

# Reporting and Interpretation of Randomized Controlled Trials With Statistically Nonsignificant Results for Primary Outcomes

Isabelle Boutron, MD, PhD

Susan Dutton, MSc

Philippe Ravaud, MD, PhD

Douglas G. Altman, DSc

**A**CCURATE PRESENTATION OF the results of a randomized controlled trial (RCT) is the cornerstone of the dissemination of the results and their implementation in clinical practice. The Declaration of Helsinki states that “Authors have a duty to make publicly available the results of their research on human subjects and are accountable for the completeness and accuracy of their reports.” To help enforce this principle, trial registration is required,<sup>1</sup> and reporting guidelines are available.<sup>2</sup> However, investigators usually have broad latitude in writing their articles<sup>3</sup>; they can choose which data to report and how to report them.

Consequently, scientific articles are not simply reports of facts, and authors have many opportunities to consciously or subconsciously shape the impression of their results for readers, that is, to add “spin” to their scientific report.<sup>4</sup> Spin can be defined as specific reporting that could distort the interpretation of results and mislead readers.<sup>3,5,6</sup> The use of spin in scientific writing can result from ignorance of the scientific issue, unconscious bias, or willful intent to deceive.<sup>3</sup> Such distorted presentation and interpretation of trial results in published articles has

**Context** Previous studies indicate that the interpretation of trial results can be distorted by authors of published reports.

**Objective** To identify the nature and frequency of distorted presentation or “spin” (ie, 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) in published reports of randomized controlled trials (RCTs) with statistically nonsignificant results for primary outcomes.

**Data Sources** March 2007 search of MEDLINE via PubMed using the Cochrane Highly Sensitive Search Strategy to identify reports of RCTs published in December 2006.

**Study Selection** Articles were included if they were parallel-group RCTs with a clearly identified primary outcome showing statistically nonsignificant results (ie,  $P \geq .05$ ).

**Data Extraction** Two readers appraised each selected article using a pretested, standardized data abstraction form developed in a pilot test.

**Results** From the 616 published reports of RCTs examined, 72 were eligible and appraised. The title was reported with spin in 13 articles (18.0%; 95% confidence interval [CI], 10.0%-28.9%). Spin was identified in the Results and Conclusions sections of the abstracts of 27 (37.5%; 95% CI, 26.4%-49.7%) and 42 (58.3%; 95% CI, 46.1%-69.8%) reports, respectively, with the conclusions of 17 (23.6%; 95% CI, 14.4%-35.1%) focusing only on treatment effectiveness. Spin was identified in the main-text Results, Discussion, and Conclusions sections of 21 (29.2%; 95% CI, 19.0%-41.1%), 31 (43.1%; 95% CI, 31.4%-55.3%), and 36 (50.0%; 95% CI, 38.0%-62.0%) reports, respectively. More than 40% of the reports had spin in at least 2 of these sections in the main text.

**Conclusion** In this representative sample of RCTs published in 2006 with statistically nonsignificant primary outcomes, the reporting and interpretation of findings was frequently inconsistent with the results.

*JAMA.* 2010;303(20):2058-2064

www.jama.com

been highlighted in letters to editors criticizing the interpretation of results<sup>7</sup> and in methodological reviews evaluating misleading claims in published reports of RCTs<sup>8,9</sup> or systematic reviews.<sup>10</sup> However, to our knowledge, the strategies used to create spin in published articles have never been systematically assessed.

**Author Affiliations:** Centre for Statistics in Medicine, University of Oxford, Oxford, United Kingdom (Drs Boutron and Altman and Ms Dutton); INSERM, U738, Paris, France (Drs Boutron and Ravaud); Assistance Publique des Hôpitaux de Paris, Hôpital Hôtel Dieu, Centre d’Épidémiologie Clinique, Paris (Drs Boutron and Ravaud); and Université Paris Descartes, Faculté de Médecine, Paris (Drs Boutron and Ravaud).

**Corresponding Author:** Isabelle Boutron, MD, PhD, Centre d’Épidémiologie Clinique, Hôpital Hôtel Dieu, 1, Place du Parvis Notre-Dame, 75181 Paris CEDEX 4, France (isabelle.boutron@htd.aphp.fr).

We aimed to identify spin in reports of parallel-group RCTs with statistically nonsignificant results for the primary outcome and to develop a scheme for classification of spin strategies. We focused on trials with statistically nonsignificant primary outcomes because the interpretation of these results are more likely to be affected by a preconceived notion of effectiveness, resulting in a biased interpretation.<sup>9</sup>

## METHODS

### Selection of Articles

The articles were screened from a representative cohort of articles of RCTs indexed in PubMed. The search strategy and eligibility criteria for this cohort have been described elsewhere.<sup>11</sup> Randomized controlled trials were defined as prospective studies assessing health care interventions in human participants randomly allocated to study groups. Reports of cost-effectiveness studies, reports of diagnostic test accuracy, and non-English-language reports were excluded.

In brief, the Cochrane Highly Sensitive Search Strategy,<sup>12</sup> performed in PubMed to identify primary reports of RCTs published in December 2006 and indexed in PubMed by March 22, 2007, yielded 1735 PubMed citations. After reading the titles and abstracts of retrieved citations, reports of obviously noneligible trials were excluded, and the full-text article and any online appendices were obtained and evaluated for 879 selected citations. Of these, 263 citations were excluded after the full text was read; the remaining 616 were included in this representative sample of RCTs.

