

# Bite-Size Science and Its Undesired Side Effects

Marco Bertamini<sup>1</sup> and Marcus R. Munafò<sup>2</sup>

<sup>1</sup>University of Liverpool and <sup>2</sup>University of Bristol

## Abstract

Short and rapid publication of research findings has many advantages. However, there is another side of the coin that needs careful consideration. We argue that the most dangerous aspect of a shift toward “bite-size” publishing is the relationship between study size and publication bias. Findings based on a single study or a study based on a limited sample size are more likely to be false positive, because the false positive rate remains constant, whereas the true positive rate (the power) declines as sample size declines. Pressure on productivity and on novelty value further exacerbates the problem.

## Keywords

allied field, behavioral, cognition, perception, publishing

“Delay is preferable to error”

Thomas Jefferson

## “The Culture Demands Shorter Reports”

In journals from all areas of psychology, there is a trend toward short and rapid publications. A recent example is a new type of article in the *Journal of Experimental Psychology: Human Perception and Performance* called “observations” (2,000 words maximum). In the words of the incoming Editor, James Enns, “The culture demands shorter reports and rapid publication” (Price, 2011, p. 68).

Those in the field have claimed that there are many advantages for short reports: faster communication of results; ease of assimilation; ease of access for people outside the field; ease of processing for editors and reviewers; and more dynamic exchange of fresh ideas, even if they may turn out to be wrong. In addition to these advantages, there are practical considerations affecting this trend. The main one is an increased pressure on researchers to produce quantifiable output. Despite an ostensible emphasis on quality, quantity is still important, as it contributes to résumé length, exposure, name recognition, and summary statistics of productivity; after all, the units used in most citation analyses are individual articles. In their article, Ledgerwood and Sherman (2012, this issue) have provided a table of pros and cons of the short format, and we share some of their concerns.

In this article, we first address the issue of whether article impact should be adjusted for number of pages. We then

discuss some of the more general problems of the short article format. After that, we turn to the link between study size and publication bias, which is, in our opinion, the single most serious source of danger. Finally, in our conclusions, we suggest that evaluating journal impact in terms of citation should be balanced with some measure of publication bias and that replication of unexpected findings should be encouraged during the editorial process.

## Haslam’s Citation Analysis

Some empirical evidence seems to suggest that the trend toward short articles is beneficial. In particular, Haslam (2010) has shown that, when adjusted for length, short articles are cited more frequently than other articles. Based on this, Haslam concluded that “short articles appear to be somewhat more efficient in generating scientific influence” and that “journals [...] might be well advised to add a short-article format” (p. 264).

Short articles are often, though not always, single study articles and tend to include smaller samples. This single-study model is broadly consistent with the neuroscience model advocated also by Park (2009). Park also touches on forcing people to be concise and on an increased use of the Internet to store data and stimuli. These are separate issues, but the idea that “bite-size” publishing results in greater impact is more problematic.

## Corresponding Author:

Marco Bertamini, University of Liverpool, Bedford Street South, School of Psychology, Liverpool, L697ZA, UK  
E-mail: M.Bertamini@liv.ac.uk

There are some technical problems with a citation impact adjusted for length, and it should not be taken as a bona fide superior measure of impact. If the same findings can be written in either a short or long format, and assuming that the two articles would get cited equally, the impact per page would be higher for the short article, but it would be misleading to say that the short article has achieved any greater impact than the long one. Moreover, suppose I conduct two studies providing converging evidence for the same conclusion and I can publish them in one long article or in two short articles. My colleagues A. Friend and A. Foe always cite all my work because it is relevant for what they do. They will cite either one long or two short articles in all their publications. Based on their citations, each of the three articles would have the same impact, but on a per-page measure, the shorter articles are more influential. This would be purely because of how we measure impact, not because of a difference in influence.

As an example of a short but highly influential article, Haslam (2010) cites Watson and Crick's (1953) report on the structure of DNA. However, that was a theoretical piece proposing a specific model. The debate about format in psychology is focused on empirical reports, and different considerations may apply in the absence of data. This issue could be explored by looking at impact for nonempirical articles. However, the three journals analyzed by Haslam (*Psychological Science*, *Cognition*, *Journal of Experimental Social Psychology*) publish mainly empirical findings and, therefore, do not lend themselves to this analysis.

