Introduction to Sample Size Determination and Power Analysis for Clinical Trials

John M. Lachin

From the Biostatistics Center, George Washington University, Bethesda, Maryland

ABSTRACT: The importance of sample size evaluation in clinical trials is reviewed and a general method is presented from which specific equations are derived for sample size determination or the analysis of power for a wide variety of statistical procedures. The method is discussed and illustrated in relation to the t test, tests for proportions, tests of survival time, and tests for correlations as they commonly occur in clinical trials. Most of the specific equations reduce to a simple general form for which tables are presented.

KEY WORDS: *sample size determination, statistical power, survival analysis, tests for correlations, tests for proportions, t tests*

INTRODUCTION

It is widely recognized among statisticians that the evaluation of sample size and power is a crucial element in the planning of any research venture. Often it becomes necessary for the statistician to introduce these basic concepts to collaborators who may be aware of the problem but who do not understand the basic statistical logic. In this paper a simple expression is presented that can be used for sample size evaluation for a wide variety of statistical procedures and that has often been employed in collaboration with medical researchers in the conduct of clinical trials [1]. The method presented is quite general and, it is hoped, may be applied by clinician and statistician alike in a variety of research settings.

When conducting a statistical test, two types of error must be considered: Type I (false positive) and Type II (false negative), with probabilities α and β, respectively. In the following we will consider the general family of statistics, say X, that are normally distributed under a null hypothesis (H_0) as $N(\mu_0, \Sigma_0^2)$ and under an alternative hypothesis (H_1) as $N(\mu_1, \Sigma_1^2)$; where $\mu_1 > \mu_0$ or $\mu_1 < \mu_0$ and where Σ_0^2 and Σ_1^2 are some function of the variance σ^2
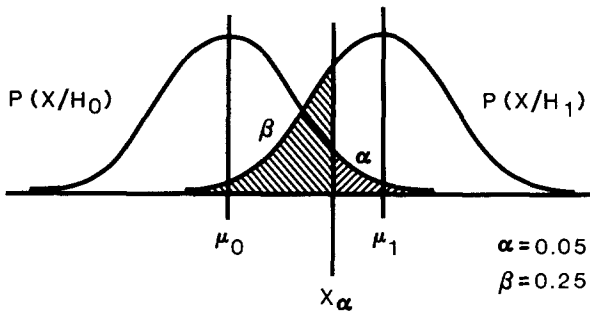
Figure 1 The distribution of a statistic X with variance Σ^2 under the null hypothesis H_0: $\mu = \mu_0$, i.e., the curve $P(X|H_0)$, and that under the alternative hypothesis H_1: $\mu = \mu_1$ or the curve $P(X|H_1)$, and the probabilities of Type I error (α) and Type II error (β), where $X_\alpha = \mu_0 + Z_\alpha \Sigma$.

of the individual observations and the sample size N.[1] Given these distributions one can then determine α and β as shown in Figure 1.

In a clinical trial the parameter μ is the treatment-control difference in the outcome measure of interest, e.g., the mean difference on some measurable pharmacologic effect such as serum cholesterol, or the difference in the proportion displaying an event such as healing. In such cases μ_0 is usually zero and μ_1 is specified as the minimal clinically relevant therapeutic difference. (For a more basic introduction to these concepts, see [2–4]).

When the statistical test is conducted, the probability of Type I error, α, is specified by the investigator. However, the probability that a significant result will be obtained if a real difference (μ_1) exists (i.e., the power of the test, $1 - \beta$) depends largely on the total sample size N. As one increases N the spread of the distributions in Figure 1 decreases, i.e., the curves tighten; thus β decreases (power increases). Thus if the statistical test fails to reach significance, the power of the test becomes a critical factor in reaching an inference. It is not widely appreciated that the failure to achieve statistical significance may often be related more to the low power of the trial than to an actual lack of difference between the competing therapies. Clinical trials with inadequate sample size are thus doomed to failure before they begin and serve only to confuse the issue of determining the most effective therapy for a given condition. Thus one should take steps to ensure that the power of the clinical trial is sufficient to justify the effort involved.

Conversely, if the power of the trial in detecting a specified clinically relevant difference (μ_1) is sufficiently high, say 0.95, failure to achieve significance may properly be interpreted as probably indicating negligible relevant difference between the competing therapies. Thus the proper interpretation of a "negative" result is based largely upon a consideration of the power of the experiment.

---

[1]The notation $N(\mu, \Sigma^2)$ denotes the normal distribution with mean μ and variance Σ^2. If $X \sim N(\mu, \Sigma^2)$ then $Z = (X - \mu)\Sigma^{-1}$ is distributed as $N(0, 1)$, the latter being widely tabulated as the standard normal distribution.

These points have been illustrated by Freiman et al. [5] who showed that of 71 recent clinical trials that reached a negative result, 67 had power less than 0.90 in detecting a moderate (25%) therapeutic improvement. Their conclusion is that many of the therapies studied were not given a fair test simply due to inadequate sample sizes and, thus, inadequate power.

SAMPLE SIZE AND POWER

The problem in planning a clinical trial is to determine the sample size N required such that in testing H_0 with stated probability of Type I error α, the probability of Type II error is a desired small level β. The parameters of the problem are α, β, μ_0, μ_1, Σ_0^2, and Σ_1^2.

Since the variances Σ^2 are functions of N, the sample size required is that which simultaneously satisfies the equalities $\Pr(Z > Z_\alpha) = \alpha$ if H_0 is true and $\Pr(Z > Z_\alpha) = 1 - \beta$ if H_1 is true; where Z_α is the standard normal deviate at the α significance level (e.g., $Z_\alpha = 1.645$ for $\alpha = 0.05$, one-sided) and where $Z = (X - \mu_0)\Sigma^{-1}$ is the simple statistic one would use in testing H_0; where $Z \sim N(0, 1)$ if H_0 is true (see footnote 1).

It can easily be shown, however, that the sample size that satisfies these equalities also satisfies the equality

$$|\mu_1 - \mu_0| = Z_\alpha \Sigma_0 + Z_\beta \Sigma_1 \tag{1}$$

The term $Z_{\alpha/2}$ is employed for a two-tailed test. Derivations for particular cases are given in Snedecor and Cochran [6, p. 111] and Fleiss [7, p. 29], among others.

This basic relationship can readily be grasped from Figure 1. To relate equation (1) to Figure 1, note that the critical value of X at the α level of significance is $X_\alpha = \mu_0 + Z_\alpha \Sigma_0$. One can thus readily derive equation (1) from Figure 1 where the "distance" $|\mu_1 - \mu_0|$ is the sum of two parts, $|X_\alpha - \mu_0| = Z_\alpha \Sigma_0$ and $|\mu_1 - X_\alpha| = Z_\beta \Sigma_1$, where Z_β results from the specification X_α, μ_1, and Σ_1.

This equation can then be used to evaluate sample size or power for the most commonly used statistical tests once Σ_0, Σ_1, and α have been specified. The three basic questions one can ask are

1. What *sample size* is required to ensure power $1 - \beta$ of detecting a relevant difference μ_1?
2. What is the *power* (Z_β) of the experiment in detecting a relevant difference μ_1 when a specific sample size N is employed?
3. What *difference* μ_1 can be detected with power $1 - \beta$ if the experiment is conducted with a specified sample size N?

Usually, question 1 is employed in planning an experiment and question 2 is employed in evaluating the results of an experiment. Question 3 can be employed in either case.

For the determination of sample size (question 1) one simply solves equation (1) for N once the expression for the variances Σ^2 have been obtained. In many cases, Σ^2 will be a function of the form $\Sigma^2 = \sigma^2/N$, where

σ^2 is the variance of the individual measurements and N is the *total* sample size. In this case

$$|\mu_1 - \mu_0| = (Z_\alpha \sigma_0 / \sqrt{N}) + (Z_\beta \sigma_1 / \sqrt{N}) \tag{2}$$

Solving for total sample size N one obtains simply

$$N = \left[\frac{Z_\alpha \sigma_0 + Z_\beta \sigma_1}{\mu_1 - \mu_0} \right]^2 \tag{3}$$

Likewise, to determine power (question 2) one solves for Z_β to obtain from (2)

$$Z_\beta = \frac{\sqrt{N}\,|\mu_1 - \mu_0| - Z_\alpha \sigma_0}{\sigma_1} \tag{4} \tag{4}$$

Power $(1 - \beta)$ can then be determined from the value Z_β by referring Z_β to tables of the normal distribution where values $Z_\beta < 0.0$ indicate power $<$ 0.50.

