

A new classification of spin in systematic reviews and meta-analyses was developed and ranked according to the severity

Amélie Yavchitz^{a,b,c,*}, Philippe Ravaud^{a,b,c,d}, Douglas G. Altman^e, David Moher^{f,g},
Asbjørn Hrobjartsson^{h,i}, Toby Lasserson^j, Isabelle Boutron^{a,b,c}

^aCentre de Recherche Épidémiologie et Statistique Sorbonne Paris Cité (CRESS-UMR1153), Inserm/Université Paris Descartes, 1 place du Parvis Notre Dame, Paris 75004, France

^bCentre d'Épidémiologie Clinique, AP-HP (Assistance Publique des Hôpitaux de Paris), Hôpital Hôtel Dieu, 1 place du Parvis Notre Dame, Paris, France

^cFrench Cochrane Center, Paris Descartes University, Sorbonne Paris Cité, Faculté de Médecine, 1 place du Parvis Notre Dame, Paris 75004, France

^dDepartment of Epidemiology, Columbia University Mailman School of Public Health, New York, NY, USA

^eCentre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK

^fClinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, Canada

^gSchool of Epidemiology, Public health and Preventive Medicine, Faculty of Medicine, University of Ottawa, Canada

^hNordic Cochrane Centre, Rigshospitalet Department 7811, Copenhagen, Denmark

ⁱResearch Unit for Evidence-Based Medicine, Odense University Hospital/University of Southern Denmark, Odense, Denmark

^jCochrane Editorial Unit, London, UK

Accepted 19 January 2016; Published online 2 February 2016

Abstract

Objectives: We aimed to (1) identify and classify spin (i.e., a description that overstates efficacy and/or understates harm) in systematic reviews and (2) rank spin in abstracts of systematic reviews according to their severity (i.e., the likelihood of distorting readers' interpretation of the results).

Study Design: First, we used a four-phase consensus process to develop a classification of different types of spin. Second, we ranked the types of spin in abstracts according to their severity using a Q-sort survey with members of the Cochrane Collaboration.

Results: We identified 39 types of spin, 28 from the main text and 21 from the abstract; 13 were specific to the systematic review design. Spin was classified into three categories: (1) misleading reporting, (2) misleading interpretation, and (3) inappropriate extrapolation. Spin ranked as the most severe by the 122 people who participated in the survey were (1) recommendations for clinical practice not supported by findings in the conclusion, (2) misleading title, and (3) selective reporting.

Conclusion: This study allowed for identifying spin that is likely to distort interpretation. Our classification could help authors, editors, and reviewers avoid spin in reports of systematic reviews. © 2016 Elsevier Inc. All rights reserved.

Keywords: Bias; Data interpretation; Distortion; Spin; Systematic reviews; Classification

1. Background

Clinical and health-policy decision making should be based on the best evidence. Systematic reviews and meta-analyses are the cornerstones of therapeutic evaluation because they summarize all available evidence. Clinicians, decision makers, and researchers often use systematic reviews and meta-analyses (henceforth referred to as systematic reviews) to keep up to date with the medical literature, develop clinical practice guidelines, and sometimes plan future research [1,2]. Patients can also use systematic reviews to help inform treatment options. An accurate presentation and interpretation of results of systematic reviews is therefore essential for appropriate dissemination and application of results. However, authors

Conflict of interest: None.

Funding: A.Y. was funded by a grant from the Fondation pour la Recherche Médicale (FDT20140930875). D.M. is funded through a University Research Chair.

Transparency declaration: A.Y. affirms that this article is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned have been explained [55].

* Corresponding author. 1, Place du Parvis Notre-Dame, Paris Cedex 4 75181, France. Tel.: +1-33-0-1-42-34-78-33; fax: +1-33-0-1-42-34-87-90.

E-mail address: amelie.yavchitz@aphp.fr (A. Yavchitz).

What is new?**Key findings**

- We identified 39 types of spin: 28 could occur in the main text and 21 in abstracts; 13 were specific to the systematic review design.

What this adds to what was known?

- The ranking of spin in abstracts according to their severity highlighted that the most severe spin were (1) recommendations for clinical practice not supported by findings in the conclusion, (2) misleading title, and (3) selective reporting.

What is the implication and what should change now?

- This classification of spin could help editors and reviewers identify spin and reduce misinterpretation of review findings and also help authors avoid spin when reporting a systematic review.

usually have broad latitude in writing their reports of scientific results [3], with the possibility of adding “spin” [4].

The term spin was used in the medical literature for the first time by Horton [5] about the rhetoric of research. Horton defined spin as “the conscious and unconscious tricks of authorial rhetoric.” Then, the term spin was used in the first study that systematically assessed distorted interpretation in published articles [4]. In this study, spin included intentional or nonintentional misrepresentation of study findings.

The presence of spin in scientific reports was systematically studied initially in reports of randomized controlled trials (RCTs) [4,6,7]. Spin has been noted in about 60% of abstract conclusions in RCT reports [4], and its presence was found to affect clinicians’ interpretation of the research findings [8]: clinicians were more likely to rate a treatment as beneficial despite the primary outcome being statistically nonsignificant. Recent studies have described and proposed classifications of spin for other study designs, such as diagnostic accuracy studies [9], nonrandomized studies, or observational studies [10–14]. These studies have shown that the use of spin by authors is diverse and very prevalent and common, particularly in the abstract conclusions [4,7,9,15,16]. Furthermore, spin in published articles is disseminated to the public through press releases and news items [6,17] and leads to misinterpretation of the study findings by readers [8].

Authors have highlighted and described some cases of the distorted interpretation of review findings [18] or misleading reporting that could bias readers’ interpretation of the results [19,20]. Despite the importance of systematic reviews [21,22], the interpretation of their findings is still

problematic [23–27]. Identifying spin in reports is a subjective task, and a classification of spin with a clear definition of the different types of spin would be useful for authors and reviewers to aid in identifying spin in reports. We are not aware of a previously published classification scheme for spin for systematic reviews.

