Distortion in the Communication of Nonsignificant Primary Outcomes: The Spin Strategy in Multiple Sclerosis Trials

Marta Mascareñas-García, MD, MPH [1,2] Alejandro Rivero-de-Aguilar, MD [1,3] Cristina Candal-Pedreira, PhD [1,4,5] Guadalupe García, MPH [1] Carla Guerra-Tort, MPH [1] Lucía Martín-Gisbert, PhD [1] Julia Rey-Brandariz, PhD [1,4,5] Mónica Pérez-Ríos Ríos, PhD [1,4,5] Beatriz Casal-Acción, BA [6] María Isolina Santiago-Pérez, BS [7] and Leonor Varela-Lema, PhD [1,4,5]

Objective: Spin refers to reporting strategies that highlight the benefits of an experimental treatment or divert attention from nonsignificant primary outcomes. We aimed to assess spin in randomized controlled trials (RCTs) on pharmaceutical efficacy in multiple sclerosis (MS) and explore associated factors.

Methods: A systematic literature search was conducted in MedLine (PubMed), EMBASE, and Cochrane using database-specific thesauri ("Multiple Sclerosis" and "Drug Therapy") to identify relevant studies. We included multiple sclerosis phase 3 and 4 randomized controlled trials with parallel, superiority designs that were published between 2013 and 2024 reporting nonsignificant primary outcomes. Spin was assessed in title, abstract conclusion, results, discussion, and conclusions. A descriptive analysis was followed by exploratory bivariate logistic regression. Independent variables included trial phase, sample size, drug type, comparison, follow-up time, registration, Consolidated Standards of Reporting Trials (CONSORT) mention, risk of bias (RoB2), journal quartile, first author affiliation, and conflict of interest.

Results: Forty articles met inclusion criteria. Spin appeared in at least one section in 25 articles (62.5%) and in 3 or more in 19 articles (47.5%). The most frequent locations were abstract conclusions, discussion, and conclusions. Spin was significantly associated with smaller sample size (odds ratio [OR] = 7.00, 95% confidence interval [CI] = 1.29–37.91, $p = 0.024$), non-Q1 journals (OR = 4.38, 95% CI = 1.03–18.63, $p = 0.046$), and first author affiliation outside Europe or the United States (OR = 5.09, 95% CI = 1.15–22.62, $p = 0.032$).

Interpretation: Spin is common in MS randomized controlled trials with nonsignificant primary outcomes and may mislead clinical decisions.

ANN NEUROL 2026;99:316–327

Randomized controlled trials (RCTs) are the gold standard of evidence-based medicine for assessment of drug benefits. However, suboptimal methodology can introduce bias and flaw the risk–benefit ratio.[1] Incomplete

View this article online at wileyonlinelibrary.com. DOI: 10.1002/ana.78125

Received Jun 22, 2025, and in revised form Nov 17, 2025. Accepted for publication Nov 21, 2025.

Address correspondence to Dr Rivero-de-Aguilar, Department of Preventive Medicine and Public Health, Universidade de Santiago de Compostela. Rúa San Francisco s/n, Santiago de Compostela, Spain. E-mail: alejandro.riverodeaguilar@gmail.com

From the [1]Department of Preventive Medicine and Public Health, Universidade de Santiago de Compostela. Rúa San Francisco s/n, Santiago de Compostela, Spain; [2]Department of Preventive Medicine, University Hospital of Santiago de Compostela. Rúa da Choupana, Santiago de Compostela, Spain; [3]Department of Neurology Calle Mourente, Montecelo Hospital, Pontevedra, Spain; [4]Health Research Institute of Santiago de Compostela, Santiago de Compostela, Spain; [5]CIBERESP, Consortium for Biomedical Research in Epidemiology and Public Health, Madrid, Spain; [6]Scientific Advice Unit, Avalia-t. Galician Health Knowledge Agency, ACIS, Santiago de Compostela, Spain; and [7]Directorate-General of Public Health, Galician Regional Health Authority, Epidemiology Department, Santiago de Compostela, Spain

Additional supporting information can be found in the online version of this article.

or inaccurate reporting also contributes to result misinterpretation and raises concerns about efficacy and safety reliability.

Several approaches aim to improve RCT reporting quality. Clinical trial registries help reduce selective reporting by discouraging incomplete results or changes in primary outcomes.[2] Peer-reviewed journals requiring adherence to the Consolidated Standards of Reporting Trials (CONSORT) statement further enhance comprehensive and transparent RCT communication.[3] Nevertheless, these initiatives do not prevent authors from manipulating information or constructing the narrative of their outcomes to deliver misleading messages on the beneficial effects.[4]

Manipulation of information takes various forms and is difficult to assess. One way that information can be distorted is the so-called spin, a term coined by Boutron et al[5] to describe the "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." Their investigation involving a representative cohort of RCTs-based articles indexed in PubMed showed that approximately 60% used this technique.

Since then, several investigations have shown spin in RCTs across the health care landscape.[6–11] Notwithstanding, to our knowledge, no specific research on spin has been conducted for multiple sclerosis (MS) RCTs. Previous studies carried out by our group have demonstrated publication bias and transparency and reporting limitations in the field of MS.[12,13] In the current study, we aimed to assess the presence of spin in published RCTs that tested the efficacy of pharmaceuticals for any type of MS and explore associated factors.

Methods

Bibliographic Search

In July 2024, a systematic search was conducted in MedLine (PubMed), EMBASE, and Cochrane to identify phase 3 and 4 RCTs in MS published since 2013. The search strategy was developed by an information specialist (author B.C.A.) using database specific thesauri ("Multiple Sclerosis" and "Drug Therapy") and word variations applying Boolean operators. No language nor study design restrictions were applied. The search strategy is provided in Supplementary Material Data S1.