Similarly, the minimal detectable difference with power $1 - \beta$ given a sample size N (question 3) is obtained by solving equation (2) for μ_1. In some cases an explicit solution is obtained, whereas in others, e.g., for proportions, an iterative procedure is required since both μ_1 and σ will be functions of the same parameters. Thus we shall only consider questions 1 and 2 in the following.

Since equations (3) and (4) follow obviously from equation (2), it is often convenient to express the basic relationship in the form

$$\sqrt{N}|\mu_1 - \mu_0| = Z_\alpha \sigma_0 + Z_\beta \sigma_1 \tag{5}$$

which can then be solved for N or Z_β. This form will be employed in cases where the basic equations equivalent to equations (3) and (4) become cumbersome.

The following sections demonstrate how these simple relationships can be employed with Student's t tests, chi-square tests for proportions, analyses of survival time, and tests for correlations. In each case, the explicit equations for sample size and power computation are presented with examples.

Most procedures allow unequal group sizes reflected in the *sample fractions* Q_e and Q_c where $n_e = Q_e N$, $n_c = Q_c N$ and $Q_e + Q_c = 1$. In the following, the subscripts e and c are used to denote the experimental and control groups where the total sample size then is $N = n_e + n_c$. Obviously, $Q_e = Q_c = 0.5$ for equal-sized groups. For these procedures it is well known that power is maximized and total sample size minimized for equal-sized groups, but due to ethical considerations, unequal-sized groups are at times desirable.

Virtually all these methods can be used with a simple calculator. To use these methods one must first specify the parameters of the problem. In addition to Z_α and Z_β, and the sample fractions Q_e and Q_c, the specific parameters of that test must be specified. For example, for the t test of two independent groups, the group means ν_e and ν_c are required as well as the standard deviation of the measures, σ. For some other statistical procedures

a separate standard deviation is not required since the variance will be a function of the expectation and/or the sample size alone.

ADDITIONAL CONSIDERATIONS

Sample size evaluation for a clinical trial is almost always a matter of compromise between the available resources and the various objectives, such as safety with small effects desirable and efficacy with large effects desirable, [1, 2]. This leads to a recursive process whereby one cycles through various specifications of the desired detectable effects and considers the resulting sample size in relation to the objectives and the resources available. Eventually one reaches a sample size and statement of objectives that are consistent with each other and the available resources.

In this process, however, attention should also be given to the operational aspects of the trial. Foremost among these are the factors related to the administration of the program of therapy and the evaluation of outcome. As a simple example, consider a clinical trial of an ulcer healing agent with healing assessed endoscopically after 4 weeks. Noncompliance, dropouts, and lack of control of other factors such as diet, drinking, and smoking may all combine to reduce the observed healing rate and thus reduce the statistical power of the trial. Likewise, failure to set uniform standards for endoscopic examination and criteria for healing will increase the variability of the outcome measurement, again leading to reduced power.

Among these, a major consideration is the rate of dropouts, patients who terminate therapy for reasons related neither to the disease under treatment nor the therapy. If an R dropout rate is expected, a simple but adequate adjustment is provided by $N_d = N/(1 - R)^2$ where N is the sample size calculated assuming no dropouts and N_d that required with dropouts [1]. Likewise, to evaluate power one would use equations and tables with $N = N_d(1 - R)^2$ where N_d is the observed or expected sample size. Additional procedures are described in [8] and [9].

STUDENT'S t TEST

In its most general form, Student's t is used to test the hypothesis that the mean of a normal variable, v, equals some specified value H_0: $\mu_0 = v_0$ against some alternative H_1: $\mu_1 = v_1$, $v_1 \neq v_0$, when the variance is unknown. The test statistic is of the form $t = \sqrt{N}(x - \mu_0)/S$ where x is the sample mean with standard error S^2/N, S^2 being the unbiased sample estimate of the variance σ^2 on $N - 1$ degrees of freedom (df). The distribution of t becomes increasingly close to that of a standard normal variable as df increases, at least 30 df being required for the approximation to be adequate [10]. Thus equations (1) through (4) can be employed to yield an approximate evaluation of sample size and power.

This approach, however, will tend to overestimate power for given S^2 and N, and thus it will tend to underestimate the required sample size, although this effect is increasingly negligible for increasing df. An adequate adjustment is obtained by the correction factor f = (df + 3)/(df + 1), where

fN patients are actually employed after N is obtained from equation (3), or, alternately, by N/f used in equation (4) when solving for power [6, p. 114].

For those who desire an exact solution, an iterative procedure is required and is described, with brief tables, in Cochran and Cox [11, p. 19]. Under this procedure, one obtains a trial value for N that is then adjusted in light of the resulting degrees of freedom.

In sample size or power evaluation for the t test a critical feature is the specification of σ^2. For the other tests considered later (proportions, etc.) the variances are not specified separately. Usually a value for σ^2 can be specified based on prior experiments using the same measurements; in these cases it is best to use the largest value σ^2 expected. Often a pilot study is helpful to provide an estimate of σ^2 under the conditions to be used in the experiment to be conducted. Of course, if power is to be evaluated *after* an experiment was conducted with a given N, then the obtained estimate S^2 of the variance should be employed.

A Single Mean

The most basic form of the t test is the test of H_0: $\mu_0 = \nu_0$ for some a priori specified mean value ν_0 with variance σ_0^2, against an alternative H_1: $\mu_1 = \nu_1 \neq \nu_0$ with variance σ_1^2. The test statistic is as presented where x is the mean of a single sample of observations with sample variance S^2. Given specified α, β, μ_0, μ_1, σ_0, and σ_1, the equations for sample size N or power (Z_β) are exactly as presented in equations (3) and (4), respectively.

Two Independent Groups

The t test is most widely used to test the null hypothesis that the means of two independent groups are equal, H_0: $\mu_0 = (\nu_e - \nu_c) = 0$, based on two separate samples of sizes $n_e = Q_e N$ and $n_c = Q_c N$, $Q_e + Q_c = 1$. The fractions Q_e and Q_c are the sample fractions and refer simply to the proportion of patients in each group, N being the total sample size. The test statistic employs the pooled estimate $\bar{S}^2$ of the common variance $\sigma_e^2 = \sigma_c^2 = \sigma^2$ with $N - 2$ df. It is well known that power is maximized for $Q_e = Q_c = 0.5$.

Under H_1, μ_1 is specified as the minimal relevant difference to be detected, $\mu_1 = |\nu_e - \nu_c| \neq 0$, and it follows that $\Sigma_0^2 = \Sigma_1^2 = \sigma^2(Q_e^{-1} + Q_c^{-1})/N$. Using equation (1) the equations for total N and Z_β are

$$N = \frac{\sigma^2(Q_e^{-1} + Q_c^{-1})(Z_\alpha + Z_\beta)^2}{\mu_1^2} \qquad (6)$$

$$Z_\beta = \frac{|\mu_1|\sqrt{N} - Z_\alpha\sigma\sqrt{Q_e^{-1} + Q_c^{-1}}}{\sigma\sqrt{Q_e^{-1} + Q_c^{-1}}} \qquad (7)$$

where for equal sample sizes $(Q_e^{-1} + Q_c^{-1}) = 4.0$.