Abstracts are a primary source of dissemination of research findings. Readers often base their initial assessment of research on the information reported in an abstract, and in some geographic areas, the abstract may be all that health professionals have easy access to [28]. The title and abstract conclusions deliver the simplest take-home message that is easy to memorize. The presentation and interpretation of study results specifically in the title and abstract is therefore essential.

We aimed to (1) develop a classification scheme for spin in reports of systematic reviews and (2) rank types of spin that could occur in abstracts of systematic reviews according to their perceived severity (i.e., the likelihood of distorting readers’ interpretation).

2. Methods

First, in a four-phase consensus process, we identified, by literature review and interviews, potential types of spin in systematic reviews of the effects of therapeutic interventions and developed a classification scheme of spin. Second, we ranked types of spin in abstracts according to their severity by using a Q-sort survey with invited members of the Cochrane Collaboration.

2.1. Definition

We defined spin as a specific way of reporting, intentional or not, to highlight that the beneficial effect of the experimental treatment in terms of efficacy or safety is greater than that shown by the results (i.e., overstate efficacy and/or understate harm). In this definition, we did not consider spin that understated the beneficial effect of the experimental intervention in terms of safety and efficacy (i.e., understate efficacy and/or overstate harm) because these types of spin were not considered in previous works about spin for other study designs [4,7,9,11]. Furthermore, to allow an accurate assessment of the severity of the spin by participants, it was important to focus on a homogeneous definition. Spin is a part of a larger concept of biased reporting where some important elements are omitted, modified, or reported in a misleading manner by authors.

We defined the severity of a spin as the perceived likelihood of distorting readers’ interpretation when reading the abstract. The different types of spin could be more or less likely to distort readers’ interpretation. For example, a selective reporting of outcomes favoring the beneficial effect of a treatment could be more misleading for readers than the use of linguistic spin.

2.2. Development of a classification scheme of spin in systematic reviews and meta-analyses

We developed a classification scheme of spin in systematic reviews of therapeutic interventions in a four-phase consensus process.

All the authors of this study had expertise in conducting and reporting systematic reviews and participated in the development of the classification scheme of spin in systematic reviews.

2.2.1. Phase 1

We performed a literature review about misleading, biased, misconducted, or misinterpreted systematic reviews to identify potential types of spin that could overstate efficacy or understate harm. We searched (1) important methodological work on spin and extracted potential types of spin that could be apply to systematic reviews, (2) recommendations on how a systematic review should be reported and interpreted and determined how this could be spin (For this purpose, we searched the Cochrane Handbook, part 2: “general methods for Cochrane review” [29] and the Methodological standards for the conduct of Cochrane Intervention Review version 2 [30]), (3) letters, comments, or editorials about the interpretation of systematic review findings, which could be a potential strategy of spin, in two databases, MEDLINE and the Cochrane Methodology Register, and (4) main articles with recommendations about systematic reviews and meta-analysis interpretation [31–37]. More details are available in the Appendix at www.jclinepi.com.

2.2.2. Phase 2

Two authors (A.Y. and I.B.) classified potential types of spin retrieved in phase 1 according to their location in the article (i.e., spin that could occur in the main text of the report or that could occur in the abstract or the title of the report).

2.2.3. Phase 3

Interviews face to face or by conference call were organized with all the authors. After a presentation of the concept of spin, the authors (A.Y. and I.B.) presented the list of potential spin types developed in phase 2. Then, in an open exchange, we discussed each potential type of spin, during which clarifications, rewording, opinions, justifications, or new ideas were expressed.

2.2.4. Phase 4

The list of spin types was refined, taking into account the comments made in phase 3. Then, we obtained a classification scheme of spin for systematic reviews and meta-analyses of therapeutic interventions. We resent the classification scheme for approval to all the authors.

2.3. Ranking spin in the abstract according to its perceived severity: an online survey

2.3.1. Study design

We planned an online survey in which participants were asked to rank different types of spin in the article abstract (including the title) according to their perceived severity.

The survey focused on spin in the title and abstract because these sections are the most widely read and often the only freely available part of the report. They are the first and sometimes the only source of information for readers [28,38].

2.3.2. Participants

We invited the following members of the Cochrane Collaboration to participate in the survey: (1) the coordinating editors in any review or method group, (2) the members of the Methodology Review Group and Bias Method Group, (3) all editors of the six review groups that published the largest number of reviews between 2010 and 2013: musculoskeletal, pregnancy and childbirth, anesthesia, neonatal, airway and pain, palliative and supportive care groups. The survey occurred between February 27 and April 10, 2015, and 476 people were invited to take part in the survey.

2.3.3. Survey

A standardized and personalized e-mail was used to invite all potential participants to participate in a survey to sort spin in article abstracts according to the perceived severity of the spin defined as the likelihood that the spin distorted the interpretation of the results. The e-mail is available in the Appendix at www.jclinepi.com. We sent one or two reminders in absence of a response to the previous e-mail.

Participants ranked spin by using the Q-sort survey. This research method, first described by William Stephenson in 1935, combines qualitative and quantitative methods and provides a scientific foundation for studying subjectivity or a person's viewpoint [39–42]. In a Q-sort survey, people are shown a sample of statements on a topic and asked to rank-order them from their point of view according to some preference, judgment, or feeling about each statement.

Ranking involved three steps. Each type of spin was reported on a virtual “card.” First, participants had to read the cards carefully and divide them into three groups according to their severity: (1) less severe, (2) medium severity, and (3) more severe. Second, participants were asked to take the cards from each group (less severe/medium severity/more severe), read them again, and place each of them in one of seven bins on a score sheet ranging from -3 , less severe, to $+3$, more severe. Once they placed all cards on the score sheet, they could review the distribution once more and shift the cards one final time. Before distribution, the survey was pilot-tested with a convenience sample of 20 epidemiologists to determine whether the process and cards were

understandable. We performed the survey using FlashQ open source Software 1.0, the Flash application for performing Q-sorting online [43]. The Fig. 1 shows an example of sorting.

