SIXTEEN S-SQUARED OVER D-SQUARED: A RELATIONFOR CRUDE SAMPLE SIZE ESTIMATES

ROBERT LEHR, STATISTICS IN MEDICINE, VOL. 11, 1099-1102 (1992)

SUMMARY

I suggest for memorization an equation for calculating approximate sample size requirements intended only for a specific set of values (80 per cent power for a two-tailed alpha = 0.05 test) which seems to occur often in biopharmaceutical research. After presenting the formula in terms of variance estimate s^2 and effect size d, I derive a few alternative forms and then discuss the accuracy of the approximation and other properties as well as examples of its use.

INTRODUCTION

A common question posed to a pharmaceutical biostatistician is 'How many subjects will I need for this study?'. Often the biostatistician defers an answer until he or she receives a protocol, reviews any relevant sample data (if available),has clear definitions of a primary endpoint and corresponding clinically significant result, and has applied some 'powerful' software or statistical tables (or even some approximation formula).

The purpose of this note is not to try to replace the usual process, but rather to state a crude but easily remembered equation that applies to many of the sample size equations that confront a pharmaceutical biostatistician. It can help in arriving at ballpark figures at multicentre meetings and similar settings where computer terminals and appropriate tables are often unavailable. The relation that follows, therefore, is worth committing to memory.

For a two-sample, two-tailed t-test with alpha level 0.05, the following relation yields an approximate power of 0.80:

$$\boldsymbol{n = 16s^2 / d^2} \qquad \textbf{(1)}$$

where n is the size of each sample (treatment group) s^2 is an estimate of the population variance, and d is the difference to be detected. Conveniently, this inverts to:

$$\mathbf{d = 4s / \sqrt{\mathit{n}}} \qquad \textbf{(2)}$$

for quickly approximating detectable differences given sample size and variance estimates.

Another form of relationship (1) is

$$\boldsymbol{n = 16\,(cv)^2 / p^2} \qquad \textbf{(3)}$$

where cv = 100 s / $\bar{x}$ and p is the percentage of $\bar{x}$ (a sample mean) to be detected.

Author's address: Berlex Laboratories, 300 Fairfield Road, Wayne, NJ 07470, U.S.A.

For a test of the equality of two independent **proportions** p_1 and p_2, the use of equation(l) with $s^2 = pq$, [where $p = (p_1 + p_2)/2$ and $q = 1–p$] and $d = |p_1 - p_2|$ results in a reasonably good sample size approximation for Fisher's exact test or a chi-square test. This result follows from the fact that a binomially distributed variable is the sum of a fixed number of independent identically distributed Bernoulli variables, each therefore having thc same variance *(pq)*. Note that the symbol 'p' rather than the more conventional '$\overline{p}$' is being used to represent the mean of two proportions.

For **crossover studies and other related designs**, the same relations apply but the s^2 used is the mean square error term from an appropriate ANOVA, or 1/2 times an estimate of the variance of the difference (paired by subjects) of the two treatments, that is, $var(x_1 - x_2)$ where x_1 is the subject's response on treatment 1 and x_2 is his response on treatment 2. The resulting value for *n* will represent the total sample size. In conjunction with equation (3) this is useful for bioequivalence studies.

DISCUSSION

1 Accuracy - Table I gives powers calculated for a two-tailed, two-sample t-test given several values and ranges for *n*. As *n* gets larger, the power approaches 0 807.

2 The accuracy does not depend upon the actual magnitudes of s and *d;* only on their ratio.

3 For a one-sample test, we obtain the proper relation by replacing the 16 with 8.

4 We can make adjustments to improve the approximation for smaller *n*, but these would spoil the simplicity of thc relation.

5. Many sources (for example, References 1, 2, and 3) provide more general equations that lead to similar approximations, but the sheer simplicity of the basic equation (1) coupled with the frequency with which the specific value of 0.80 for a two-tailed alpha = 0.05 test seems to occur makes the relation well worth the effort of memorization.

6 Certain catch phrases such as 'to detect half as big a difference you need four times the sample size' are easily illustrated using equation (1). Also, the effect of the magnitude of the variance upon the sample size estimate is readily apparent.