For example, consider an experiment where σ is known not to exceed or known to be $\sigma = 1.0$ and it is desired to detect a difference $\mu_1 = (\nu_e - \nu_c) = 0.20$. From equation (6), to ensure a 90% chance of detecting this difference ($Z_\beta = 1.282$) with $\alpha = 0.05$ (one-sided, $Z_\alpha = 1.645$), $N = 858$ is

required. This would yield a t test on 856 ($N - 2$) df, and thus with the correction factor $f = 1.006$, the final sample size desired is $fN = 860$.

Suppose, however, that the experiment was actually conducted with only 102 patients. The correction factor is 1.02 and equation (7) is employed with $N = (102/f) = 100$, to yield $Z_\beta = -0.645$, thus indicating that for $N = 100$ the experiment had about 26% power. If the experiment produced a negative result, however, equation (6) could also be used to show that with $N = 100$ there was almost 100% power in detecting a difference $\mu_1 = 1.0$ ($Z_\beta = 3.355$). Thus one could safely rule out a difference on the order of $\mu_1 = 1.0$.

Paired Observations

In the event that the observations in the two groups are linked together by pairing or repeated measures at times a and b on the same patient, the t test is conducted using the mean difference $\overline{d} = \overline{X}_b - \overline{X}_a$ with a standard error $\Sigma^2 = \sigma_d^2/N$ where $\sigma_d^2 = 2\sigma^2(1 - \rho)$ if $\sigma_a^2 = \sigma_b^2 = \sigma^2$, ρ being the prepost correlation. From equation (1) the equations for N and Z_β for detecting a true difference $\mu_1 = \nu_b - \nu_a$ are:

$$N = \frac{(Z_\alpha + Z_\beta)^2 \sigma_d^2}{\mu_1^2} \tag{8}$$

$$Z_\beta = \frac{|\mu_1|\sqrt{N} - Z_\alpha \sigma_d}{\sigma_d} \tag{9}$$

which are equivalent to using equations (3) and (4) with σ_d in place of σ_0 and σ_1.

In this instance, often an estimate of σ_d^2 is available from prior experience. If not, an estimate of σ^2 can be used with an estimate of the correlation ρ. Note that pairing is only efficient if $\rho > 0$, i.e., there is positive correlation between the a and b measurements. If no estimate of ρ is available, it is safest to assume $\rho = 0$ or, nominally, $\rho = 0.10$.

Two Independent Groups with Paired Observations

A common related design is to employ two treatments in samples of sizes n_e and n_c where each patient also serves as his own control with measures at times a and b such as before and after treatment. In this case the test statistic is the same as for two independent groups where the prepost differences for each patient are used as the individual observations; i.e., $\overline{X}_e = \overline{d}_e$ and $\overline{X}_c = \overline{d}_c$; and where the pooled estimate of the variance of the differences (S_d^2) is employed. The problem is formulated as $\mu_1 = |\delta_e - \delta_c|$, $\delta_e = (\nu_{eb} - \nu_{ea})$, $\delta_c = (\nu_{cb} - \nu_{ca})$, with $\Sigma^2 = \sigma_d^2(Q_e^{-1} + Q_c^{-1})/N$. This yields equations equivalent to equations (6) and (7) with σ_d substituted for σ.

PROPORTIONS

In experiments where the basic outcome is a qualitative variable, such as success versus failure, the data are usually expressed as a proportion, e.g., the proportion of successes, or simply p. The exact probability distribution

of such a proportion is the binomial distribution that has parameters N (sample size) and π (the true population proportion). For large N (i.e., asymptotically) the binomial distribution may be approximated by a normal distribution with mean $\mu = \pi$ and variance $\Sigma^2 = \pi(1 - \pi)/N$. Thus, in experiments involving tests for proportions, the basic equations may be used for the determination of sample size and power.

A Single Proportion

In one-sample problems that yield a single proportion, the hypothesis $H_0:\mu_0 = \pi_0$ is tested where one wishes to detect a clinically relevant alternative $H_1:\mu_1 = \pi_1$ where $\pi_1 > \pi_0$ or $\pi_1 < \pi_0$. Given a proportion p from a sample of size N, the test statistic employed is $Z = (p - \pi_0)/\Sigma_0$ where $\Sigma_0^2 = \pi_0(1 - \pi_0)/N$ and where $Z \sim N(0, 1)$ if H_0 is true. As an example, in a cohort follow-up study, one might test that the k year mortality equals that obtained in a previous (and comparable) cohort, where π_0 is that observed in this latter cohort.

For the determination of sample size or power one substitutes $\sigma_0^2 = \pi_0(1 - \pi_0)$ and $\sigma_1^2 = \pi_1(1 - \pi_1)$ into equation (3) or (4); the equations for sample size N and power Z_β being

$$N = \left[\frac{Z_\alpha\sqrt{\pi_0(1 - \pi_0)} + Z_\beta\sqrt{\pi_1(1 - \pi_1)}}{\pi_1 - \pi_0} \right]^2 \tag{10}$$

$$Z_\beta = \frac{\sqrt{N}|\pi_1 - \pi_0| - Z_\alpha\sqrt{\pi_0(1 - \pi_0)}}{\sqrt{\pi_1(1 - \pi_1)}} \tag{11}$$

Two Independent Proportions

In the case of two independent samples of sizes n_e and n_c, the null hypothesis $H_0:\mu_0 = (\pi_e - \pi_c) = 0$ is tested with the statistic $Z = (p_e - p_c)/S$ where p_e and p_c are the proportions of events in the two samples, σ_0^2 is estimated as $S^2 = (n_e^{-1} + n_c^{-1})\bar{p}(1 - \bar{p})$, $\bar{p} = Q_e p_e + Q_c p_c$, and where under H_0, $Z \sim N(0, 1)$.

For the determination of sample size and power, the minimal relevant difference $\mu_1 = |\pi_e - \pi_c|$ is then specified. Since the variance will depend on the values specified and not on the absolute difference, both π_e and π_c must be specified. This yields $\sigma_1^2 = [\pi_e(1 - \pi_e) Q_e^{-1} + \pi_c(1 - \pi_c) Q_c^{-1}]$. Under the null hypothesis $H_0:\pi_e = \pi_c = \pi$, $\mu_0 = 0$ and π is specified as $\pi = Q_e\pi_e + Q_c\pi_c$. This then yields $\sigma_0^2 = (Q_e^{-1} + Q_c^{-1})\pi(1 - \pi)$. Substituting into equation (5) yields the well-known formula

$$\sqrt{N}|\pi_e - \pi_c| = Z_\alpha\sqrt{\bar{\pi}(1 - \bar{\pi})(Q_e^{-1} + Q_c^{-1})}$$
$$+ Z_\beta\sqrt{\pi_e(1 - \pi_e)Q_e^{-1} + \pi_c(1 - \pi_c)Q_c^{-1}} \tag{12}$$

which can then be solved for N or Z_β.

This expression can be simplified, however, by noting that for equal sample sizes $\sigma_0^2 = 4\bar{\pi}(1 - \bar{\pi})$ is always greater than or equal to $\sigma_1^2 = 2\pi_e(1$

$-\pi_e) + 2\pi_c(1 - \pi_c)$. This then allows use of the simpler equations

$$N = \frac{(Z_\alpha + Z_\beta)^2 4\bar{\pi}(1 - \bar{\pi})}{(\pi_e - \pi_e)^2} \tag{13}$$

$$Z_\beta = \frac{\sqrt{N}|\pi_e - \pi_c|}{2\sqrt{\bar{\pi}(1 - \bar{\pi})}} - Z_\alpha \tag{14}$$

Halperin (personal communication to Paul Canner) has shown that the approximation (13) will yield total sample sizes no greater than $Z_\beta^2 + 2Z_\alpha Z_\beta$ above that obtained from equation (12); i.e., to within 5.86 units for $\alpha = 0.05$ (one-sided), $\beta = 0.10$.

In using these formulas, note that π depends on the actual values π_e and π_c specified under H_1 and not just on the relevant difference $\mu_1 = |\pi_e - \pi_c|$. Also, since $\bar{\pi}(1 - \bar{\pi})$ is at a maximum for $\bar{\pi} = 0.50$, it then follows that for fixed positive μ_1, as π_c gets smaller, the required sample size also gets smaller and power increases assuming $\pi_c < \pi_e$). In such problems, therefore, it is safest to specify the largest realistic value for π_c (where $\pi_e > \pi_c$ and, $\pi_c < 0.50$) so as not to underestimate sample size or overestimate power.