Details, with screenshots of the online survey, are available in the Appendix at www.jclinepi.com.

2.4. Analysis

Means and standard deviations were used to describe quantitative variables and frequencies and percentages to describe categorical variables. Spin was classified and ranked in descending order by the mean score given by participants. Statistical analysis involved use of SAS, version 9.3 (SAS Inst. Inc., Cary, NC, USA).

3. Results

3.1. Classification scheme of spin in systematic reviews and meta-analyses

3.1.1. Identification of types of spin in systematic reviews

From our literature review and consensus process, we identified 39 different types of spin: 28 relevant to the main text of a report and 21 in the report's abstract and title. From the 39 types of spin, 13 (33%) were specific to systematic reviews (Table 1), and 26 (67%) could occur in other study designs. For example, the extrapolation of study findings from a surrogate marker to the global improvement of the disease could occur in systematic reviews as well as in RCTs, but the inadequate focus on primary study results instead of meta-analysis results is specific to systematic reviews that include meta-analysis.

3.1.2. Classification of spin

We classified the types of spin into three main categories: (1) misleading reporting, (2) misleading interpretation, and (3) inappropriate extrapolation (Table 1).

- 1) Misleading reporting was defined as incomplete or inadequate reporting of the methods, study analysis, study results, or any important information that could be misleading to the reader. The misleading reporting of systematic reviews included 10 items in the main text and eight items in the abstract (Table 1), such as the selective reporting of efficacy outcome favoring the beneficial effect of the experimental intervention or the lack of clarity over the amount of information summarized across outcomes in a review. For example, in a systematic review published in 2014 [44], the abstract failed to mention that among the 36 trials included in the systematic review, only eight studies (22%) contributed to the meta-analysis of the primary outcome.
- 2) Misleading interpretation was defined as an interpretation of the study results that could be misleading to

the reader. This category contained 13 types of spin for the main text and 10 for the abstract. This category included the interpretation by authors of a non-statistically significant result with a wide confidence interval for harm outcomes as being safe. This category also included the inadequate claim of efficacy of an intervention, despite the high risk of publication bias or heterogeneity. For example, in the systematic review [45] comparing albumin and crystalloids in seriously ill patients, the authors concluded that “albumin is a safe treatment,” but the pooled relative risk for death was 1.11 with a confidence interval from 0.95 to 1.28 for all patients receiving albumin, 1.12 (95% CI, 0.85–1.46) for surgery or trauma patients, and 1.76 (95% CI, 0.91–2.78) for patients with burns. The authors claimed the safety of albumin based on nonstatistically significant results with a wide confidence interval.

- 3) Inappropriate extrapolation was defined as an inappropriate generalization of study results. This category contained five types of spin for the main text and three for the abstract, including, for example, the extrapolation of the results for a surrogate outcome, such as a biological outcome, to the global improvement of the disease or extrapolating the review's findings to a different intervention (i.e., claiming efficacy of one specific intervention although the review covers a class of several interventions). For example, in the systematic review assessing the efficacy of biologics vs. placebo for treatment of moderate-to-severe plaque psoriasis, the authors conclude on the efficacy of one specific drug instead of a class of intervention—biologic treatments [46].

3.2. Ranking of spin in abstracts of systematic reviews and meta-analysis according to their severity

3.2.1. Participants

Among the 476 people invited to take part in the survey, 169 (35%) clicked the link at least once and 123 (73%) of these participated in the Q-sort survey. We excluded one participant who acknowledged a lack of expertise or particular knowledge about systematic reviews and their interpretation. Half of the participants were located in Europe; 89 (73%) were affiliated with a review group of the Cochrane Collaboration and 51 (42%) with a methods group. The participants had a high level of expertise in systematic reviews; most ($n = 107$, 90%) had been involved at least once in a systematic review as an author (Table 2).

3.2.2. The most severe types of spin

Among the 21 types of spin that could occur in the abstract of the systematic reviews (Table 3), the seven types rated as the most severe were (1) the conclusion formulates recommendations for clinical practice not supported by the

Table 1. Classification of spin in the main text and abstracts of systematic reviews and meta-analyses

