Tidsskriftet Tidsskrift for Den norske legeforening

What proportion of published research findings are false?

MEDICINE AND NUMBERS

STIAN LYDERSEN

stian.lydersen@ntnu.no
Stian Lydersen, PhD, professor of medical statistics at the Regional Centre for Child and Youth Mental Health and Child Welfare, Department of Mental Health, Norwegian University of Science and Technology.
The author has completed the ICMJE form and declares no conflicts of interest.

METTE LANGAAS

Mette Langaas, PhD, professor of statistics at the Department of Mathematical Sciences, Norwegian University of Science and Technology, and professor II at the Department of Statistical Analysis, Machine Learning and Image Analysis (SAMBA) at the Norwegian Computing Center.
The author has completed the ICMJE form and declares no conflicts of interest.

Not all published research findings are reproducible — some because the findings are incorrect. What is the extent of the problem?

A number of researchers have attempted to estimate how often published findings are false. They have used widely different approaches.

Different methods

The article 'Why most published research findings are false' by John Ioannidis attracted considerable attention when it was published in 2005 (1). The article was not based on data, but postulated a model for the proportion of false positive findings among published positive findings based on the following four quantities: the proportion of actually true hypotheses of all the hypotheses tested, statistical power, significance level (5 %) and bias. In this context, bias means the proportion of the studies in which the hypothesis would appear to be true although it is not, for example because of publication bias or poor study design. Ioannidis estimated the positive predictive value, i.e. the proportion of true findings among all positive findings, for a series of different combinations of these four quantities. In large-scale randomised controlled trials with adequate power (80 %) he considered it to be realistic that the proportion of true null hypotheses could be 50 % and that the bias was only 10 %. This gives an estimate of positive predictive value of 85 %. For exploratory observational studies with an adequate power of 80 %, a proportion of true null hypotheses of 9 % and a bias of 30 % we obtain a positive predictive value of 20 %. Studies with a lower proportion of true null hypotheses or less power result in an even lower positive predictive value ((1), Table 4).

In 2014, Jager and Leek estimated the proportion of false positive findings based on data (2). They read electronically all 77 430 publications in The Lancet, Journal of the American Medical Association, New England Journal of Medicine, BMJ and American Journal of Epidemiology from 2000, 2005 and 2010. The analyses were based on the fact that when the null hypotheses are true, the *p*-values will be evenly distributed between 0 and 1, but when the alternative hypotheses are true, the *p*-values will be skewed towards 0. This is illustrated in Figure 1.

