Why the 'new statistics' isn't new

Smith and Morris ('Building confidence in confidence intervals', June 2015) and Cumming (2014) encourage psychologists to move away from null hypothesis significance testing and to report effect sizes and confidence intervals instead without invoking significance (i.e. the 'new' statistics). We offer three points of contention.

First, confidence intervals (CIs) are based on null hypothesis significance testing (NHST). Calculating a CI as suggested by Smith and Morris in *The Psychologist* and by Cumming (2014) does not in any way differ from NHST. Indeed, the CI is obtained using the same underlying statistical method as the p value. For example, consider that the lower and upper bounds of the 95 per cent confidence interval are simply a function of the standard error of the estimate multiplied by the Z-score (i.e. the critical value for significance or 'alpha'). Changing the alpha level (e.g. from 5% to 1%) will change the Z-score (from 1.96 to 2.575) and hence the value range of the CI. Thus, the CI is directly based on a null hypothesis significance test. Using the CI will not help psychologists move away from this practice. In fact, researchers often treat the CI as a p value simply by judging whether or not the interval includes a '0' (Hoekstra et al., 2014).

Second, the true definition of a confidence interval is not very intuitive or very practical. Contrary to popular opinion, the values contained within the CI aren't necessarily that interesting. Why? Because they don't mean what most psychologists think they mean. For example, the 95 per cent CI does *not* mean that '95 per cent of the sample data lie within the interval' (a CI is not a range of plausible values for the sample mean). A CI also does *not* mean that 'there is a 95 per cent probability that the true population parameter lies within the interval'. In fact, the 95 per cent probability does not refer to the interval itself but rather to the reliability/confidence in the method that is used to obtain the interval. In other words, the CI does not refer to a specific sample but rather to the notion of repeated sampling (method).

A correct definition: *If this procedure were repeated on many samples (e.g. conduct an experiment over and over again), and a hypothetical confidence interval would be calculated for each individual sample, then 95 per cent of the time, the CI would contain the true population parameter.*

Note: this means that we have no idea if the CI calculated for a specific sample actually contains the true population value! The hypothetical nature of the interpretation of the CI makes it not very intuitive to meaningfully use or apply (hence the widespread misinterpretation). Also note that although it sounds like they might, exotic procedures such as 'bootstrapping' or 'jack-knifing' do not substantively change the interpretation of the CI.

Third, the real new statistics: beyond the confidence interval. We are not suggesting that reporting effect sizes and confidence intervals is bad practice. On the contrary, confidence intervals and effect sizes often contain more detailed and useful information than p values alone. However, there are other ways to construct confidence intervals that do not rely on NHST. One example is the 'credibility interval', which is based on Bayesian procedures and allows a researcher to actually express the obtained interval in terms of a real probability. The larger point is that if we honestly want psychologists to adopt a 'new statistics', we need to build confidence (not in) but outside the confidence interval.

Sander van der Linden
Department of Psychology, Princeton University
Breanne Chryst
Department of Statistics, Yale University

References
Cumming, G. (2014). The new statistics: Why and how. *Psychological Science, 25*, 7–29.
Hoekstra, R., Morey, R.D., Rouder, J.N. & Wagenmakers, E.J. (2014). Robust misinterpretation of confidence intervals. *Psychonomic Bulletin & Review, 21*(5), 1157–1164.