a small sample size, they risk falsely accepting the null hypothesis (or a Type II error).

CASE 2:  $n_1 = n_2, \sigma_1 \neq \sigma_2$

When the standard deviations of the two groups differ, a hybrid of the two standard deviations known as the root mean square ( $\sigma'$ ), should be used to estimate the ES as follows:<sup>8</sup>

$$\sigma' = \sqrt{\frac{\sigma_1^2 + \sigma_2^2}{2}} \tag{Eq. 5}$$

where  $\sigma_1$  and  $\sigma_2$  are the standard deviations of the two treatments and  $\sigma'$  is the root mean square of the standard deviations. The value,  $\sigma'$ , is substituted for  $\sigma$  in equation 1 to calculate the ES.

CASE 3: Paired data

For paired data, the ES is first calculated assuming the data are unpaired. Second, the calculated ES is increased by a factor of  $\sqrt{2}$  to account for the homogeneity of the paired measurements:<sup>8</sup>

$$ES_{\text{paired}} = ES_{\text{unpaired}} \times \sqrt{2} \tag{Eq. 6}$$

The increase in ES for paired data decreases the sample size as predicted by equations 3 and 4. Thus, repeated measurements in subjects decrease the sample size (and increase the power) compared with measurements in two different groups of subjects.

## (b) PROPORTIONAL OR NOMINAL DATA –

For proportional data, the estimate of the sample size is similar to that for parametric data with two exceptions. First, proportional data must be transformed (using the arcsine transformation) before simple mathematical operations can be performed. Second, the constant in equations 3 and 4, is omitted. Thus, the expression used to estimate the sample size for proportional data is:<sup>8</sup>

$$n = \frac{15.7}{h^2} \quad (\text{Eq. 7})$$

for an  $\alpha_2 = 0.05$  and a  $\beta_1 = 0.2$ . The sample size may also be calculated using the mathematical expression in Appendix B. In addition, both the Primer of Biostatistics and Instat™ compute sample sizes for proportional data. However, all three of these techniques yield larger sample sizes than the equations of Cohen<sup>8</sup> (equation 7) as discussed below.

**Example 3:** Suppose the incidence of two events under consideration are: P1 (control) = 0.45 and P2 = 0.25. The arcsine transformation of P1 is 1.471 and of P2 is 1.0478. Using equation 2, “h” is 0.424. When this value is substituted into equation 7 together with an  $\alpha_2$  of 0.05 and a  $\beta_1$  of 0.2, a minimum of 88 subjects are required per group. Some statistics packages (Primer of Biostatistics and Instat™) may overestimate the sample size (by approximately 10%) compared with the size based on equation 7. However, the sample size estimates for proportions after Cohen<sup>8</sup> appear to be accurate, reliable and slightly smaller than values based on these other approaches.

Small differences in the sample size between treatments do not substantially decrease the power of the study, but large differences may.<sup>8</sup> When the size of one group is fixed (for example, only a limited number of subjects can receive an expensive new treatment) at a value that is far less than the sample size calculated for same size treatments, the size of the unfixed group must be increased disproportionately such that the total sample size exceeds the total for equal sized groups.<sup>8,10</sup> Calculation of the sample size for unequal size groups uses equation 8 as follows:

$$n_u = \frac{n_f \times n}{2n_f - n} \quad (\text{Eq. 8})$$

where  $n_u$  is the unfixed sample size,  $n_f$  is the fixed sample size and  $n$  is the sample size estimate for same size groups (as per equations 4 or 7).  $n_f$  must be  $>0.5n$ , otherwise the denominator in equation 8 becomes negative or zero, thereby making equation 8 insoluble. If  $n_f$  cannot be increased to solve equation 8, then  $\alpha$ ,  $\beta$  or the ES