Main category	Description of spin in the main text	Description of spin in abstract
Misleading reporting	<ol style="list-style-type: none"> (1) Failure to acknowledge a departure from protocol that could modify the interpretation of results (2) Selective reporting of or overemphasis on efficacy outcomes favoring the beneficial effect of the experimental intervention (e.g., secondary outcomes, subgroup analyses) (3) Selective reporting of or overemphasis on harm outcomes favoring the safety of the experimental intervention (4) No or inadequate reporting of the limitations of the systematic review (5) Selective citation of articles in favor of the beneficial effect of the experimental intervention (6) Authors hide or do not present any conflict of interest (7) Conclusion focusing selectively on statistically significant efficacy outcome (8) <i>Selective reporting of analysis favoring the beneficial effect of the experimental intervention (e.g., selective analysis using random or fixed effect according the results)</i> (9) <i>Inadequate focus on the results of primary studies favoring the beneficial effect of the experimental intervention instead of the meta-analysis results</i> (10) <i>Changing the scale of the forest plot to magnify the results (diamond size)</i> 	<ol style="list-style-type: none"> (1) Selective reporting of or overemphasis on efficacy outcomes favoring the beneficial effect of the experimental intervention (2) Selective reporting of or overemphasis on harm outcomes favoring the safety of the experimental intervention (3) Failure to report a wide confidence interval of estimates (4) Authors hide or do not present any conflict of interest (5) Inadequate focus on the results of primary studies favoring the beneficial effect of the experimental intervention instead of the meta-analysis results (6) Conclusion focusing selectively on statistically significant efficacy outcome (7) <i>Failure to report the number of studies/patients actually contributing to the analysis for main outcomes</i> (8) <i>Failure to specify the direction of the effect when it favors the control intervention</i>
Misleading interpretation	<ol style="list-style-type: none"> (11) Title claims or suggests a beneficial effect of the experimental intervention not supported by the findings (12) Inadequate interpretation of nonstatistically significant results (with a wide confidence interval) as a lack of effect or an equivalent effect for efficacy outcomes (13) Inadequate interpretation of nonstatistically significant results (with a wide confidence interval) as demonstrating safety for harm outcome (14) Inadequate focus on <i>P</i>-value instead of magnitude of the effect estimates for harm or efficacy outcome (15) Focus on relative effect when the absolute effect is small (16) Misleading interpretation of cited articles, favoring the beneficial effect of the experimental intervention (17) Conclusion claiming equivalence or comparable effectiveness for nonstatistically significant results with a wide confidence interval (18) Conclusion formulating recommendations for clinical practice not supported by the findings (19) Conclusion claiming safety based on nonstatistically significant results with a wide confidence interval (20) <i>Conclusion ignoring the high risk of bias of the studies, the heterogeneity or the reporting bias (i.e., a low level of evidence) in the interpretation of the results</i> (21) <i>No or inadequate consideration of heterogeneity in results interpretation (i.e., no assessment of heterogeneity reported, claiming the absence of heterogeneity not supported by the data, claiming the beneficial effect of the treatment despite high heterogeneity, no downgrading the evidence in cases of high heterogeneity, interpreting nonstatistical significant results for the test of heterogeneity as an evidence of no heterogeneity etc)</i> (22) <i>No or inadequate consideration of the risk of bias of primary studies included in results interpretation</i> 	<ol style="list-style-type: none"> (9) Title claims or suggests a beneficial effect of the experimental intervention not supported by the findings (10) Inadequate focus on <i>P</i>-value instead of magnitude of the effect estimates for harm or efficacy outcome (11) Focus on relative effect when the absolute effect is small (12) Conclusion claiming equivalence or comparable effectiveness for nonstatistically significant results with a wide confidence interval (13) Conclusion formulating recommendations for clinical practice not supported by the findings (14) Conclusion claiming safety based on nonstatistically significant results with a wide confidence interval (15) <i>Conclusion claiming the beneficial effect of the experimental treatment despite high risk of bias in primary studies</i> (16) <i>Conclusion claiming the beneficial effect of the experimental treatment despite reporting bias</i> (17) <i>Conclusion claiming the beneficial effect of the experimental treatment despite high heterogeneity</i> (18) <i>Ignoring that the review included different study design (e.g., controlled trial or observational studies)</i>

(Continued)

Table 1. Continued

Main category	Description of spin in the main text	Description of spin in abstract
	<i>(i.e., no reporting of the risk of bias of the primary studies, claiming the low risk of bias of studies included not supported by the data, no downgrading the evidence despite several high risk of bias studies)</i>	
	(23) <i>No or inadequate consideration of reporting bias in results interpretation (i.e., no reporting of an assessment of reporting bias, claiming efficacy despite an evidence of reporting bias, claiming the absence of reporting bias not supported by the data, negative test result interpreted as absence of publication bias, use of the test without the condition of validity, inadequate interpretation of a funnel plot, etc)</i>	
Inappropriate extrapolation	(24) Inadequate extrapolation of the results from surrogate markers or specific outcome to the global improvement of the disease	(19) Conclusion extrapolating the review's findings to a different population or setting
	(25) Inadequate extrapolation of the results to a larger population, a larger setting or a wider set of interventions (e.g., from a specific rehabilitation program to all rehabilitation programs, to a specialized unit to a nonspecialized medical unit etc)	(20) Conclusion extrapolating the review's findings to a different intervention (i.e., claiming efficacy of one specific intervention although the review covers a class of several interventions)
	(26) Conclusion extrapolating the review's findings to a different population or setting	(21) Conclusion extrapolating the review's findings from a surrogate marker or a specific outcome to the global improvement of the disease
	(27) Conclusion extrapolating the review's findings to a different intervention (i.e., claiming efficacy of one specific intervention although the review covers a class of several interventions)	
	(28) Conclusion extrapolating the review's findings from a surrogate marker or a specific outcome to the global improvement of the disease	

Spin retained in italics is specific to systematic review and meta-analysis designs.

Table 2. General characteristics of the expert participants in the survey

Characteristics	Total, N = 122
Location	
Europe	58 (49)
Canada	24 (20)
Oceania	15 (13)
United States	12 (10)
Asia	7 (6)
South America	2 (2)
Cochrane group member	
Primary review group	69 (56)
Methods group	51 (42)
Both	20 (16)
No answer	2 (2)
No. of systematic reviews involved in	
0	12 (10)
1–3	54 (45)
4–6	97 (31)
7–10	5 (4)
>10	11 (9)
No. of systematic reviews peer reviewed	
0	16 (13)
1–3	46 (39)
4–6	29 (24)
7–10	12 (10)
>10	16 (13)

Data are no. (%).

findings, (2) the title claims or suggests a beneficial effect of the experimental intervention not supported by the findings, (3) selective reporting of or overemphasis on efficacy outcomes or analysis favoring the beneficial effect of the experimental intervention, (4) the conclusion claims safety based on nonstatistically significant results with a wide confidence interval, (5) the conclusion claims the beneficial effect of the experimental treatment despite a high risk of bias in primary studies, (6) selective reporting of or overemphasis on harm outcomes or analysis favoring the safety of the experimental intervention, and (7) the conclusion extrapolates the review findings to a different intervention (e.g., claiming efficacy of one specific intervention although the review covered a class of several interventions). The distribution of responses for the Q-sort survey for each type of spin is available in the [Appendix at www.jclinepi.com](http://www.jclinepi.com).