From this sample, we selected parallel-group RCTs with clearly identified primary outcomes. We excluded equivalence or noninferiority trials, crossover trials, cluster trials, factorial and split-body designs, trials with more than 2 groups, and phase 2 trials. Primary outcomes were those explicitly reported as such in the published article. If none was explicitly reported, we considered the outcomes stated in the sample size estimation; if outcomes were not stated in the sample size estimation, we took the

outcomes in the primary study objectives, if available. If no primary outcome was clearly identified (ie, explicitly specified in the article, in a sample size calculation, or in the primary study objectives), the article was excluded.

One reviewer (I.B.) screened the full-text articles and determined results for all primary outcomes according to statistical significance: results statistically significant (ie,  $P < .05$ ), results that did not reach statistical significance (ie,  $P \geq .05$ ), or unclear results. We included only trials with nonsignificant results (ie,  $P \geq .05$ ) for all primary outcomes. When no formal statistical analyses were reported for the primary outcomes, we attempted to calculate the effect size and confidence interval for the primary outcomes, and the article was included if the estimated treatment effect was not statistically significant. If we could not calculate the effect size using the published data, the article was excluded.

### Assessment of Selected Articles

For each selected article, 2 readers (I.B., S.D.) independently read the title, abstract, and Methods, Results, Discussion, and Conclusions sections, as well as online appendices referenced in the articles, when available. The reviewers independently appraised the content of the article using a pretested and standardized data abstraction form; then they met to compare results. All discrepancies were discussed to obtain consensus; if needed, the article was discussed with a third reader (D.G.A.). The reproducibility was moderate, with a  $\kappa$  of 0.47 (95% confidence interval [CI], 0.27-0.67) for presence of spin in the abstract Conclusions and of 0.64 (95% CI, 0.47-0.82) for spin in the article Conclusions.

### General Characteristics of Selected Articles

For each selected article, we recorded the funding source (ie, for-profit, non-profit, or both; not reported, no funding), 2007 journal impact factor, number of citations in 2008, the experimental intervention, comparator, sample size,

and type of primary outcomes (safety, efficacy, both).

### Reporting the Primary Outcomes in Abstract and Main Text

We checked whether the primary outcomes were clearly identified in the abstract. We also recorded the reporting of results for the primary outcomes both in the abstract and in the article (ie, reporting of estimated effect size with or without precision and reporting of summary statistics [eg, proportion of event, mean] for each group with or without precision).

### Definition of Spin

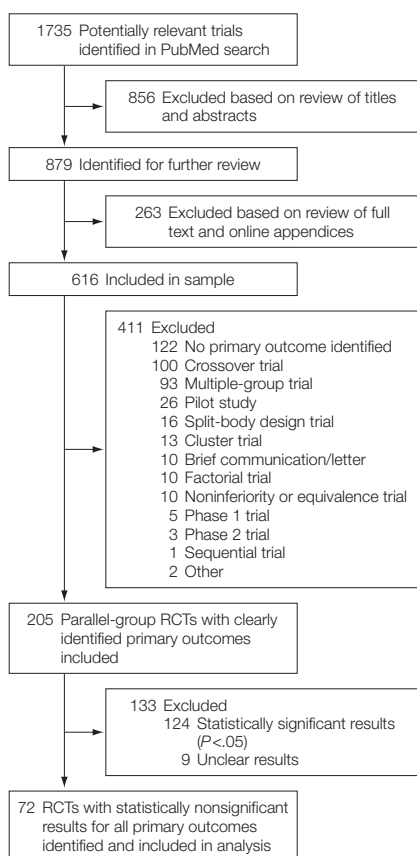
In the context of a trial with statistically nonsignificant primary outcomes, spin was defined as use of specific reporting strategies, from whatever motive, to highlight that the experimental treatment is beneficial, despite a statistically nonsignificant difference for the primary outcome, or to distract the reader from statistically nonsignificant results.

### Development of Classification Scheme

All of the authors participated in the development of a classification scheme to standardize the collection of the strategies used for spin in the selected reports. For this purpose, in a first step, we reviewed the literature published on this topic.<sup>3,6,13-22</sup> We also contacted by e-mail all the members of the Cochrane Statistical Method Group and invited them to send us any examples of published RCTs with spin, in any medical field, and with any publication date. Lastly, we reviewed a sample of trials with statistically nonsignificant results published in general medical journals with high impact factors or in specialist journals.<sup>23</sup> The classification scheme was developed following discussion and agreement among the authors.

### Strategies of Spin

Using the developed classification scheme, we searched for spin in each section of the manuscript in our sample, ie, abstract Results; abstract Conclu-

**Figure.** Study Selection

RCT indicates randomized controlled trial.

sions; and main-text Results, Discussion, and Conclusions (ie, last paragraph of the manuscript when this paragraph summarized the results) sections. We then determined whether authors had used a spin strategy. The strategies of spin considered were (1) a focus on statistically significant results (within-group comparison, secondary outcomes, subgroup analyses, modified population of analyses); (2) interpreting statistically nonsignificant results for the primary outcomes as showing treatment equivalence or comparable effectiveness; and (3) claiming or emphasizing the beneficial effect of the treatment despite statistically nonsignificant results. All other spin strategies that could not be classified according to this scheme were systematically recorded and secondarily classified.