For example, suppose we wished to conduct a controlled clinical trial of a new therapy and the rate of successes in the control group is not expected to be greater than $\pi_c = 0.05$. Further, we would consider the new therapy to be superior—cost, risks and other factors considered—if $\pi_e = 0.15$, thus yielding $\mu_1 = 0.10$, $\bar{\pi} = 0.10$, and $4\bar{\pi}(1 - \bar{\pi}) = 0.36$. Using equation (13) with $\alpha = 0.05$ (one-sided) and $\beta = 0.10$ yields $N = 310$ (rounded up from 308.4); the more precise formula (12) yields $N = 306$ (rounded from 304.6).

Suppose, however, that the experiment was conducted with only $N = 100$. Using equation (14) indicates that the power of the experiment in detecting $\mu_1 = 0.10$ with $\pi_c = 0.05$ is only 51% ($Z_\beta = 0.022$). If a negative result was obtained, however, one might wish to determine the power of having detected larger differences, say $\mu_1 = 0.40$ for $\pi_c = 0.05$. This yields $\pi_e = 0.45$, $\bar{\pi} = 0.25$, and $2\sqrt{\bar{\pi}(1 - \bar{\pi})} = 0.886$. From equation (14) we find that $N = 100$ yields 99.9% power ($Z_\beta = 2.87$). Thus a true difference of this magnitude could confidently be ruled out.

For further illustration, Lachin [2] used these procedures to discuss sample size considerations for FDA Phase II and III clinical trials of new drugs, and these methods have been used in a variety of clinical trials. Additional references include [1, 5, 7, and 8].

The Angular Transformation

The procedures just described are usually preferred since the tests for proportions using the normal approximation to the binomial are equivalent to the usual χ^2 tests (see under Discussion following). Others [12], however, have employed the angular transformation $A(p) = 2 \arcsin \sqrt{p}$, where $A(p)$ is expressed in radians, not degrees.[2] Given a proportion p with binomial

---

[2]For those whose calculators provide the sin function in degrees, the conversion factor is arcsin (radians) = (0.017453) arcsin (degrees).

expectation π, then $A(p)$ is approximately normally distributed as $N[A(\pi), N^{-1}]$. Since the variance $(\Sigma^2 = 1/N)$ is now independent of the expectation, the resulting sample size and power equations are further simplified. This approach, however, is not as accurate as that described herein.

As an illustration, again consider the example presented earlier under Two Independent Proportions with $\pi_c = 0.05$ and $\pi_e = 0.15$. The equation based on the arcsin transformation with equal sample sizes is

$$N = \frac{2(Z_\alpha + Z_\beta)^2}{[A(\pi_e) - A(\pi_c)]^2} \tag{15}$$

and for $\alpha = 0.05$ (one-sided) and $\beta = 0.10$ we find $N = 290$. This is somewhat less than the $N = 310$ estimated from the approximate equation (13) and the more precise formula (12), which yields $N = 306$. In general the angular transformation procedure yields N about 3–5% less than that from equation (12), with $\alpha = 0.05$ (one-sided), $\beta = 0.10$.

Paired Observations

Now consider the problem where two groups of observations are linked together in some way such as through matching or repeated measures on the same individuals at times a and b. This is exactly analogous to the problem of the t test for paired observations except that the outcome is now qualitative rather than quantitative. In this case, the basic data are expressed as

where m_{+-}, for example, is the number of pairs with $(+)$ for observation a (time a or the a pair member) and $(-)$ for observation b. For the a observations m_a is the total number $(+)$ and likewise m_b for the b observations. The frequencies (m's) are then converted to proportions (p's) by dividing by the total number of pairs, N.

In such problems, one wishes to test the null hypothesis $H_0:\mu_0 = (\pi_b - \pi_a) = 0$. Note, however, that $\pi_b - \pi_a = \pi_{-+} - \pi_{+-}$; thus the problem can then be expressed solely in terms of the discordant proportions π_{-+} and π_{+-} where H_0 implies that $\pi_{-+} = \pi_{+-} = \pi$. The test statistic employed is $Z = (p_{-+} - p_{+-})/S$ where $S^2 = 2\bar{p}/N$, $\bar{p} = \frac{1}{2}(p_{-+} + p_{+-})$ is the sample estimate of π and where under H_0, $Z \sim N(0, 1)$. Note that Z^2 is equivalent to the McNemar χ^2 statistic usually employed (see under Discussion following).

For sample size or power determination the clinically relevant difference $\mu_1 = |\pi_{-+} - \pi_{+-}|$ is specified. The corresponding variance has been shown by Miettinen [13] to be $\sigma_1^2 = 2\pi_{-+} \pi_{+-}/\bar{\pi}$ where $\bar{\pi} = \frac{1}{2}(\pi_{-+} + \pi_{+-})$. Under

$H_0: \mu_0 = (\pi_{-+} - \pi_{+-}) = 0$, which implies $\sigma_0^2 = 2\bar{\pi}$. Substituting into equation (5) yields the basic relationship

$$\sqrt{N} |\pi_{-+} - \pi_{+-}| = Z_\alpha \sqrt{2\bar{\pi}} + Z_\beta \sqrt{2\overline{\pi_{-+} \pi_{+-}}/\bar{\pi}} \tag{16}$$

which can be solved for the total number of paired observations (N) or power (from Z_β).

For example, consider that we wish to detect a difference $\mu_1 = 0.15$ where $\pi_{+-} = 0.05$, (implying $\pi_{-+} = 0.20$), using equation (16) for $\alpha = 0.05$ (one-sided), $\beta = 0.10$ yields $N = 80$.

Two Independent Groups with Paired Observations

As with the t test, this can be expanded to the problem of two independent groups of patients with paired observations on each patient.[3] Under this design, repeated observations ($+$ or $-$) are obtained at times a and b on two independent groups of sizes $n_e = Q_e N$ and $n_c = Q_c N$. The null hypothesis of no treatment by time interaction $H_0: \mu_0 = \delta_e - \delta_c = 0$ is to be tested, where $\delta_e = \pi_{eb} - \pi_{ea}$ is the change over time in the treated group and $\delta_c = \pi_{cb} - \pi_{ca}$ is that for controls.

As shown under Paired Observations, the problem can be expressed solely in terms of the discordant observations within the two-way table for each group, denoted as π_{e+-}, π_{e+-}, π_{c+-}, and π_{c-+}, which in turn define the degree of interaction $\mu_1 = |\delta_e - \delta_c|$ to be detected. Under H_1 the sample statistic $D = d_e - d_c = p_{e-+} - p_{e+-} - p_{c-+} + p_{c+-}$ is normally distributed with $\mu_1 = |\Delta|$ where

$$\Delta = \pi_{e-+} - \pi_{e+-} - \pi_{c-+} + \pi_{c+-} \tag{17}$$

and

$$\sigma_1^2 = \frac{4\pi_{e-+}\pi_{e+-}}{Q_e(\pi_{e-+} + \pi_{e+-})} + \frac{4\pi_{c-+}\pi_{c+-}}{Q_c(\pi_{c-+} + \pi_{c+-})} \tag{18}$$

A sufficient condition for H_0 to be true is the assumption of homogeneity wherein the treated and control groups are assumed to be drawn from the same population with common parameters $\pi_{+-} = Q_e\pi_{e+-} + Q_c\pi_{c+-}$ and $\pi_{-+} = Q_e\pi_{e-+} + Q_c\pi_{c-+}$ yielding $\mu_0 = 0$. Alternatively, such a severe assumption may not be required and one might fit a no-interaction model to the interaction parameters to obtain the set of no-interaction parameters, π', as $\pi'_{e+-} = \pi_{e+-} + \gamma$, $\pi'_{e-+} = \pi_{e-+} - \gamma$, $\pi'_{c+-} = \pi_{c+-} - \gamma$ and $\pi'_{c-+} = \pi_{c-+} + \gamma$, where $\gamma = \Delta/4$ and Δ is defined as in equation (17). Alternatively, one might employ the same parameters π_{c+-}, π_{e+-} under H_0 and H_1 and then complete the no-interaction model with parameters $\pi'_{c+-} = \pi_{c+-}$, $\pi'_{c-+} = \pi_{c-+} + \frac{1}{2}\Delta$, $\pi'_{e+-} = \pi_{e+-}$ and $\pi'_{e-+} = \pi_{e-+} - \frac{1}{2}\Delta$. In each case D is normally distributed with $\mu_0 = 0$ and variance σ_0^2 of the same form as equation (18) but with π' substituted for the π.