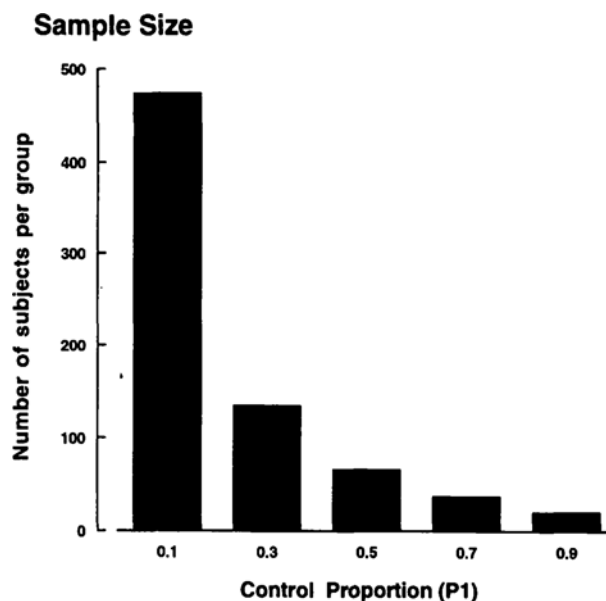


FIGURE 2 The sample size (ordinate axis) in a study in which the probability of an event in the treatment group (P2) is 50% that in the control group (P1). The control proportion (P1) is the probability that the event occurs in the control group, shown on the abscissa. The sample size estimates were based on  $\alpha_2 = 0.05$ , and  $\beta = 0.20$  and equation 7.

should be adjusted to decrease  $n$ . In example 2 with equal sample sizes, 64 subjects were required for each group or 128 subjects in total. However, if one sample size had been fixed ( $n_f$ ) to 35 subjects, then the unfixed sample size ( $n_u$ ) would require 375 subjects for the same  $\alpha$ ,  $\beta$  and ES. In this case, the number of subjects increases to 410 or 3 fold the number with equal sample sizes.

When summarizing the sample size estimation, *all* of the assumptions used should be reported including the  $\alpha$  and  $\beta$  values, and the means and standard deviations or the estimated probabilities of the outcome events (i.e., proportions) (with the actual data or the sources of the data). When incomplete information is provided, it may be difficult to determine the assumptions used to calculate the sample size. For example, when the sample size is reported to be based on a 50% decrease in the incidence of an event in the treatment group compared with the control group, there is a whole range of sample sizes possible as shown in Figure 2. In this case, the actual incidence used in the calculation of the sample size should be reported. The summary of this information usually involves only a statement or two at the conclusion of the methods section to outline the details of how the sample size was estimated.

## 2 Power analysis

If, after completion of a study, the null hypothesis is accepted, two possible scenarios may exist: (1) that the treatments have equipose (or are similar) or (2) that the power of the study was inadequate to prove the treatments differed (Type II error). Before concluding that the treatments have equipose, it is important to determine whether a statistically significant difference between the treatments could have been detected if the treatments truly differed.

The power of a study is the probability of correctly accepting the null hypothesis ( $1-\beta$ ). Biological scientists have accepted a maximum value for  $\beta$  of 0.2 or a power of 80%. Thus, when treatments are found to have equipose, the null hypothesis may be falsely accepted if the power of the study is <80% or correctly accepted if the power is  $\geq 80\%$ . The power of a study is determined after completion of the study, using the actual sample size and ES and the  $\alpha$  value. Power may be determined for the same types of data as for the sample size calculation. Whenever the null hypothesis is accepted, a power analysis is warranted to validate the conclusions of the study. Power tables using the sample size and ES (calculated from the results of the study) as well as the  $\alpha$  value are available in the standard textbooks.<sup>8</sup>

*Example 4:* Upon completion of a study in which 20 subjects were enrolled in each of two groups, the investigators concluded that two treatments had equipose. The investigators had not performed a sample size calculation before undertaking the study. In this example, the ES was 0.5,  $\alpha_2$  was 0.05 and the sample size was 20. Using power tables,<sup>8</sup> the power of the study was only 33%, less than the minimum power of 80% required to correctly accept the null hypothesis. Based on this low power, the investigators could not conclude that the two treatments were similar, but rather that they were unable to detect a statistically significant difference between the treatments. On the basis of the results of this study, the investigators would have required 64 subjects in each treatment group to accept the null hypothesis with a power of 80%.

Several strategies may be considered in order to maximize the power of a study.<sup>8,10</sup> First, the sample size in each treatment group should be similar. When the sample sizes are equal, the power of the study is maximal; as the difference between sample sizes increases, the power of the study decreases. Other strategies that increase the power include repeat measurements in the same subjects rather than singular measurements in several cohorts of subjects and measurement techniques that minimize the variability in the outcome variable

(i.e., minimizes the standard deviation of the measured variable). These strategies must be considered during the design phase of the study to maximize the power of the study.