### Extent of Spin

We determined the extent of spin across the whole report, defined as the number of sections with spin in the abstract (spin in the Results section only, in the Conclusions section only, or in both sections) and in the main text (spin in one section other than the Conclusions section only, in 2 sections, or in all 3 sections). The assessment of the extent of spin is exploratory and should not be considered a scoring system. This classification scheme was developed by consensus among the authors for a pragmatic purpose: to be able to capture the diversity of spin in terms of volume (ie, whether spin concerned only a small part or most of the article).

### Level of Spin in Conclusions

We also classified the level of spin in the Conclusions sections of the abstract and the main text as follows. *High spin* was defined as no uncertainty in the framing, no recommendations for further trials, and no acknowledgment of the statistically nonsignificant results for the primary outcomes; in addition, when the Conclusions section reported recommendations to use the treatment in clinical practice, we classified this section as having a high level of spin. *Moderate spin* was defined as some uncertainty in the framing or recommendations for further trials but no acknowledgment of the statistically nonsignificant results for the primary outcomes. *Low spin* was defined as uncertainty in the framing and recommendations for further trials or acknowledgment of the statistically nonsignificant results for the primary outcomes. This classification of the level of spin is exploratory and not validated and should not be considered a scoring system. The level of spin was used to explore the heterogeneity of spin in the reporting of conclusions.

### Statistical Analysis

Medians and interquartile ranges for continuous variables and number (%) of articles for categorical variables were calculated. Statistical analyses were per-

formed using SAS version 9.1 (SAS Institute Inc, Cary, North Carolina).

## RESULTS

### General Characteristics of Selected Articles

Of the 616 PubMed citations retrieved, 205 reports of parallel-group RCTs were identified. Among these reports, we identified and appraised 72 reports with statistically nonsignificant results for the primary outcomes (FIGURE). Characteristics of the included reports are presented in TABLE 1. Most reports evaluated efficacy ( $n=63$  [87.5%; 95% CI, 77.6%-94.1%]), and half evaluated pharmacological treatments. The funding source was for-profit (only or with a nonprofit source) in one-third of the reports and was not stated in 27 (37.5%).

### Reporting of Primary Outcomes in Abstract and Main Text

Primary outcomes were clearly identified in 44 of the 72 report abstracts (61.1%; 95% CI, 48.9%-72.4%). In 3 abstracts (4.2%; 95% CI, 0.9%-11.7%), a secondary outcome was reported as being the primary outcome. Only 9 abstracts (12.5%; 95% CI, 5.9%-22.4%) reported the effect size and 95% confidence interval, and 28 (38.9%; 95% CI, 27.6%-51.1%) did not report any numerical results for primary outcomes. In only 16 articles (22.2%; 95% CI, 13.3%-33.6%) did the main text describe the effect size and its precision for primary outcomes; in 21 (29.2%; 95% CI, 19.0%-41.1%), the main text reported only summary statistics for each group, without precision.

### Spin Strategies

The strategies of spin in each article section are shown in TABLE 2. The title was reported with spin in 13 of the 72 articles (18.0%; 95% CI, 10.0%-28.9%). Spin was identified in 27 (37.5%; 95% CI, 26.4%-49.7%) and 42 (58.3%; 95% CI, 46.1%-69.8%) of the abstract Results and Conclusions sections, respectively. We identified spin in 21 (29.2%; 95% CI, 19.0%-41.1%), 31 (43.1%; 95% CI, 31.4%-55.3%), and 36 (50.0%; 95% CI, 38.0%-62.0%) of the main-text Re-

sults, Discussion, and Conclusions sections, respectively.

The strategies of spin were also diverse (Table 2). Examples are provided in eTable 1, available at <http://www.jama.com>. In abstracts, spin consisted mainly of focusing on within-group comparison and subgroup analyses in the Results section. One-quarter of the abstract Conclusions sections focused on only the beneficial effect of treatment, claiming equivalence or comparable effectiveness (n=10 [13.9%; 95% CI, 6.9%-24.1%]), claiming efficacy (n=4 [5.6%; 95% CI, 1.5%-13.6%]), or focusing on only statistically significant results such as within-group, secondary outcome, or subgroup analyses (n=3 [4.2%; 95% CI, 0.9%-11.7%]). Furthermore, 9 abstract Conclusions sections (12.5%; 95% CI, 5.9%-22.4%) acknowledged statistically nonsignificant primary outcomes but focused on or emphasized statistically significant results.

Other specific strategies of spin were identified. In some reports in which primary outcomes concerned safety, authors interpreted statistically nonsignificant results as demonstrating lack of any difference in adverse events. As an example, the authors of one study concluded that “we have demonstrated (for the first time) that [with the treatment], embryo implantation is unaltered.” Some reports focused on an overall within-group comparison as if the trial planned was a before-after study, concluding, for example, that “the mean improvement . . . was clinically relevant in both treatment groups.”

Some authors focused on another objective to distract the reader from the statistically nonsignificant results, such as identifying a genetic prognostic factor of improvement.

### Extent of Spin

As shown in TABLE 3, the extent of spin varied. In total, 49 of the 72 abstracts (68.1%; 95% CI, 56.0%-78.6%) and 44 main texts (61.1%; 95% CI, 48.9%-72.4%) were classified as having spin in at least 1 section. More than 40% of the articles had spin in at least 2 sections of the main text. Spin was identified in all

sections of 20 abstracts (27.8%; 95% CI, 17.9%-39.6%) and 14 articles (19.4%; 95% CI, 11.1%-30.5%).