4. Discussion

To our knowledge, our study is the first to provide a classification scheme of spin in systematic reviews. We identified 28 types of spin in the main text and 21 in the abstract that were classified in three categories: (1) misleading reporting, (2) misleading interpretation, and (3) inappropriate extrapolation. The ranking of spin in abstract according to their severity highlighted that the most severe were

Table 3. Spin in abstracts ranked by severity, Q-sort results

Label	Mean rating of severity	Rank
Conclusion contains recommendations for clinical practice not supported by the findings	1.83 (1.18)	1
Title claims or suggests a beneficial effect of the experimental intervention not supported by the findings	1.43 (1.49)	2
Selective reporting of or overemphasis on efficacy outcomes or analysis favoring the beneficial effect of the experimental intervention	0.75 (1.20)	3
Conclusion claims safety based on nonstatistically significant results with a wide confidence interval	0.46 (1.18)	4
Conclusion claims the beneficial effect of the experimental treatment despite high risk of bias in primary studies	0.43 (1.32)	5
Selective reporting of or overemphasis on harm outcomes or analysis favoring the safety of the experimental intervention	0.31 (1.45)	6
Conclusion extrapolates the review's findings to a different intervention (i.e., claiming efficacy of one specific intervention although the review covers a class of several interventions)	0.19 (1.27)	7
Conclusion extrapolates the review's findings from a surrogate marker or a specific outcome to the global improvement of the disease	0.16 (1.42)	8
Conclusion claims the beneficial effect of the experimental treatment despite reporting bias	−0.12 (1.33)	9
Authors hide or do not present any conflict of interest	−0.16 (1.78)	10
Conclusion focuses selectively on statistically significant efficacy outcome	−0.17 (1.54)	11
Conclusion claims equivalence or comparable effectiveness for nonstatistically significant results with a wide confidence interval	−0.19 (1.40)	12
Failure to specify the direction of the effect when it favors the control intervention	−0.2 (1.62)	13
Failure to report a wide confidence interval of estimates	−0.29 (1.28)	14
Conclusion extrapolates the review's findings to a different population or setting	−0.43 (1.43)	15
Ignores that the review included different study design (e.g., controlled trial or observational studies)	−0.44 (1.40)	16
Conclusion claims the beneficial effect of the experimental treatment despite high heterogeneity	−0.51 (1.24)	17
Inadequate focus on the results of primary studies favoring the beneficial effect of the experimental intervention instead of the meta-analysis results	−0.63 (1.55)	18
Ignores the number of studies/patients actually contributing to the analysis for main outcomes	−0.69 (1.39)	19
Focus on <i>P</i> -value instead of magnitude of the effect estimates for harm or efficacy outcome	−0.78 (1.20)	20
Focus on relative effect when the absolute effect is small	−0.96 (1.34)	21

Data are mean (SD) rating, ranging from −3, less severe, to +3, more severe.

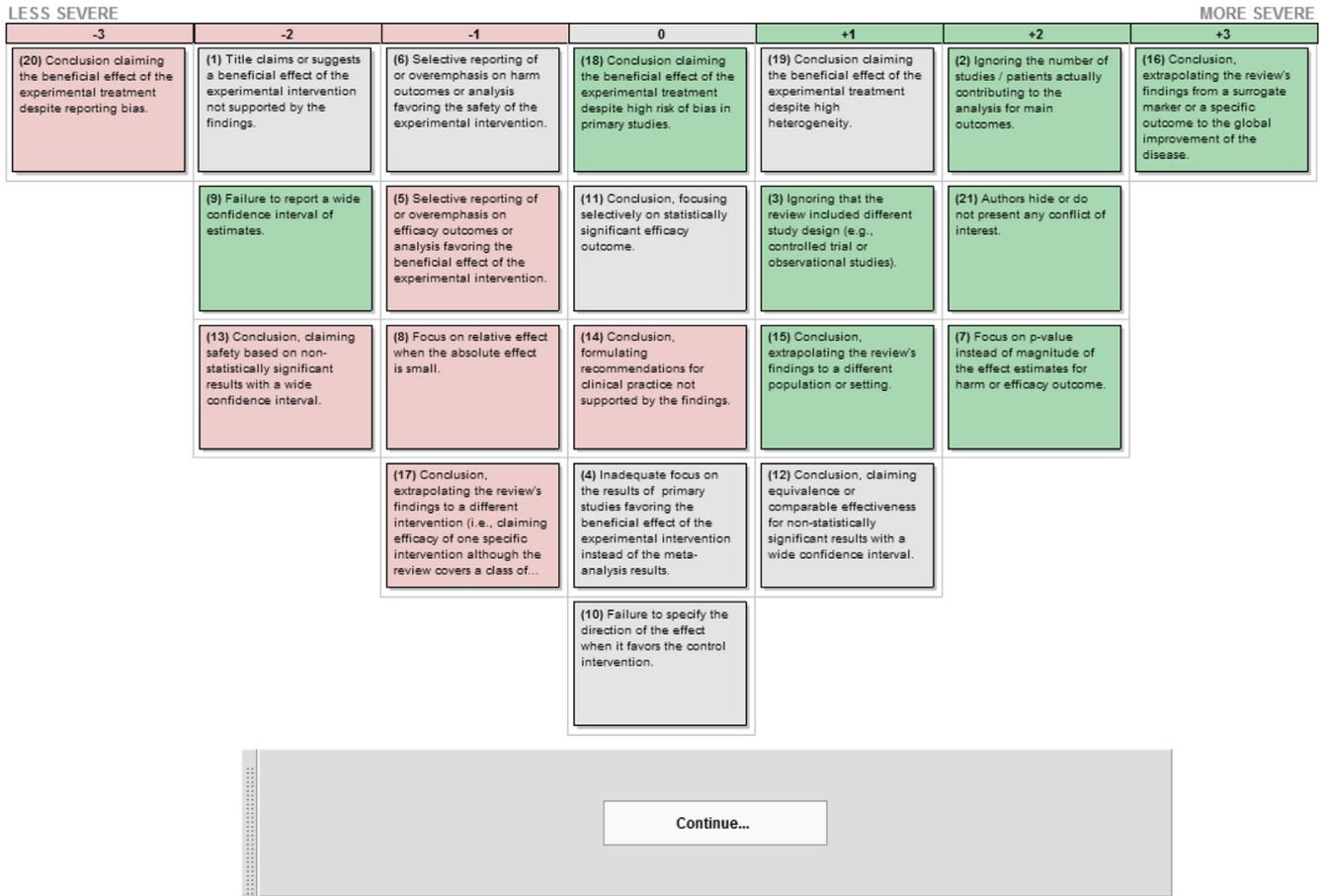
(1) recommendations for clinical practice not supported by findings in the conclusion, (2) misleading title, and (3) selective reporting.

