

# Publication bias: What are the challenges and can they be overcome?

Ridha Joober, MD, PhD; Norbert Schmitz, PhD; Lawrence Annable, Dipstat;  
Patricia Boksa, PhD

Joober, Boksa — Douglas Mental Health University Institute and Department of Psychiatry, McGill University, Montréal, Que.;  
Schmitz, Annable — Department of Psychiatry, McGill University Health Centre, Montréal, Que.

Appearances to the mind are of four kinds.  
Things either are what they appear to be;  
Or they neither are, nor appear to be;  
Or they are, and do not appear to be;  
Or they are not, and yet appear to be.  
Rightly to aim in all these cases  
Is the wise man's task.  
Epictetus, 2nd century AD

In the last few years, several meta-analyses<sup>1-4</sup> have reappraised the efficacy and safety of antidepressants and concluded that the therapeutic value of these drugs may have been significantly overestimated (see Ioannidis<sup>5</sup>). In some instances, the authors of these meta-analyses resorted to the United States' Freedom of Information Act to obtain unpublished data that, when included in meta-analyses with previously published data, reduced significantly the purported therapeutic value of selective serotonin reuptake inhibitors.<sup>6</sup>

In the case of clinical trials, withholding negative results from publication — publication bias — could have major consequences for the health of millions. In preclinical and experimental research, this bias may seriously distort the literature, drain scarce resources by undertaking research in futile quests, and lead to misguided research and teaching practices. Over and above scientific considerations, research participants consent to participate in research on the understanding that they are contributing to advances in treatment and scientific knowledge. Our ethical duty as researchers and editors is to honour this engagement and publish both positive and negative outcomes in an equitable manner. Animals do not give consent, but the research community is ethically bound to make the best use of the results, which is not the case when negative results are not publicized.

In large part, it is the highly competitive environment for funding and career promotion that incites researchers to submit predominantly positive results for publication, knowing

that they are more likely to be considered for publication by editors, more favourably reviewed by peers and, once published, more likely to be cited. For editors, it is the competition for citation index and the financial survival of journals that makes it more attractive to publish positive findings.

Although publication bias has been documented in the literature for decades and its origins and consequences debated extensively, there is evidence suggesting that this bias is increasing. A recent investigation covering more than 4600 publications from different countries and disciplines found strong evidence for a steady and significant increase in publication bias over the years. The frequency of papers declaring significant statistical support for their a priori formulated hypotheses increased by 22% between 1990 and 2007 ( $n = 4656$ ,  $p < 0.001$ ). Psychology and psychiatry are among the disciplines in which this increase is highest ( $p < 0.001$ ).<sup>7</sup> A case in point is the field of biomedical research in autism-spectrum disorder (ASD), which suggests that in some areas negative results are completely absent; among 4 fields of research that emerged in the last 10 years (immune dysregulation/inflammation, oxidative stress, mitochondrial dysfunction and toxicant exposure<sup>8</sup>), more than 89% of 437 studies reported a significant association between ASD and 1 or more parameters investigated, with 100% of 115 studies on oxidative stress reporting positive results.

It might be argued that this very high proportion of positive findings and almost complete absence of negative results reflect improvements in formulating and testing hypotheses. It is, for example, reasonable to suspect that strong competition for funding and publishing would filter out poorly formulated and poorly tested studies, which are more likely to yield negative results. However, the poor replicability of research results, an endemic problem in biomedical research,<sup>9</sup> does not support this hypothesis. Further, it has been shown that studies emerging from countries where there is greater

**Correspondence to:** R. Joober, Douglas Hospital Research Centre, Frank B. Common Pavilion, 6875 blvd. LaSalle, Verdun QC H4H 1R3; ridha.joober@douglas.mcgill.ca

*J Psychiatry Neurosci* 2012;37(3):149-52.

DOI: 10.1503/jpn.120065

competition for funding tend to overestimate the true effect sizes (as estimated from adequate meta-analyses) compared with studies emerging from countries with less competition for funding.<sup>10</sup> Similarly, a highly significant correlation ( $R^2 = 0.13$ ,  $p < 0.001$ ) between impact factor and overestimation of effect sizes<sup>11</sup> has been reported.

Publication bias has an escalating and damaging effect on the integrity of knowledge. The research process usually starts by conjecturing a relationship between an independent variable and a dependent variable, and the purpose of hypothesis testing is to determine how the belief in the relationship will change compared with its a priori credibility. This a priori credibility is constructed on the basis of an analysis of the available literature. When published, the results of this testing should contribute to an unbiased update of the credibility of the relationship. In the presence of publication bias, belief in the relationship increases artificially and iteratively with each positive publication. This, in turn, diminishes the credibility of hypothesis testing because it is based on biased information, and calls into question the integrity of the entire experimental framework. The above analysis is based on relatively simple statistical reasoning that will be explained in the next few paragraphs.

