Error in sample size formula

Sir,

Re: Suresh K, Chandrashekara S. Sample size estimation and power analysis for clinical research studies. J Hum Reprod Sci 2012;5:7-13.

Although informative and useful Suresh and Chandrashekara's article on sample size estimation and power analysis contains a serious error (Suresh and Chandrashekara, 2012). In the section titled "sample size estimation with two means" they state the minimum required sample size for detecting a mean difference between two groups is:

$$N=\frac{(r+1)\left(Z_{\alpha/2}+Z_{1-\beta}\right)^2\sigma^2}{rd^2}$$

Where

α is the false positive rate

β is the false negative rate

N is the sample size required to detect an inter-group mean difference of d with specified α and power of $1-\beta$

σ^2 is the variance in each group (both groups having the same variance)

r is the ratio of size (n_1 and n_2) of the two groups, that is, $r = n_1/n_2$

Z is the standard normal distribution deviate, note this is the absolute of the z-score, as in (Suresh and Chandrashekara, 2012) Tables 2 and 3.

The formula as stated cannot be correct as relabeling of the two groups results in different values of N.

Example:

If $n_1 = 100$ and $n_2 = 200$ then $r = 1/2$ and $\frac{(r+1)}{r} = 3$

Swapping the two groups around:

Then $n_1 = 200$ and $n_2 = 100$ then $r = 2$ and $\frac{(r+1)}{r} = 1.5$

The N calculated for the first case is twice that of the second; they should be identical.

In fact, the formula given is the formula for n_2, which I prove thus.

In the 'Sample Size estimation with two means' case, the z-score of the test statistic d is related to the required false positive rate and power by[1]

$$\frac{d}{\text{stdErr}(d)} = Z_{\alpha/2} + Z_{1-\beta}$$

Where the standard error of d is

$$\text{stdErr}(d) = \sqrt{\frac{\sigma^2}{n_1}+\frac{\sigma^2}{n_2}} = \sqrt{\frac{\sigma^2}{rn_2}+\frac{\sigma^2}{n_2}} = \sqrt{\frac{(1+r)\sigma^2}{rn_2}}$$

Substituting the expression for stdErr (d) into the first equation and rearranging gives:

$$n_2 = \frac{(r+1)\left(Z_{\alpha/2}+Z_{1-\beta}\right)^2\sigma^2}{rd^2}$$

The formula for N is then

$$N = n_1 + n_2 = rn_2 + n_2 = (r+1)n_2$$

$$N = \frac{(r+1)^2}{r}\frac{\left(Z_{\alpha/2}+Z_{1-\beta}\right)^2\sigma^2}{d^2}$$

With the new formula group labels can be swapped without changing the value calculate for N.

The original erroneous formula could result in studies seriously underestimating their required sample size. For instance, the required sample size (as calculated by the current formula) is half that truly required, given equal numbers in the two groups. I therefore draw this error to your attention. The illustrative

examples that follow the formula presentation are also in error.

Desmond Dedalus Campbell[1,2]
[1]Department of Psychiatry,
[2]Centre for Genomic Sciences,
University of Hong Kong,
Hong Kong, China

Address for correspondence:
Dr. Desmond Dedalus Campbell,
L10-64 Lab Block, Faculty of Medicine,
21 Sassoon Road, Central and Western,
Hong Kong, China.
E-mail: ddc123@hku.hk

REFERENCE

1. Wittes J. Sample size calculations for randomized controlled trials. Epidemiol Rev 2002;24:39-53.

Access this article online	
Quick Response Code:	Website: www.jhrsonline.org
	DOI: 10.4103/0974-1208.138878