### Problems Associated With Short Reports

Part of the argument for the short format is that, other things being equal, shorter articles are easier and quicker to read. But assuming a fixed amount of studies are carried out, shorter articles may mean more articles. If a journal were to switch to the short-article format while keeping the yearly number of articles fixed, then that might shift articles onto other journals, and onto new journals. The growth in the number of peer-reviewed articles and in the number of scientific journals over time is well documented (e.g., Larsen & von Ins, 2010; Neff & Olden, 2010). The increase in articles creates more work for the editors, for the reviewers, and perhaps most importantly, for anyone researching the topic. The change to the short format may be meant to help reviewers, but it may backfire as there could be more requests to review (Huth, 1986). The same can be said about the way short articles speed up processing of a manuscript by allowing editors to decide quickly about acceptance or rejection. If rejected papers are resubmitted after minor changes to a different journal, and if we consider the community of reviewers as a whole, then this quick turnaround strategy will not lighten their workload.

As authors, we may all get mildly annoyed at reviewers asking for an extra experiment or just more evidence to be convinced (Ploegh, 2011). On the other hand, this kind of back and forth correspondence between peers has a long and

illustrious tradition in the history of science. Often an extra experiment does change the conclusions of a paper, therefore avoiding the publication of a partial or distorted result.

If replication is a cornerstone of the scientific method, then an advantage of multiexperiment papers is that replication is inherent and usually rather stringently defined (because the experiments come from the same lab). This latter point is important—replication is no “silver bullet,” and loose definitions of replication can result in initial false positive findings being propagated in the literature (Sullivan, 2007).

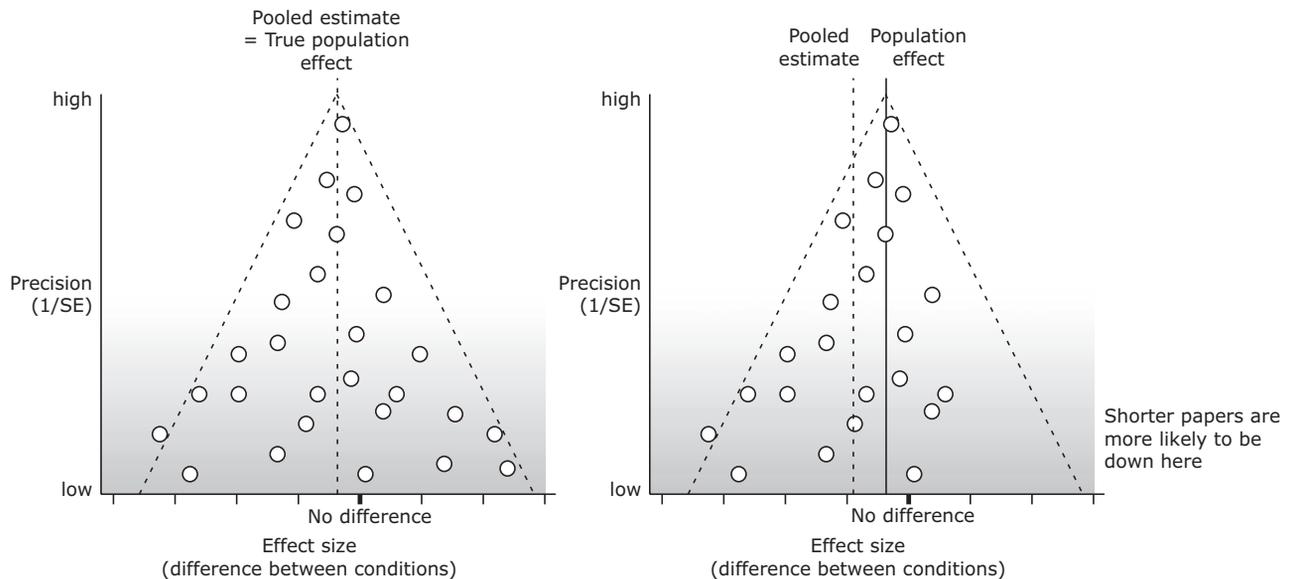
Short articles may also be more prone to what has been described as “citation amnesia” (Garfield, 1982; Robinson & Goodman, 2011), especially when the authors have to meet tight word count criteria. A finding is bound to sound more newsworthy when the discussion of previous relevant work is less detailed, therefore, there is pressure on authors not to go into great depth when researching and discussing previous work. A bit of ignorance helps in discovering “new” things.