7. A simple and intuitively pleasing derivation obtains from the fact that approximately 80 per cent of the area under the standard normal curve lies to the right of z = - 0 84, giving rise to $(d—2s.e.)/s.e. = 0.84$, where s.e. = standard error = the square root of $2s^2/n$. Solving this equation for *n* yields a relation close to the stated approximation. Alternately, the relation derives from a more general formula.

8. We can obtain similar equations for tests with different alpha levels (0.10,0.01) etc. and/or different powers; unfortunately the resulting coefficients are not as 'nice' as 16, as Table II illustrates.

EXAMPLES

1 Find the approximate number of subjects necessary to provide 80 per cent power to detect a difference of 3 between the mean response of treatments a and b with a two-sample, two-tailed alpha = 0.05 t-test given a pooled variance estimate of 28.

Using equation (1), $n = 16s^2/d^2 = 16 \times 28/9 = 49.7$ subjects per group. (The power obtained by a non-central t approximation using $n = 50$ and the given variance and difference is 0.81)

2 A 25 subject crossover bioequivalence study (13 subjects in one sequence and 12 in the other) produces an ANOVA with mean square error 900 and mean AUC for the standard treatment of 125. Approximately what per cent difference of the standard mean can we detect with 80 per cent *post hoc* power if we use a 0.05 level two-tailed comparison based on the t-distribution?

Using note 4 and an inverted form of equation (3),

$p = 4(cv) / \sqrt{n} = 4 \times 24/5$ = a 19.2 per cent mean difference.

3 How many patients do we need to detect (with 0.80 power) a difference in success rates of an active drug and placebo of 0.45 and 0.35 using a two-tailed corrected chi-square test?

Using equation (1) with $s^2 = pq$, $p = (0.35 + 0.45)/2 = 0.4$, $q = 1 - p = 0.6$, and $d = 0.45 - 0.35 = 0.10$, so $n = 16 \times (0.4)(0\ 6)/(0.01) = 384$ in each treatment group. (Numbers generated by this relation are slightly less than those projected by standard computer routines for a Fisher exact test, and slightly more than those determined for an uncorrected chi-square test.)

4 How many patients per group do we need to detect a difference of 0.20 in the proportions of success for two treatments?

Although p_1 and p_2 are not given, the product pq has an upper bound of 0.25, so a conservative estimate is n = 16(0.25)/(0.04), or 100 per treatment group. The farther from 0.5 that $(p_1 + p_2)/2$ lies, the smaller is the required sample size.

The number of equations that scientists have committed to memory either intentionally or because of repeated use is most likely minimal. Standard, more general and accurate sample-size equations such as

$$n = \frac{2\sigma^2 \, [Z_{1-\alpha} + Z_{1-\beta}]^2}{[\mu_1 - \mu_2]^2}$$

for t-tests, or

$$n = \frac{\left[c_{\alpha/2} \sqrt{2PQ} - c_{1-\beta} \sqrt{P_1Q_1 + P_2Q_2} \right]^2}{[P_2 - P_1]^2}$$

(where Z_x and c_x both represent inverse normal values) for chi-square tests are not remembered by most, and even if they are, one needs an inverse normal table to accompany them. The utility of $16s^2/d^2$ together with its simplicity warrants its memorization by pharmaceutical statisticians. The fact that the same relation applies both to trials with continuous measures and to trials with resulting dichotomous variables adds to its appeal. The primary cautions or reminders to convey along with the equation are that (1) it is a rough approximation and (2) it only applies to the specific power and alpha values of 0.80 and 0.05 for two-tailed tests.

Table I

n determined to provide 80 per cent power computed using $16s^2 / d^2$	Power computed for given value of *n* using non-central t distribution
6	0.72
10	0.76
16	0.78
>20	0.79 - 0.81

Table II. Coefficients to substitute for '16' for different size and power specifications

Power	Alpha level		
	0.01	0.05	0.10
0.80	23.5	16	12.5
0.90	30	21	17.5
0.95	36	26	22

ACKNOWLEDGEMENTS

The author is grateful for the suggestions of Dr. Harji Patel, the referees, and the editors of Statistics in Medicine.

REFERENCES

1. Fleiss J. L. *Statistical Methods for Rates and Proportions*, Wiley, 1981.

2. Snedecor, G. W. and Cochran W. G. *Statistical Methods*, Iowa University Press, 1980.

3. Woolson, R. F. *Statistical Methods for the Analysis of Biomedical Data, Wiley*, 1987.