When the null hypothesis is accepted at the conclusion of a study, some investigators dismiss the study as a failure because a statistically significant difference had not been achieved. However, such studies should not be dismissed frivolously. If the power of the study was sufficient to detect a clinically relevant difference between the treatments, then the treatments were similar and the results relevant to clinical practice. This study would merit publication. If the study had insufficient power to detect a clinically relevant difference, then the flawed study design also merits publication to serve as a guide to other investigators for future study designs. As the application of power analysis becomes widespread, the clinical relevance of studies in which the null hypothesis is accepted, will be enhanced.

## Appendix A

### *An alternative approach to estimating the sample size for parametric data<sup>10</sup>*

Equation 9 may be used to estimate the sample size. This equation requires an iterative process as shown in example 5.

$$n \geq \frac{2s_p^2}{\delta^2} (t_{\alpha(2),v} + t_{\beta(1),v})^2 \quad (\text{Eq. 9})$$

where  $n$  is the sample size,  $s_p$  is the pooled standard deviation,\*  $\delta$  is the difference between the two means and  $t_{\alpha(2),v}$  and  $t_{\beta(1),v}$  are the  $t$  values<sup>10</sup> that correspond to  $\alpha_2$ , one-tailed  $\beta$  (or  $\beta_1$ ) and  $v$  degrees of freedom (in this case, the degrees of freedom for two unrelated groups =  $(2 \times n) - 2$ ).

Equation 9 may be simplified to equation 3 as follows: first, the ratio,  $\delta/s_p$ , is replaced by ES in equation 4 (assuming  $\sigma_1 = \sigma_2$ ). Second, for large  $v$  values,  $t_{\alpha(2),v}$  approaches  $t_{\alpha(2),\infty}$ , which is also the standard normal deviate for  $\alpha_2$ ,  $Z_{\alpha(2)}$ . Using the same approach for  $t_{\beta(1),v}$ , equation 9 becomes:

$$n = \frac{2(Z_{\alpha(2)} + Z_{\beta(1)})^2}{\text{ES}^2} \quad (\text{Eq. 10})$$

\*Where  $s_p^2 = \frac{SS_1 + SS_2}{v_1 + v_2}$ . Here, SS is the sum of squares of

the data in treatments 1 and 2 and  $v$  are the degrees of freedom ( $n-1$ ) in treatments 1 and 2.<sup>10</sup>

Third, substituting 1.96 for  $Z_{0.05(2)}$  and 0.8416 for  $Z_{0.2(1)}$  into the numerator of equation 10 yields the numerator of equation 4, 15.7.<sup>10,11</sup> Cohen also noted that for these  $\alpha_2$  and  $\beta_1$  values, the sample size must be adjusted upwards slightly to yield accurate estimates of the sample size. He increased the sample size by the addition of a constant, in this case, 1, to give equation 4.

**Example 5:** To estimate the sample size ( $n$ ) using equation 9, a series of iterations (or repeated calculations) is required because the sample size estimate is also one of the assumptions that determines the  $t$  values on the right side of the equation. We begin by guessing a sample size (any guess is acceptable, although the most efficient solution is achieved by overestimating the first sample size) and determine the two  $t$  values ( $t_{\alpha(2)}$  and  $t_{\beta(1)}$ ) that correspond to this first sample size using  $t$  tables.<sup>10</sup> These values, together with the  $s_p$  and  $\delta$  are substituted into the right side of equation 9 to yield the first calculated sample size. The second estimate of the sample size is a value that lies between the first estimated and calculated sample sizes. Using this new estimate of  $n$ , two new  $t$  values are generated and substituted into equation 9. The second calculated sample size is compared with the second estimated sample size and the process continues until the estimated and calculated sample sizes converge.

Using the data from the premedication example,  $\delta$  is 25,  $\sigma$  is 50,  $\alpha_2 = 0.05$  and  $\beta_1 = 0.2$ . Our first estimate of the sample size is 20 subjects per group. This corresponds to a  $\nu = 38$ ,  $t_{0.05(2),38} = 2.024$  and  $t_{0.2(1),38} = 0.851$ .<sup>11</sup> Substituting these values into equation 9 yields

$$n = \frac{2x(50)^2}{(25)^2} (2.024 + 0.851)^2 = 63.6.$$

Our second estimate of the sample size is 60 subjects which corresponds to a  $\nu = 118$ . In this case,  $t_{0.05(2),118} = 1.975$  and  $t_{0.2(1),118} = 0.845$ . Substituting these values yields an  $n = 63.8$ . Thus, our sample size should be at least 64 subjects. This value is consistent with the results of the example for CASE 1.

