In the last decade, guidelines for the presentation of statistical results in medical journals have emphasized confidence intervals (CIs) as an adjunct to, or even a replacement for, statistical tests and p values. Because of the intimate links between the 2 concepts, authors now use statements like “the 95% CI overlaps 0” where they would formerly have stated “the difference is not statistically significant at the 5% level.” Although this interchangeability is technically correct in 1-sample situations, it does not carry over fully to comparisons involving 2 samples. A frequently encountered misconception is that if 2 independent 95% CIs overlap each other, as they do in Fig. 1, then a statistical test of the difference will not be statistically significant at the 5% level.

Why is this not necessarily so? Consider the means in 2 independent groups, meanA and meanB, with for simplicity meanA being the smaller of the 2. The 95% CI for the mean in group A is approximately given by meanA plus or minus twice the standard error of the mean for that group, SEA, and correspondingly for group B. A mathematical check for whether these CIs overlap is given by adding the distance 2SEA (from meanA to the upper bound of the CI) to 2SEB and comparing this sum with the distance between the 2 means, that is, meanB minus meanA (Fig. 2). The CIs overlap when

$$\text{mean}_B - \text{mean}_A < 2\, \text{SE}_A + 2\, \text{SE}_B$$  

But overlapping confidence intervals do not demonstrate that group means are not statistically significantly different from each other. In a 2-sample t-test to compare 2 means, significance is attained at the 0.05 level if the t statistic exceeds the critical value of about 2, which occurs when the difference between the means exceeds twice its standard error, namely, if

$$\text{mean}_B - \text{mean}_A > 2\sqrt{\text{SE}_A^2 + \text{SE}_B^2}$$

This standard error reflects the fact that the standard error of a difference involves summing the standard error of each estimate, but doing so by “adding in quadrature,” for example,

$$1 + 1 = \sqrt{1^2 + 1^2} = 1.414$$

Thus, to evaluate the overlap of 2 95% CIs and to determine whether at the same time the difference between the means is not statistically significant at the 5% level, one must first check that the confidence intervals overlap (as in Fig. 1), and then check that the difference between the means exceeds 2 × standard error. This is illustrated in Fig. 2, where the 2 means are not statistically significantly different, but the confidence intervals almost overlap.
means is significant at the 0.05 level, the following rough rule can be used:

\[ 2 \sqrt{(\text{SEA}^2 + \text{SEB}^2)} < \text{mean}_B - \text{mean}_A < 2 \text{SEA} + 2 \text{SEB} \]

If \(\text{SEA}\) and \(\text{SEB}\) are equal, the condition is as follows:

\[ 2.83 \text{SE} < \text{mean}_B - \text{mean}_A < 4 \text{SE} \]

When one SE is 25% larger than the other, the boundaries are 3.2 and 4.5 times the smaller SE. As the lower boundary remains close to 3, Moses\(^1\) was prompted to display group means with error bars that were 1.5 SE around the mean in order to have a “by eye” test of significance between the 2 group means while presenting the information in the 2 groups separately.

---

**Reference**


**Correspondence to:** Dr. Rory Wolfe, Department of Epidemiology and Preventive Medicine, Central and Eastern Clinical School, Monash University and the Alfred Hospital, Commercial Rd., Prahran, Victoria 3183, Australia; fax 61 3 9903 0556; rory.wolfe@med.monash.edu.au

---

**BOO K S RECEIVED**


