



UvA-DARE (Digital Academic Repository)

Retire significance, but still test hypotheses

Haaf, J.M.; Ly, A.; Wagenmakers, E.-J.

DOI

[10.1038/d41586-019-00972-7](https://doi.org/10.1038/d41586-019-00972-7)

Publication date

2019

Document Version

Final published version

Published in

Nature

[Link to publication](#)

Citation for published version (APA):

Haaf, J. M., Ly, A., & Wagenmakers, E.-J. (2019). Retire significance, but still test hypotheses. *Nature*, 567(7749), 461. <https://doi.org/10.1038/d41586-019-00972-7>

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

Retiring significance: keep hypothesis tests

We agree that arbitrarily branding experimental findings as significant or non-significant generates a false sense of certainty (V. Amrhein *et al. Nature* 567, 305–307; 2019). However, when done properly, hypothesis testing is an important precondition for estimating an effect size.

In his 1928 book *Statistical Methods for Research Workers*, British statistician Ronald Fisher remarked that “it is a useful preliminary before making a statistical estimate ... to test if there is anything to justify estimation at all”. And British polymath Harold Jeffreys declared in *Theory of Probability* in 1939 that “variation must be taken as random until there is positive evidence to the contrary”. Hence, testing and estimation are complementary. Testing establishes whether there is an effect, and that helps to determine whether or not the magnitude needs to be estimated.

What happens when statistical testing is skipped and the null hypothesis is ignored? Well, noise would be interpreted as structural, and any differences between observations would be considered meaningful. Parameters would need to be estimated for all these differences, resulting in a “mere catalogue” of data “without any summaries at all”, as Jeffreys put it.

Without the restraint provided by testing, an estimation-only approach will lead to overfitting of research results, poor predictions and overconfident claims.

Julia M. Haaf, Alexander Ly, Eric-Jan Wagenmakers
University of Amsterdam, The Netherlands.
j.m.haaf@uva.nl