---

[3]Lachin et al. [1] also present extensions to analyses across independent subgroups within two independent primary groups.

The null hypothesis $H_0: \Delta = 0$ is then tested using $Z = (d_e - d_c)/\hat{\sigma}_0$, usually with $\hat{\sigma}_0$ defined from the sample p's under the assumption of homogeneity. In the latter case $\pi'_{e+-} = \pi'_{c+-} = \pi_{+-}$ and $\pi'_{e-+} = \pi'_{c-+} = \pi_{-+}$.

Substituting into equation (5) yields the equation

$$
\sqrt{N}|\pi_{e-+} - \pi_{e+-} - \pi_{c-+} + \pi_{c+-}| =
$$

$$
Z_\alpha \sqrt{\frac{4\pi'_{e-+}\pi'_{e+-}}{Q_e\,(\pi'_{e-+} + \pi'_{e+-})} + \frac{4\pi'_{c-+}\pi'_{c+-}}{Q_c\,(\pi'_{c-+} + \pi'_{c+-})}}
$$

$$
+ Z_\beta \sqrt{\frac{4\pi_{e-+}\,\pi_{e+-}}{Q_e\,(\pi_{e-+} + \pi_{e+-})} + \frac{4\pi_{c-+}\,\pi_{c+-}}{Q_c(\pi_{c-+} + \pi_{c+-})}} \quad (19)
$$

with the π' defined under the assumption of homogeneity or after fitting one of the no-interaction models. This can then be solved for total sample size N or power.

For example, consider a clinical trial in which 100 patients, 50 in each group, are to undergo evaluation before and after treatment and we desire the power of the study to detect group differences. The parameters of the problem may be specified as π_{c+-}, δ_c (which yields π_{c-+}), π_{e+-}, and Δ (which then yields π_{e-+}). Assume we are interested in detecting moderate differences such as $\pi_{c+-} = 0.03$, $\delta_c = 0.05$, $\pi_{e+-} = 0.03$, and $\mu_1 = \Delta = 0.15$. Using $\pi'_{c+-} = \pi_{c+-}$ and $\pi'_{e+-} = \pi_{e+-}$, fitting a no-interaction model and then using equation (19) with $\alpha = 0.05$, (one-sided), we find that $Z_\beta = 0.734$ and power = 77% ($\beta = 0.23$). Solving for sample size in equation (19) with $\beta = 0.10$ indicates that $N = 151$ yields 90% power of detecting these same effects.

Discussion

Although the problems just given are presented in terms of the normal approximation to the binomial, a *two-tailed* test using each of the statistics, Z, presented under A Single Proportion, Two Independent Proportions, and Paired Observations yields the same p value as the one df chi-square test usually employed in the same situation. For each of these Z and chi-square (χ^2) tests, it is easily shown that $\chi^2 = Z^2$ and thus that the p values for the two tests are the same. For example, the 1 df chi-square critical value at the 0.05 level is $\chi^2_{0.05} = 3.841$, which equals $(1.96)^2$, where $Z_{0.025} = 1.96$ (the *two-tailed* critical value at the 0.05 level). Thus, if one intends to use the inherently two-tailed chi-square test, two-tailed sample size or power determination should be employed (i.e., using $Z_{\alpha/2}$ rather than Z_α). Otherwise, sample size may be severely underestimated.

When a two-tailed test is to be conducted, however, one must carefully consider each of the two possible alternatives. For example, in tests of a single proportion, $H_0: \pi = \pi_0$ is tested against an alternative, which for a two-sided test is specified as $H_1: \mu_1 = \delta = |\pi_1 - \pi_0| \neq 0$. The two-sided test thus implies two alternative values for $\pi_1: \pi_{1u} = \pi_0 + \delta$ and $\pi_{1\ell} = \pi_0 - \delta$. Obviously, since the variances depend on π_1, the estimated sample size will be greater, and power smaller, for the alternative (π_{1u} or $\pi_{1\ell}$) closest to 0.50. [Note that $\pi(1 - \pi)$ is maximized at $\pi = 0.50$]. In fact, the larger of the two resulting sample size estimates may be as much as 4.64 times the smaller

estimate. Thus, if a two-tailed analysis is to be conducted, one should consider the two implied alternatives (e.g., π_{1u} and $\pi_{1\ell}$) and use whichever is closest to 0.50.

An alternative would be to employ sample size procedures using the power function of the chi-square test itself, which is inherently two-tailed. Lachin [14] discusses this procedure for the general r × c contingency table and showed that the use of the limiting chi-square power approach and the two-tailed asymptotic normal equation (11) were in close agreement for the 2 × 2 contingency table.

SURVIVAL ANALYSIS

In many clinical trials, simple proportions as described in the last section will be used to evaluate the outcome, such as to evaluate the healing or improvement rate in an acute condition with a short-term therapy. In many other cases, however, the important feature is not only the outcome event, such as death, but the time to the terminal event, the *survival time*. In these trials, the data is analyzed using life-table methods that consider the time to the terminal event for each patient and that provide a more powerful estimate of the, say, T year survival than is obtained from the crude proportion of survivors after T years. Some basic references on this procedure are [15–17].

The basic life-table method of analysis is distribution free in that no underlying assumptions about the distribution of time to event need be specified. For sample size evaluation, however, some such assumption must be made. The most common assumption is that time to survival is exponentially distributed with hazard rate λ, where at any time t the proportion of survivors, $P_s(t)$, is given as $P_s(t) = e^{-\lambda t}$. Under this model $\log [P_s(t)]$ is linearly decreasing in time with slope λ. In a cohort of N patients, all followed to the terminal event with mean survival time M, the hazard rate is estimated as $L = M^{-1}$ and asymptotically $L \sim N(\lambda, \lambda^2/N)$, [18].

Two Independent Groups

Consider that there are two independent groups of sizes n_e and n_c all followed to the terminal event where time t is measured from the time of entry into the study. The null hypothesis of equality of survival is equivalent under exponential survival to H_0: $(\lambda_e - \lambda_c) = 0$, which can be tested using the statistic $Z = (L_e - L_c)/S$ where L_e and L_c are the estimated hazard rates, $L_e = M_e^{-1}$, $L_c = M_c^{-1}$, $S^2 = (n_e^{-1} + n_c^{-1})\bar{L}^2$, $\bar{L} = (Q_e L_e + Q_c L_c)$, and where under H_0, $Z \sim N(0, 1)$.

For the determination of sample size and power one specifies the minimal relevant difference $\mu_1 = |\lambda_e - \lambda_c|$, which yields $\sigma_1^2 = (\lambda_e^2 Q_e^{-1} + \lambda_c^2 Q_c^{-1})$. Under the null hypothesis $\mu_0 = 0$ and $\sigma_0^2 = \bar{\lambda}^2 (Q_e^{-1} + Q_c^{-1})$ where $\bar{\lambda} = Q_e \lambda_e + Q_c \lambda_c$. Substituting into equation (5) yields

$$\sqrt{N}|\lambda_e - \lambda_c| = Z_\alpha \sqrt{\bar{\lambda}^2 (Q_e^{-1} + Q_c^{-1})} + Z_\beta \sqrt{\lambda_e^2 Q_e^{-1} + \lambda_c^2 Q_c^{-1}} \qquad (20)$$

which can then be solved for N or Z_β.