### Level of Spin in Conclusions

The level of spin in Conclusions sections is illustrated in Table 3, and examples are provided in eTable 2. We identified spin in more than half of the Conclusions sections; the level of spin was high (ie, no uncertainty in the framing, no recommendations for further trials, and no acknowledgment of the statistically nonsignificant results for the primary outcomes or recommendations to use the treatment in clinical practice) in 24 abstracts Conclusions sections (33.3%; 95% CI, 22.7%-45.4%) and 19 main-text Conclusions sections (26.4%; 95% CI, 16.7%-38.1%).

Examples of spin identified are presented in the eAppendix.

### COMMENT

This study appraised the strategies of spin used in reports of RCTs with statistically nonsignificant results for primary outcomes. We evaluated 72 reports selected from all reports of RCTs published in December 2006.<sup>11</sup> Spin used in the articles and their abstracts was common, but strategies used for spin varied. Furthermore, spin seemed more prevalent in article abstracts than in the main texts of articles.

Our results are consistent with those of other related studies showing a positive relation between financial ties and favorable conclusions stated in trial reports.<sup>24,25</sup> Other studies assessed discrepancies between results and their interpretation in the Conclusions sections.<sup>10,26</sup> Yank and colleagues<sup>10</sup> found that for-profit funding of meta-analyses was associated with favorable conclusions but not favorable results. Other studies have shown that the Discussion sections of articles often lacked a discussion of limitations.<sup>27</sup>

Our results add to these previous methodological reviews<sup>10,24-26</sup> in that ours was a systematic study of the use of inappropriate presentation in published trial reports, for which we propose a classification of the strategies authors use for

spin in their reports. Furthermore, unlike other studies<sup>10,24-26</sup> that investigated a specific category of journals, medical area, or category of treatment, ours took a representative sample.

**Table 1.** Report Characteristics

Characteristic	No. (%) [95% CI] (N = 72)
Journal type	
Specialty	69 (95.8) [88.3-99.1]
General medical	3 (4.2) [0.9-11.7]
Top 5 medical area	
Cardiology	9 (12.5) [5.9-22.4]
Obstetrics/gynecology	9 (12.5) [5.9-22.4]
Surgery	9 (12.5) [5.9-22.4]
Psychiatry/psychology	7 (9.7) [4.0-19.0]
Anesthesia	5 (6.9) [2.3-15.5]
Primary outcome	
Safety of treatment	7 (9.7) [4.0-19.0]
Efficacy of treatment	63 (87.5) [77.6-94.1]
Safety and efficacy	1 (1.4) [0.0-7.5]
Other (diagnosis accuracy)	1 (1.4) [0.0-7.5]
Primary outcome	
Binary	25 (34.7) [23.9-46.9]
Other	43 (59.7) [47.5-71.1]
Binary and other	4 (5.6) [1.5-13.6]
Sample size, median (IQR) [range]	84 (46-206) [4-6848]
Journal impact factor, median (IQR) [range]	2.9 (2.3-4.5) [0-52.6]
Number of citations in 2008, median (IQR) [range]	4 (1-7) [0-98]
Experimental treatment	
Drug	37 (51.4) [39.3-63.3]
Surgery/procedure	11 (15.3) [7.9-25.7]
Participative intervention (eg, rehabilitation, education)	7 (9.7) [4.0-19.0]
Device	7 (9.7) [4.0-19.0]
Other	10 (13.9) [6.9-24.1]
Comparator(s)	
Placebo	19 (26.4) [16.7-38.1]
Usual care	14 (19.4) [11.1-30.5]
Drug	18 (25.0) [15.5-36.6]
Surgery/procedure	7 (9.7) [4.0-19.0]
Participative intervention (eg, rehabilitation, education)	5 (6.9) [2.3-15.5]
Device	4 (5.6) [1.5-13.6]
Other	5 (6.9) [2.3-15.5]
Funding source	
None	2 (2.8) [0.3-9.7]
For-profit	16 (22.2) [13.3-33.6]
Nonprofit	19 (26.4) [16.7-38.1]
For-profit and nonprofit	8 (11.1) [4.9-20.7]
Not reported	27 (37.5) [26.4-49.7]

Abbreviations: CI, confidence interval; IQR, interquartile range.

We identified many strategies of spin. The most familiar and common approach was to focus on statistically significant results for other analyses, such as within-group comparisons, secondary outcomes, or subgroup analyses. An-

other common strategy was to interpret  $P > .05$  as demonstrating a similar effect when the study was not designed to assess equivalence or noninferiority (such trials require specific design and conduct, as well as a larger sample size, than

superiority trials). This dubious interpretation was used only when the comparator was an active treatment.

Some authors interpreted the trial results as being from a before-after study; they focused on within-group comparisons that were statistically significant for the experimental treatment but not for the comparator, which they incorrectly interpreted as demonstrating the beneficial effect of the treatment.<sup>28</sup> Some authors reported that they had demonstrated the beneficial effect of both treatments when the results showed a statistically significant change from baseline for each group or for both groups combined.<sup>20,29</sup> Some reports of safety trials provided an inadequate interpretation of the nonsignificant results by concluding lack of harm of the experimental treatment. Other methods relied on masking the nonsignificant results by focusing on other objectives. In one report, the authors statistically compared the experimental group, also not with the comparator in that trial, but rather with the placebo group of another trial to conclude that the treatment was better than placebo.

