Rigor and Transparency Index, a new metric of quality for assessing biological and medical science methods.

Joe Menke1, Martijn Roelandse2, Burak Ozyurt3, Maryann Martone4, Anita Bandrowski5

Affiliations:

1 Center for Research in Biological Systems, UCSD; SciCrunch Inc https://orcid.org/0000-0002-7374-7549

2 Independent consultant at martijnroelandse.dev, http://orcid.org/0000-0001-5596-8084

3 Department of Neuroscience, UCSD http://orcid.org/0000-0003-3944-1893

4 Department of Neuroscience, UCSD; SciCrunch Inc https://orcid.org/0000-0002-8406-3871

5 Department of Neuroscience, UCSD; SciCrunch Inc https://orcid.org/0000-0002-5497-0243

Abstract

The reproducibility crisis in science is a multifaceted problem involving practices and incentives, both in the laboratory and in publication. Fortunately, some of the root causes are known and can be addressed by scientists and authors alike. After careful consideration of the available literature, the National Institutes of Health identified several key problems with the way that scientists conduct and report their research and introduced guidelines to improve the rigor and reproducibility of pre-clinical studies. Many journals have implemented policies addressing these same criteria. We currently have, however, no comprehensive data on how these guidelines are impacting the reporting of research. Using SciScore, an automated tool developed to review the methods sections of manuscripts for the presence of criteria associated with the NIH and other reporting guidelines, e.g., ARRIVE, RRIDs, we have analyzed ~1.6 million PubMed Central papers to determine the degree to which articles were addressing these criteria. The tool scores each paper on a ten point scale identifying sentences that are associated with compliance with criteria associated with increased rigor (5 pts) and those associated with key resource identification and authentication (5 pts). From these data, we have built the Rigor and Transparency Index, which is the average score for analyzed papers in a particular journal. Our analyses show that the average score over all journals has increased since 1997, but remains below five, indicating that less than half of the rigor and reproducibility criteria are routinely addressed by authors. To analyze the data further, we examined the prevalence of individual criteria across the literature, e.g., the reporting of a subject's sex (21-37% of studies between 1997 and 2019), the inclusion of sample size calculations (2-10%), whether the study addressed blinding (3-9%), or the identifiability of key biological resources such as antibodies (11-43%), transgenic organisms (14-22%), and cell lines (33-39%). The greatest increase in prevalence for rigor criteria was seen in the use of randomization of subjects (10-30%), while software tool identifiability improved the most among key resource types (42-87%). We further analyzed individual journals over time that had implemented specific author guidelines covering rigor criteria, and found that in some journals, they had a big impact, whereas in others they did not. We speculate that unless they are enforced, author guidelines alone do little to improve the number of criteria addressed by authors. Our Rigor and Transparency Index did not correlate with the impact factors of journals.