The R-Factor: A Measure of Scientific Veracity

Joshua Nicholson[1] and Yuri Lazebnik[2]

[1]Department of Biological Sciences, Virginia Tech, Blacksburg, VA 24061
[2]Yale University Cardiovascular Research Center, New Haven, CT 06511

April 17, 2023

Abstract

Scientists, institutions and journals have been increasingly evaluated statistically, by metrics that focus on the number of published reports rather than on their content, raising a concern that this approach interferes with the progress of biomedical research. To offset this effect, we propose to use the R-factor, a metric that indicates whether a report or its conclusions have been verified.

SCIENCE AND SOCIETY

The R-Factor: A Measure of Scientific Veracity

JOSHUA NICHOLSON [1], YURI LAZEBNIK [2]

1. Department of Biological Sciences, Virginia Tech, Blacksburg, VA 24061
2. Yale University Cardiovascular Research Center, New Haven, CT 06511

ABSTRACT

Scientists, institutions and journals have been increasingly evaluated statistically, by metrics that focus on the number of published reports rather than on their content, raising a concern that this approach interferes with the progress of biomedical research. To offset this effect, we propose to use the R-factor, a metric that indicates whether a report or its conclusions have been verified.

The ability of academic scientists to keep their job, be promoted, or receive funding has become increasingly dependent on three statistical parameters: the number of their publications, how often these publications have been cited, and the impact factor of the journals in which these publications appeared (Abbott et al. 2010, Hall 2012, Van Noorden 2010, Sahel 2011). The reliance on these parameters varies among countries and institutions (Abbott et al. 2010, Hall 2012, Sahel 2011, Van Noorden 2010), but the administrative convenience of the statistical approach suggests that it will continue to spread (Abbott et al. 2010, Van Noorden 2010). A growing concern is that this approach interferes with the progress of biomedical research by forcing publication prematurely, before the veracity of the findings has been verified (Abbott et al. 2010, Fang, Steen, and Casadevall 2012, Lawrence 2007, Ioannidis 2005b, Ioannidis 2005a, Young, Ioannidis, and Al-Ubaydli 2008). As a result, the number of reports that are irreproducible and thus potentially misleading, especially to non-experts in the field, has grown sufficiently large (Begley and Ellis 2012, Ioannidis 2005b, Ioannidis 2005a) to call for action to solve this problem (Couzin-Frankel 2012, https://www.scienceexchange.com/reproducibility , http://openscienceframework.org/project/EZcUj/wiki/home).

A systemic solution would be to offset the parameters that encourage publication with a parameter (s) that evaluate what is reported. Currently, this function is served by the citation index of a report, and the impact factor of the journal in which this report appeared. However, the citation index can be misleading, if only because it increases even if the report is cited as being irreproducible or wrong (Lawrence 2007). The utility of the impact factors, which are average citation indexes for the papers published by the journals over the last two years, has also been questioned, especially if used as a tool to evaluate individual scientists (Lawrence 2007, Sahel 2011, Editorial 2013).

We propose to use a measure termed the R-factor, which would indicate how many studies attempted to verify a given article - that is to determine whether the results can be reproduced or the main conclusions confirmed-and what was the outcome. A newly published article would have the R-factor of 0. If another article finds that the experiments described in the article can be repeated with similar results, and/or the main conclusions or predictions are correct, then the R-factor becomes 1. If either of these conditions are not met, the R-factor would be 0. As more studies attempt to verify the article, The R-factor would change to a value between 0 and 1. For example, if ten studies attempt to verify a

report and all successfully do so, its R-factor would be 1 (10/10). If two of them fail, the R-factor would be 0.8 (8/10) and if all find it irreproducible, then the R-factor would be 0 (0/10). The number of studies used to calculate the R-factor would be indicated in brackets next to it, such as 0.8 (10). The R-factor is applicable to any report that makes a testable conclusion, whether the study is experimental or theoretical and would not punish the authors that conducted rigorous research but made wrong interpretations, nor the authors who made right conclusions for a wrong reason. The R-factor of scientists, institutions, or journals would be the average of the R-factors of the papers they have published.

We suggest that by giving an explicit numerical value to the veracity of scientific reports the R-factor would make biomedical research more rigorous and efficient, and its results and conclusions more accessible and transparent outside of a specific research field. For example, the need to explain a low R-factor at the next evaluation would make a scientist think twice before publishing a study that calls for further verification. Having an R-factor assigned to each publication would bring the discussion about the veracity of studies from the grapevine to the public view and for the public benefit. An outsider to a field could use the R-factor as a guide to focus on more reliable publications without the need to seek the opinions of the insiders. The possibility of receiving an R-factor of 0 (n) could be used as a deterrent against an overly enthusiastic colleague or advisor who pushes for publishing the results before they are verified. Science journals would also be more attentive to the content of manuscripts to avoid hurting their R-factor, while individuals and institutions would pride themselves on the quality of their research by citing the R-factor along with their citations indexes.

Our optimistic view raises three practical questions: How feasible is it to determine the R-factors, who would do that and keep the scores, and would the R-factor cause more harm than good?