Lastly, our results highlight the important prevalence of spin in the abstract as compared with the main text of an article. These results have important implications, because readers often base their initial assessment of a trial on the information reported in an abstract. They may then use this information to decide whether to read the full report, if available. Furthermore, abstracts are freely available, and in some situations, clinical decisions might be made on the basis of the abstract alone.<sup>30</sup>

Our study has several limitations. First, the assessment of spin necessarily involved some subjectivity, because the strategies used for spin were highly variable and interpretation depended on the context. Interpretation of trial results is not a straightforward process, and some disagreement may arise, even among authors.<sup>31</sup> We attempted to limit this subjectivity by having 2 reviewers extract the data independently using a standardized data abstraction form, with any disagreements resolved by consensus. How-

**Table 2.** Spin in the Title, Abstract, and Main Text of Articles

Spin	No. (%) [95% CI]	
	Abstract (n = 72)	Main Text (n = 72)
Spin in the title	NA	13 (18.0) [10.0-28.9]
Spin in Results section	27 (37.5) [26.4-49.7]	21 (29.2) [19.0-41.1]
Focus on statistically significant within-group comparison <sup>a</sup>	8 (11.1) [4.9-20.7]	10 (13.9) [6.9-24.1]
Focus on statistically significant secondary outcomes	3 (4.2) [0.9-11.7]	0
Focus on statistically significant subgroup analyses	6 (8.3) [3.1-17.3]	4 (5.6) [1.5-13.6]
Focus on statistically significant modified population of analyses (eg, per-protocol analyses)	1 (1.4) [0.0-7.5]	1 (1.4) [0.0-7.5]
Focus on statistically significant within- and between-group comparisons for secondary outcomes	1 (1.4) [0.0-7.5]	0
Other spin only in the Results section	8 (11.1) [4.9-20.7]	6 (8.3) [3.1-17.3]
Spin in the synthesis of the results in Discussion section	NA	31 (43.1) [31.4-55.3]
Reporting of statistically nonsignificant outcome as if the trial were an equivalence trial	NA	6 (8.3) [3.1-17.3]
Focus on statistically significant secondary outcomes <sup>b</sup>	NA	3 (4.2) [0.9-11.7]
Focus on statistically significant subgroup analyses <sup>c</sup>	NA	4 (5.6) [1.5-13.6]
Focus on statistically significant modified population of analyses	NA	1 (1.4) [0.0-7.5]
Focus on overall within-group improvement	NA	9 (12.5) [5.9-22.4]
Ruling out adverse event	NA	3 (4.2) [0.9-11.7]
Other spin in the synthesis of the results	NA	7 (9.7) [4.0-19.0]
Spin in Conclusions section	42 (58.3) [46.1-69.8]	36 (50.0) [38.0-62.0]
Focus only on treatment effectiveness	17 (23.6) [14.4-35.1]	14 (19.4) [11.1-30.5]
Claiming equivalence for statistically nonsignificant results	10 (13.9) [6.9-24.1]	6 (8.3) [3.1-17.3]
Claiming efficacy with no consideration of the statistically nonsignificant primary outcome	4 (5.6) [1.5-13.6]	4 (5.6) [1.5-13.6] <sup>d</sup>
Focusing only on statistically significant results	3 (4.2) [0.9-11.7]	4 (5.6) [1.5-13.6]
Acknowledge statistically nonsignificant results for the primary outcome but emphasize the beneficial effect of treatment	3 (4.2) [0.9-11.7]	1 (1.4) [0.0-7.5]
Acknowledge statistically nonsignificant results for the primary outcome but emphasize other statistically significant results	6 (8.3) [3.1-17.3]	4 (5.6) [1.5-13.6]
Other spin in Conclusions section	16 (22.2) [13.3-33.6]	21 (29.2) [19.0-41.1]
Conclusion ruling out an adverse event on statistically nonsignificant results	3 (4.2) [0.9-11.7]	3 (4.2) [0.9-11.7]
Conclusion focusing on within-group assessment (both treatments are effective/treatment administered in both groups is effective (eg, add-on studies))	6 (8.3) [3.1-17.3]	9 (12.5) [5.9-22.4]
Recommendation to use the treatment	3 (4.2) [0.9-11.7]	3 (4.2) [0.9-11.7] <sup>e</sup>
Focus on another objective	2 (2.8) [0.3-9.7]	3 (4.2) [0.9-11.7]
Comparison with placebo group of another trial	1 (1.4) [0.0-7.5]	0
Statistically nonsignificant subgroup results reported as beneficial	1 (1.4) [0.0-7.5]	1 (1.4) [0.0-7.5]
Other	0	2 (2.8) [0.3-9.7]

Abbreviations: CI, confidence interval; NA, not applicable.

<sup>a</sup>All these articles also reported the statistically nonsignificant results for the primary outcome in the abstract and in the main text.

<sup>b</sup>One article also associated with other spin strategies.

<sup>c</sup>One article associated with particular focus on overall within-group analyses.

<sup>d</sup>One article associated with other spin (focuses on overall within-group assessment).

