

Reproducibility of scientific results

In the article *Threads of A Replication Crisis In Imperial Computer Science*ⁱ, Andy Cockburn et al. Analyzed the top ten literary theories downloaded from 41 journals and found that 15% of them used statistical significance as evidence, and proposed that computer science was indeed in replication Crisis and describes how the crisis entered the field of computer science. After that, he came up with a solution.

Andy Cockburn noted that computer science faces the problem of low repeatability, including the following points:

- Publication bias: The acceptance rates of papers that supported their hypothesis were much higher than those that did not.

- File drawer effect: Null discoveries are often not made public and therefore hidden from the scientific community.

- P-hacking: researchers intentionally design experimental methods and statistical methods to get their desired conclusions.

- p-fishing: researcher look for statistically significant effects beyond the original hypothesis.

It can be seen that many published studies contain intentional or unintentional dishonesty. From data processing to experimental models to inductive conclusions, there is a lot of room for fraud. If there is no guarantee of repeatability, these published results will lose credibility. It is also mentioned in the article *Why Most Published Research Findings Are False ?*ⁱⁱ that most scientific research achievements, including medical research, are biased to a certain extent. That is to say, in order to pursue the publication of papers, researchers sometimes publish misleading or useless work results and ignore valuable research work. The risks and rewards of academic career, the structure and habits of peer-reviewed journals, and the way universities and research institutions work will affect scientists' scientific research work, interfere with their topics, working methods, and final results.

As mentioned in the previous paragraph, only 15% of the results used statistical significance, of which all but two met the criteria of $P < 0.05$. This suggests that computer science is also affected by publication bias: findings that support this hypothesis are more likely to be accepted for publication than those that do not.

In order to make more research results more authentic, we should use more credible and effective methods in the field of computer science. These include large-scale collaborative research, repeatable, replicable, and reproducible respectively culture, sharing, better statistical methods, standardized definitions and analysis, more appropriate (and often more stringent) statistical thresholds, and improvements in research design standards, peer reviews, reporting and science communication, and training of researchers.

Generally speaking, computer science is divided into theoretical computer science and applied computer science. In the former, research does not rely on experience and experimental methods, but mainly relies on logic and mathematical derivation. However, in the latter, computer science, like all other sciences, is facing the problem of replication crisis. The computer science community needs to strengthen and select appropriate intervention measures, take scientific research practice more seriously, and adopt more stringent inspection measures and experimental tests when all feasible.

Word count: 484

Reference

- ⁱ Cockburn A. et. al. (2020) Threats of a Replication Crisis in Empirical Computer Science. ACM doi:10.1145/3360311
- ⁱⁱ Ioannidis J (2005) Why Most Published Research Findings are False, PloS Medicine. 2(8): e124.