### Publication Bias

We want to turn to what, in our opinion, is a crucial issue of the implications of short-article publishing: the link between study size and publication bias.

This bias is nicely captured by a funnel plot (Fig. 1), which has become an important tool in biomedical research (Sterne & Egger, 2001). A short article tends to contain less data than a long article. The smaller dataset (in particular, small sample size) implies more variability in the results, and this is the variability that is shown in a funnel plot: Precision (which is usually measured as the inverse of the standard error and is directly related to study size) is on the  $y$  axis and effect size is on the  $x$  axis. Larger studies will be more accurate and will show a narrow range of estimates around the true population effect. Smaller, less accurate studies will give a wider distribution of estimates, although if all are published, these should still center on the true population effect. Therefore, this funnel should be symmetrical if there is no bias operating. Often it is not. The asymmetry implies a disproportionately large number of published effects in one direction. These deviations from the true effect are concentrated at the bottom of the funnel because it is more likely to obtain large deviations from small studies (because of the law of large numbers). Small studies that generate either null results or results that are the opposite of what the author(s) predicted will be systematically censored (i.e., there will be publication bias), and this will be reflected in an asymmetrical funnel. Formal regression-based tests exist to assess this (Egger, Smith, Schneider, & Minder, 1997).

A more subtle problem is that results in small studies are more likely to be false positives if they are statistically significant. This is because the false positive rate remains constant at 5%, whereas the true positive rate (the power) depends on sample size. Therefore, as power/sample size decreases, the ratio of true positives to false positives among those studies that achieve statistical significance also decreases (Sterne &



**Fig. 1.** A funnel plot is a scatterplot where each datapoint represents a single published study, with effect size plotted against precision (typically the inverse of the individual study standard error). Effect size may be a difference ( $d$ ), a correlation coefficient ( $r$ ), or other metric such as an odds ratio. The effect size may reflect either a positive or negative association, on either side of the point where there is no difference. The dashed vertical line shows the pooled effect size estimate when all studies are combined (e.g., in a meta-analysis). In the example shown, the pooled effect size estimate is not null, and therefore the line is to the left of the location where the difference is zero. More precise studies (e.g., larger) give a narrower distribution of effect size estimates, while less precise studies (e.g., smaller) give a wider distribution of estimates, with both centered on the true population effect. The theoretical distribution of effect sizes, therefore, should be a symmetrical funnel, centered on the true population effect (left panel). The pooled estimate will differ from the population effect if it is affected by publication bias (right panel). Some small studies that show either no effect or an effect opposite to that anticipated are missing, and the shape of the funnel is distorted. The presence of this asymmetry can be tested formally using regression-based methods (Egger et al., 1997).

Smith, 2001). We know that studies in psychology are endemically underpowered (Sedlmeier & Gigerenzer, 1989), and this, therefore, increases the risk of false positives. A culture of brief, single-study reports based on small datasets will inevitably lead to even greater contamination of the literature by false positive findings.

False positives are a problem because “erroneous ideas much more easily enter the literature than leave it” (McManus, 2002, p. 322). (McManus was writing about the myths that left-handers die younger and suffer more from immune disorders.)

A publication bias may originate from several sources: dishonesty is probably the least likely. The main issue is how journals lead to selective publication: significant effects and, in particular, results that are perceived as “newsworthy” are more likely to be submitted and more likely to be accepted for publication (Dickersin, Chan, Chalmers, Sacks, & Smith, 1987). That is, extreme values have greater novelty value. Use of multiple analyses on the same dataset and selective reporting of statistics can also produce an inflated effect (Nieuwenhuis, Forstmann, & Wagenmakers, 2011).

The existence of publication biases can lead to an artificially inflated sense of the strength (or even presence) of an effect from a casual inspection of the literature. Moreover, the extent to which an individual study overestimates any underlying true effect correlates with the impact factor of a journal (Munafò, Stothart, & Flint, 2009). This may seem astonishing and hard to explain, but one should consider that high impact

factor journals do seek novel and ground-breaking findings and that they often publish short articles. In the words of Munafò et al. (2009), “We also note that journals with high impact factors tend to publish studies with high bias scores and small sample sizes” (p. 120). Pressure to achieve “impact” may also be stronger in some countries. There is a stronger tendency for individual studies to overestimate the true effect in articles from authors based in North America than in articles from authors based in Europe; this also seems to correlate with the amount of government R&D funding in individual countries (Munafò, Attwood, & Flint, 2008).

