Statcheck scans full-text articles for statistical errors and publishes the results on *pubpeer*

(Statement released by the Executive Board of the German Psychological Society)

October 20, 2016

Since the end of August, 2016, fifty thousand articles published in psychological journals were scanned for statistical errors by *statcheck*, a program developed at the University of Tilburg (The Netherlands). *Statcheck* is freely downloadable as a stand-alone program and as an App. In a nutshell, the algorithm implemented in *statcheck* tests whether statistical information reported in an article (e.g., *p*-values according to conventional null-hypothesis significance testing, degrees of freedom, etc.) is correct and plausible. The results of this screening are then posted on *pubpeer*, a platform allowing an interactive exchange about published scientific articles.

The developers of *statcheck* have repeatedly stressed that their intention is not to shame individual authors by flagging their papers, "...but to help them avoid making these mistakes in the future" (see, e.g., the interview on Vox). Nevertheless, many researchers – especially those whose papers are among the 50.000 that were automatically screened – are worried about the fact that the screening of their article occurred (1) without the authors' awareness, (2) without being able to actually verify whether the results of this screening are actually correct, and (3) without the opportunity to comment on the screening of their paper before the results were published on *pubpeer*. In addition, many colleagues are deeply concerned about the fact that it is obviously difficult to remove an entry on *pubpeer* after an error that had been "detected" by *statcheck* turned out to be a false positive.

The German Psychological Society shares these concerns. Of course, programs such as *statcheck* can, in principle, help improve the quality of reported research – if they yield valid results. And statistical errors do occur in articles, as prior research has shown (see Nuijten et al., 2015). Authors could therefore use *statcheck* to scan their paper for such errors before they submit it to a journal.

However, the detection of an alleged error necessarily requires a high level of sensitivity and cooperative intentions among all parties. Before a paper is publicly flagged for alleged statistical errors (on *pubpeer* or elsewhere), the authors of this paper should be given the opportunity to double-check and comment on the results of the screening. If an alleged error then turns out to be a false positive, any posts or comments in which the articles is flagged need to be removed or revoked at once.

The necessity to double-check the results obtained by *statcheck* has recently been demonstrated by Thomas Schmidt, professor of experimental psychology at the Technical University Kaiserslautern (Germany). Schmidt noted that *statcheck* is unable to deal with "corrected" *p*-values (which could, for instance, result from correcting degrees of freedom of the respective test statistic or from correcting for type-1 error inflations). The developers of *statcheck* are well aware of these problems and stress that the algorithm is still "work in progress". But the problem is: once an alleged error has been "detected" and the respective paper has been publicly flagged, the damage to the researcher's reputation has been done and is no longer controllable. Again, reports about flagged papers can obviously not be revoked on *pubpeer* even if the alleged error eventually turns out to be a false positive.

The German Psychological Society supports any attempt to improve the quality of psychological research as long as they yield valid results, but is concerned about the automatic publication of alleged errors without double-checking with the original authors. As long as it is unclear how many false positives (and how many false negatives) *statcheck* actually produces, the algorithm should not be used, neither in scientific articles nor (especially not) in public postings on *pubpeer*. Moreover, false positives should immediately be removed from the platform.

References:

Epskamp, S. & Nuijten, M. B. (2016). *Statcheck: Extract statistics from articles and recompute p values.* [Online Dokument]. URL: http://CRAN.R-project.org/package=statcheck (R package version 1.2.2)

Nuijten, M. B., Hartgerink, C. H. J., van Assen, M. A. L. M., Epskamp, S., & Wicherts, J. M. (2015). The prevalence of statistical reporting errors in psychology (1985-2013). *Behavior Research Methods.* [Online First Publication]. DOI: 10.3758/s13428-015-0664-2