COLLOQUIUM PAPER

Misrepresentation and distortion of research in biomedical literature

Isabelle Boutron[a,b,c,1] and Philippe Ravaud[a,b,c,d]

[a]Methods of Therapeutic Evaluation Of Chronic Diseases (METHODS) team, INSERM, UMR 1153, Epidemiology and Biostatistics Sorbonne Paris Cité Research Center (CRESS), F-75014 Paris, France; [b]Faculté de Médicine, Paris Descartes University, 75006 Paris, France; [c]Centre d'Épidémiologie Clinique, Hôpital Hôtel Dieu, Assistance Publique des Hôpitaux de Paris, 75004 Paris, France; and [d]Department of Epidemiology, Columbia University Mailman School of Public Health, New York, NY 10032

Edited by David B. Allison, Indiana University Bloomington, Bloomington, IN, and accepted by Editorial Board Member Susan T. Fiske November 14, 2017 (received for review June 14, 2017)

Publication in peer-reviewed journals is an essential step in the scientific process. However, publication is not simply the reporting of facts arising from a straightforward analysis thereof. Authors have broad latitude when writing their reports and may be tempted to consciously or unconsciously "spin" their study findings. Spin has been defined as a specific intentional or unintentional reporting that fails to faithfully reflect the nature and range of findings and that could affect the impression the results produce in readers. This article, based on a literature review, reports the various practices of spin from misreporting by "beautification" of methods to misreporting by misinterpreting the results. It provides data on the prevalence of some forms of spin in specific fields and the possible effects of some types of spin on readers' interpretation and research dissemination. We also discuss why researchers would spin their reports and possible ways to avoid it.

misinterpretation | bias | misreporting | misrepresentation | detrimental research practice

Publication in peer-reviewed journals is an essential step in the scientific process. It generates knowledge, influences future experiments, and may impact clinical practice and public health. Ethically, research results must be reported completely, transparently, and accurately (1, 2). However, publication is not simply the reporting of facts arising from a straightforward and objective analysis of those facts (3). When writing a manuscript reporting the results of an experiment, investigators usually have broad latitude in the choice, representation, and interpretation of the data. They may be tempted consciously or unconsciously to shape the impression that the results will have on readers and consequently "spin" their study results.

In this article, we will explain the concept of spin, explore why and how investigators distort the results of their studies, and describe the impact of spin in reports and possible ways to avoid generating it. This article reflects our knowledge and opinion on this topic and is informed by a literature review. Furthermore, the scope of this study is limited to the occurrence of this phenomenon exclusively within the field of biomedicine.

Methods

We systematically searched MEDLINE via PubMed for articles on spin with an abstract. We searched the entire database, which begins with 1966. We used the following search strategy: (distorted[Title] AND interpretation[Title]) OR (detrimental[Title] AND research[Title] AND practice[Title]) OR (questionable[Title] AND research[Title] AND practice[Title]) OR (questionable[Title] AND reporting[Title]) OR (misleading[Title] AND reporting[Title]) OR "misleading representation" [Title] OR beautification[Title] OR misrepresentation[Title] OR "interpretive bias" [Title] OR (misrepresent[Title] OR misrepresentation[Title] OR misrepresentation[Title] OR misrepresentations[Title] OR misrepresentative[Title] OR misrepresented[Title] OR misrepresenting[Title] OR misrepresents[Title]) OR (overstate[Title] OR overstated[Title] OR overstated[Title] OR overstatement[Title] OR overstatements [Title] OR overstates[Title] OR overstating[Title]) AND has abstract[text] (search date May 23, 2017). We also searched Google Scholar for all articles citing key articles in the field of biomedicine (4–6). One researcher screened all titles and abstracts, retrieved the full text when appropriate, and extracted information on the type of spin, the prevalence of spin, the factors associated with spin, the impact of spin on readers' interpretation of the results, and the possible ways to reduce spin. We considered articles published in English or French, whatever their study designs: systematic assessment, methodological systematic reviews, consensus methods to develop classification of spin, randomized controlled trials evaluating the impact of spin, and so forth. The search retrieved 592 citations, of which 49 were relevant. We relied not only on this literature search but also on a personal collection of articles on spin that fulfill these eligibility criteria. This search has some limitations, as only a single researcher screened citations, abstracts, and full texts. We cannot rule out the possibility that we missed some relevant reports.

Definition of the Concept of Spin

Spin has become a standard concept in public relations and politics in recent decades. It is "a form of propaganda, achieved by providing a biased interpretation of an event or campaigning to persuade public opinion in favor of or against some organization or public figure" (https://en.wikipedia.org/w/index.php?title=Spin_(propaganda)&oldid=793952705). "Spin doctors" modify the perception of an event to reduce any negative impact or to increase any positive impact it might have on public opinion. For this purpose, spin doctors could attempt to bury potentially negative information or selectively "cherry-pick" specific information or quotes.

The concept of spin can also be applied to scientific communications. Spin can also be defined as a specific reporting that fails to faithfully reflect the nature and range of findings and that could affect the impression that the results produce in readers, a way to distort science reporting without actually lying (7). Spin could be unconscious and unintentional. Reporting results in a manuscript implies some choices about which data analyses are reported, how data are reported, how they should be interpreted, and what rhetoric is used. These choices, which can be legitimate in some contexts, in another context can create an inaccurate impression of the study results (3). It is almost impossible to determine whether spin is the consequence of a lack of understanding of methodologic principles, a parroting of common practices, a form of unconscious behavior, or an actual willingness to mislead the reader. However, spin, when it occurs, often favors the author's vested interest (financial, intellectual, academic, and so forth) (3).