The statistic most often used to support the post hoc credibility of the relationship is the probability of observing a relationship when in fact there isn't one:  $\Pr(T^+/\text{no R})$ . This probability is called type I error and represents the false-positive rate. If this probability is low, this will often, but not necessarily, strengthen the belief in the relationship that was hypothesized. Another important statistic that needs to be considered in evaluating the credibility of the relationship is statistical power (i.e., the probability of declaring a relationship as true when indeed it is). It is equivalent to the sensitivity of the test and is the complement ( $1-\beta$ ) of type II error ( $\beta$ ), or the false-negative rate. The statistical power depends on the real strength of the tested relationship, the sample size used in the experiment and type I error that is set by the researcher to declare the observed results unlikely to be due to sampling variation (significance level). The most critical element in this process is the a priori credibility of the relationship. Both the a priori probability and statistical power encapsulate to a large extent the care and skills of the researcher to formulate hypotheses and to design the best possible experiment to test them within the limits imposed by feasibility. If the power or the a priori credibility of the tested hypothesis is poor, belief in the hypothesis will not be strengthened after testing, no matter how low the type I error.

When they submit their work for publication, authors claim to have done their best to select credible hypotheses, that they used the best possible methods to test them and, in most instances, that they concluded that their observations were compatible with their hypotheses as testified by low type I error. The task of editors, reviewers and readers is then to assess these claims by evaluating the posttesting credibility of the hypotheses. Here, the main question is to determine the probability that the tested relationship is true given that the test presented by the author came out statistically signifi-

cant:  $\Pr(R/T^+)$ . This probability is analogous to the predictive value of a positive test (PVP) in epidemiology and depends on type I error, statistical power and the a priori probability of the hypothesis. It is here that the task of editors and reviewers is most critical. Using their expertise and knowledge of the field of research, they should ask the following series of questions. What would be the most likely effect size expected for the kind of hypothesis tested and, given that effect size, what should be the power of the study? What is the prior credibility of the hypothesis? How clear is the experimental design, and how open is it to data dredging (i.e., seeking significant results through extensive and unplanned explorations)? What is the track record in this specific research field for replicating findings? How much is fashion driving the field? What are the biases that could be present in the study?

In a thought-provoking paper, "Why most published research findings are false," Ioannidis has argued that when considering all these factors, it is possible that most of what is published in the current biomedical literature is false.<sup>12</sup> To support this claim, Ioannidis has shown that in a given scientific field, in order for the PVP to be greater than 50% (i.e., relationships have a > 50% chance of being true when the test is positive), the following relation must hold:  $(1-\beta)\rho > \alpha$ , with  $\rho$  being a very close approximation of the a priori probability of the tested relationships in that field. In the case that  $(1-\beta)\rho \leq \alpha$ , relationships published in that field are more likely to be false, and no knowledge can be gained even if type I error is set at a very low level. Thus, even in cases where a study concludes that a relationship is true based on a very low  $p$  value, this conclusion can be questioned if the statistical power and/or the a priori probabilities are very low. Ioannidis claims that most published research is false mainly because the a priori probabilities of most tested relationships in most research fields are very low. He illustrates this by candidate gene testing in complex disorders, where thousands of positive results were poorly or not replicated.<sup>13</sup> Of course, low statistical power ( $1-\beta$ ) contributes substantially to this problem. Many other factors contribute to reduce the PVP, the chief of which are various kinds of bias. If these assumptions are true, which we believe to be the case, then negative results are seriously under-represented in the literature.

Remarkably, under these same assumptions of low a priori probabilities and low statistical power, the probability that a relationship is not real when the test result is negative, or  $\Pr(\text{nonR}/T^-)$ , is in fact high. This probability is the predictive value of a negative result (PVN). In other words, a negative result is the most likely one to be expected and the most likely true result given the conditions under which we usually perform research (with low a priori probabilities and/or low statistical power). This is similar to gambling on an outcome where the prior odds are very low and the players have little information on which to base their bets. Under this scenario, it is not surprising that most of the players will lose.