Inclusion and Exclusion Criteria

We considered all phase 3 and 4 RCTs with non-statistically significant results on the primary outcome. Studies were eligible for inclusion if they had parallel and superiority study design and aimed to assess the efficacy of pharmaceuticals on health-related outcomes in MS. Studies on disease modifying treatments (DMTs), symptomatic and relapse treatments for any type of MS, as well as studies including Clinically Isolated Syndrome (CIS) were included. Studies were excluded if they were pilot or exploratory trials, used crossover or noninferiority or equivalence study designs, or had laboratory or economic outcomes as primary end points. We also excluded RCTs evaluating vaccines or non-pharmacological interventions (eg, traditional herbs).

Study Selection

Titles and abstracts of all records identified were screened independently by two investigators (authors M.M.-G. and L.V.-L.) for eligibility, followed by a full-text review of all potentially relevant articles. Discrepancies in eligibility or data interpretation were resolved by consensus.

Data Extraction and Study Variables

Peer extraction of the data was conducted manually by a pair of investigators using an ad hoc designed template (authors M.M.-G., C.C.-P., G.G., C.G.-T., L.M.-G., J.R.-B., M.P.-R., and L.V.-L.). In case of disagreement, data were reviewed by a third investigator (author A.R.-d-A.). We retrieved information regarding the methodological characteristics (trial phase, number of participant centers and randomized patients, MS type, drug use, comparison group, and primary outcome definition), quality of reporting and bias (registration in a public database, mentioning CONSORT statement adherence and risk of bias), the characteristics and metrics of the journal and the article itself (journal thematic—neurology, multiple sclerosis, or other—Journal Citation Reports [JCR] and impact of the article), and the affiliation region of the first author as well as the financial and conflict of interest (COI) declarations (retrieved from the COI statement, affiliations, funding, acknowledgments, or any other section of the article).

When the RCT phase was not reported in the article, we retrieved the information from Clinical Trial Registries (ClinicalTrials.gov, ISRCTN registry or Iranian Registry of Clinical Trials), when available. If the study was not registered and the phase not mentioned in the article, the RCT phase was established by 2 of the principal investigators (authors M.M.G. and A.R.A.) attending to the description of the study.[14] The risk of bias for each of the included articles was judged by a pair of independent reviewers (authors M.M.-G., C.C.-P., G.G., J.R.-B., and A.R.-d-A.) and discrepancies resolved by a third experienced reviewer (author L.V.-L.) using the Cochrane Risk of Bias Tool (RoB2). We report the overall risk of bias

determined by the algorithm (high, some concerns, and low).[15] The JCR data —including Journal Impact Factor (JIF) and quartiles (Q)—were extracted from Clarivate Analytics for the year in which the article was published. The impact of each article was assessed using the Field-Weighted Citation Impact (FWCI) from Scopus.

Spin Assessment

Assessment of spin was performed independently by a pair of investigators (authors M.M.G. and L.V.L.). In cases of disagreement, data were reviewed by a third investigator (author A.R.A.) and the 3 discussed the study until consensus was reached. To ensure consistency with the established methodology, we evaluated 4 articles and discussed the assessments. To assess the inter-rater agreement between the 2 investigators, we calculated the Cohen's Kappa test. Spin for the primary outcomes was assessed following the methodology proposed by Boutron et al.[5] When the primary outcomes were not explicitly reported in the article, we considered as the primary outcome the outcome used for the sample size calculation, if provided. If not, it was established based on the primary aim of the RCT as reported in the article.

Spin was evaluated in the following sections: Title, Abstract (conclusion), Results, Discussion, and Conclusions/Last paragraph. The following strategies of spin were considered:

- Type (A) focusing on statistically significant results other than the nonsignificant primary outcome (such as within-group comparison, secondary outcomes, subgroup populations, or in the case of multiple primary outcomes, focusing only on the one being statistically significant). Articles were classified under this type of spin if they reported a positive result of the trial based on secondary analyses while overlooking, downplaying, or understating the negative result of the primary outcome. In cases of multiple outcomes, we also checked whether the multiplicity effect was considered in the analysis;
- Type (B) interpreting nonsignificant results as treatment equivalence or noninferiority, even though the study was not being designed to assess equivalence or noninferiority;
- Type (C) claiming benefits of the trial despite statistically nonsignificant results on the primary outcomes;
- Type (D) other (strategies that could not be classified in any of the other categories).

For each article, the total number of article sections having spin was determined.

Statistical Analysis

Firstly, we performed a descriptive analysis. Quantitative variables are presented as median accompanied by interquartile range (IQR). Categorical variables are presented as absolute frequency (n) and relative frequency (%).

Afterward, we conducted a bivariate analysis using non-adjusted logistic regression models to explore the relationship between the presence of spin in any section of the article (dependent variable) and different independent variables related to the study and publication (trial phase, randomized patients, drug use, comparison, time of follow-up, trial registration, mention of CONSORT adherence, RoB2, quartile, first author affiliation, and potential COI with the industry). We did not adjust for multiplicity in the analysis as they were exploratory, meaning hypothesis-generating and not definitive. However, to limit the effect of multiple comparisons, we selected just one independent variable when 2 or more were strongly correlated to each other (ie. randomized sample size and number of participant centers), prioritizing variables without missing data. Due to the low number of events in some categories we regrouped when appropriate (follow-up, quartile group, and author affiliation region) to ensure reliable estimates. When regrouping was not possible, we excluded those variables with less than 10 events in all the comparative categories. The significance threshold was set at $p < 0.05$. All statistical analyses were performed using SPSS (version 29).