This paper results from the Arthur M. Sackler Colloquium of the National Academy of Sciences, "Reproducibility of Research: Issues and Proposed Remedies," held March 8–10, 2017, at the National Academy of Sciences in Washington, DC. The complete program and video recordings of most presentations are available on the NAS website at www.nasonline.org/Reproducibility.

Author contributions: I.B. and P.R. designed research, performed research, analyzed data, and wrote the paper.

Conflict of interest statement: P.R. is director of the French EQUATOR (Enhancing the Quality and Transparency of Health Research) Center. I.B. is deputy director of the French EQUATOR Center.

This article is a PNAS Direct Submission. D.B.A. is a guest editor invited by the Editorial Board.

[1]To whom correspondence should be addressed. Email: isabelle.boutron@aphp.fr.

Published online March 12, 2018.

MEDICAL SCIENCES

Practices of Spin

There are several ways to spin a report (4, 6, 8–10). These different practices are usually interrelated, and the amount of spin in published reports varies (Fig. 1). Specific classifications of spin have been developed for different study designs and contexts [randomized controlled trials with nonstatistically significant results (4), observational studies evaluating an intervention (10), diagnostic accuracy studies (8), systematic reviews (9)]. Here, we report practices of spin organized under the following categories: misreporting the methods, misreporting the results, misinterpretation, and other types of spin. The classification of the practices reported here represents our chosen approach, but several different approaches are possible. Future work based on systems to inductively code and classify data such as spin would help provide a rigorous and exhaustive analysis of spin that is generalizable across manuscripts.

Misreporting the Methods. Authors could intentionally or unintentionally misrepresent the methods they used. This type of spin will alter the readers' critical appraisal of the study and could impact the interpretation of evidence synthesis. It could consist of changing objectives, reporting post hoc hypotheses as if they were prespecified, switching outcomes and analysis, or masking protocol deviations. Scientists could also engage in what we characterize as "beautification" of the methods, when they report the methods as if they were complying with the highest standards when in fact they were not. For example, some studies report "double-blind" methods, but the blinding is not credible (11, 12), or report an intent-to-treat analysis, but some patients are excluded from the analysis (13, 14). The term "randomized controlled trial" (RCT) can also be used erroneously. A survey of authors of 2,235 reports of RCTs published in Chinese medical journals showed that only about 7% met the methodological criteria for authentic randomization; 93% were falsely reported as RCTs (15). Finally, authors could claim adherence to quality standards such as reporting guidelines (e.g., the CONSORT Statements), when in reality, the adherence of their reports to these standards is far from perfect.

Misreporting Results. Misreporting of results is defined as an incomplete or inadequate reporting of results in a way that could mislead the reader. This type of spin particularly involves selective reporting of statistically significant results, ignoring results that contradict or counterbalance the initial hypothesis, and misleading display of results through choice of metrics and figures. Undesirable consequences include wasted time and resources on misdirected research and ill-founded actions by health providers misled by partial results.

Selective reporting of outcomes and analysis. Selective reporting of outcomes and analysis is defined as the reporting of some outcomes or analysis but not others, depending on the nature and direction of the results. The literature contains evidence of researchers favoring statistically significant results. A comparison of outcomes reported in protocols of RCTs submitted to ethics committees or registered in trial registries showed that scientists selectively report statistically significant outcomes (16–18). An automated text-mining analysis of P values reported in more than 12 million MEDLINE abstracts over the course of 25 y showed an increase in the reporting of P values in abstracts and a strong clustering at P values of 0.05 and of 0.01 or lower, which could suggest "P-value hacking" (19, 20). P-hacking, a detrimental practice, is defined as the misreporting of true effect sizes in published reports after researchers perform several statistical analyses and selectively choose to report or focus on those that produce statistically significant results (20). Practices that can lead to P-hacking include an interim analysis to decide whether an experiment or a study should be stopped prematurely (21), as well as post hoc excluding of outliers from the analysis, deciding to combine or split groups, adjusting covariates, performing subgroup analysis (22), or choosing the threshold for dichotomizing continuous outcomes. Cherry-picking can be particularly problematic in this era of massive observational data (23).