In recent years, the biomedical research community has widely focused on transparency and reporting quality. The development of reporting guidelines such as the CONSORT [47] Statement for randomized trials and PRISMA Statement [31,48] for systematic reviews has been a major step forward in the improvement of scientific reporting. Implementation of these guidelines by journals should have an important impact on the quality of reporting, although evidence for this remains slight [49]. These guidelines are evidence based and provide a minimum set of recommendations to facilitate clarity, completeness, and transparency of reports of systematic reviews. However, the PRISMA Statement provides limited guidance on interpreting results. Most systematic reviews include many analyses. The authors must select which outcomes and analyses to emphasize in the text and summarize in the abstract. When those outcomes have been prioritized after protocol, their use to support the conclusions could bias readers' interpretation of the results. Reviewers and editors may not be sufficiently aware or trained to identify spin.

Moreover, review findings are difficult to interpret [50–52]. Systematic reviews are supposed to overcome personal opinions inherent to narrative reviews by setting out the ground rules for the study and then following them, but in fact, they leave much room for opinion and

misleading presentation or interpretation of findings. Previous studies showed that generating a conclusion is difficult, whatever the expertise of the authors. For example, in the study by Lai et al. [52], participants (hospital practitioners attending an evidence-based medicine course and final-year medical students) were asked to choose an appropriate conclusion for the abstracts of four Cochrane reviews. Most participants were unable to generate the appropriate conclusion in terms of the direction of the effect and strength of evidence. The interpretation of review findings is even more difficult for a consumer audience. Glenton et al. described the difficulties in generating a plain-language summary understandable by the public [53]. Therefore, systematic review findings are particularly subject to misinterpretation.

Previous studies have developed classifications for spin and showed that the prevalence is troubling [4,9]. However, we are not aware of previous research on the mechanism of spin. Spin in scientific publications may arise from (1) ignorance of scientific standards, (2) young researchers' imitation of previous practice, (3) unconscious prejudice, or (4) willful intent to influence readers [3]. This type of practice can be further enhanced by the current reward system, which unfortunately focuses on the number of publications (“publish or perish”) [54] rather than their quality. We need research to understand why and how authors use spin in scientific reporting (i.e., the mechanism of spin) and to assess its impact on readers' interpretation.



Pink, -3 to -1 severity; grey, 0 severity; green, +1 to +3 severity.

Fig. 1. Example of sorting. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Our study has limitations. First, the ranking of spin in the report abstract reflects the severity perceived by individuals with expertise in the field of systematic reviews; we did not evaluate the real impact of spin on the interpretation of findings by readers. However, the participants, all editors of the Cochrane Collaboration, were experts. Second, the participants ranked the spin according to their severity in theory but not in the context of a real systematic review. Thus, this ranking did not include the fact that the severity of a spin depends on not just the strategy used by the authors but also how far they depart from a fair presentation of the study as well as the consequences to patients. For example, concealing the cardiovascular adverse effects of a drug seems more severe than omitting a side effect such as nausea or constipation. Third, we focused on spin in systematic reviews of therapeutic interventions. Spin may differ in other systematic reviews such as reviews of diagnostic or prognostic studies. Fourth, the ranking of spin is not the fruit of consensus, and there is some disagreement among participants on how severe spin is. Fifth, the response rate to the survey (26%) was low. Finally, our

classification is not exhaustive. We may not have identified all types of spin.

Our classification of spin may have a role in preventing and detecting spin that could occur in systematic reviews. This classification of spin could raise awareness among authors and help journal editors and peer reviewers focus attention on identifying spin in key parts of systematic reviews and therefore reduce the potential to misrepresent research findings. Future research is needed to assess the frequency of spin in systematic reviews and meta-analyses and its impact on readers' interpretation of results.

Acknowledgments

The authors thank Laura Smales (BioMedEditing, Toronto, Canada) for English language proofreading.

The authors thank all participants of the survey who agreed to be named: Roland Bachter, Ruth Brassington, Mohammed T. Ansari, Evan Mayo-Wilson, Yoon K. Loke, Wilco Jacobs, Mariska M.G. Leeflang, Lucy Turner,