One of the authors has received correspondence from *Psychological Science* saying that they hope to find manuscripts that “report new discoveries that will make our readership sit up and take notice.” These, a priori, are more likely to be false positives (Ioannidis, 2005). Flukes tend to meet this criterion by their nature (they are surprising and different from what other people tend to find). In part this is unavoidable, but what is relevant here is that bite-size articles make this problem worse. We are all aware of the need for results to be replicated. Long articles with multiple experiments show whether an effect can be replicated and supported by converging evidence.

Using simulations, Ioannidis (2005) has demonstrated a number of interesting aspects of the problem. For example, he found that the smaller the studies; the “hotter” a scientific field; and the greater the flexibility in designs, definitions, outcomes, and analyses used, the less likely the published research findings are true. He concluded that “statistical significance

testing in the report of *a single study* gives only a partial picture, without knowing how much testing has been done outside the report and in the relevant field at large” (p. 701, italics ours). Indeed, there are far more statistically significant findings present in most literatures than we would expect even if the effects reported are indeed real (Ioannidis, 2011).

We assumed that short articles include less data than long articles. However, one could simply force more data into a smaller space. It is not controversial to say that articles should be concise. To support our assumption of a correlation between sample size and article length, we looked at all articles published in *Cognition* in 2007 (65 regular articles and 35 brief articles). The individuals in the samples were adults, infants, or animals of other species.

The total sample size was, on average, 122 for regular articles and 95 for brief articles. Medians were 80 and 52, respectively. Of the 35 brief articles, 20 reported a single study. The total sample size was, on average, 80 for the single-study brief articles and 102 for the others. Therefore, the link between length and amount of data seems to exist. Brief articles in *Cognition* often include multiple studies, and at more than 11 pages on average, they would probably be full articles in some other journal. What we call brief articles may, therefore, be relative. Nevertheless the correlation between article length and study size seems to exist.

## Conclusions

By far the most popular and influential measure of quality for journals is their Impact Factor. This is computed on the basis of citations, and therefore should reflect influence in the field. As critics have often pointed out, it does not distinguish between citations that confirm and extend the original findings and citations that criticize and debunk them. In this sense, journals have little formal incentive to minimize bias in the effects reported and to minimize false alarms, although we have no doubt that every good editor is trying to do that. Including a measure of publication bias in the evaluation of a journal would provide a more direct incentive. By combining different measures, such as number of articles, Impact Factor, and publication bias, we could arrive at better measures of the quality of a journal in reporting interesting but replicable and valid results. At the moment, there are sophisticated tools to count articles published and number of citations. However, it is much more difficult to measure bias and the frequency of false alarms. Systematic and large-scale meta-analyses are the standard tools, but they will not be enough without some increase in published replication studies, and they are not feasible if there are insufficient comparable studies on a particular topic. Unfortunately, the culture of mere replication (and publication of failures to replicate) is not well established in psychology. Although psychology is not unique, a few other disciplines fare better. These are important differences between fields, because we can estimate false alarms (imperfectly) after a field has matured to a necessary degree, but not before.

Despite increased interest in bibliometrics, there is also a growing consciousness of the limitations of any individual index, especially because as soon as one index achieves the status of the measure of choice, with practical implications, authors and institutions will start to adapt and play the system. In the United Kingdom, after an initial intention to rely heavily on metrics for the research exercise that compares all Departments in the country (the Research Assessment Exercise now renamed Research Excellence Framework), the project has been revised to focus on quality and not only on quantity ([www.hefce.ac.uk/research/ref/](http://www.hefce.ac.uk/research/ref/)). Each author submits only four publications for evaluation by a panel of peers, which partially removes the incentive to seek quantity at the (potential) expense of quality.

In summary, the shift toward shorter articles including fewer studies, the reduced space for locating results in the context of previous findings, and the pressure on productivity and novelty all combine to create a potentially damaging effect on the scientific literature. It is interesting that in fields such as genetic association studies, in which the false-positives problem is well-recognized, independent replication in the same study is now a requirement for publication in many high-quality journals (Munafò, 2010). Perhaps this is an option that should be at least considered in psychological science.

## Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

## References

- Dickersin, K., Chan, S., Chalmers, T., Sacks, H., & Smith, H., Jr. (1987). Publication bias and clinical trials. *Controlled Clinical Trials*, 8, 343–353. doi:10.1016/0197-2456(87)90155-3
- Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, 315, 629–634.
- Garfield, E. (1982). More on the ethics of scientific publication: Abuses of authorship attribution and citation amnesia undermine the reward system of science. *Current Contents*, 30, 5–10.
- Haslam, N. (2010). Bite-size science: Relative impact of short article formats. *Perspectives on Psychological Science*, 5, 263–264. doi:10.1177/1745691610369466
- Huth, E. J. (1986). Irresponsible authorship and wasteful publication. *Annals of Internal Medicine*, 104, 257–259.
- Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine*, 2, e124.
- Ioannidis, J. P. A. (2011). Excess significance bias in the literature on brain volume abnormalities. *Archives of General Psychiatry*, 68, 773–801. doi:10.1001/archgenpsychiatry.2011.28
- Larsen, P. O., & von Ins, M. (2010). The rate of growth in scientific publication and the decline in coverage provided by Science Citation Index. *Scientometrics*, 84, 575–603. doi:10.1007/s11192-010-0202-z
- Ledgerwood, A., & Sherman, J. W. (2012). Short, sweet, and problematic? The rise of the short report in psychological science. *Perspectives on Psychological Science*, 7, xxx–xxx.

- McManus, I. C. (2002). *Right hand, left hand*. London, England: Weidenfeld & Nicolson.
- Munafò, M. R. (2010). Credible genetic associations? *International Journal of Molecular Epidemiology and Genetics, 1*, 31–34.
- Munafò, M. R., Attwood, A. S., & Flint, J. (2008). Relationship between geographical location of research group and effect size estimate. *Psychological Medicine, 38*, 1213–1214. doi:10.1017/S003329170800353X
- Munafò, M. R., Stohart, G., & Flint, J. (2009). Bias in genetic association studies and impact factor. *Molecular Psychiatry, 14*, 119–120. doi:10.1038/mp.2008.77
- Neff, B. D., & Olden, J. D. (2010). Not so fast: Inflation in impact factors contributes to apparent improvements in journal quality. *BioScience, 60*, 455. doi:10.1525/bio.2010.60.6.9
- Nieuwenhuis, S., Forstmann, B. U., & Wagenmakers, E. J. (2011). Erroneous analyses of interactions in neuroscience: A problem of significance. *Nature Neuroscience, 14*, 1105–1107. doi:10.1038/nn.2886
- Park, D. C. (2009). Publishing in the psychological sciences. *Perspectives on Psychological Science, 4*, 36–37. doi:10.1111/j.1745-6924.2009.01103.x
- Ploegh, H. (2011). End the wasteful tyranny of reviewer experiments. *Nature, 472*, 391.
- Price, M. (2011). Restoring glory. *Monitor, 42*, 68. doi:10.1038/472391a
- Robinson, K. A., & Goodman, S. N. (2011). A systematic examination of the citation of prior research in reports of randomized, controlled trials. *Annals of Internal Medicine, 154*, 50–55.
- Sedlmeier, P., & Gigerenzer, G. (1989). Do studies of statistical power have an effect on the power of studies? *Psychological Bulletin, 105*, 309–316. doi:10.1037/0033-2909.105.2.309
- Sterne, J. A. C., & Egger, M. (2001). Funnel plots for detecting bias in meta-analysis: Guidelines on choice of axis. *Journal of Clinical Epidemiology, 54*, 1046–1104.
- Sterne, J. A. C., & Smith, G. D. (2001). Sifting the evidence—What’s wrong with significance tests? *British Medical Journal, 322*, 226–231. doi:10.1136/bmj.322.7280.226
- Sullivan, P. F. (2007). Spurious genetic associations. *Biological Psychiatry, 61*, 1121–1126. doi:10.1016/j.biopsych.2006.11.010
- Watson, J. D., & Crick, F. H. C. (1953). Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. *Nature, 171*, 737–738.