<sup>e</sup>Three articles associated respectively with particular focus on overall within-group analyses, acknowledgment of negative primary outcome plus focus on positive secondary outcomes, and focus only on statistically significant results.

ever, to our knowledge, no objective measure exists for the subjective component of interpretation.<sup>32</sup> Consequently, to be completely transparent, a detailed summary of all the examples of spin we classified is available in the eAppendix.

We dichotomized trial findings as positive or negative using an arbitrary value ( $P = .05$ ) as a significance threshold. However, we acknowledge that the interpretation of RCTs should not be based solely on the arbitrary  $P$  value of .05 dichotomizing findings as positive or negative.

We focused on spin only in trials for which the primary outcomes were clearly defined and results for the primary outcomes were not statistically significant. This focus implies that the strategies identified may not be applicable to all reports of RCTs and that other strategies of spin may not have been identified. Furthermore, when the results of an RCT are not statistically significant, the risk of spin may be increased. Trialists and sponsors are rarely neutral regarding the results of their trial. They may have invested considerable time, energy, and money in developing the experimental intervention and expended much effort in planning and conducting the trial. Therefore, they may have a strong preconception about the beneficial effect of the experimental intervention. Furthermore, the results of the trial could have important implications at different levels, eg, for the publication of the trial results in terms of delay and type of journal<sup>33</sup>; for the use of the experimental treatment in clinical practice; and, consequently, for future career advancement or profit.<sup>34,35</sup> A trial with statistically nonsignificant results will thus frequently be a disappointment and could lead to subconscious or even deliberate intent to mislead the reader when presenting and interpreting the trial results.<sup>32,36</sup> Few authors have studied this phenomenon, but Hewitt and colleagues reviewed a panel of 17 trial reports with nonsignificant results published in *BMJ*. They found that, despite evidence that the treatment might be ineffective, in 3 trials the authors seemed to support the experimental intervention.<sup>9</sup>

**Table 3.** Extent of Spin in the Abstract and Main Text of Articles and Level of Spin in Conclusions Sections

Spin	No. (%) [95% CI]	
	Abstract (n = 72)	Main Text (n = 72)
Extent of spin		
None	23 (31.9) [21.4-44.0]	28 (38.9) [27.6-51.1]
In 1 section other than Conclusions section	7 (9.7) [0.4-19.0]	5 (6.9) [2.3-15.5]
In the Conclusions section only	22 (30.6) [20.2-42.5]	9 (12.5) [5.9-22.4]
In 2 sections	NA	16 (22.2) [13.3-33.6]
In all sections	20 (27.8) [17.9-39.6]	14 (19.4) [11.1-30.5]
Level of spin in Conclusions section		
None	30 (41.7) [30.2-53.9]	36 (50.0) [38.0-62.0]
Low <sup>a</sup>	9 (12.5) [5.9-22.4]	8 (11.1) [4.9-20.7]
Moderate <sup>b</sup>	9 (12.5) [5.9-22.4]	9 (12.5) [5.9-22.4]
High <sup>c</sup>	24 (33.3) [22.7-45.4]	19 (26.4) [16.7-38.1]

Abbreviations: CI, confidence interval; NA, not applicable.

<sup>a</sup>Spin in the Conclusions section with acknowledgment of the statistically nonsignificant results for the primary outcome or spin in the Conclusions section with no acknowledgment of the statistically nonsignificant results for the primary outcome but reported with uncertainty and recommendations for further trials.

<sup>b</sup>Spin in the Conclusions section with no acknowledgment of the statistically nonsignificant results for the primary outcome but reported with uncertainty or recommendations for further trials.

<sup>c</sup>Spin in the Conclusions section with no uncertainty, no recommendations for further trials, and no acknowledgment of the statistically nonsignificant results for the primary outcome or spin with recommendation to use the treatment in clinical practice.

We focused on only some categories of spin, and other forms of spin may not have been identified. For example, we did not consider some specific strategies of spin, such as authors obscuring the risk associated with the experimental treatment, as reported in the published Vioxx GI Outcomes Research (VIGOR) study. That report concealed the cardiovascular risk by presenting the hazard of myocardial infarction as if the comparator (ie, naproxen) were the intervention group, concluding on the protective effect of the comparator (relative risk, 0.2; 95% CI, 0.1-0.7)<sup>37</sup> instead of the harmful effect of the experimental treatment (ie, rofecoxib) (relative risk, 5.00; 95% CI, 1.68-20.13).<sup>38</sup>

We cannot say to what extent the spin we identified might have been deliberately misleading, the result of lack of knowledge, or both. Nor are we able to draw conclusions about the possible effect of the spin on peer reviewers' and readers' interpretations. Studies evaluating the effect of framing on clinical practice have focused on the reporting of treatment-effect estimates and showed inconsistent results.<sup>39,40</sup>

Our study has identified many different strategies that authors use to provide a biased interpretation of results

of RCTs with statistically nonsignificant results for primary outcomes. Peer and editorial reviewers must be aware of the different strategies of spin used to temper the article text. The choice of analyses reported (statistically significant analyses such as subgroup analyses or within-group analyses) and the terms used to report and interpret results are important in a scientific article. Special attention should be paid to inadequate interpretation of the trial results, particularly when authors conclude on efficacy from secondary outcomes, subgroup analyses, or within-group comparisons or when the authors inadequately interpret lack of difference as demonstrating equivalence in terms of safety or efficacy. The publication process in biomedical research tends to favor statistically significant results and to be responsible for "optimism bias" (ie, unwarranted belief in the efficacy of a new therapy).<sup>41</sup> Reports of RCTs with statistically significant results for outcomes are published more often and more rapidly than are those of trials with statistically nonsignificant results.<sup>34,42</sup> Good evidence exists of selective reporting of statistically significant results for outcomes in published articles.<sup>33,43-46</sup>