Jennifer Petkovic, Andrea Tricco, Veronica Pitt, Marialena Trivella, Donna Coffin, Sean Grant, Barbara Mintzes, Jarn Wetterslev, Jan Odgaard-Jensen, Julian Higgins, Emma Sydenham, Roberta Scherer, Sunita Vohra, Jordi Pardo Pardo, Matthew Page, Henrica C.W. de Vet, Willi Sauerbrei, Carlos Tajar, Ian Shemilt, Angela Webster, Sarah Nolan, Nicky Cullum, Marie-Pierre Gagnon, Emma Welsh, Tammy Clifford, Steve McDonald, Gustavo Zanoli, Larissa Shamseer, Jelena Savovic, Jennifer Evans, Ian Shrier, Nicole Skoetz, Hayley Jones, Chris Cates, P. Lina Santaguida, M. Willson, Katrina Williams, Philippa Middleton, Bernd Richter, Jo Weldon, Norma Terrin, Fulvia Baldassarre, Miranda Cumpston, Patrick Archambault, Joseph Mathew, Renea Johnston, Vivian Welch, Wieland, Deborah Caldwell, Kieran, Jessie McGowan, Steff Lewis, Yali Liu, An-Wen Chan, Nuala Livingstone, Eccleston, Karmela Krleza-Jeric, Merisih Mahmić-Kaknjo, Rebecca Turner, Nader Shaikh, Jasvinder Singh, Elizabeth Tanjong Ghogomu, Harald Herkner, Nathan Pace, Declan Devane, Rachele Buchbinder, Barney Reeves, Matthias Briel, Miranda Langendam, Guido Schwarzer, Martin Taphoorn, Jani Ruotsalainen, Stephen Morley, Raymond Daniel, Jaime Miranda, Neil O'Connell, Peter Herbison, M.J. Van Den Bent, Amy Drahota, Kerry Dwan, Michael Bracken, Mia Schmidt-Hansen, Jim Neilson, Kristin Danko, David Haas, Mathew Zacharias, Susan Chang, Janet Wale, Westeel, Gill Gyte, Khadra Galaal, Ewan McNicol, Michael Bennett, Lara Maxwell, Raj Naik, Lee Hooper, Anne Lethaby, Kristin Carson, Erika Ota, Amanda C. de C. Williams, William B. Hood, Lelia Duley, Sally Spencer, Jesus Lopez Alcalde.

Authors' contributions: A.Y., I.B., D.M., D.G.A., T.L., A.H., and P.R. conceived and designed the study and contributed to the experiments performed. A.Y. and I.B. wrote the first draft. A.Y., I.B., D.M., D.G.A., T.L., A.H., and P.R. contributed to the writing of the article. Isabelle Pane, senior computer engineer, Center de recherche Épidémiologies et Biostatistiques, INSERM U1153, Paris, France, contributed to online survey design and data management. Gabriel Baron, senior statistician, Center de recherche Épidémiologies et Biostatistiques, INSERM U1153, Paris, France, analyzed the data. A.Y. and I.B. had access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Supplementary data

Supplementary data related to this article can be found at <http://dx.doi.org/10.1016/j.jclinepi.2016.01.020>.

References

- [1] Cook DJ, Mulrow CD, Haynes RB. Systematic reviews: synthesis of best evidence for clinical decisions. *Ann Intern Med* 1997;126:376–80.
- [2] Bastian H, Glasziou P, Chalmers I. Seventy-five trials and eleven systematic reviews a day: how will we ever keep up? *PLoS Med* 2010;7:e1000326.
- [3] Fletcher RH, Black B. “Spin” in scientific writing: scientific mischief and legal jeopardy. *Med Law* 2007;26:511–25.
- [4] Boutron I, Dutton S, Ravaud P, Altman DG. Reporting and interpretation of randomized controlled trials with statistically nonsignificant results for primary outcomes. *JAMA* 2010;303:2058–64.
- [5] Horton R. The rhetoric of research. *BMJ* 1995;310:985–7.
- [6] Yavchitz A, Boutron I, Bafeta A, Marroun I, Charles P, Mantz J, et al. Misrepresentation of randomized controlled trials in press releases and news coverage: a cohort study. *PLoS Med* 2012;9:e1001308.
- [7] Lockyer S, Hodgson R, Dumville JC, Cullum N. “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* 2013;14:371.
- [8] Boutron I, Altman DG, Hopewell S, Vera-Badillo F, Tannock I, Ravaud P. 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* 2014;32:4120–6.
- [9] Ochodo EA, de Haan MC, Reitsma JB, Hooft L, Bossuyt PM, Loeffel MMG. Overinterpretation and misreporting of diagnostic accuracy studies: evidence of “spin”. *Radiology* 2013;267:581–8.
- [10] Prasad V, Jorgenson J, Ioannidis JPA, Cifu A. Observational studies often make clinical practice recommendations: an empirical evaluation of authors' attitudes. *J Clin Epidemiol* 2013;66:361–366.e4.
- [11] Lazarus C, Haneef R, Ravaud P, Boutron I. Classification and prevalence of spin in abstracts of non-randomized studies evaluating an intervention. *BMC Med Res Methodol* 2015;15:85.
- [12] Cofield SS, Corona RV, Allison DB. Use of causal language in observational studies of obesity and nutrition. *Obes Facts* 2010;3:353–6.
- [13] Brown AW, Brown MMB, Allison DB. Belief beyond the evidence: using the proposed effect of breakfast on obesity to show 2 practices that distort scientific evidence. *Am J Clin Nutr* 2013;98:1298–308.
- [14] Tzoulaki I, Liberopoulos G, Ioannidis JPA. Assessment of claims of improved prediction beyond the Framingham risk score. *JAMA* 2009;302:2345–52.
- [15] Cummings P, Rivara FP. Spin and boasting in research articles. *Arch Pediatr Adolesc Med* 2012;166:1099–100.
- [16] Hróbjartsson A, Gøtzsche PC, 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;63:373–7.
- [17] Sumner P, Vivian-Griffiths S, Boivin J, Williams A, Venetis CA, Davies A, et al. The association between exaggeration in health related science news and academic press releases: retrospective observational study. *BMJ* 2014;349:g7015.
- [18] Ezzo J, Bausell B, Moerman DE, Berman B, Hadhazy V. Reviewing the reviews. How strong is the evidence? How clear are the conclusions? *Intl J Technol Assess Health Care* 2001;17:457–66.
- [19] Lau J, Ioannidis JPA, Terrin N, Schmid CH, Olkin I. The case of the misleading funnel plot. *BMJ* 2006;333:597–600.
- [20] Cook D, Guyatt G. Colloid use for fluid resuscitation: evidence and spin. *Ann Intern Med* 2001;135:205–8.
- [21] Clarke M, Hopewell S, Chalmers I. Clinical trials should begin and end with systematic reviews of relevant evidence: 12 years and waiting. *Lancet* 2010;376:20–1.
- [22] Chalmers I, Bracken MB, Djulbegovic B, Garattini S, Grant J, Gülmezoglu AM, et al. How to increase value and reduce waste when research priorities are set. *Lancet* 2014;383:156–65.
- [23] Weed DL. Meta-analysis and causal inference: a case study of benzene and non-Hodgkin lymphoma. *Ann Epidemiol* 2010;20:347–55.