The PVN is often overlooked when editors evaluate negative results, although it represents the complementary question to that which editors and reviewers ask when they evaluate

manuscripts with positive results. Of course, negative results are most interesting when they refute hypotheses that received strong prior confirmation in the literature. These refutations are also most credible when they are methodologically sound and well conducted. Here the issue of statistical power becomes crucial, as negative results could be due to small sample sizes with insufficient statistical power to detect significant effects. However, studies with sound hypotheses that are well conducted are still worth publishing even if they are suspected of being not adequately powered. Indeed, as long as this shortcoming is discussed in the paper, this kind of negative study is more likely to be a true negative than a false-negative result (high PVN) and will ultimately contribute to an accurate estimation of the true effect size of the tested relationship in future meta-analyses.

From these considerations, it appears that the literature is predominantly biased toward positive results, of which many are likely to be false, whereas negative results that are more likely to be true-negative results are disappearing.<sup>9</sup> This may explain why, despite thousands of published papers in psychiatry, it is sometimes very difficult to identify solid facts beyond some fundamental observations. We believe that the scientific community has the responsibility to change this situation.

In the literature, there are at least 3 different categories of negative results.<sup>1</sup>

1. Conclusive negative results: clear evidence of an opposite effect (e.g., treatment harms when benefit was expected) or a neutral effect (no effect of a new treatment) in a well-designed study. A well-designed study must include a compelling rationale, explicit formulations of its hypotheses and primary outcome measures and clear a priori estimation of its statistical power.<sup>2</sup>
2. Exploratory negative results: well-designed and adequately powered study with neutral or opposite results based on exploratory data analysis. These include results emerging from secondary hypotheses or exploration of the data with post hoc hypotheses.<sup>3</sup>
3. Inconclusive negative results: no evidence of an effect in a study that was too small and inadequately powered (e.g., no treatment effect due to small sample size).

Currently, papers in category 1 are sometimes successful in finding their way to publication, particularly in the case of human clinical trials. Registration of clinical trials, which is required by regulatory agencies and some journals, to a certain extent increased public access to negative results from clinical trials. However, registration of clinical trials has not, so far, been an unqualified success.<sup>14</sup> Registration of planned research does not apply to human experimental, as opposed to clinical, trials or animal studies. Whereas registration might improve access to negative results in the latter 2 categories, requiring the registration of such research may be very difficult, if not impossible, given that the acute consequences of not publishing negative results in these cases may be less harmful than those associated with not publishing negative results in human clinical trials.

Access to data arising from categories 2 and 3 remains limited. As we see it, a major impediment to publication of nega-

tive findings is the current structure in publishing, which relies heavily on citation rates and impact factors as metrics for quality. Impact factors also give an indication of how much a journal is read and noticed. Thus publishers and advertisers will generally support a high-impact journal, whereas a journal with a waning impact factor will have difficulty finding financial support, even in a nonprofit model. Moreover, authors, whose career success is measured to a large extent by publication in high-impact journals, will submit their most significant results to journals with the highest impact factors. Thus continued submission of high-quality work to a journal will depend, to some extent, on its impact factor. In addition, finding expert external reviewers who will devote time to reviewing journal articles is increasingly difficult, possibly more so in the case of studies with negative results, which are considered less interesting.

As outlined earlier, publication of negative findings is essential to interpreting the overall significance of a field of research. However, papers with negative findings are less likely to be highly cited than papers with positive findings and less likely overall to be noticed in the scientific community. Unless all journals make a concerted effort to promote publication of high-quality negative studies, it will be very difficult for any one journal or only a few journals to spearhead such a movement and change the current climate. Nonetheless, many open-access journals state their commitment to publish manuscripts regardless of whether they report positive or negative results as long as their methodology is sound. This is certainly a move in the right direction, although it does not necessarily guarantee the eradication of the problem. One major problem with most current open-access models is their necessary reliance on publication charges, often assumed by authors. This may deter authors from submitting manuscripts with negative results, particularly since these manuscripts may not be highly rewarded in terms of recognition and citation when they are published. In addition, it has been suggested that open-access journals with relatively high publication charges might introduce a new bias (e-publication bias). For example, in a recent study examining papers published in the *Annals of the Rheumatic Diseases* either with open access or with subscription access (depending on whether authors chose to pay publication charges), Jakobsen and colleagues<sup>15</sup> found that a significantly greater proportion of studies published in the open-access section were industry funded, which could lead to preferential publication of results supporting industry products. Thus high open-access charges combined with the low incentive that authors may derive from negative publications may not solve the publication bias problem.

A possible, although ambitious avenue, would be to launch an online electronic series of Journals of Negative/Neutral Results in various research fields to be funded by public or charitable support. Divorcing the publication process from all financial constraints (and hence the tyranny of the impact factor) would go a long way to help negative findings emerge from the dark recesses of researchers' data books into the light of publication.

