In defence of inferential statistics

In February Basic and Applied Social Psychology (an American Psychological Association journal) announced that it was banning the reporting of null hypothesis significance testing procedures (NHSTP) and confidence intervals (CI) (Trafimow & Marks, 2015). We are writing to express our hope that the journals published by the British Psychological Society will not be lured into similarly banning CIs and distancing psychology from medical research in which CIs are routinely employed. We believe that CIs offer an as yet undeveloped but potentially very valuable tool for psychologists to interpret their data (see e.g. Smith & Morris, in press). Any ban that involves throwing out the CI baby with the NHSTP bathwater should be avoided.

Trafimow and Marks (2015) condemn CIs because, they say, 'A 95% confidence interval does not indicate that the parameter of interest has a 95% probability of being within the interval. Rather, it means merely that if an infinite number of samples were taken and confidence intervals computed, 95% of the confidence intervals would capture the population parameter.'

It is true, as Cumming (2012) points out (p.79), that a 95 per cent confidence interval refers to the whole process of taking a sample and calculating a CI, 95 per cent of which will capture the population mean. However, it follows that the 95 per cent CIs that you calculate will most likely capture the population parameter.

The great value of CIs is that they provide valuable probabilistic information about the true location of the population mean. NHSTP deals with the normally uninteresting null hypothesis: the probability of the data if the difference or relationship is zero, or some other specific value. CIs help us conceptualise the plausible locations of the parameter (e.g. population mean or effect size), and the variability or precision of that estimate. As Smith and Morris (in press) point out, when we know both an effect size and its CI we can make a much more useful interpretation of the results of our research than when we have an effect size alone. We know of no alternative to standard errors in some form, such as CIs, for describing the likely variability in our effect size if we repeat our research. Given the relatively small sample sizes of much psychology research, the CIs of the effect sizes can be disconcertingly large and remind researchers that a simple effect size, or other point estimate, can suggest a precision that is not justified. Failure to report this variability does not make it go away but does expose those following up the research to dangers of misinterpretation.

Trafimow and Marks's (2015) solution to the banning of NHSTP and CIs is to require bigger sample sizes and the reporting of descriptive statistics with frequency and distributional data. In general, such information is welcome. However, the reason for the original development of NHSTP was that it is always necessary to decide whether or not to act in the future as if a real effect is likely. CIs of effect sizes give good guidance to such decisions, but it is not clear upon what evidence these fundamental decisions will be based if CIs and NHSTP are banned.

Another issue with demands for larger samples is that psychology researchers are inevitably faced with limitations through cost and time upon the number of participants that they can test. Resources devoted to doubling sample sizes for one study are not then available for new research questions. If the original sample size was, in fact, sufficient, there is a serious ethical and practical question of whether an unnecessary increase in sample sizes will do more harm than good to the future of psychology. How will one decide if the sample is large enough? Given that the purpose of larger samples is to increase the precision of the estimates, reporting that precision should be required, rather than forbidden. Until there are alternative and generally accepted means of answering the question 'Could the effects have arisen by chance?', we recommend reporting CIs and, where researchers find them helpful, NHSTP.

Peter Morris, Catherine Fritz, Graham Smith, Amar Cherchar, Robin Crockett, Chris Roe, Roz

Collings, Kimberley Hill, David Saunders, Martin Anderson and Lucy Atkinson University of Northampton