Results

The systematic search identified 3,791 articles. After applying selection criteria, 40 articles were included in the analysis. The flow diagram is shown in the Figure.

Methodological Characteristics

Tables 1 and 2 summarizes the methodological characteristics of the analyzed RCTs and associated articles. Of the included articles, 24 were phase 3 RCTs (60.0%). The median number of randomized patients was 93 (IQR = 57–303). Included RCTs mostly targeted active forms of MS (n = 26, 65.0%). The type of drug most frequently assessed in included RCTs were DMTs (n = 21, 52.5%), followed by symptomatic drugs (n = 15, 37.5%). Placebo was used as the comparator in most of the RCTs included (n = 25, 62.5%). The primary outcome measure was clinical in 34 (85.0%) of the RCTs. The follow-up period regarding the primary outcome measure was over 1 year in 55.0% of the articles (n = 22), in 30% of the articles (n = 12) it was less than 6 months.

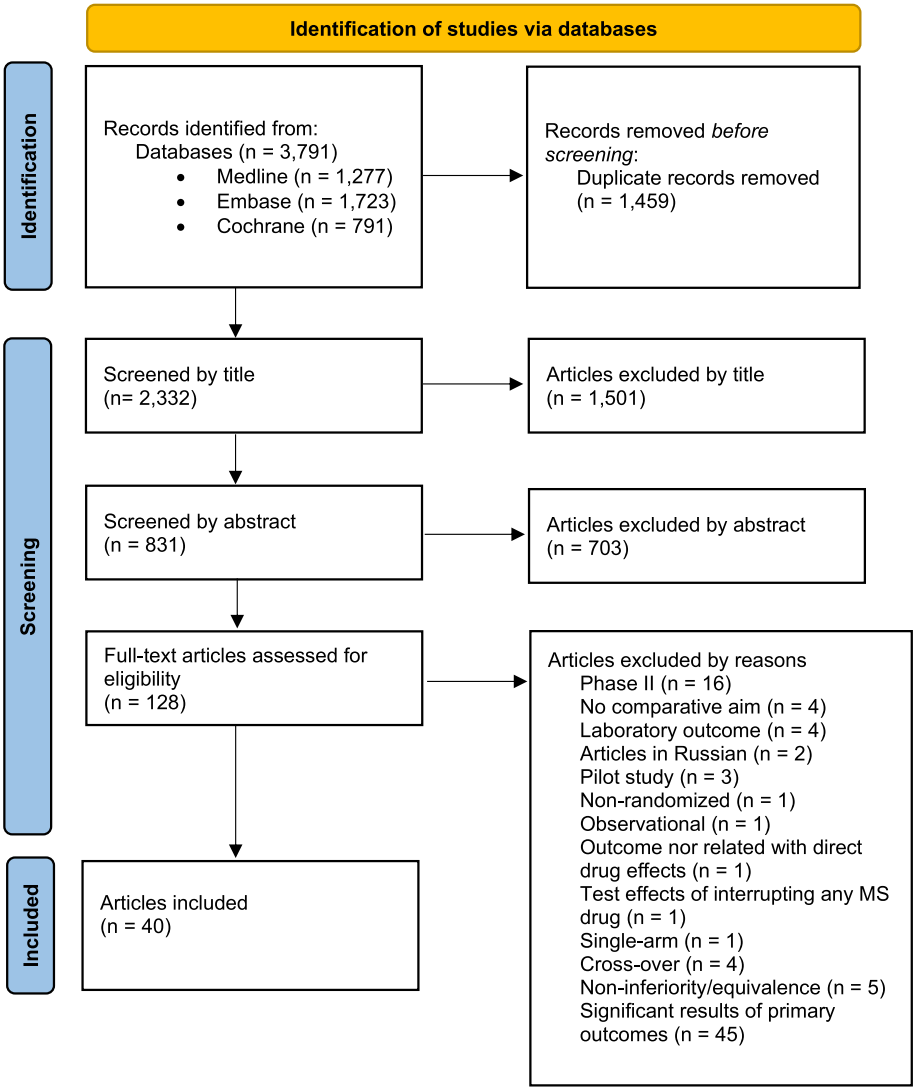

FIGURE: PRISMA flow diagram. PRISMA = Preferred Reporting Items for Systematic Reviews and Meta-Analyses. [Color figure can be viewed at www.annalsofneurology.org]

Quality of Reporting and Risk of Bias of Included Articles

A fifth of the articles reported CONSORT adherence (n = 8, 20.0%). The risk of bias, assessed using the RoB2 tool, was judged "high" in 21 articles (52.5%) and "low" in 5 articles (12.5%).

Journal and Article Metrics and Affiliations

Twenty-one articles (52.5%) were published in neurology journals. Overall, 17 (42.5%) articles were published in first quartile journals, with a median JIF of 4.0 (IQR = 2.3–10.8), whereas 4 were published in non-indexed journals. The articles had a median FWCI of 1.9 (IQR = 0.7–4.6); this metric was not available for 6 articles. The first author's affiliation was based in Europe or the United States in 23 (57.5%) articles.

Funding and COI

Private funding was the most frequently reported source: either alone (n = 14, 35.0%) or coexisting with public funding (n = 4, 10.0%). Four articles (10.0%) did not report funding information in the article and trials were not registered, making it impossible to investigate the funding source. Thirty-five articles (87.5%) had a COI statement, of which 18 declared COI. In affiliations, COI was detected in 10 (25.0%) of the articles. Overall, potential COI with the industry was detected in 22 (55.0%) of the articles.

