Response to Statement by the Executive Board of the German Psychological Society on the Use of statcheck

October 25, 2016

Dear colleagues,

Thank you for sending us your statement concerning the detection of inconsistencies in statistical results reported in psychological articles. We agree with you that we need to promote accurate reporting of statistical results and have developed statcheck (Epskamp & Nuijten, 2014) to assist in this effort. Specifically, statcheck extracts statistics reported in the APA style from documents, and checks whether the reported *p*-value is consistent with the reported test statistic and degrees of freedom.

As we understand it, your main critique is twofold. Firstly, you are worried that the sensitivity and specificity of statcheck are unknown, and secondly, that the statcheck reports uploaded on PubPeer contain false positives that can lead to reputational damage. Your advice is to refrain from using statcheck in scientific articles or in public postings as on PubPeer.

Before we respond to your critique concerning the diagnostic value of statcheck, we would like to clarify some aspects of the PubPeer project. First, the PubPeer project is an independent project by Chris Hartgerink (Hartgerink, 2016) that is unrelated to the paper of Nuijten et al. (in press). Second, we would like to stress that statcheck does *not* automatically upload anything to PubPeer, as the first paragraph of your letter seems to imply. Third, we see value in openly discussing inconsistencies in published articles in an impersonal and factual manner, given that our own experiences in corresponding directly with authors about errors have not led to any documented corrections (Bakker & Wicherts, 2011; Wicherts et al., 2011). Fourth, please note that anyone, including the original author, is allowed to comment on PubPeer reports if they wish to do so. Also, we are not in a position to disallow anyone from using statcheck before or after publication to detect inconsistencies in articles.

The specificity and sensitivity of statcheck were investigated in detail in Nuijten et al. (in press) by comparing its results to manual checks (Wicherts et al., 2011). We clearly noted statcheck's shortcomings in our publications. We continue to further refine statcheck and investigate the influence of possible bugs or other problems on our estimates of the prevalence of inconsistencies in psychology (see e.g., Nuijten, 2016). We therefore welcome all researchers' comments on the performance of statcheck. So far, no bugs have been found that noticeably affect estimates of inconsistencies in statistical results in psychology. Hence we see no reason to adapt our initial estimates, or to discourage using statcheck in scientific articles, given that researchers take into account the program's limitations.

We as scientists have the obligation to correct reporting errors even if the tools we use are not 100% accurate. Actually we are not aware of any diagnostic tool that has this property, but that should not discourage us from using them to improve psychological science.

Kind regards,
Michèle Nuijten, Chris Hartgerink, Marcel van Assen, Sacha Epskamp, & Jelte Wicherts

References

Bakker, M., & Wicherts, J. M. (2011). The (mis)reporting of statistical results in psychology journals. *Behavior Research Methods, 43*, 666-678. doi: 10.3758/s13428-011-0089-5

Epskamp, S. & Nuijten, M. B. (2014). statcheck: Extract statistics from articles and recompute p values. Retrieved from http://CRAN.R-project.org/package=statcheck. (R package version 1.0.1)

Hartgerink, C. H. J. (2016). 688,112 Statistical results: Content mining psychology articles for statistical test results. *Data*, 1(3), 14. DOI: 10.3390/data1030014

Nuijten, M. B. (2016). Error prevalence without t2 and F2: Extra analysis for Nuijten et al. (in press). Retrieved from http://rpubs.com/michelenuijten/ExtraAnalysisT2F2.

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

Wicherts, J. M., Bakker, M., & Molenaar, D. (2011). Willingness to share research data is related to the strength of the evidence and the quality of reporting of statistical results. *PLoS One, 6*(11), e26828.