

## 8: Everybody Makes Mistakes

Until now, I have presumed that scientists are capable of making statistical computations with perfect accuracy, and only err in their choice of appropriate numbers to compute. Scientists may misuse the results of statistical tests or fail to make relevant computations, but they can at least calculate a  $p$  value, right?

Perhaps not.

Surveys of statistically significant results reported in medical and psychological trials suggest that many  $p$  values are wrong, and some statistically insignificant results are actually significant when computed correctly.<sup>25, 2</sup> Other reviews find examples of misclassified data, erroneous duplication of data, inclusion of the wrong dataset entirely, and other mixups, all concealed by papers which did not describe their analysis in enough detail for the errors to be easily noticed.<sup>1, 26</sup>

Sunshine is the best disinfectant, and many scientists have called for experimental data to be made available through the Internet. In some fields, this is now commonplace: there exist gene sequencing databases, protein structure databanks, astronomical observation databases, and earth observation collections containing the contributions of thousands of scientists. Many other fields, however, can't share their data due to impracticality (particle physics data can include many terabytes of information), privacy issues (in medical trials), a lack of funding or technological support, or just a desire to keep proprietary control of the data and all the discoveries which result from it. And even if the data were all available, would anyone analyze it all to spot errors?

Similarly, scientists in some fields have pushed towards making their statistical analyses available through clever technological tools. A tool called Sweave, for instance, makes it easy to embed statistical analyses performed using the popular R programming language inside papers written in LaTeX, the standard for scientific and mathematical publications. The result looks just like any scientific paper, but another scientist reading the paper and curious about its methods can download the source code, which shows exactly how all the numbers were calculated. But would scientists avail themselves of the opportunity? Nobody gets scientific glory by checking code for typos.

Another solution might be replication. If scientists carefully recreate the experiments of other scientists and validate their results, it is much easier to rule out the possibility of a typo causing an errant result. Replication also weeds out fluke false positives. Many scientists claim that experimental replication is the heart of science: no new idea is accepted until it has been independently tested and retested around the world and found to hold water.

That's not entirely true; scientists often take previous studies for granted, though occasionally scientists decide to systematically re-test earlier works. One new project, for example, aims to reproduce papers in major psychology journals to determine just how many papers hold up over time – and what attributes of a paper predict how likely it is to stand up to retesting.<sup>[1]</sup> In another example, cancer researchers at Amgen retested 53 landmark preclinical studies in cancer research. (By “preclinical” I mean the studies did not involve human patients, as they were testing new and unproven ideas.) Despite working in collaboration with the authors of the original papers, the Amgen researchers could only reproduce six of the studies.<sup>5</sup> Bayer researchers have reported similar difficulties when testing potential new drugs found in published papers.<sup>49</sup>

This is worrisome. Does the trend hold true for less speculative kinds of medical research? Apparently so: of the top-cited research articles in medicine, a quarter have gone untested after their publication, and a third have been found to be exaggerated or wrong by later research.<sup>32</sup> That's not as extreme as the Amgen result, but it makes you wonder what important errors still lurk unnoticed in important research. Replication is not as prevalent as we would like it to be, and the results are not always favorable.

### Footnotes

[1] The Reproducibility Project, at <http://openscienceframework.org/reproducibility/>

This page titled 8: Everybody Makes Mistakes is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Alex Reinhart via [source content](#) that was edited to the style and standards of the LibreTexts platform.