Frequency of Spin

Table 3 provides the frequency of spin by type and article section. Twenty-five articles (62.5%) presented spin in at least one article section; of them, 10 articles (25.0%) had spin in 4 sections. In 3 articles, spin was present in all the

TABLE 1. Characteristics of Included Articles

Characteristics	Total (n = 40)	No spin (n = 15)	Spin (n = 25)
Trial phase; n (%)			
Phase III	24 (60.0)	12 (80.0%)	12 (48.0%)
Phase IV	16 (40.0)	3 (20.0%)	13 (52.0%)
Number of participant centers[a]; median (IQR)	9 (1–56)	30 (17–118)	1 (1–21)
Randomized patients; median (IQR)	93 (57–303)	172 (60–970)	84 (54–152)
MS Type; n (%)			
CIS and/or RR and/or SP active	26 (65.0)	9 (60.0)	17 (68.0)
SP inactive and/or PP	5 (12.5)	3 (20.0)	2 (8.0)
Both	9 (22.5)	3 (20.0)	6 (24.0)
Drug use; n (%)			
DMT	21 (52.5)	9 (60.0)	12 (48.0)
Relapses	4 (10.0)	2 (13.3)	2 (8.0)
Symptomatic	15 (37.5)	4 (26.7)	11 (44.0)
Comparison; n (%)			
Placebo	25 (62.5)	11 (73.3)	14 (56.0)
Active drug	9 (22.5)	3 (20.0)	6 (24.0)
Other[b]	6 (15.0)	1 (6.7)	5 (20.0)
Exclusively clinical primary outcome[c], n (%)	34 (85.0)	14 (93.3)	20 (80.0)
Follow-up; n (%)			
<6 months	12 (30.0)	2 (13.3)	10 (40.0)
≥ 6 months-1 year	6 (15.0)	2 (26.7)	2 (8.0)
≥ 1–2 years	12 (30.0)	4 (26.7)	8 (32.0)
≥ 2 years	10 (25.0)	5 (33.3)	5 (20.0)
Trial registered in a public database; n (%)	27 (67.5)	11 (73.3)	16 (64.0)
Prospective registration[d], n (%)	13 (32.5)	4 (26.7)	9 (36.0)
Mentioned CONSORT adherence; n (%)	8 (20.0)	4 (26.7)	4 (16.0)
RoB2 assessment; n (%)			
Low	5 (12.5)	4 (26.7)	1 (4.0)
Some concerns	14 (35.0)	5 (33.3)	9 (36.0)
High	21 (52.5)	6 (40.0)	15 (60.0)

Abbreviations: CIS = clinically isolated syndrome; CONSORT = consolidated standards of reporting trials; DMT = disease modifying treatment; MS = multiple sclerosis; PP = primary progressive; RR = relapsing–remitting; SP = secondary progressive.

[a]8 missing.

[b]Other types of comparisons were established for trials assessing posology, discontinuation regimes, trials without a comparator drug or placebo, or different formulations of the same drug.

[c]Exclusively clinical outcomes were those that did not have radiological outcomes, alone or with clinical outcomes. For example: annualized relapse rate, disability (ie, EDSS) or quality of life.

[d]Prospective registration was established when the registration date preceded start of recruitment.

TABLE 2. Characteristics of Included Articles

Characteristics	Total (n = 40)	No spin (n = 15)	Spin (n = 25)
Journal thematic; n (%)			
Neurology	21 (52.5)	9 (60.0)	12 (48.0)
Multiple sclerosis	6 (15.0)	2 (13.3)	4 (16.0)
Other	13 (32.5)	4 (26.7)	9 (36.0)
Journal Impact Factor[a]; median (IQR)	4.0 (2.3–10.8)	8.9 (3.5–26.2)	3.0 (1.9–5.1)
Quartile of the journal; n (%)			
Q1	17 (42.5)	10 (66.7)	7 (28.0)
Q2	7 (17.5)	3 (13.3)	5 (20.0)
Q3	7 (17.5)	1 (6.7)	6 (24.0)
Q4	4 (10.0)	1 (6.7)	3 (12.0)
Not indexed	5 (12.5)	1 (6.7)	4 (16.0)
FWCI; median (IQR)[b]	1.9 (0.7–4.7)	3.4 (1.5–5.8)	1.1 (0.6–2.4)
First author affiliation region; n (%)			
Europe	13 (32.5)	7 (46.7)	6 (24.0)
USA[c]	10 (25.0)	5 (33.3)	5 (20.0)
Other	17 (42.5)	3 (20.0)	14 (56.0)
Declared funding; n (%)			
Public	14 (35.0)	2 (13.3)	12 (48.0)
Private	14 (35.0)	6 (40.0)	8 (32.0)
Mixed	4 (10.0)	4 (26.7)	0 (0)
No funding	4 (10.0)	1 (6.7)	3 (12.0)
No information	4 (10.0)	2 (13.3)	2 (8.0)
Conflicts on COI statement; n (%)			
No	17 (42.5)	5 (33.3)	12 (48.0)
Yes	18 (45.5)	9 (60.0)	9 (36.0)
No COI statement	5 (12.5)	1 (6.7)	4 (16.0)
COI in Affiliations; n (%)	10 (25.0)	5 (33.3)	5 (20.0)
COI in Acknowledgements; n (%)			
No	21 (52.5)	6 (40.0)	15 (60.0)
Yes	11 (27.5)	6 (40.0)	5 (20.0)
No acknowledgments statement	8 (20.0)	3 (20.0)	5 (20.0)
COI in other parts of the article	5 (12.5)	3 (20.0)	2 (8.0)
Potential COI with the industry; n (%)	22 (55.0)	11 (73.3)	11 (44.0)

Abbreviations: COI = Conflict of Interest; FWCI = Field-Weighted Citation Impact.
[a]4 missing.
[b]6 missing.
[c]We did not include Canada because no articles with this characteristic were included.