In conclusion, in this representative sample of RCTs indexed in PubMed and published in December 2006 with statistically nonsignificant primary outcomes, the reporting and interpretation of findings was frequently inconsistent with the results. However, this work is only a first step, and future research is needed. Determining which category and level of spin affect readers' interpretation is important. Future research on the reasons for and the mechanisms of spin would also be useful. We hope that highlighting this issue may lead to more vigilance by peer reviewers and editors to reduce the use of these questionable strategies, which can distort the interpretation of research findings.

**Author Contributions:** Dr Boutron had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

**Study concept and design:** Boutron, Dutton, Ravaud, Altman.

**Acquisition of data:** Boutron, Dutton.

**Analysis and interpretation of data:** Boutron, Dutton, Ravaud, Altman.

**Drafting of the manuscript:** Boutron.

**Critical revision of the manuscript for important intellectual content:** Dutton, Ravaud, Altman.

**Statistical analysis:** Boutron.

**Financial Disclosures:** None reported.

**Funding/Support:** Dr Boutron was supported by a grant from the Societe Francaise de Rhumatologie (SFR) and that Lavoisier Program (Ministère des Affaires étrangères et européennes).

**Role of the Sponsors:** The SFR and the Lavoisier Program (Ministère des Affaires étrangères et européennes) had no role in the design and conduct of the study; the collection, management, analysis, and interpretation of the data; or the preparation, review, or approval of the manuscript.

**Online-Only Material:** eTable 1, eTable 2, and the eAppendix are available at <http://www.jama.com>.

**Additional Contributions:** We are very grateful to Ly-Mee Yu, MSc (Center for Statistics in Medicine, University of Oxford, Oxford, United Kingdom), for her important work in developing the database of the representative reports of randomized controlled trials indexed in PubMed. Ms Yu received no compensation for her contributions.