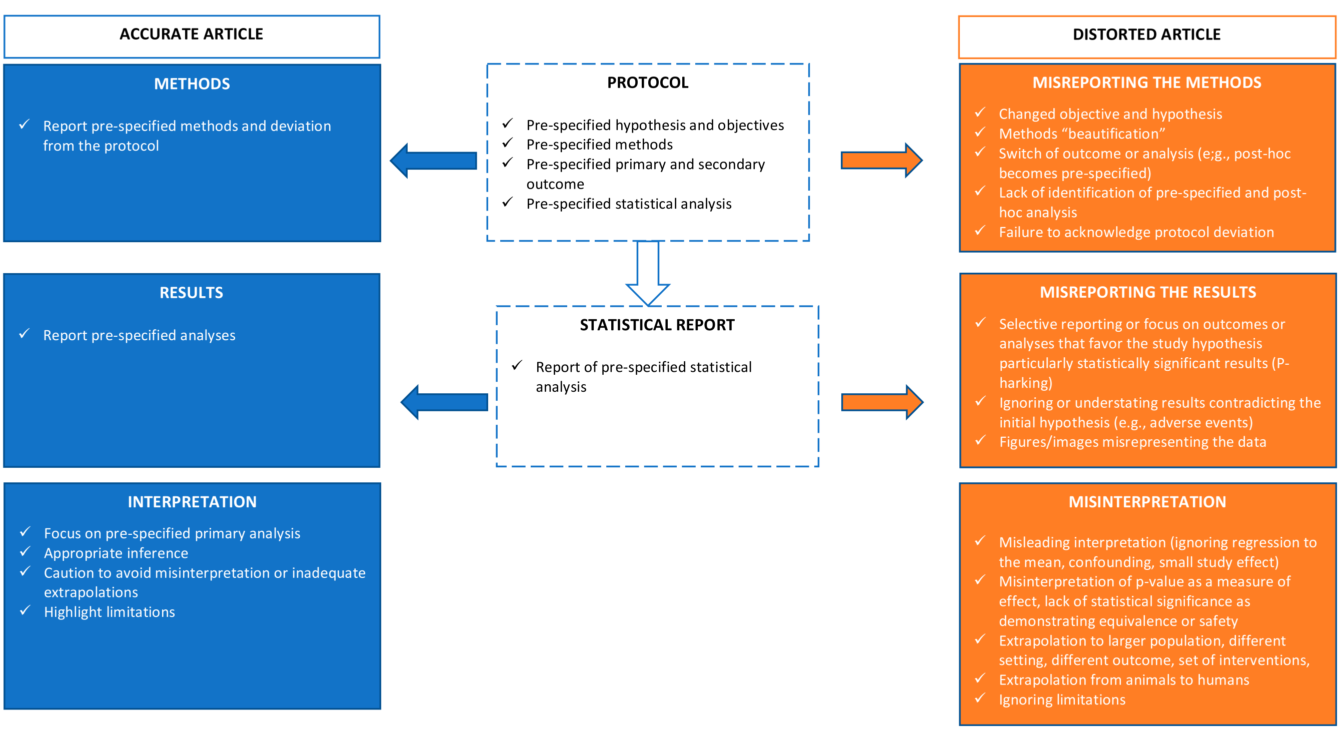

Fig. 1. Practices of spin in published reports.

Ignoring or understating results that contradict or counterbalance the initial hypothesis. Authors may be tempted to consciously or unconsciously mask or understate some troublesome results, such as nonstatistically significant outcomes or statistically significant harm. For example, the reporting and interpretation of the risk of all-cause mortality in the DAPT (dual antiplatelet therapy) study randomizing 9,961 patients to continue DAPT beyond 1 y after stent placement or receive a placebo for 30 mo, raised some concerns (24, 25).

Misreporting results and figures. The presentation of results can affect their interpretation. For example, choosing to report the results as either relative risk reduction or absolute risk reduction can substantially impact readers' interpretation and understanding, particularly when the baseline risk is low. Similarly, reporting odds ratios (ORs) instead of risk ratios (RRs) when the baseline risk is high can easily be misinterpreted (26).

The graphical display of data is a very powerful tool for disseminating and communicating the results of a study. Researchers are continually working on the best way to represent their data and be informative to the reader by using increasingly innovative methods. Some figures, such as the CONSORT flow diagram, were so informative that they are now required by most journals. However, figures can be misleading. For example, a break in the *y* axis, failure to represent the statistical uncertainty with the confidence interval (CI), scaling the figure on the results, and extending survival curves to the right without representing the statistical uncertainty can create the false impression that a treatment is beneficial. A study of 288 articles in the field of neuroscience, displaying a total of 1,451 figures, showed that important information was missing from 3D graphs; particularly, uncertainty of reported effects was reported in only 20% of 3D graphics, a practice bound to mislead the reader (27).

In the field of basic science, use of image-processing programs, routinely used to improve the quality of images, can actually shape the impression that results will make on readers. An assessment of the images of 20,642 published articles in 40 journals over the course of 20 y found that 3.8% of published articles contained questionable images, and that the number was increasing (28). Some modifications are obvious, such as the deletion or addition of a band from or to the visualization of a blot, whereas others are subtler, such as adjusting the brightness and contrast of a single band in a blot, cleaning up unwanted background of an image in a blot, splicing lanes together without clearly indicating the splicing, enhancing a specific feature on a micrograph by image processing, or adjusting the brightness of only a specific part of an image (29). Drawing the line between appropriate image manipulation and detrimental practice is difficult. Editors have developed specific guidelines to encourage transparency and avoid distortion in the manipulation of images. They have estimated that about 20% of the accepted papers contained at least one figure that did not comply with accepted practices (29, 30). In addition, the publishing of images presumes a choice of the images that will be presented in the articles from among all images available. Obviously, this choice can be influenced by the message the researcher wants to convey.

Misinterpretation. Misinterpretation refers to an interpretation of the results that is not consistent with the actual results of the study. In the Discussion section of a paper, authors may take a strong position that relies more on their opinion than on the study results. Interpretation of results is misleading when researchers focus on a within-group comparison; when they ignore regression to the mean and confounding; when they inappropriately posit causality (31); when they draw an inappropriate inference from a composite outcome (32); or report *P* values as a measure of an effect whereas, in reality, it is only a measure of how likely it is that a result occurs by chance. A systematic methodologic review of 51 RCTs assessing complex interventions with statistically significant small effects showed that authors exercised no caution in their interpretation of results in about half of the reports (33). For example, in one study with RR = 0.95 (95% CI 0.93–0.97), the authors concluded that "Complex interventions can help elderly people to live safely and independently, and could be tailored to meet individuals' needs and preferences" (34).

Inadequate interpretation of the *P* value as a measure of the strength of a relationship occurs also in the field of genetics. For example, the effect of a single gene is usually very small, with RRs ranging from 1.1 to 1.4 (35), but a focus on the *P* value (low if the sample size is high) could be misinterpreted as showing a strong relationship. Furthermore, for diagnostic, prognostic, or screening markers in epidemiologic studies, the limited magnitude of the OR considered meaningful (i.e., about or >70) is rarely discussed (36). Nonstatistically significant results could also be misinterpreted as demonstrating equivalence or safety despite lack of power.