TABLE 3. Presence of Spin, Strategies, and Extension

Spin assessment	n (%)
Any spin	25 (62.5)
Spin in the title	5 (12.5)
Spin in the abstract (conclusions)	22 (55.0)
Type A	13 (32.5)
Type B	6 (15.0)
Type C	3 (7.5)
Spin in results	13 (32.5)
Type A	12 (30.0)
Type B	1 (2.5)
Type C	0
Spin in discussion	21 (52.5)
Type A	15 (37.5)
Type B	3 (7.5)
Type C	3 (7.5)
Spin in conclusion/last paragraph	22 (55.0)
Type A	13 (32.5)
Type B	5 (12.5)
Type C	4 (10.0)
Extent of spin (number of sections with spin)	
0	15 (37.5)
1	2 (5.0)
2	4 (10.0)
3	6 (15.0)
4	10 (25.0)
5	3 (7.5)

5 sections considered (7.5%). Spin was identified in the abstract (n = 22, 55.0%), discussion (n = 21, 52.5%), and conclusions sections (n = 22, 55.0%). In the results section, spin was less common, being detected in 13 articles (32.5%). The most predominant type of spin in every assessed section was focusing on significant results (spin type a). All instances of spin identified could be classified into type A, B, or C category. No type D spin was observed. Table 4 provides an example of each type of spin.

Supplementary Material Data S2 contains a full list of included articles summarizing the assessment of spin by section, as well as the agreement for each assessment. Overall, the inter-rater agreement was 0.77 for the title, 0.70 for the abstract, 0.80 for the results, 0.80 for the discussion, and 0.85 for the conclusions.

Association of Spin With Study and Article Characteristics

In the bivariate non-adjusted logistic regression analysis (Table 5), those articles with a sample size under the 50th percentile of the randomized patients in the included studies were more likely to present spin (odds ratio [OR] = 7.00, 95% confidence interval [CI] = 1.29–37.91, p value = 0.024). In addition, articles that were not published in Q1 journals presented over 4 times more spin than those published in Q1 journals (OR = 4.38, 95% CI = 1.03–18.63, p value = 0.046). Last, those articles where the first author presented an affiliation different than Europe or the United States were 5 times more likely to present spin (OR = 5.09, 95% CI = 1.15–22.62, p value = 0.032). No significant association was found for any of the other variables assessed.

Discussion

The present study shows that spin, a form of reporting bias, was present in nearly 2 out of 3 MS phase 3 or 4 RCTs published since 2013 with nonsignificant results for primary outcomes. Spin was most found in the discussion and conclusion sections (both in the abstract and main text). Our analysis suggests that the presence of spin in any of the sections may be related to a smaller sample size, publication in journals with an impact factor below Q1, or a first author affiliated outside Europe or the United States.

Spin has also been found in other similar investigations carried out in many other medical fields such as urology (76%),[6] obstetrics and gynecology (53%),[7] otorhinolaryngology (70%),[8] vascular surgery (72%),[9] acupuncture (56%),[10] or cardiology (57–67%).[11] In the field of neurology and neurosurgery, spin has been analyzed over tinnitus[16] and traumatic brain injury RCTs.[17] However, these trials used slightly different methodology and included both trials with significant and nonsignificant results, potentially underestimating spin and making comparisons difficult.

To our knowledge, this is the first study assessing spin in MS RCTs. These findings are highly relevant as MS is a chronic, incurable disease with rising incidence and multiple treatment options to manage its course and symptoms. Clinicians face challenges in selecting the most suitable treatment, and spin can hinder accurate interpretation and application of evidence.[18] Previous investigations have also shown that MS RCTs are also liable to

TABLE 4. Examples of Types of Spin

Primary outcome	Result	Excerpt from the manuscript	Type of spin	Rationale
Time to first confirmed clinical relapse by the end of the double-blind period	Hazard ratio 0.66, 95% CI = 0.39–1.11, $p = 0.29$	"Analysis of the primary endpoint did not show a significant difference in time to first confirmed clinical relapse in the intervention 1 group compared with intervention 2. However […] in retrospect, time to first clinical relapse or high MRI activity would have been preferable as the primary endpoint, because this composite outcome better reflects the study design and clinical practice. The results of the prespecified sensitivity analysis, in which intervention 1 was associated with a statistically significant reduction in the combined risk of clinical relapse or high MRI activity relative to intervention 2, support the efficacy of intervention 1."	A	The authors acknowledge a nonsignificant primary outcome but redirect attention to a secondary analysis showing statistical significance. By reframing the main result and emphasizing this finding, they ultimately claim efficacy for the intervention despite the trial not meeting its primary objective for which it was designed.
Mean scores of EDSS in follow-up visits	$p = 0.78$	"Intervention 1 and intervention 2 showed similar efficacy and safety outcome in patients with RRMS."	B	The authors assume equivalence/non-inferiority but the trial was designed to demonstrate superiority
Mean change from baseline pain intensity to mean weekly pain scores within a maximum of 16 weeks	Intervention 1 (1.92 ± 2.01; 30%) and intervention 2 (1.81 ± 1.94; 27%); $p = 0.6760$	"Overall, this trial demonstrated the long-lasting therapeutic potential, the good tolerability and favorable safety profile of intervention 1 – especially in terms of drug abuse and dependency. Based on the presented results, there is no special focus on the harm caused by intervention 1 treatment. Although the statistical proof of efficacy for intervention 1 versus intervention 2 treatment is pending, physicians should consider the potential benefits of the multifactorial effects of intervention 1."	C	The authors acknowledge that statistical proof of efficacy is lacking but emphasize the drug's therapeutic potential, safety, and tolerability, and even suggest clinical consideration. This framing conveys a positive interpretation and implies benefits despite nonsignificant primary results.

