The ongoing discussion regarding standard deviation and standard error

Bartholomew Kay
Department of Sport and Exercise Science, University of Auckland, Auckland, New Zealand

As a nonstatistician, I have watched the discussion on this topic with interest (1–8). Dr. Curren-Everett and others make a clear statement that variability of results should always be reported using the standard deviation (SD) and not the standard error (SE). Others cited above variously agree and disagree with the guidelines published by Dr. Curren-Everett and others. The reasoning provided in support of the guideline is that the SE is an “inappropriate estimate of variability amongst observations.” Indeed, the SE is determined by dividing the SD by the square root of the number of observations, meaning that the larger the number of observations, the lower the SE becomes. The intended use for the SE is in providing the SD of means that would be returned if the researcher repeatedly collected their number of observations (n) with a new cohort selected randomly from the population concerned. It also allows the calculation of confidence intervals and the like.

I would like to add to the debate by stating the following. First, the statistic reported in any given paper should be used to make the correct inference. If the researcher wants to make assertions about the variability of his/her particular results about the mean of those results, or the difference between means in that cohort, then the SD is the appropriate statistic. If the researcher wants to make assertions about the likely variability of means that could be collected from a random cohort derived from the population concerned, or likely value of any difference between such hypothetical means that may be collected in future, then the SE is absolutely the correct statistic upon which to rely for that inference. Second, readers of published research are generally interested in the applicability of results to the population concerned, and not about the particular results obtained by the researchers in that particular cohort; therefore, I believe researchers should absolutely report SEs and discuss their relevance with respect to the population, and not the cohort concerned. Finally, the SD of results from the particular cohort concerned in published literature can be derived by simply multiplying the SE by the square root of n. It would be strange indeed to see published research that does not report the value of n. It seems to me that the important issue here is not which statistic to report but rather that the correct inference is made. Again, when I read published research, I am interested in the general applicability/value of the results to the population concerned. I care exactly zero about the results obtained from those particular subjects. If I really want to know how variable results were about the mean in the cohort concerned, I can do some pretty simple math to derive the SD where it is not reported.

By way of a suggestion, it might be deemed appropriate to say the following: “from our cohort of n, the mean of condition 1 was x and the mean of condition 2 was y. The SD of results about x was SDx and the SD about y was SDy. The SE of the difference between x and y was the absolute value of SE of the difference (x − y), with P = exact P value. This suggests that if we repeated our study again and again with a new random sample, the likelihood of the means being statistically different from one another on each occasion is . . . .” This is as opposed to the following: “our condition 1 was statistically different from our condition 2,” about which most educated readers would care very little; after all, it is only a small sample derived from the population about which most educated readers are concerned in any case.

REFERENCES

Address for reprint requests and other correspondence: B. Kay, Dept. of Sport and Exercise Science, Univ. of Auckland, Private Bag 92019, Auckland Mail Centre, Auckland 1142, New Zealand (e-mail: b.kay@auckland.ac.nz).