HARKing, or hypothesizing after results are known (37), or JARKing, justifying after results are known (38), are also inappropriate practices. For example, in the DAPT study, the authors proposed a post hoc explanation for the increased rate of death in the experimental group based on a post hoc analysis to mitigate the role of prolonged treatment on this increased risk of mortality (25). Finally, authors can be tempted to extrapolate their results beyond the data to a larger population, setting, or intervention, and even provide recommendations for clinical practice (39). One extrapolation is the projection of results from an animal experiment to an application in humans.

Other Types of Spin. Rhetoric, defined as language designed to have a persuasive or impressive effect, can be used by authors to interest and convince the readers (5). Any author can exaggerate the importance of the topic, unfairly dismiss previous work on it, or use persuasive words to convince the reader of a specific point of view (40, 41). Based on our and others' experience (40, 41), a typical article might declare that a certain disease is a "critical public health priority" and that previous work on the topic showed "inconsistent results" or had "methodologic flaws." In such cases, the discussion will inevitably claim that "this is the first study showing" that the new research provides "strong" evidence or "a clear answer"; the list of adjectives and amplifiers is large. Some of these strategies are actually taught to early career researchers. A retrospective analysis of positive and negative words in abstracts indexed in PubMed from 1974 to 2014 showed an increase of 880% in positive words used over the four decades (from 2% in 1974–1980 to 17.5% in 2014) (42). There is also a website that features a collection of the rhetoric used for nonstatistically significant results (https://mchankins.wordpress.com/2013/04/21/still-not-significant-2).

Even the references cited in a manuscript can be selected according to their results to convey a desired message. For example, an analysis of the patterns of knowledge generation surrounding the controversy between proponents and opponents of a population-wide reduction in salt consumption showed that reports were more likely to cite studies that had conclusions similar to rather than different from those of the author doing the citing (43).

Prevalence of Some Forms of Spin in Published Reports

Evidence of discrepancies between the study results and the conclusions of published reports in specific fields has been reported in case studies and in systematic assessments of cohorts (31, 44, 45). A comparison of published findings and Food and Drug Administration reviews of the underlying data revealed publication bias (i.e., studies with nonstatistically significant results omitted from published piece) as well as spin in the conclusion (i.e., the conclusion was biased in favor of a beneficial effect of the experimental treatment despite nonstatistically significant results) (46). A Delphi consensus survey of expert opinion identified some types of spin (overinterpretation of significant findings in small trials, elective reporting based on *P* values, and selective reporting of outcomes in the abstract) found among the questionable research practices most likely to occur (47).

Biomedical spin was first systematically investigated with a representative sample of two-arm parallel-group RCTs with nonstatistically significant primary outcomes indexed in PubMed in 2006 (4). In this study, spin was defined as "specific reporting strategies, whatever their motive, to highlight that the experimental treatment is beneficial, despite a statistically nonsignificant difference for the primary outcome, or to distract the reader from statistically nonsignificant results" (4). The study showed a high prevalence of spin particularly in the abstract's conclusions, which for more than half of the reports contained examples of spin. Other methodological systematic reviews focusing on two-arm parallel-group RCTs with nonstatistically significant primary outcomes in specific medical fields found consistent results (48–54). Spin has also been assessed in different study designs. One study in the field of HIV assessed the interpretation of noninferiority trial results and showed spin in two-thirds of the studies with inconclusive results (55). In diagnosis-accuracy studies, spin was identified in one-third of the articles published in high-impact factor journals (8), and in the field of molecular diagnostic tests, more than half of the reports overinterpreted the clinical applicability of the findings (39). In observational studies evaluating an intervention, spin was identified in the abstract's conclusions in more than 80% of reports, the most frequent type of spin being the use of causal language (10). To our knowledge, no systematic assessment of spin in systematic reviews and meta-analyses has been reported. However, a classification of spin was developed that particularly allowed for the identification of the most severe types of spin in such reports (9).

These methodological systematic reviews evaluated only a specific body of literature in the field of life sciences and, more specifically, biomedicine. To our knowledge, there are no data on the prevalence of researchers using spin, but we suspect that it is a quite common practice among researchers (56).

Impact of Spin

One important question is whether spin matters and can actually impact readers' interpretations of study results. Spin can affect researchers, physicians, and even journalists who are disseminating the results, but also the general public, who might be more vulnerable because they are less likely to disentangle the truth. Patients who are desperately seeking a new treatment could change their behavior after reading distorted reporting and interpretations of research findings.

An RCT evaluated the impact of spin found in abstracts of reports of cancer RCTs on researchers' interpretation (57). A sample was selected of 30 reports of RCTs with a nonstatistically significant primary outcome that also had some kind of spin in the abstract's conclusions. All abstracts were rewritten to be reported without spin. Overall, 300 corresponding authors and investigators of RCTs were randomized to read either an abstract with spin or one without spin and assess whether the experimental treatment would be beneficial to patients on a scale of 0 (very unlikely) to 10 (very likely). After reading the abstract with spin, readers were more likely to believe the treatment would be beneficial to patients [mean difference 0.71 95% (95% CI 0.07–1.35), $P = 0.030$]. The presence of spin in abstracts may also affect the content of stories disseminated in news items. A study assessing the diffusion of spin from published articles to press releases and the news showed that spin in press releases and the mass media was related to the presence of spin in the abstracts of peer-reviewed reports of RCTs (58). Furthermore, interpretation of RCTs based solely on press releases or media coverage could distort the interpretation of research findings in a way that favors the experimental treatment (58). This study highlighted the significant role of researchers, editors, and peer-reviewers in the dissemination of distorted research findings (58). This distorted dissemination can have serious consequences. A study comparing the number of citations of articles published in the *New England Journal of Medicine* showed that the articles that garnered media attention received 73% more citations than did control articles (59). This issue is all the more significant because media coverage can affect future research as well as clinical practice. For example, a study entitled "Lithium delays progression of amyotrophic lateral sclerosis" (ALS), involving mice and tested in a small sample of patients, concluded that "these results offer a promising perspective for the treatment of human patients affected by ALS" (60). This was rapidly followed by an uptick in the use of this treatment by patients with ALS. Two controversial articles on statins followed by great debate in the media (61, 62) were associated with an 11% and 12% increase in the likelihood of existing users stopping their treatment for primary and secondary prevention, respectively (63). Such effects could result in more than 2,000 extra cardiovascular events across the United Kingdom over a 10-y period.