95% CI = 95% confidence interval; EDSS = Expanded Disability Status Scale; MRI = magnetic resonance imaging; RRMS = relapsing-remitting multiple sclerosis.

present methodological and reporting limitations which could further compromise the reliability of the results.[13,19]

Similarly to Boutron et al,[5] we found that the most common type of spin in all sections was spin focusing only on statistically significant results. It is important to note that we did not penalize articles for merely communicating positive secondary analyses but rather for overlooking, diluting, or downplaying nonsignificant primary

TABLE 5. Bivariate Logistic Regression Analysis With Presence of Spin in Any of the Assessed Sections as Dependent Variable

Covariate	Spin, n (%)	OR	95% CI	*p*
Phase 3	12 (50.0)	Ref	Ref	Ref
Phase 4	13 (81.2)	4.33	0.98–19.20	0.054
Randomized sample > P_{75}	7 (70.0)	Ref	Ref	Ref
Randomized sample ≥ P_{50} and < P_{75}	15 (75.0)	5.44	0.80–36.87	0.082
Randomized sample < P_{50}	3 (30.0)	7.00	1.29–37.91	**0.024**
DMT	12 (57.1)	Ref	Ref	Ref
Relapses	2 (50.0)	0.75	0.09–6.39	0.79
Symptomatic	11 (73.3)	2.06	0.49–8.65	0.32
Active drug	6 (66.7)	Ref	Ref	Ref
Placebo	14 (56.0)	0.64	0.13–3.14	0.58
Other	5 (83.3)	2.50	0.19–32.19	0.48
Follow-up < 6 mo	10 (83.3)	Ref	Ref	Ref
Follow-up ≥ 6 mo	15 (53.6)	0.23	0.04–1.25	0.089
Trial registered in a public database	16 (59.3)	Ref	Ref	Ref
Trial not registered in a public database	9 (69.2)	1.55	0.38–6.31	0.543
Not mentioned CONSORT adherence	21 (65.6)	Ref	Ref	Ref
Mention CONSORT adherence	4 (50.0)	0.52	0.11–2.51	0.418
RoB2 Low	1 (20.0)	Ref	Ref	Ref
RoB2 Some concerns	9 (64.3)	7.20	0.62–83.34	0.154
RoB2 High	15 (71.4)	10.00	0.92–108.82	0.059
Q1	18 (78.3)	Ref	Ref	Ref
Other than Q1	7 (41.2)	4.38	1.03–18.63	**0.046**
First author: Europe and the United States	11 (47.8)	Ref	Ref	Ref
First author: other regions	14 (82.4)	5.09	1.15–22.62	**0.032**
Not potential COI with industry	14 (77.8)	Ref	Ref	Ref
Potential COI with industry	11 (50.0)	0.29	0.07–1.15	0.077

95% CI = 95% confidence interval; COI = conflict of interest; CONSORT = Consolidated Standards of Reporting Trials; DMT = disease-modifying therapy; OR = odds ratio; P = Percentile; RoB2 = Cochrane Risk of Bias tool.
The figures in bold in Table 5 represent significant p-values.

outcomes and not acknowledging the absence of an appropriate methodology to determine effectiveness in other outcomes. Whether intentional or not intentional is not rare that authors omit information on the primary outcome and/or just emphasize those significant results which could lead to overestimations of positive findings by readers less trained in critical scientific reading. For these readers, the efficacy of secondary outcomes could be interpreted as demonstrated efficacy in those domains. A study assessing the impact spin in oncology abstracts had on clinicians[20] showed that those physicians randomized to reading the abstract version with spin were more likely to evaluate the treatment as beneficial compared to those abstracts without spin.

Some articles included in our analysis claimed treatment equivalence, although studies were not designed for this purpose (spin type B). Conducting equivalence or noninferiority trials in MS has been proven useful to try to achieve a better resource allocation and avoid the potential detrimental effects of existing drugs while maintaining effectiveness. However, to prove equivalence, a specific statistical design including a pre-established margin of clinical significance, among other characteristics, is necessary. A superiority design is not suitable to prove equivalence and could distort study results.[21]

Regarding spin type C, it was the least prevalent in our sample. This might be because such an explicit claim of benefit without even referring to, for example, a secondary outcome, is more likely to be filtered out during author or editorial review. However, we can argue that what ultimately matters is not only the type of spin itself, but the message that reaches the reader. Even subtler forms of spin, although less direct, may still lead to an interpretation of efficacy.

The main reasons that have been cited for the high prevalence of spin have been lack of knowledge of reporting standards, intention to influence readers, and intention to increase publication in journals.[22] In a previous study carried out by our group comparing published and unpublished MS RCTs registered in ClinicalTrials.gov, it was shown that 1 out of 3 phase 3 or 4 trials in MS remained unpublished after their completion, and that disclosing a favorable primary outcome result was associated with more chances of having RCT results published.[23]

In the present study, we aimed to explore the association between the presence of spin and different study characteristics. Among them, the prevalence of spin was significantly associated with small sample size. Small sample sized studies often yield nonsignificant results due to low statistical power. In addition, they are more prone to biases like selective data analysis and reporting, potentially driven by publication pressure.[24] Spin further amplifies this issue, distorting scientific knowledge, misleading practitioners, and potentially leading to poor therapeutic decisions.