## REFERENCES

- DeAngelis CD, Drazen JM, Frizelle FA, et al; International Committee of Medical Journal Editors. Clinical trial registration: a statement from the International Committee of Medical Journal Editors. *JAMA*. 2004;292(11):1363-1364.
- Altman DG, Schulz KF, Moher D, et al; CONSORT GROUP (Consolidated Standards of Reporting Trials). The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med*. 2001;134(8):663-694.
- Fletcher RH, Black B. "Spin" in scientific writing: scientific mischief and legal jeopardy. *Med Law*. 2007;26(3):511-525.
- Junger D. The rhetoric of research: embrace scientific rhetoric for its power. *BMJ*. 1995;311(6996):61.
- Bailar JC. How to distort the scientific record without actually lying: truth, and arts of science. *Eur J Oncol*. 2006;11(4):217-224.
- Marco CA, Larkin GL. Research ethics: ethical issues of data reporting and the quest for authenticity. *Acad Emerg Med*. 2000;7(6):691-694.
- Hróbjartsson A, Gøtzsche PC. Powerful spin in the conclusion of Wampold et al.'s re-analysis of placebo versus no-treatment trials despite similar results as in original review. *J Clin Psychol*. 2007;63(4):373-377.
- Jefferson T, Di Pietrantonj C, Debalini MG, Rivetti A, Demicheli V. Relation of study quality, concordance, take home message, funding, and impact in studies of influenza vaccines: systematic review. *BMJ*. 2009;338:b354.
- Hewitt CE, Mitchell N, Torgerson DJ. Listen to the data when results are not significant. *BMJ*. 2008;336(7634):23-25.
- Yank V, Rennie D, Bero LA. Financial ties and concordance between results and conclusions in meta-analyses: retrospective cohort study. *BMJ*. 2007;335(7631):1202-1205.
- Hopewell S, Dutton S, Yu LM, Chan AW, Altman DG. The quality of reports of randomised trials in 2000 and 2006: comparative study of articles indexed in PubMed. *BMJ*. 2010;340:c723.
- Robinson KA, Dickersin K. Development of a highly sensitive search strategy for the retrieval of reports of controlled trials using PubMed. *Int J Epidemiol*. 2002;31(1):150-153.
- Horton R. The rhetoric of research. *BMJ*. 1995;310(6985):985-987.
- Boutron I, Guittet L, Estellat C, Moher D, Hróbjartsson A, Ravaud P. Reporting methods of blinding in randomized trials assessing nonpharmacological treatments. *PLoS Med*. 2007;4(2):e61.
- Blader JC. Can keeping clinical trial participants blind to their study treatment adversely affect subsequent care? [published online ahead of print March 3, 2005]. *Contemp Clin Trials*. 2005;26(3):290-299. doi:10.1016/j.cct.2005.01.003.
- Al-Marzouki S, Roberts I, Marshall T, Evans S. The effect of scientific misconduct on the results of clinical trials: a Delphi survey. *Contemp Clin Trials*. 2005;26(3):331-337.
- Gøtzsche PC. Believability of relative risks and odds ratios in abstracts: cross sectional study. *BMJ*. 2006;333(7561):231-234.
- Jørgensen KJ, Johansen HK, Gøtzsche PC. Flaws in design, analysis and interpretation of Pfizer's antifungal trials of voriconazole and uncritical subsequent quotations. *Trials*. 2006;7:3.
- Hoekstra R, Finch S, Kiers HA, Johnson A. Probability as certainty: dichotomous thinking and the misuse of p values. *Psychon Bull Rev*. 2006;13(6):1033-1037.
- Zinsmeister AR, Connor JT. Ten common statistical errors and how to avoid them. *Am J Gastroenterol*. 2008;103(2):262-266.
- Pocock SJ, Ware JH. Translating statistical findings into plain English. *Lancet*. 2009;373(9679):1926-1928.
- Wang R, Lagakos SW, Ware JH, Hunter DJ, Drazen JM. Statistics in medicine—reporting of subgroup analyses in clinical trials. *N Engl J Med*. 2007;357(21):2189-2194.
- Mathieu S, Boutron I, Moher D, Altman DG, Ravaud P. Comparison of registered and published primary outcomes in randomized controlled trials. *JAMA*. 2009;302(9):977-984.
- Rattinger G, Bero L. Factors associated with results and conclusions of trials of thiazolidinediones. *PLoS One*. 2009;4(6):e5826.
- Als-Nielsen B, Chen W, Gluud C, Kjaergard LL. Association of funding and conclusions in randomized drug trials: a reflection of treatment effect or adverse events? *JAMA*. 2003;290(7):921-928.
- Jørgensen AW, Hilden J, Gøtzsche PC. Cochrane reviews compared with industry supported meta-analyses and other meta-analyses of the same drugs: systematic review. *BMJ*. 2006;333(7572):782.
- Ioannidis JP. Limitations are not properly acknowledged in the scientific literature. *J Clin Epidemiol*. 2007;60(4):324-329.
- Matthews JN, Altman DG. Interaction 2: compare effect sizes not P values. *BMJ*. 1996;313(7060):808.
- Moyer CA. Between-groups study designs demand between-groups analyses: a response to Hernandez-Reif, Shor-Posner, Baez, Soto, Mendoza, Castillo, Quintero, Perez, and Zhang. *Evid Based Complement Alternat Med*. 2009;6(1):49-50.
- Hopewell S, Clarke M, Moher D, et al; CONSORT Group. CONSORT for reporting randomized controlled trials in journal and conference abstracts: explanation and elaboration. *PLoS Med*. 2008;5(1):e20.
- Horton R. The hidden research paper. *JAMA*. 2002;287(21):2775-2778.
- Kaptchuk TJ. Effect of interpretive bias on research evidence. *BMJ*. 2003;326(7404):1453-1455.
- Dwan K, Altman DG, Arnaiz JA, et al. Systematic review of the empirical evidence of study publication bias and outcome reporting bias. *PLoS One*. 2008;3(8):e3081.
- Rising K, Bacchetti P, Bero L. Reporting bias in drug trials submitted to the Food and Drug Administration: review of publication and presentation. *PLoS Med*. 2008;5(11):e217.
- Bero L, Oostvogel F, Bacchetti P, Lee K. Factors associated with findings of published trials of drug-drug comparisons: why some statins appear more efficacious than others. *PLoS Med*. 2007;4(6):e184.
- Woloshin S, Schwartz LM, Casella SL, Kennedy AT, Larson RJ. Press releases by academic medical centers: not so academic? *Ann Intern Med*. 2009;150(9):613-618.
- Bombardier C, Laine L, Reicin A, et al; VIGOR Study Group. Comparison of upper gastrointestinal toxicity of rofecoxib and naproxen in patients with rheumatoid arthritis. *N Engl J Med*. 2000;343(21):1520-1528.
- Krumholz HM, Ross JS, Presler AH, Egilman DS. What have we learnt from Vioxx? *BMJ*. 2007;334(7585):120-123.
- McGettigan P, Sly K, O'Connell D, Hill S, Henry D. The effects of information framing on the practices of physicians. *J Gen Intern Med*. 1999;14(10):633-642.
- Bucher HC, Weinbacher M, Gyr K. Influence of method of reporting study results on decision of physicians to prescribe drugs to lower cholesterol concentration. *BMJ*. 1994;309(6957):761-764.
- Chalmers I, Matthews R. What are the implications of optimism bias in clinical research? *Lancet*. 2006;367(9509):449-450.
- Turner EH, Matthews AM, Linardatos E, Tell RA, Rosenthal R. Selective publication of antidepressant trials and its influence on apparent efficacy. *N Engl J Med*. 2008;358(3):252-260.
- Chan AW, Altman DG. Identifying outcome reporting bias in randomised trials on PubMed: review of publications and survey of authors. *BMJ*. 2005;330(7494):753.
- Chan AW, Hróbjartsson A, Haahr MT, Gøtzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA*. 2004;291(20):2457-2465.
- Al-Marzouki S, Roberts I, Evans S, Marshall T. Selective reporting in clinical trials: analysis of trial protocols accepted by *The Lancet*. *Lancet*. 2008;372(9634):201.
- Song F, Parekh S, Hooper L, et al. Dissemination and publication of research findings: an updated review of related biases. *Health Technol Assess*. 2010;14(8):iii, ix-xi, 1-193.