Why Researchers Add Spin to Their Reports

Competitive Environment and Importance of Positive Findings. Scientists are under pressure to publish, particularly in high-impact factor journals. Publication metrics, such as the number of publications, number of citations, journal impact factor, and h-index are used to measure academic productivity and scientists' influence (64).

However, we have some evidence that editors, peer-reviewers, and researchers are more interested in statistically significant effects. An RCT comparing the assessment of two versions of a well-designed RCT that differed only by the findings (positive vs. negative primary endpoint) showed that peer-reviewers were more likely to recommend the positive version of the manuscript's findings for publication. They were also more likely to detect errors in and award a low score to the methods of the negative version of the same manuscript, even though the Methods sections in both versions were identical (65). In the field of basic science, negative studies can be considered failures and useless.

This highly competitive "publish or perish" environment may favor detrimental research practices (66); thus, spinning the study results and a "spun" interpretation could be an easy way to confer a more positive result and increase the interest of reviewers and editors. A study of more than 4,600 articles published in all disciplines between 1990 and 2007 showed an increase in statistically significant results by more than 22%, with 86% of articles reporting a statistically significant result (67).

Lack of Guidelines to Interpret Results and Avoid Spin. To improve transparency, authors are encouraged to report their studies according to reporting guidelines, such as the ARRIVE (68) or CONSORT 2010 (69) guidelines. There is some evidence that editors' endorsement and implementation of these guidelines improves the completeness of reporting. However, no guidelines on avoiding spin in published reports are either available for public consumption or requested by editors. Furthermore, in some quarters, adding spin may actually be considered usual practice to "interest" the reader, and researchers may even be trained to add spin, particularly in their grant proposals. The Introduction and Discussion sections of papers are often used to tell a story. Some researchers argue that the use of linguistic spin and rhetoric is "an essential element of science communication" and that "scientific papers stripped of spin will be science without its buzz" (70).

How Can We Reduce the Use of Spin?

Change the Perception of Spin from "Commonly Accepted Practice" to "Seriously Detrimental Research Practice." Editors, funders, institutions, and researchers take very seriously such research misconduct as data falsification or fabrication and plagiarism. They are developing specific guidelines and procedures to avoid these forms of misconduct, although such malpractice is probably very rare. In contrast, misrepresentation or distortion of research in published reports is underrecognized, despite its possible impact on research, clinical practice, and public health. Worse, these forms of malpractice may be considered acceptable by the scientific community. A survey of researchers in psychology showed that more than 50% admitted not reporting all measures and deciding to stop collecting data after seeing the results. Overall, they did not regard these practices as malpractice (71). Researchers should be specifically trained to detect and avoid spin in published reports.

Require and Enforce Protocol Registration. To detect spin, essential information in the protocol and statistical analysis plan, such as the prespecified primary outcome and prespecified analysis, must be accessible. Registration of the protocol before the conduct of the experiment has been an important step forward in clinical research. Access to the statistical analysis plan and raw data could also facilitate the detection and elimination of spin. However, there is a general feeling among researchers, particularly in the field of basic science, that prespecifying all methods and analysis in a protocol and focusing the results interpretation and conclusion only on the prespecified analyses would reduce creativity (72). Although we must be open to new, unexpected results, we must be aware of the risk of apophia (the tendency to see patterns in random data), confirmation bias (the tendency to focus on evidence in line with expectations), and hindsight bias (the tendency to see an event as being predictable only after it has occurred) (73).

Reporting Guidelines and New Processes of Reporting. The development of reporting guidelines was a very important step toward achieving complete, accurate, and transparent reporting. These guidelines are endorsed by editors who require adherence to the guidelines in their instructions to authors. These guidelines indicate the minimum set of information that should be systematically reported by authors for specific studies. However, they do not provide recommendations on how results should be interpreted, how the conclusions should be reported, and how to avoid spin. Nevertheless, summarizing the results of a study into a succinct sentence in the conclusion is challenging and—inevitably—will not capture every nuance of the methodology, data, or clinical relevance of a study (74). We probably need to expand these guidelines to improve the presentation and interpretation of results. Some editors have proposed initiatives that could reduce spin. For example, the *Annals of Internal Medicine* requires the reporting of a limitation in the abstract (75, 76). In 2016, the American Statistical Association released a statement on statistical significance and the *P* value with six principles underlying the proper use and interpretation of this statistical tool (77).

We should also question the current process in which the interpretation of study results is reported by the researchers who performed the experiment. Results may be more accurate with the interpretation and conclusions reported by dispassionate researchers who would offer inferences based only on the Methods and Results sections. One approach would be based on collective intelligence, with results interpreted by several researchers—content experts, methodologists, statisticians—who would confer with each other to provide the most consensual interpretation of the study results.

Editors, Peer-Review, and Postpublication Monitoring/Feedback. In theory, peer-reviewers and editors should determine whether the conclusions match the results. However, a systematic assessment of peer-reviewers' reports showed that even when they identify some spin in reports, only two-thirds of the spin is completely deleted by the authors. Furthermore, some peer-reviewers are actually requesting the addition of spin, and one study found that they failed to even identify spin in the abstract's conclusion in 76% of the reports (78). We need to provide specific training and tools to peer-reviewers and editors to facilitate the detection of spin. A user's guide to detect misleading claims in clinical research reports (79) and tips for interpreting claims (6) are available, but should be more widely used. Additionally, editors should be held clearly accountable for the content of a published manuscript. Regular monitoring of the content of research publications, which has been successfully implemented for the detection of selective reporting of outcomes, could be an effective method to change the practices of researchers and editors alike (80).