We did not find a statistically significant difference in the prevalence of spin between the phase 4 and phase 3 trials (81% vs 50%, $p = 0.054$). However, this observed difference may warrant consideration. Differences in methodological rigor between phases —particularly in sample size— could partly account for this. In our sample, phase 4 trials had significantly smaller sample sizes than phase 3 trials. Because pivotal phase 3 trials are designed to meet regulatory standards, they are generally expected to follow stricter methodological criteria, including more robust sample size planning.

Similarly, no significant association was found between risk of bias and the presence of spin ($p = 0.059$) although spin prevalence increased progressively with higher levels of bias. A previous publication[7] reported similar results. In our sample, the reference category ("low" risk) included fewer than 10 events, potentially limiting the reliability. Additional research is needed to clarify this possible association.

On the contrary, RCTs published in Q1 quartile journals presented less risk of spin than articles published in lower impact journals within the same category. Another study assessing spin in urology RCTs also reported a negative, although weak, association between the severity of spin and the impact factor of the journal, although the median impact factor of the journals they assessed was significantly higher than that of ours.[6] As discussed in that article, high impact journals evaluation criteria are stricter in terms of methodology and validity assessment.[25] This is partly reflected in our findings, where the median impact factor of journals without spin was higher than that of journals with spin (8.9 vs 3.0), although there was overlap in their IQR (3.5–26.2 vs 1.9–5.1, respectively). These results suggest that the impact factor might also serve as an indicator of a lower likelihood of spin. In our regression analysis, we chose journal quartiles over raw impact factors because quartiles provide a relative ranking within each subject category, making them easier to interpret and more transparent. Furthermore, impact factor values were unavailable for 6 articles, and imputing those missing values could have introduced bias. Using quartiles therefore allowed us to maintain methodological rigor while preserving completeness of the dataset.

That said, it is important to note that spin was still present in a substantial proportion of articles published in high-impact journals (41%). Additionally, 59% of articles in neurology/multiple sclerosis journals contained spin, compared with 69% in other journals. These findings indicate that publication in top-tier journals cannot be considered a guarantee of freedom from spin.

Another key finding was the association between first-author affiliation outside Europe or the United States and higher odds of spin. As spin is a narrative bias, and the first author is responsible for drafting the manuscript, this position likely has the greatest influence on how results are presented.[26] No previous studies have evaluated this association; nonetheless, higher scientific misconduct rates have been reported in countries such as China, Malaysia, Mexico, Taiwan, Pakistan, and Iran.[27] A study examining factors that contribute to scientific misconduct in Asian countries pointed out reasons such as insufficient

training in scientific writing and ethics or inadequate regulatory measures.[28]

Our study did not detect significant differences between registered and non-registered RCTs. In another study assessing spin, this was also not deemed significant.[10] Whereas registration might be indicative of good scientific conduct, it does not entirely avoid reporting bias.[29] The same occurs with CONSORT statement adherence. Articles published in CONSORT-endorsing journals present in general better reporting quality. However, inadequacies and missing information are still common.[30]

Similarly to another study carried out in cardiology RCTs, no association was found with type of pharmaceutical, indication, or funding source of conflict. In both our study and the one mentioned, the prevalence of spin was lower in RCTs presenting COI with the industry.[11] Caution should be taken with these results given that COIs are commonly under-reported in the literature.

This study presents limitations. Assessing spin implies subjectivity that could lead to variations in the retrieved information. In addition, the methodology proposed by Boutron et al has not been formally adopted or validated by any institution. Nevertheless, we chose to use their definition because it was developed through a literature review and in consultation with the Cochrane Statistical Methods Group, an international forum with vast experience in methodology issues. Moreover, as it has been widely applied in previous studies assessing spin in other medical fields,[6–11] it provides a useful reference framework and facilitates comparison across research. Although the articles included are reflective of the existing MS RCTs, the number of articles in the analysis was low. This particularly affects the power of our association analysis, which can only be considered exploratory and highlights the need for additional research. Furthermore, not including other types of interventions such as traditional medicines or psychosocial interventions could reduce the generalizability of the findings. However, these last types of interventions are of a different nature and differ in terms of regulatory needs.

This study has several strengths. One of the strengths is the comprehensive search strategy, which was designed and conducted by an information specialist to maximize sensitivity and minimize the risk of missing relevant studies. By not applying filters for RCTs, we ensured the inclusion of potentially eligible articles that may not have been indexed as such. Although this approach resulted in a high proportion of exclusions during abstract screening, it reflects a deliberate methodological decision aimed at ensuring completeness and rigor. In the study, we explored the association of spin with factors related to the characteristics and metrics of the journal and the publication that had not been previously investigated. Second, the investigators involved in this study are very familiar with quality evaluations of RCTs, including risk of bias assessments. The findings could help guide future research and support the development of measures to prevent this practice.

Our findings highlight the need for readers to be critical when reading and drawing conclusions from MS phase 3 and 4 articles with nonsignificant primary outcomes. More importantly, we consider that journals should create awareness of this practice among their editors and reviewers, who are the ultimate guardians of the quality of the published papers. Finally, it would be interesting to conduct further research in the field to understand the underlying factors as well as to review existing methodologies and reporting standards to accommodate for these practices.

Acknowledgement

There are no funders to report.

Author Contributions

M.M.-G., A.R.-d.-A., L.V.-L., B.C.-A., and M.P.-R. contributed to the conception and design of the study; M.M.-G., A.R.-d.-A., L.V.-L., C.C.-P., G.G., C.G.-T., L.M.-G., J.R.-B., M.P.-R., and M.I.S.P. contributed to the acquisition and analysis of the data; M.M.-G., A.R.-d.-A., and L.V.-L. contributed to drafting the text or preparing the figures. [Correction added on 09 February 2026, after first online publication: Author contribution text has been revised in this version.]