This equation was also presented by Pasternack and Gilbert [19] and was

shown by George and Desu [20] to be slightly conservative in comparison to the exact distribution of the ratio L_e/L_c, which has an F distribution. George and Desu also present the following approximation

$$\sqrt{N}[\tfrac{1}{2}\ell n(\lambda_c/\lambda_e)] = Z_\alpha + Z_\beta \tag{21}$$

which they show to be accurate to within two sample units of the exact solution.

Another approximation can be obtained directly from equation (20) by noting that for equal sample sizes $4\lambda^2$ is always less than or equal to $2(\lambda_e^2 + \lambda_c^2)$. This then yields

$$\sqrt{N}|\lambda_e - \lambda_c|/(\lambda_e + \lambda_c) = Z_\alpha + Z_\beta \tag{22}$$

when using equal sample sizes. This approximation will yield values between those from equation (20) and the approximation (21) of George and Desu and can be shown to be within Z_β^2 less than that obtained from equation (20).

Two Independent Groups with Censoring

The formulation just presented will rarely be applicable because it assumes that all N patients will be followed to the terminal event no matter how much time is required for the last patient to reach that event. This is rarely practicable. A more realistic approach is to allow for the trial to be terminated at time T. Assume that the patients enter the trial at a uniform rate over the interval 0 to T and that exponential survival applies, as earlier. If we denote

$$\phi(\lambda) = \lambda^3 T/(\lambda T - 1 + e^{-\lambda T}) \tag{23}$$

then it can be shown that $\sigma_0^2 = \phi(\lambda)(Q_e^{-1} + Q_c^{-1})$ and $\sigma_1^2 = \phi(\lambda_e)Q_e^{-1} + \phi(\lambda_c)Q_c^{-1}$ where $\lambda = Q_e\lambda_e + Q_c\lambda_c$ [18]. Substituting $\mu_1 = |\lambda_e - \lambda_c|$, $\mu_0 = 0$, σ_0^2, and σ_1^2 into equation (5) yields

$$\sqrt{N}|\lambda_e - \lambda_c| = Z_\alpha\sqrt{\phi(\overline{\lambda})(Q_e^{-1} + Q_c^{-1})} + Z_\beta\sqrt{\phi(\lambda_e)Q_e^{-1} + \phi(\lambda_c)Q_c^{-1}} \tag{24}$$

with the $\phi(\lambda)$ as defined in equation (23). This can then be solved for sample size N or power Z_β.

These expressions can be simplified, however, since empirically for $Q_e = Q_c$, $\sigma_1^2 > \sigma_0^2$, and as employed by Gross and Clark [18, p. 264] we can use the simple equation

$$\sqrt{N}|\lambda_e - \lambda_c| = (Z_\alpha + Z_\beta)\sqrt{\phi(\lambda_e)Q_e^{-1} + \phi(\lambda_c)Q_c^{-1}} \tag{25}$$

Again this approximation is highly accurate.

In the event that all patients enter the trial at the same point, or if each patient enters the trial at random but each is only followed up to T years after entry, the resulting equations are identical except that $\phi(\lambda)$ in equation (23) becomes simply $\lambda^2/(1 - e^{-\lambda t})$.

At this point it should be noted that the sample size obtained with a

study of T years duration is that required to yield the same number of deaths (events) as obtained from equation (21), allowing for the fact that not all patients will have died when the study is terminated. This is also true of the following procedure.

Two Independent Groups with Limited Recruitment and Censoring

In the formulation just presented, note that patients are eligible to enter the trial up to the trial end date, time T. Usually, however, it will be desired to recruit patients for study over an interval 0 to T_0 and then to follow all recruited patients to the time of the terminal event, or to time T where $T > T_0$. Based on the developments in [18, pp. 66–67], it can readily be shown that the variances σ_0^2 and σ_1^2 are as in the previous section but with $\phi(\lambda)$ now defined as

$$\phi^*(\lambda) = \lambda^2 \left[1 - \frac{e^{-\lambda (T - T_0)} - e^{-\lambda T}}{\lambda T_0} \right]^{-1} \tag{26}$$

The desired sample size or power is obtained on substituting $\phi^*(\lambda)$ for $\phi(\lambda)$ in equation (24), or in equation (25) to yield an accurate approximation, and solving for N or Z_β.

For example, consider that a clinical trial is to be conducted for a disease with moderate levels of mortality with hazard rate $\lambda = 0.30$, yielding 50% survivors after 2.3 years. Suppose that with treatment we are interested in a reduction in hazard to $\lambda = 0.2$, i.e., an increase in survival to 64% at 2.3 years. With equal-sized groups, $\alpha = 0.05$ (one-sided) and $\beta = 0.10$, equation (20) yields $N = 218$ deaths are required, i.e., 218 patients all followed to time of death. The approximation (22) yields $N = 216$ and the equation (21) of George and Desu yields $N = 210$. If the study was to be terminated after 5 years, then using equation (24) with equation (23) yields $N = 504$ patients; the approximation yields $N = 508$. Finally, assume that recruitment was to be terminated after the first 3 years of a 5-year study, then using equation (24) with equation (26) yields $N = 378$.

Note that under all these plans the sample size requirements are based on the need to accrue approximately 210 deaths during the study. Also note that if a fixed number of patients is to be studied, it is better for those patients to be recruited quickly and followed for a longer period of time than to extend the period of study and reduce the rate at which the patients enter the study. This example shows that 504 patients would be needed for a 5-year study where the patients can enter the study evenly during the full 5-year period, whereas 378 patients would be needed if recruitment was compressed into a 3-year period with total study duration again 5 years. The reason for this quite simply is related to the total patient months of experience of the cohort, i.e., the elapsed time from the time of randomization to time T summed over all patients. For a 5-year study with recruitment compressed into the initial 3 years, the average patient months of exposure would be 3.5 years, whereas for a 5-year study with recruitment spanning the total 5 years, the average exposure to treatment would be 2.5 years.

CORRELATIONS

In observational studies that involve correlations as the principal form of analysis, two types of hypotheses are usually tested: (1) whether a true correlation actually exists using H_0: $\rho = 0$ versus H_1: $\rho = \rho_1 \neq 0$; and (2) whether two correlations are significantly different using H_0: $(\rho_e - \rho_c) = 0$ versus H_1: $\mu_1 = (\rho_e - \rho_c) \neq 0$. The simplest approach to such problems is to employ Fisher's arctanh transformation [5]:

$$C(r) = \frac{1}{2}\log_e \frac{(1 + r)}{(1 - r)}$$

Given a sample correlation r based on N observations that is distributed about an actual correlation value (parameter) ρ, then $C(r)$ is normally distributed with mean $C(\rho)$ and variance $\sigma^2 = 1/(N - 3)$. The transformation of r to C (and vice versa) is widely tabulated. (Note that this is usually termed Fisher's Z transformation, but we here use C to avoid conflict in notation.)

A Single Correlation

In detecting a relevant simple correlation of degree H_1: $\mu_1 = \rho_1$, one tests the null hypothesis H_0: $\rho = 0$ using the test statistic $Z = C(r)\sqrt{N - 3}$ where $Z \sim N(0, 1)$. Substituting into equation (5) yields

$$\sqrt{N - 3}C(\rho_1) = Z_\alpha + Z_\beta \tag{27}$$

from which the required sample size or power may be obtained. Obviously, to detect a true correlation ρ_1 greater than 0.50 [$C(\rho_1) = 0.549$], a small N would suffice. Note that H_0: $\rho = 0$ is equivalent to a null hypothesis that the regression coefficient is also zero.