Changing the Reward System and Developing Collaborative Research. The current reward system for scientists, based mainly on the number of publications and the journal impact factor, could be aiding and abetting the misleading behavior (81). Some researchers engaged in various aspects of biomedical science have been working on the future of the research enterprise, tackling its systemic flaws (82–84). They are particularly questioning the expectation that this enterprise should continue expanding (83, 84). These researchers argue that the highly competitive environment compresses the time dedicated to thinking and the willingness to engage in high-risk projects (82–84). A 2-d workshop bringing together 30 senior researchers engaged in various aspects of biomedical science proposed specific remedies to improve the system and create a "sustainable" system in the future (84). Others proposed replacing the current system with a new system that would reward research that is productive, high-quality, reproducible, shareable, and translatable (85). The use of a badge to acknowledge open practices has been effective in changing researchers' behavior (86).

The use of new forms of research based on collective intelligence via the massive open laboratory could also be a way to reduce the risk of spin. Such research imposes rigorous adherence to scientific methods, with a clear statement of hypothesis systematically preceding experiments; hence, cherry-picking would be caught easily because the data and hypothesis are open to all and fully searchable (87).

Conclusions

Spin in published reports is a significant detrimental research practice (4, 57). However, the general scientific audience may not be fully aware of this. For example, spin is frequently not detected, even by readers with a high level of expertise and awareness, such as peer-reviewers (78). We need to raise awareness among the general scientific audience about the issues related to the presence of spin in published reports. Our proposals on ways to move forward should be food-for thought for researchers, editors, and funders.

ACKNOWLEDGMENTS. We thank Scott J. Harvey and Lina El Chall, who helped with the literature search.

1. World Medical Association (2013) World Medical Association Declaration of Helsinki: Ethical principles for medical research involving human subjects. *JAMA* 310:2191–2194.
2. Wager E, Kleinert S (2011) Responsible research publication: International standards for authors. A position statement developed at the 2nd World Conference on Research Integrity, Singapore, July 22–24, 2010. *Promoting Research Integrity in a Global Environment*, eds Mayer T, Steneck N (Imperial College Press/World Scientific Publishing, Singapore), pp 309–316.
3. Fletcher RH, Black B (2007) "Spin" in scientific writing: Scientific mischief and legal jeopardy. *Med Law* 26:511–525.