## Appendix B

*An alternative approach to estimating the sample size for proportional data*<sup>10</sup>

Equation 11 may be used to estimate the sample size for proportional data as follows:

$$n = \frac{A \left[ 1 + \sqrt{1 + \frac{4\delta_h}{A}} \right]^2}{4\delta_h^2} \quad (\text{Eq. 11})$$

where:

$$A = \left[ Z_{\alpha(2)} \sqrt{\frac{(p_1 + p_2)(q_1 + q_2)}{2}} + Z_{\beta(1)} \sqrt{p_1 q_1 + p_2 q_2} \right]^2$$

where  $Z_\alpha$  and  $Z_\beta$  are the normal standard deviates or  $t_{\alpha(2),\infty}$  and  $t_{\beta(1),\infty}$  respectively,<sup>10</sup>  $p_1$  and  $p_2$  are the proportions of the events and  $q_1$  and  $q_2$  are the complementary values, i.e.,  $q_i = 1 - p_i$ :

$$\text{and } \delta_h = |p_1 - p_2|$$

**Example 6:** Solving the example for proportions,  $p_1 = 0.45$ ,  $p_2 = 0.25$ . The corresponding  $q_1$  and  $q_2$  values are 0.55 and 0.75.  $\delta_h = 0.45 - 0.25 = 0.20$ .  $Z_{0.05(2)} = 1.96$  and  $Z_{0.2(1)} = 0.8416$ .<sup>11</sup>

Hence:

$$A = \left[ 1.96 \sqrt{\frac{(0.45 + 0.25)(0.55 + 0.75)}{2}} + 0.8416 \sqrt{(0.45)(0.55) + (0.25)(0.75)} \right]^2$$

$$= 3.52$$

Now:

$$n = \frac{3.52 \left[ 1 + \sqrt{1 + \frac{4(0.2)}{3.52}} \right]^2}{4(0.2)^2} = 97.7$$

This result is similar to the estimate in example 3.

## References

- 1 Moodie PF, Craig DB. Experimental design and statistical analysis. *Can Anaesth Soc J* 1986; 33: 63-5.
- 2 Villeneuve E, Mathieu A, Goldsmith CH. Power and sample size calculations in clinical trials from anesthesia journals. *Anesth Analg* 1992; 74: S337.
- 3 Mathieu A, Villeneuve E, Goldsmith CH. Critical appraisal of methodological reporting in the anaesthesia literature. *Anesth Analg* 1992; 74: S195.
- 4 Freiman JA, Chalmers TC, Smith H Jr., Keubler RR. The importance of beta, the type II error and sample size in the design and interpretation of the randomized control trial.

- Survey of 71 "negative" trials. *N Engl J Med* 1978; 299: 690-4.
- 5 *Gardner MJ, Bond J.* An exploratory study of statistical assessment of papers published in the British Medical Journal. *JAMA* 1990; 263: 1355-7
  - 6 *Fisher DM.* Statistics in Anesthesia. *In: Anesthesia.* Miller RD (ed.). 4th ed. New York: Churchill Livingstone, 1994: 782-5
  - 7 *Cohen J.* The statistical power of abnormal social psychological research: a review. *J Abnorm Soc Psychol* 1962; 65: 145-53.
  - 8 *Cohen J.* Statistical Power Analysis for the Behavioural Sciences. 2nd ed. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1988: 1-74, 179-213
  - 9 *Glantz SA.* Primer of Biostatistics. 3rd ed. New York: McGraw-Hill, 1992: 91-4, 133-8, 387
  - 10 *Zar JH.* Biostatistical Analysis. 2nd ed. Englewood Cliffs: Prentice-Hall, Inc., 1984: 134-7, 171-6, 397-400, 484-5, 586-8
  - 11 *Snedecor GW, Cochran WG.* Statistical Methods. 7th ed. Ames, Iowa: The Iowa State University Press, 1980: 469.