Two Independent Correlations

In detecting a relevant difference in correlations H_1: $\mu_1 = |C(\rho_e) - C(\rho_c)| \neq 0$ obtained from two *independent samples*, the null hypothesis H_0: $\mu_0 = 0$ is tested using the statistic $Z = C(r_e) - C(r_c)/\Sigma_0$ where $\Sigma_0^2 = N^{-1}(Q_e^{-1} + Q_c^{-1})$, $n_e - 3 = Q_eN$, $n_c - 3 = Q_cN$, and where under H_0, $Z \sim N(0, 1)$. The correlations r_e and r_c are obtained from *two* samples of sizes n_e and n_c such as $r_e = r_{e(uv)}$ and $r_c = r_{c(uv)}$ for variables u and v in groups e and c.

Substituting $\mu_0 = 0$, $\mu_1 = |C(\rho_e) - C(\rho_c)|$, and $\Sigma_1^2 = \Sigma_0^2$ into equation (5) yields

$$\frac{\sqrt{N}|C(\rho_e) - C(\rho_c)|}{\sqrt{Q_e^{-1} + Q_c^{-1}}} = Z_\alpha + Z_\beta \tag{28}$$

which can then be solved for total sample size (N) or power (Z_β). Note that N from equation (28) will actually be six units less than that actually needed since $n_e + n_c - 6 = N$.

Table 1   Total Sample Size *(N)* from Equation (29)[a] as a Function of K (Where $\mu_1 = K\sigma$) for Various α (One-sided) and β[b]

K	$\alpha = 0.05$ $\beta = 0.20$	$\alpha = 0.05$ $\beta = 0.10$	$\alpha = 0.025$ $\beta = 0.10$	$\alpha = 0.025$ $\beta = 0.05$	$\alpha = 0.01$ $\beta = 0.05$
0.01	61,852	85,674	105,106	129,962	157,690
0.025	9,898	13,708	16,818	20,794	25,232
0.05	2,476	3,428	4,206	5,200	6,308
0.075	1,100	1,524	1,870	2,312	2,804
0.10	620	858	1,052	1,300	1,578
0.125	396	550	674	832	1,010
0.15	276	382	468	578	702
0.175	202	280	344	426	516
0.20	156	216	264	326	396
0.25	100	138	170	208	254
0.3	70	96	118	146	176
0.4	40	54	66	82	100
0.5	26	36	44	52	64
0.6	18	24	30	38	44
0.7	14	18	22	28	34
0.8	10	14	18	22	26
1.0	8	10	12	14	16

[a]Rounded to the next highest even number
[b]For a two-sided determination at level α, the table should be used with the value $\alpha/2$.

Table 2   Power $(1 - \beta)$ from Equation (30) as a Function of K and Total Sample Size N (where $\mu_1 = K\sigma$) with $\alpha = 0.05$ (one-sided)

					K					
N	0.05	0.10	0.15	0.20	0.25	0.30	0.40	0.50	0.75	1.00
10	0.0685	0.0920	0.1209	0.1556	0.1964	0.2431	0.3519	0.4745	0.7663	0.9354
20	0.0776	0.1155	0.1650	0.2265	0.2991	0.3808	0.5572	0.7228	0.9563	0.9977
30	0.0852	0.1363	0.2051	0.2913	0.3914	0.4993	0.7074	0.8629	0.9931	0.9999
40	0.0920	0.1556	0.2431	0.3519	0.4745	0.5996	0.8119	0.9354	0.9990	0.9999
50	0.0983	0.1741	0.2795	0.4087	0.5489	0.6831	0.8817	0.9707	0.9999	0.9999
60	0.1042	0.1920	0.3145	0.4618	0.6147	0.7514	0.9269	0.9871	0.9999	0.9999
70	0.1100	0.2094	0.3483	0.5113	0.6724	0.8065	0.9556	0.9944	0.9999	0.9999
80	0.1155	0.2265	0.3808	0.5572	0.7228	0.8504	0.9734	0.9977	0.9999	0.9999
90	0.1209	0.2431	0.4122	0.5996	0.7663	0.8851	0.9842	0.9990	0.9999	0.9999
100	0.1261	0.2595	0.4424	0.6387	0.8037	0.9123	0.9907	0.9996	0.9999	0.9999
200	0.1741	0.4087	0.6831	0.8817	0.9707	0.9953	0.9999	0.9999	0.9999	0.9999
300	0.2180	0.5347	0.8297	0.9656	0.9964	0.9998	0.9999	0.9999	0.9999	0.9999
400	0.2595	0.6387	0.9123	0.9907	0.9996	0.9999	0.9999	0.9999	0.9999	0.9999
500	0.2991	0.7228	0.9563	0.9977	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
750	0.3914	0.8629	0.9931	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
1000	0.4745	0.9354	0.9990	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
2500	0.8037	0.9996	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
5000	0.9707	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999

Table 3 K Values for Use with Tables 1 and 2 or Equations (29) and (30)[a]

Statistical test	μ_0 Under the null hypothesis (H_0)	μ_1 under H_1	Equation in text	K as a function of $\theta = (\mu_1 - \mu_0)$	
				K for equal sample sizes	K for unequal samples with sample fractions Q_e and Q_c where $Q^* = (Q_e^{-1} + Q_c^{-1})$
t Tests for Means (See Student's t test for correction factor)					
1. A single sample mean	ν_0	ν_1	3, 4	$|\nu_1 - \nu_0|/\sigma$	N.A.
2. Two independent groups e and c	$(\nu_e - \nu_c) = 0$	$\nu_e - \nu_c$	6, 7	$|\nu_e - \nu_c|/2\sigma$	$|\nu_e - \nu_c|/\sigma\sqrt{Q^*}$
				(σ^2 = variance of the observations)	
3. Correlated observations at times a and b	$(\nu_b - \nu_a) = 0$	$\nu_b - \nu_a$	8, 9	$|\nu_b - \nu_a|/\sigma_d$	N.A.
				(σ_d^2 = variance of the differences $x_b - x_a$)	
4. Two independent groups with paired observations	$(\delta_e - \delta_c) = 0$ $\delta_e = \nu_{eb} - \nu_{ea}$ $\delta_c = \nu_{cb} - \nu_{ca}$	$(\delta_e - \delta_c)$	6, 7	$|\delta_e - \delta_c|/2\sigma_d$	$|\delta_e - \delta_c|/\sigma_d\sqrt{Q^*}$
Tests for Proportions					
1. A single proportion					
Normal approximation	π_0	$\pi_1 = \pi_0$	10, 11	See Proportions. A Single Proportion	
Angular transformation	$A(\pi_0)$	$A(\pi_1)$		$|A(\pi_1) - A(\pi_0)|$	N.A.
	$A(\pi) = 2 \arcsin \sqrt{\pi}$ in radians				
2. Two independent groups e and c					
Normal approximation	$(\pi_e - \pi_c) = 0$	$\pi_e - \pi_c$	13, 14	$\frac{1}{2}|\pi_e - \pi_c|[\bar{\pi}(1 - \bar{\pi})]^{-1}$ $\bar{\pi} = (\pi_e + \pi_c)/2$	See Proportions, Two Independent Proportions
Angular transformation	$|A(\pi_e) - A(\pi_c)| = 0$	$A(\pi_e) - A(\pi_c)$	15	$\frac{1}{2}|A(\pi_e) - A(\pi_c)|$	$|A(\pi_e) - A(\pi_c)|/\sqrt{Q^*}$