4. Boutron I, Dutton S, Ravaud P, Altman DG (2010) Reporting and interpretation of randomized controlled trials with statistically nonsignificant results for primary outcomes. *JAMA* 303:2058–2064.
5. Horton R (1995) The rhetoric of research. *BMJ* 310:985–987.
6. Sutherland WJ, Spiegelhalter D, Burgman MA (2013) Policy: Twenty tips for interpreting scientific claims. *Nature* 503:335–337.
7. Bailar JC (2006) How to distort the scientific record without actually lying: Truth, and arts of science. *Eur J Oncol* 11:217–224.
8. Ochodo EA, et al. (2013) Overinterpretation and misreporting of diagnostic accuracy studies: Evidence of “spin”. *Radiology* 267:581–588.
9. Yavchitz A, et al. (2016) A new classification of spin in systematic reviews and meta-analyses was developed and ranked according to the severity. *J Clin Epidemiol* 75: 56–65.
10. Lazarus C, Haneef R, Ravaud P, Boutron I (2015) Classification and prevalence of spin in abstracts of non-randomized studies evaluating an intervention. *BMC Med Res Methodol* 15:85.
11. Desbiens NA (2000) Lessons learned from attempts to establish the blind in placebo-controlled trials of zinc for the common cold. *Ann Intern Med* 133:302–303.
12. Farr BM, Gwaltney JM, Jr (1987) The problems of taste in placebo matching: An evaluation of zinc gluconate for the common cold. *J Chronic Dis* 40:875–879.
13. Schulz KF, Grimes DA, Altman DG, Hayes RJ (1996) Blinding and exclusions after allocation in randomised controlled trials: Survey of published parallel group trials in obstetrics and gynaecology. *BMJ* 312:742–744.
14. Thabut G, Estellat C, Boutron I, Marc Samama C, Ravaud P (2006) Methodological issues in trials assessing primary prophylaxis of venous thrombo-embolism. *Eur Heart J* 27:227–236.
15. Wu T, Li Y, Bian Z, Liu G, Moher D (2009) Randomized trials published in some Chinese journals: How many are randomized? *Trials* 10:46.
16. Chan AW, Hróbjartsson A, Haahr MT, Gøtzsche PC, Altman DG (2004) Empirical evidence for selective reporting of outcomes in randomized trials: Comparison of protocols to published articles. *JAMA* 291:2457–2465.
17. Mathieu S, Boutron I, Moher D, Altman DG, Ravaud P (2009) Comparison of registered and published primary outcomes in randomized controlled trials. *JAMA* 302:977–984.
18. Dwan K, et al. (2014) Evidence for the selective reporting of analyses and discrepancies in clinical trials: A systematic review of cohort studies of clinical trials. *PLoS Med* 11:e1001666.
19. Chavalarias D, Wallach JD, Li AH, Ioannidis JP (2016) Evolution of reporting P values in the biomedical literature, 1990-2015. *JAMA* 315:1141–1148.
20. Head ML, Holman L, Lanfear R, Kahn AT, Jennions MD (2015) The extent and consequences of p-hacking in science. *PLoS Biol* 13:e1002106.
21. Montori VM, et al. (2005) Randomized trials stopped early for benefit: A systematic review. *JAMA* 294:2203–2209.
22. Schulz KF, Grimes DA (2005) Multiplicity in randomised trials II: Subgroup and interim analyses. *Lancet* 365:1657–1661.
23. Taylor J, Tibshirani RJ (2015) Statistical learning and selective inference. *Proc Natl Acad Sci USA* 112:7629–7634.
24. Mauri L, et al.; DAPT Study Investigators (2014) Twelve or 30 months of dual antiplatelet therapy after drug-eluting stents. *N Engl J Med* 371:2155–2166.
25. Sharp M, Haneef R, Ravaud P, Boutron I (2017) Dissemination of 2014 dual antiplatelet therapy (DAPT) trial results: A systematic review of scholarly and media attention over 7 months. *BMJ Open* 7:e014503.
26. Grant RL (2014) Converting an odds ratio to a range of plausible relative risks for better communication of research findings. *BMJ* 348:f7450.
27. Allen EA, Erhardt EB, Calhoun VD (2012) Data visualization in the neurosciences: Overcoming the curse of dimensionality. *Neuron* 74:603–608.
28. Bik EM, Casadevall A, Fang FC (2016) The prevalence of inappropriate image duplication in biomedical research publications. *MBio* 7:e00809-16.
29. Rossner M, Yamada KM (2004) What's in a picture? The temptation of image manipulation. *J Cell Biol* 166:11–15.
30. Pearson H (2005) Image manipulation: CSI: Cell biology. *Nature* 434:952–953.
31. Brown AW, Bohan Brown MM, Allison DB (2013) Belief beyond the evidence: Using the proposed effect of breakfast on obesity to show 2 practices that distort scientific evidence. *Am J Clin Nutr* 98:1298–1308.
32. Cordoba G, Schwartz L, Woloshin S, Bae H, Gøtzsche PC (2010) Definition, reporting, and interpretation of composite outcomes in clinical trials: Systematic review. *BMJ* 341:c3920.
33. Siontis GC, Ioannidis JP (2011) Risk factors and interventions with statistically significant tiny effects. *Int J Epidemiol* 40:1292–1307.
34. Beswick AD, et al. (2008) Complex interventions to improve physical function and maintain independent living in elderly people: A systematic review and meta-analysis. *Lancet* 371:725–735.
35. Attia J, et al. (2009) How to use an article about genetic association: C: What are the results and will they help me in caring for my patients? *JAMA* 301:304–308.
36. Pepe MS, Janes H, Longton G, Leisenring W, Newcomb P (2004) Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. *Am J Epidemiol* 159:882–890.
37. Kerr NL (1998) HARKing: Hypothesizing after the results are known. *Pers Soc Psychol Rev* 2:196–217.
38. Nuzzo R (2015) How scientists fool themselves—And how they can stop. *Nature* 526: 182–185.
39. Lumbreras B, et al. (2009) Overinterpretation of clinical applicability in molecular diagnostic research. *Clin Chem* 55:786–794.
40. Cummings P, Rivara FP (2012) Spin and boasting in research articles. *Arch Pediatr Adolesc Med* 166:1099–1100.
41. Cepeda MS, Berlin JA, Glasser SC, Battisti WP, Schuemie MJ (2015) Use of adjectives in abstracts when reporting results of randomized, controlled trials from industry and academia. *Drugs R D* 15:85–139.
42. Vinkers CH, Tijdink JK, Otte WM (2015) Use of positive and negative words in scientific PubMed abstracts between 1974 and 2014: Retrospective analysis. *BMJ* 351: h6467.
43. Trinquart L, Johns DM, Galea S (2016) Why do we think we know what we know? A metaknowledge analysis of the salt controversy. *Int J Epidemiol* 45:251–260.
44. Altwairgi AK, Booth CM, Hopman WM, Baetz TD (2012) Discordance between conclusions stated in the abstract and conclusions in the article: Analysis of published randomized controlled trials of systemic therapy in lung cancer. *J Clin Oncol* 30: 3552–3557.
45. Mathieu S, Giraudeau B, Soubrier M, Ravaud P (2012) Misleading abstract conclusions in randomized controlled trials in rheumatology: Comparison of the abstract conclusions and the results section. *Joint Bone Spine* 79:262–267.
46. Turner EH, Matthews AM, Linardatos E, Tell RA, Rosenthal R (2008) Selective publication of antidepressant trials and its influence on apparent efficacy. *N Engl J Med* 358:252–260.
47. Al-Marzouki S, Roberts I, Marshall T, Evans S (2005) The effect of scientific misconduct on the results of clinical trials: A Delphi survey. *Contemp Clin Trials* 26:331–337.
48. Vera-Badillo FE, Shapiro R, Ocana A, Amir E, Tannock IF (2013) Bias in reporting of end points of efficacy and toxicity in randomized, clinical trials for women with breast cancer. *Ann Oncol* 24:1238–1244.
49. Amos AJ (2014) A review of spin and bias use in the early intervention in psychosis literature. *Prim Care Companion CNS Disord* 16:PCC.13r01586.
50. Arunachalam L, Hunter IA, Killeen S (2017) Reporting of randomized controlled trials with statistically nonsignificant primary outcomes published in high-impact surgical journals. *Ann Surg* 265:1141–1145.
51. Gewandter JS, et al. (2015) Data interpretation in analgesic clinical trials with statistically nonsignificant primary analyses: An ACTTION systematic review. *J Pain* 16:3–10.
52. Le Fourn E, Giraudeau B, Chosidow O, Doutre MS, Lorette G (2013) Study design and quality of reporting of randomized controlled trials of chronic idiopathic or autoimmune urticaria: Review. *PLoS One* 8:e70717.
53. Lockyer S, Hodgson R, Dumville JC, Cullum N (2013) “Spin” in wound care research: The reporting and interpretation of randomized controlled trials with statistically non-significant primary outcome results or unspecified primary outcomes. *Trials* 14:371.
54. Patel SV, Van Koughnett JA, Howe B, Wexner SD (2015) Spin is common in studies assessing robotic colorectal surgery: An assessment of reporting and interpretation of study results. *Dis Colon Rectum* 58:878–884.
55. Hernandez AV, et al. (2013) Deficient reporting and interpretation of non-inferiority randomized clinical trials in HIV patients: A systematic review. *PLoS One* 8:e63272.
56. Ioannidis JP (2005) Why most published research findings are false. *PLoS Med* 2:e124.
57. Boutron I, et al. (2014) Impact of spin in the abstracts of articles reporting results of randomized controlled trials in the field of cancer: The SPIIN randomized controlled trial. *J Clin Oncol* 32:4120–4126.
58. Yavchitz A, et al. (2012) Misrepresentation of randomized controlled trials in press releases and news coverage: A cohort study. *PLoS Med* 9:e1001308.
59. Phillips DP, Kanter EJ, Bednarczyk B, Tastad PL (1991) Importance of the lay press in the transmission of medical knowledge to the scientific community. *N Engl J Med* 325: 1180–1183.
60. Fornai F, et al. (2008) Lithium delays progression of amyotrophic lateral sclerosis. *Proc Natl Acad Sci USA* 105:2052–2057.
61. Abramson JD, Rosenberg HG, Jewell N, Wright JM (2013) Should people at low risk of cardiovascular disease take a statin? *BMJ* 347:f6123.
62. Malhotra A (2013) Saturated fat is not the major issue. *BMJ* 347:f6340.
63. Matthews A, et al. (2016) Impact of statin related media coverage on use of statins: Interrupted time series analysis with UK primary care data. *BMJ* 353:i3283.
64. Boyack KW, Klavans R, Sorensen AA, Ioannidis JP (2013) A list of highly influential biomedical researchers, 1996-2011. *Eur J Clin Invest* 43:1339–1365.
65. Emerson GB, et al. (2010) Testing for the presence of positive-outcome bias in peer review: A randomized controlled trial. *Arch Intern Med* 170:1934–1939.
66. National Academies of Sciences, Engineering, and Medicine (2017) *Fostering Integrity in Research* (The National Academies Press, Washington, DC).
67. Fanelli D (2012) Negative results are disappearing from most disciplines and countries. *Scientometrics* 90:891–904.
68. Kilkenny C, Browne WJ, Cuthill IC, Emerson M, Altman DG (2010) Improving bioscience research reporting: The ARRIVE guidelines for reporting animal research. *PLoS Biol* 8:e1000412.
69. Moher D, et al. (2010) CONSORT 2010 explanation and elaboration: Updated guidelines for reporting parallel group randomised trials. *BMJ* 340:c869.
70. Greenhalgh T (1995) Commentary: Scientific heads are not turned by rhetoric. *BMJ* 310:987–988.
71. John LK, Loewenstein G, Prelec D (2012) Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychol Sci* 23:524–532.
72. Moher D, et al. (2016) Increasing value and reducing waste in biomedical research: Who's listening? *Lancet* 387:1573–1586.
73. Munafò MR, et al. (2017) A manifesto for reproducible science. *Nat Hum Behav* 1:0021.
74. Pocock SJ, Ware JH (2009) Translating statistical findings into plain English. *Lancet* 373:1926–1928.
75. Falagas ME, Vergidis PI (2004) Addressing the limitations of structured abstracts. *Ann Intern Med* 141:576–577, and erratum (2005) 142:79.
76. Yavchitz A, Ravaud P, Hopewell S, Baron G, Boutron I (2014) Impact of adding a limitations section to abstracts of systematic reviews on readers' interpretation: A randomized controlled trial. *BMC Med Res Methodol* 14:123.