In theory, since the R-factor is a simple ratio of publications that confirm or disprove the report in question, calculating it should be relatively straightforward for an expert in the research field. It would require obtaining the citation index of the report, determining which of the citing articles attempted to verify the results and how many of them were successful. Some experts would not even need to resort to the citation index, as they know the published and unpublished history of their field by heart. In practice, the ease of determining whether a study is verifiable would be true for some articles, but not the others, as it has been outlined in detail by a previous proposal to introduce a metric for evaluating reproducibility of scientific publications (Hartshorne and Schachner 2012). The ease would depend on whether the experimental procedures are described in sufficient detail to reproduce them, whether the conclusions are formulated explicitly enough to be verifiable, whether the experimental setting can be recapitulated without required expertise (Bissell 2013) and at reasonable expense, and whether the results of verification are published, which is often not the case. We suggest that the incentive to increase their R-factor would encourage scientists to describe the experimental conditions in sufficient detail and to formulate their conclusions unambiguously. The use of the R-factor in evaluating scientists and institutions would encourage authors and editors to publish reports that attempt to verify previous studies.

Who would calculate the R-factor and keep the scores? The R-factor can be calculated by individual scientists, scientific societies, bibliometric companies, such as Elsevier and Thomson Reuters, reproducibility initiatives (Couzin-Frankel 2012, https://www.scienceexchange.com/reproducibility , http://opensciencesframework.org/project/EZcUj/wiki/home) and evaluation committees. The variety of potential sources implies the need to aggregate the resulting R-factors in an accessible way, as it is currently done with citation indexes. This function can be fulfilled by an open-access resource with the required expertise (Hartshorne and Schachner 2012). For example, the NCBI, which have expertise in analyzing and annotating scientific reports can include the R-factor as a field for the papers referenced in Pubmed. A natural solution would also be to link the R-factor to the citation indexes. Introducing three types of citations - positive, if the cited report is verified, negative, if it is not, and neutral, if the report is mentioned without evaluation, which would make the citation index more meaningful and would allow the R-factor of a report to be computed in real time. We feel that once the R-factor enters

the public domain, the opportunities to keep the scores and use them would evolve beyond what we can now envision.

One concern is whether using the R-factor would do more harm than good, for example, by preventing reports of unorthodox ideas, by being used as a tool to undermine someone's reputation, or by maligning the studies after failing to reproduce them for the lack of expertise. We feel that the transparency of calculating the R-factor - the papers that will be used to calculate the R-factor are all in the public domain - would make using it for non-scientific purposes difficult. As for the new ideas, the R-factor would help a non-expert to distinguish hypotheses and ideas that have been confirmed from those that are presented or accepted as established facts without sufficient verification. We understand at the same time that science is a human activity, meaning that the R-factor can be misused as the case with other apparently benign tools, including the citation indexes and impact factors.

We hope, however, that introducing an explicit and quantitative measure that focuses on the veracity of scientific reports and the validity of their conclusions would offset at many levels - from the bench to the editorial board - the push to publish no matter what and thus would accelerate progress in biomedical research. We invite the scientific community and the institutions that evaluate the scientific literature to give the R-factor a try.

ACKNOWLEDGEMENTS

REFERENCES

Abbott, A., D. Cyranoski, N. Jones, B. Maher, Q. Schiermeier, and R. Van Noorden. 2010\n. "Metrics: Do metrics matter?" Nature no. 465 (7300):860-2. doi: 10.1038/465860a.

Begley, C. G., and L. M. Ellis. 2012. "Drug development: Raise standards for preclinic\nal cancer research." Nature no. 483 (7391):531-3. doi: 10.1038/483531a.

Bissell, M. 2013. "Reproducibility: The risks of the replication drive." Nature no. 50\n3 (7476):333-4.

Couzin-Frankel, J. 2012. "Research quality. Service offers to reproduce results for a \nfee." Science no. 337 (6098):1031. doi: 10.1126/science.337.6098.1031.

Editorial. 2013. "Beware the impact factor." *Nat Mater* no. 12 (2):89-89.

Fang, F. C., R. G. Steen, and A. Casadevall. 2012. "Misconduct accounts for the majori\nty of retracted scientific publications." Proc Natl Acad Sci U S A no. 109 (42):17028-33. doi: 10.1073/pnas.1212247109.

Hall, N. 2012. "Why science and synchronized swimming should not be Olympic sports." G\nenome Biol no. 13 (9):171. doi: 10.1186/gb4045.

Hartshorne, J. K., and A. Schachner. 2012. "Tracking replicability as a method of post\n-publication open evaluation." Front Comput Neurosci no. 6:8. doi: 10.3389/fncom.2012.00008.

http://openscienceframework.org/project/EZcUj/wiki/home. *Open Science Framework Reproducibility Project*.

http://www.scienceexchange.com/reproducibility. *Science Exchange Reproducibility Initiative*.

Ioannidis, J. P. 2005a. "Contradicted and initially stronger effects in highly cited c\nlinical research." JAMA no. 294 (2):218-28. doi: 10.1001/jama.294.2.218.

Ioannidis, John P. A. 2005b. "Why Most Published Research Findings Are False." PLoS Me\nd no. 2 (8):e124. doi: 10.1371/journal.pmed.0020124

Lawrence, P. A. 2007. "The mismeasurement of science." *Curr Biol* no. 17 (15):R583-5. doi: 10.1016/j.cub.2007.06.014.

Sahel, J. A. 2011. "Quality versus quantity: assessing individual research performance\n." Sci Transl Med no. 3 (84):84cm13. doi: 10.1126/scitranslmed.3002249.

Van Noorden, R. 2010. "Metrics: A profusion of measures." Nature no. 465 (7300):864-6\n. doi: 10.1038/465864a.

Young, Neal S., John P. A. Ioannidis, and Omar Al-Ubaydli. 2008. "Why Current Publicat\nion Practices May Distort Science." PLoS Med no. 5 (10):e201. doi: 10.1371/journal.pmed.0050201.