111

3. Correlated observations at times a and b	$(\pi_{-+} - \pi_{+-}) = 0$	$\pi_{-+} - \pi_{+-}$	16	See Proportions, Paired Observations	N.A.
4. Two independent groups with paired observations	$\delta_e - \delta_c = 0$ $\delta_e = \pi_{e-+} - \pi_{e+-}$ $\delta_c = \pi_{c-+} - \pi_{c+-}$	$\delta_e - \delta_c$	19	See Proportions, Two Independent Groups with Paired Observations	
Survival Analysis—Exponential Model					
1. Two independent groups without censoring					
Normal approximation	$(\lambda_e - \lambda_c) = 0$	$\lambda_e - \lambda_c$	22	$|\lambda_e - \lambda_c|/(\lambda_e + \lambda_c)$	See Survival Analysis, Two Independent Groups
or	$(\lambda_c/\lambda_e) = 1$	λ_c/λ_e	21	$[\ell n(\lambda_c/\lambda_e)]/2$	
2. Two independent groups with censoring at time T	$(\lambda_e - \lambda_c) = 0$	$\lambda_e - \lambda_c$	25	$|\lambda_e - \lambda_c|[2\phi(\lambda_e) + 2\phi(\lambda_c)]^{-\frac{1}{2}}$	See Survival Analysis, Two Independent Groups with Censoring
			23	$\phi(\lambda) = \lambda^3 T/ (\lambda T - 1 - e^{-\lambda T})$	
3. Two independent groups with entry to T_0 and censoring at time $T > T_0$	$(\lambda_e - \lambda_c) = 0$	$\lambda_e - \lambda_c$	25	$|\lambda_e - \lambda_c|[2\phi^*(\lambda_e) + 2\phi^*(\lambda_c)]^{-\frac{1}{2}}$	See Survival Analysis, Two Independent Groups with Limited Recruitment and Censoring
				$\phi^*(\lambda) = \lambda^2 \left[\dfrac{\exp[-\lambda(T - T_0)] - \exp(-\lambda T)}{\lambda T_0}\right]^{-1}$	
Tests for Correlations					
1. A single correlation	$C(\rho_0)$	$C(\rho_1)$	27	$|C(\rho_1) - C(\rho_0)|$	N.A.
$C(\rho)$ = Fisher's transformation and in Tables 1 or 2 or in equations (29) and (30) one uses $N - 3$.					
2. Two independent groups e and c	$[C(\rho_e) - C(\rho_c)] = 0$	$\tilde{C}(\rho_e) - C(\rho_c)$	28	$\tfrac{1}{2}|C(\rho_e) - C(\rho_c)|$	$|C(\rho_e) - C(\rho_c)|/\sqrt{Q^*}$ $N = n_e + n_c - 6,\ Q_*N = n_e^{-3}$

a A value for μ_1 is first defined in terms of the individual elements, e.g. $\mu_1 = (0.5 - 0.2) = 0.3$ for the t test. Given the specification for the other elements of K (e.g. $\sigma = 0.2$; $\theta = (\mu_1 - \mu_0) = 0.3$) one solves for K, e.g. $K = \theta/2\sigma = 0.75$. Then proceed to Table 1 or 2 or equations (29, 30) for $K = 0.75$.

Two Related Correlations

In detecting a relevant difference between two correlations r_a and r_b obtained from a *single* sample of size N, the covariance $\text{Cov}[C(r_a), C(r_b)]$ must be considered. This obviously applies when the two correlations involve a common variable, e.g., $r_a = r_{uv}$ and $r_b = r_{uw}$ for variables u, v, and w. It also applies when the two correlations do not have a variable in common, e.g., $r_a = r_{uv}$ and $r_b = r_{wx}$ for variables u, v, w, and x, due to the other intercorrelations ρ_{uw}, ρ_{ux}, ρ_{vw}, and ρ_{vx}. Due to the complexity of the covariance expressions as given in [21], the test statistics and the solutions for sample size and power will not be presented, although the latter are also obtained directly from the basic equation (1).

FURTHER SIMPLIFICATION AND TABLES

In many of the situations just described, the equations for N and Z_β resulting from equations (3) and (4) can be simplified if the difference $\theta = |\mu_1 - \mu_0|$ is presented as a function of the standard deviation of the basic observations. If $\sigma_0 = \sigma_1 = \sigma$, and θ is specified as $\theta = K\sigma$, then the equations for sample size and power simply reduce to

$$N = [(Z_\alpha + Z_\beta)/K]^2 \tag{29}$$

$$Z_\beta = K\sqrt{N} - Z_\alpha \tag{30}$$

where $K = \theta/\sigma$. Table 1 presents total N from equation (29) as a function of K for various α and β levels. Table 2 presents power obtained from Z_β using equation (30) as a function of K and total N for $\alpha = 0.05$ (one-sided). If $\mu_0 = 0$ then $\theta = |\mu_1|$ and equations (29) and (30) simply give the sample size (or power) where the minimal relevant difference is expressed as a fraction (K) of the standard deviation of the observations.

These simplified equations are applicable to most of the procedures presented in this paper. Table 3 presents the expressions for K required for these various statistical tests. This table can be used with Tables 1 and 2 or with equations (29) and (30) directly. In each case, the corresponding explicit equation in the preceding text is cited.

REFERENCES

1. Lachin J, Marks J, Schoenfield L, et al: Design and Methodological Considerations in the National Cooperative Gallstone Study: A Multi-center Clinical Trial. *Controlled Clinical Trials*, in press.

2. Lachin J: Sample size considerations for clinical trials of potentially hepatotoxic drugs. In Davidson, CS, Levy, CM, and Chamberlayne, EC, eds: *Guidelines for Detection of Hepatotoxicity Due to Drugs and Chemicals*. Washington, DC: U.S. Department of H.E.W., National Institutes of Health, NIH Publication No. 79–13, pp. 119–130, 1979.

3. Lachin JM: Statistical inference in clinical trials. In Tygstrup N, Lachin J, Juhl E, eds: *The Randomized Clinical Trial and Therapeutic Decisions*. New York: Dekker, 1981 (in press).

4. Shaw LW, Cornfield J, Cole SM: Statistical problems in the design of clinical trials and interpretation of results. In Deutsch, E, ed: *Thrombosis: Pathogenesis and Clinical Trials*. New York: Schattauer Verlag, pp. 191–202, 1973.

5. Freiman JA, Chalmers TC, Smith H, Kuebler R: The importance of Beta, the Type II error and sample size in the design and interpretation of the randomized controlled trial. *N Engl J Med* 299:690–694, 1978.

6. Snedecor GW, Cochran WG: *Statistical Methods,* 6th ed. Ames: Iowa State University Press, 1967.

7. Fleiss J: *Statistical Methods for Rates and Proportions*. New York: Wiley, 1973.

8. Halperin M, Rogot E, Gurian J, Ederer F: Sample sizes for medical trials with special reference to long term therapy. *J Chron Dis* 21:13–23, 1968.

9. Schork MA, Remington RD: The determination of sample size in treatment control comparisons for chronic disease studies in which drop-out or non-adherence is a problem. *J Chronic Dis* 20:223–239, 1967.

10. Johnson NL, Kotz S: *Distributions in Statistics: Continuous Univariate Distributions 2*. New York: Wiley, 1970.

11. Cochran WG, Cox GM: *Experimental Designs*. New York: Wiley, 1964.

12. Sokal RD, Rohlf FJ: *Biometry: The Principles and Practice of Statistics in Biometric Research*. San Francisco: Freeman, 1969.

13. Miettinen OS: The matched pairs design in the case of all-or-none responses. *Biometrics* 24:339–352, 1968.

14. Lachin J: Sample size determinations for r × c comparative trials. *Biometrics* 33:315–324, 1977.

15. Cutler SJ, Ederer F: Maximum utilization of the life table method in analyzing survival. *J Chronic Dis* 8:699–712, 1978.

16. Breslow NE: Analysis of survival data under the proportional hazards model. *Int Stat Rev* 43:45–58, 1979.

17. Peto R, Pike MC, Armitage P, Breslow NE, Cox DR, Howard SV, Mantel N, McPherson K, Peto J, Smith PG: Design and analysis of randomized clinical trials requiring prolonged observation of each patient: II. Analysis and examples. *Br J Cancer* 35:1–39, 1977.

18. Gross AJ, Clark VA: *Survival Distributions: Reliability Applications in the Biomedical Sciences*. New York: Wiley, 1975.

19. Pasternack BS, Gilbert HS: Planning the duration of long-term survival time studies designed for accrual by cohorts. *J Chronic Dis* 24:681–700, 1971.

20. George SL, Desu MM: Planning the size and duration of a clinical trial studying the time to some critical event. *J Chronic Dis* 27:15–24, 1974.

21. Dunn OJ, Clark V: Correlation coefficients measured on the same individuals. *J Am Stat Assoc* 64:366–377, 1969.