77. Wasserstein R, Lazar N (2016) The ASA's statement on p-values: Context, process, and purpose. *Am Stat* 70:129–133.
78. Lazarus C, et al. (2016) Peer reviewers identified spin in manuscripts of nonrandomized studies assessing therapeutic interventions, but their impact on spin in abstract conclusions was limited. *J Clin Epidemiol* 77:44–51.
79. Montori VM, et al. (2004) Users' guide to detecting misleading claims in clinical research reports. *BMJ* 329:1093–1096.
80. Drysdale H, Slade E, Goldacre B, Heneghan C; COMPare trials team (2016) Outcomes in the trial registry should match those in the protocol. *Lancet* 388:340–341.
81. Edwards MA, Roy S (2017) Academic research in the 21st century: Maintaining scientific integrity in a climate of perverse incentives and hypercompetition. *Environ Eng Sci* 34:51–61.
82. Alberts B, et al. (2015) SCIENTIFIC INTEGRITY. Self-correction in science at work. *Science* 348:1420–1422.
83. Alberts B, Kirschner MW, Tilghman S, Varmus H (2014) Rescuing US biomedical research from its systemic flaws. *Proc Natl Acad Sci USA* 111:5773–5777.
84. Alberts B, Kirschner MW, Tilghman S, Varmus H (2015) Opinion: Addressing systemic problems in the biomedical research enterprise. *Proc Natl Acad Sci USA* 112:1912–1913.
85. Ioannidis JP, Khoury MJ (2014) Assessing value in biomedical research: The PQRST of appraisal and reward. *JAMA* 312:483–484.
86. Kidwell MC, et al. (2016) Badges to acknowledge open practices: A simple, low-cost, effective method for increasing transparency. *PLoS Biol* 14:e1002456.
87. Treuille A, Das R (2014) Scientific rigor through videogames. *Trends Biochem Sci* 39: 507–509.