It is only through a concerted effort at different levels that

the problem of publication bias can be remedied. These efforts will not succeed unless more reward is attributed to all parties when they contribute to making negative results more accessible. Collaboration among academic institutions, journals, funding agencies, philanthropy and authors is needed. Increasing awareness among students about this problem and encouraging them to report their negative findings in their thesis dissertations and publications will be an important step toward changing the situation in the long run. Funding agencies should encourage investigators to publish both positive and negative results of the primary hypotheses tested in the research projects they have funded. It is, for example, important that funding agencies reward investigators who have made efforts to submit and publish their negative results when they are applying for renewal of their funding. Philanthropy, although mainly interested in finding cures for diseases, may be sensitized to the importance of encouraging the scientific community to publicize what does not work. Sharing data with the research community (e.g., 1 year after publication of the main results) may facilitate systematic reviews and reduce the publication of a huge number of category 2 and 3 studies. However, innovative ways to reward data sharing need to be devised if this is going to happen.

The *Journal of Psychiatry and Neuroscience* is an open-access journal committed to publishing high-quality manuscripts that will contribute to an unbiased updating of the literature. It welcomes manuscripts reporting negative results, particularly from category 1. We will also consider category 2 manuscripts favourably and encourage authors to be open about the exploratory nature of their results.

In conclusion, as Epictetus said 19 centuries ago, the wise man's task is "Rightly to aim in all these cases" of appearances/nonappearances (positive and negative tests, respectively) and what they reveal as beings/nonbeings (true-positive and false-positive results, respectively). Wise men can be easily tricked if the base of their knowledge is severely distorted. Wise men also recognize that without incentives, humans are unlikely to change their behaviour.

**Competing interests:** None declared for N. Schmitz, L. Annable and P. Boksa. R. Joober declares advisory board and speaker bureau membership with Pfizer Canada, Janssen Ortho Canada, BMS and Sunovion Canada; grant funding from them and from AstraZeneca; honoraria from Janssen Ortho Canada and from Pfizer Canada for CME presentations; and royalties for Henry Stewart talks.

## References

1. Turner EH, Matthews AM, Linardatos E, et al. Selective publication of antidepressant trials and its influence on apparent efficacy. *N Engl J Med* 2008;358:252-60.
2. Barbui C, Furukawa TA, Cipriani A. Effectiveness of paroxetine in the treatment of acute major depression in adults: a systematic re-examination of published and unpublished data from randomized trials. *CMAJ* 2008;178:296-305.
3. Eyding D, Lelgemann M, Grouven U, et al. Reboxetine for acute treatment of major depression: systematic review and meta-analysis of published and unpublished placebo and selective serotonin reuptake inhibitor controlled trials. *BMJ* 2010;341:c4737.
4. Whittington CJ, Kendall T, Fonagy P, et al. Selective serotonin reuptake inhibitors in childhood depression: systematic review of published versus unpublished data. *Lancet* 2004;363:1341-5.
5. Ioannidis JP. Effectiveness of antidepressants: An evidence myth constructed from a thousand randomized trials? *Philos Ethics Humanit Med* 2008;3:14.
6. Turner EH, Matthews AM, Linardatos E, et al. Selective publication of antidepressant trials and its influence on apparent efficacy. *N Engl J Med* 2008;358:252-60.
7. Fanelli D. Negative results are disappearing from most disciplines and countries. *Scientometrics* 2012;90:891-904.
8. Rossignol DA, Frye RE. A review of research trends in physiological abnormalities in autism spectrum disorders: immune dysregulation, inflammation, oxidative stress, mitochondrial dysfunction and environmental toxicant exposures. *Mol Psychiatry* 2012; 17:389-401.
9. Ioannidis JP. An epidemic of false claims. Competition and conflicts of interest distort too many medical findings. *Sci Am* 2011; 304:16.
10. Munafò MR, Attwood AS, Flint J. Bias in genetic association studies: effects of research location and resources. *Psychol Med* 2008;38:1213-4.
11. Munafò MR, Stohart G, Flint J. Bias in genetic association studies and impact factor. *Mol Psychiatry* 2009;14:119-20.
12. Ioannidis JP. Why most published research findings are false. *PLoS Med* 2005;2:e124.
13. Ioannidis JP, Tarone R, McLaughlin JK. The false-positive to false-negative ratio in epidemiologic studies. *Epidemiology* 2011;22:450-6.
14. Huic M, Marusic M, Marusic A. Completeness and changes in registered data and reporting bias of randomized controlled trials in ICMJE journals after trial registration policy. *PLoS ONE* 2011;6: e25258.
15. Jakobsen AK, Christensen R, Persson R, et al. Open access publishing. And now, e-publication bias. *BMJ* 2010;340:c2243.