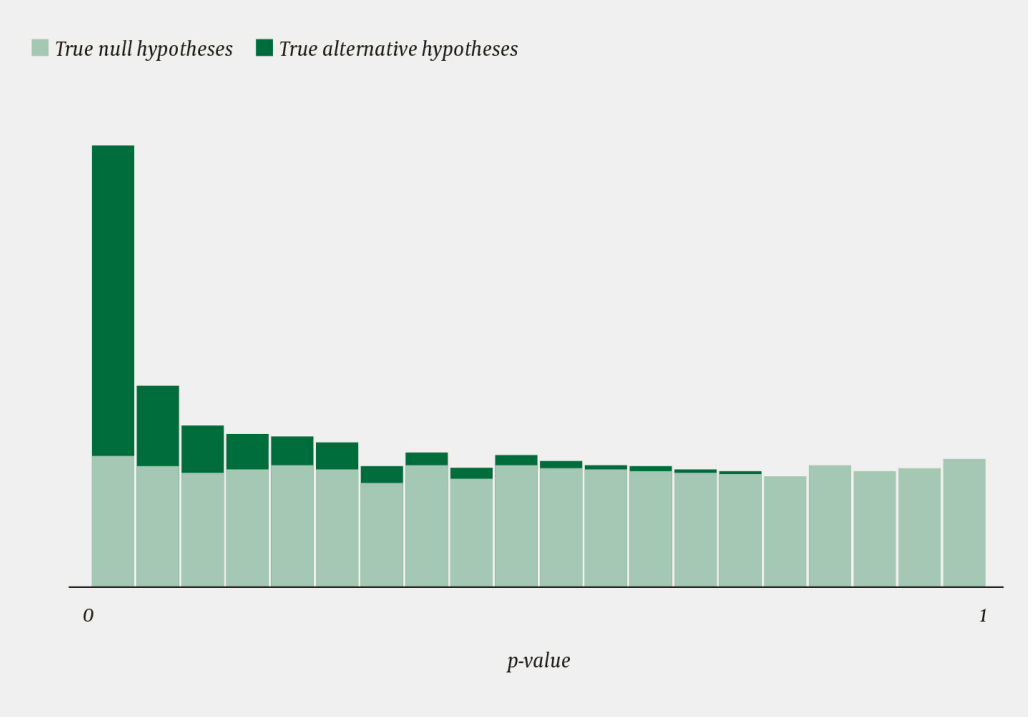

Figure 1 Typical distribution of p-values for true null hypotheses (light green) and true alternative hypotheses (dark green).

In Jager and Leek's estimate, the science-wise false discovery rate was 14 %. Their article was accompanied by discussion articles from a number of researchers. This concluded with a rejoinder from Jager and Leek (3), who wrote that the estimate of 14 % was probably optimistic, but that the rate was unlikely to exceed 50 %, at least for studies that were well planned and executed.

Researchers in the Open Science Collaboration group used another procedure to study reproducibility (4). They identified 100 studies published in three different psychology journals in 2008. These studies were replicated in new studies with new participants and a design that was as similar to the original as possible, with a planned statistical power of at least 80 %. This was a very comprehensive piece of work, and a total of 274 authors were listed. So what did they find? In the original studies, the estimated effect measured by the correlation coefficient was 0.403 on average (standard deviation 0.188), and in the replicated studies only 0.197 (0.257). Of the original studies, altogether 97 % reported a statistically significant effect (p-value < 0.05), compared to only 36 % of the replicated studies. After combining the original and the replicated studies, 68 % were statistically significant.

An admission of failure?

These three studies relied on very different methodologies. Ioannidis produced a model which was based on a number of assumptions in various study designs. The assumptions may appear realistic, but it is a weakness that they to not build on data. Jager and Leek relied on reported p-values and made estimates based on expected distributions of p-values when the null hypotheses are true or false. The Open Science Collaboration were even more thorough; they replicated 100 studies. The studies undertaken by Jager and Leek and Open Society Collaboration are based on comprehensive empirical material. Depending on how the results are emphasised, we can say that the false discovery rate was estimated at 14 % and 29 % (97 % – 68 % = 29 %) respectively in these studies. These rates must in any case be considered as high.

Is this an admission of failure for research? Absolutely not. It does, however, shed light on the importance of careful planning, implementation and reporting of studies, as well as of seeking to replicate published studies. The combined evidence from multiple studies, preferably in a systematic review or a meta-analysis, will be substantially more reliable than that of a single study.

LITERATURE

1. Ioannidis JPA. Why most published research findings are false. PLoS Med 2005; 2: e124. [PubMed][CrossRef]

2. Jager LR, Leek JT. An estimate of the science-wise false discovery rate and application to the top medical literature. Biostatistics 2014; 15: 1–12. [PubMed][CrossRef]

3. Jager LR, Leek JT. Rejoinder: An estimate of the science-wise false discovery rate and application to the top medical literature. Biostatistics 2014; 15: 39–45. [PubMed][CrossRef]

4. Open Science Collaboration. Estimating the reproducibility of psychological science. Science 2015; 349: aac4716. [PubMed][CrossRef]

Publisert: 8. December 2021. Tidsskr Nor Legeforen. DOI: 10.4045/tidsskr.21.0703
Copyright: © Tidsskriftet 2026 Downloaded from tidsskriftet.no 5 July 2026.