Potential Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability

Data are not public because it is being used for the PhD work of the first author. However, it is available upon reasonable request from the corresponding author Alejandro Rivero-de-Aguilar at alejandro.riverodeaguilar@gmail.com. <zbmrule>

References

1. Gluud LL. Bias in clinical intervention research. Am J Epidemiol 2006;163:493–501.
2. Angelis CD, Drazen JM, Frizelle FA, et al. Clinical trial registration: a statement from the International Committee of Medical Journal Editors. N Engl J Med 2004;351:1250–1251.

3. Schulz KF, Altman DG, Moher D, for the CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMJ 2010;340:c332.
4. Horton R. The rhetoric of research. BMJ 1995;310:985–987.
5. Boutron I, Dutton S, Ravaud P, et al. Reporting and interpretation of randomized controlled trials with statistically nonsignificant results for primary outcomes. JAMA 2010;303:2058–2064.
6. Wu J, Ho W, Klotz L, et al. Assessing "spin" in urology randomized controlled trials with statistically nonsignificant primary outcomes. J Urol 2023;209:494–503. https://doi.org/10.1097/JU.0000000000003105.
7. Chow R, Huang E, Fu S, et al. Spin in randomized controlled trials in obstetrics and gynecology: a systematic review. Womens Health Rep (New Rochelle) 2022;3:795–802.
8. Cooper CM, Gray HM, Ross AE, et al. Evaluation of spin in the abstracts of otolaryngology randomized controlled trials. Laryngoscope 2019;129:2036–2040.
9. Nguyen J, Li A, Tam DY, Forbes TL. Analysis of spin in vascular surgery randomized controlled trials with nonsignificant outcomes. J Vasc Surg 2022;75:1074–1080.e17.
10. Won J, Kim S, Bae I, Lee H. Trial registration as a safeguard against outcome reporting bias and spin? A case study of randomized controlled trials of acupuncture. PLoS One 2019;14:e0223305.
11. Khan MS, Lateef N, Siddiqi TJ, et al. Level and prevalence of spin in published cardiovascular randomized clinical trial reports with statistically nonsignificant primary outcomes: a systematic review. JAMA Netw Open 2019;2:e192622.
12. Rivero-de-Aguilar A, Pérez-Ríos M, Mascareñas-García M, et al. Discrepancies in the results reported for multiple sclerosis clinical trials: a comparison between ClinicalTrials.Gov and peer-reviewed journals. Mult Scler J 2024;30:1514–1524.
13. Mascarenas-Garcia M, Rivero-de-Aguilar A, Pérez-Ríos M, et al. Best practices in phase III clinical trials on DMTs for multiple sclerosis: a systematic analysis and appraisal of published trials. J Neurol Neurosurg Psychiatry 2024;95:333–341.
14. Glasser SP, Salas M, Delzell E. Importance and challenges of studying marketed drugs: what is a phase IV study? Common clinical research designs, registries, and self-reporting systems. J Clin Pharmacol 2007;47:1074–1086.
15. Sterne JAC, Savović J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019;366:l4898.
16. Velde HM, van Heteren JAA, Smit AL, Stegeman I. Spin in published reports of tinnitus randomized controlled trials: evidence of Overinterpretation of results. Front Neurol 2021;12:693937.
17. de Oliveira JVMP, Júnior ALF d O, Martins LP d F, et al. Spin in traumatic brain injury literature: prevalence and associated factors. A systematic review. J Neurosurg 2024;141:887–894.
18. Hartung DM, Bourdette DN, Ahmed SM, Whitham RH. The cost of multiple sclerosis drugs in the US and the pharmaceutical industry. Neurology 2015;84:2185–2192.
19. Gehr S, Kaiser T, Kreutz R, et al. Suggestions for improving the design of clinical trials in multiple sclerosis—results of a systematic analysis of completed phase III trials. EPMA J 2019;10:425–436.
20. Boutron I, Altman DG, Hopewell S, et al. 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;32:4120–4126.
21. Stefanos R, Graziella D'A, Giovanni T. Methodological aspects of superiority, equivalence, and non-inferiority trials. Intern Emerg Med 2020;15:1085–1091.
22. Su N, van der Linden MW, Faggion CM, et al. Assessment of spin in the abstracts of randomized controlled trials in dental caries with statistically nonsignificant results for primary outcomes: a methodological study. Caries Res 2023;57:553–562.
23. Rivero-de-Aguilar A, Pérez-Ríos M, Ruano-Raviña A, et al. Evidence of publication bias in multiple sclerosis clinical trials: a comparative analysis of published and unpublished studies registered in ClinicalTrials.gov. J Neurol Neurosurg Psychiatry 2023;94:597–604.
24. Button KS, Ioannidis JPA, Mokrysz C, et al. Power failure: why small sample size undermines the reliability of neuroscience. Nat Rev Neurosci 2013;14:365–376.
25. Seeber M. How do journals of different rank instruct peer reviewers? Reviewer guidelines in the field of management. Scientometrics 2020;122:1387–1405.
26. Riesenberg D, Lundberg GD. The order of authorship: Who's on first? JAMA 1990;264:1857.
27. Ataie-Ashtiani B. World map of scientific misconduct. Sci Eng Ethics 2018;24:1653–1656.
28. Rodrigues F, Gupta P, Khan AP, et al. The cultural context of plagiarism and research misconduct in the Asian region. J Korean Med Sci 2023;38:e88.
29. Dechartres A, Bond EG, Scheer J, et al. Reporting of statistically significant results at ClinicalTrials.gov for completed superiority randomized controlled trials. BMC Med 2016;14:192.
30. Ranganathan P. The CONSORT statement and its impact on quality of reporting of trials. Perspect Clin Res 2019;10:145–147.