- [24] Weed DL. Interpreting epidemiological evidence: how meta-analysis and causal inference methods are related. *Int J Epidemiol* 2000;29:387–90.
- [25] Bender R, Bunce C, Clarke M, Gates S, Lange S, Pace NL, et al. Attention should be given to multiplicity issues in systematic reviews. *J Clin Epidemiol* 2008;61:857–65.
- [26] 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;358:252–60.
- [27] Ioannidis JPA, Patsopoulos NA, Rothstein HR. Reasons or excuses for avoiding meta-analysis in forest plots. *BMJ* 2008;336:1413–5.
- [28] Islamaj Dogan R, Murray GC, Névéol A, Lu Z. Understanding PubMed user search behavior through log analysis. *Database (Oxford)* 2009;2009:bap018.
- [29] Higgins JPT, Green S, editors. *Cochrane handbook for systematic reviews of interventions version 5.1.0 [updated March 2011]*. The Cochrane Collaboration. Available at: www.cochrane-handbook.org.
- [30] Methodological expectations of Cochrane Intervention Reviews (MECIR): Standards for the conduct and reporting of new Cochrane Intervention Reviews 2012. Version 2.0, 2013. Available at: www.editorial-unit.cochrane.org/mecir.
- [31] Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ* 2009;339:b2700.
- [32] Sterne JA, Egger M, Smith GD. Systematic reviews in health care: investigating and dealing with publication and other biases in meta-analysis. *BMJ* 2001;323:101–5.
- [33] Langhorne P. Bias in meta-analysis detected by a simple, graphical test. Prospectively identified trials could be used for comparison with meta-analyses. *BMJ* 1998;316:471.
- [34] Gabler NB, Duan N, Liao D, Elmore JG, Ganiats TG, Kravitz RL. Dealing with heterogeneity of treatment effects: is the literature up to the challenge? *Trials* 2009;10:43.
- [35] Lijmer JG, Bossuyt PMM, Heisterkamp SH. Exploring sources of heterogeneity in systematic reviews of diagnostic tests. *Stat Med* 2002;21:1525–37.
- [36] Kent DM, Rothwell PM, Ioannidis JPA, Altman DG, Hayward RA. Assessing and reporting heterogeneity in treatment effects in clinical trials: a proposal. *Trials* 2010;11:85.
- [37] Sterne JAC, Sutton AJ, Ioannidis JPA, Terrin N, Jones DR, Lau J, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ* 2011;343:d4002.
- [38] Saint S, Christakis DA, Saha S, Elmore JG, Welsh DE, Baker P, et al. Journal reading habits of internists. *J Gen Intern Med* 2000;15:881–4.
- [39] Stephenson W. Correlating persons instead of tests. *J Pers* 1935;4:17–24.
- [40] van Exel J, de Graaf G. Q methodology—a sneak preview. *Soc Sci Med* 2005;2:1–30.
- [41] Brown SR. Q-Methodology and quality research. *Qual Health Res* 1996;6:561–7.
- [42] Dennis KE. Q methodology: relevance and application to nursing research. *ANS Adv Nurs Sci* 1986;8:6–17.
- [43] Hackert C, Braehler G. FlashQ software [Computer software] 2007. Available at: <http://www.hackert.biz/flashq/home>. Accessed February 18, 2016.
- [44] Punjasawadwong Y, Phongchiewboon A, Bunchungmongkol N. Bispectral index for improving anaesthetic delivery and postoperative recovery. *Cochrane Database Syst Rev* 2014;CD003843.
- [45] Wilkes MM, Navickis RJ. Patient survival after human albumin administration: a meta-analysis of randomized, controlled trials. *Ann Intern Med* 2001;135:149–64.
- [46] Puig L, López A, Vilarrasa E, García I. Efficacy of biologics in the treatment of moderate-to-severe plaque psoriasis: a systematic review and meta-analysis of randomized controlled trials with different time points. *J Eur Acad Dermatol Venereol* 2013;28:1633–53.
- [47] Schulz KF, Altman DG, Moher D. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *J Clin Epidemiol* 2010;63:834–40.
- [48] Beller EM, Glasziou PP, Altman DG, Hopewell S, Bastian H, Chalmers I, et al. PRISMA for abstracts: reporting systematic reviews in journal and conference abstracts. *PLoS Med* 2013;10:e1001419.
- [49] Turner L, Shamseer L, Altman DG, Schulz KF, Moher D. Does use of the CONSORT Statement impact the completeness of reporting of randomised controlled trials published in medical journals? A Cochrane review. *Syst Rev* 2012;1:60.
- [50] Henrotin YE. Interpretation of metaanalyses: pitfalls should be more widely recognized. *J Rheumatol* 2012;39:1107–9.
- [51] Ioannidis JP. Meta-research: the art of getting it wrong. *Res Synth Methods* 2010;1:169–84.
- [52] Lai NM, Teng CL, Lee ML. Interpreting systematic reviews: are we ready to make our own conclusions? A cross-sectional study. *BMC Med* 2011;9:30.
- [53] Glenton C, Santesso N, Rosenbaum S, Nilsen ES, Rader T, Ciapponi A, et al. Presenting the results of Cochrane Systematic Reviews to a consumer audience: a qualitative study. *Med Decis Making* 2010;30:566–77.
- [54] Weed DL. Preventing scientific misconduct. *Am J Public Health* 1998;88:125–9.
- [55] Altman D, Moher D. Declaration of transparency for each research article. *BMJ* 2013;347:2–4.