VU Research Portal

The Relationship Between Methodological Quality and Conclusions in Reviews of Spinal Manipulation

Assendelft, W.J.J.; Koes, B.W.; Knipschild, P.G.; Bouter, L.M.

published in

JAMA

1995

DOI (link to publisher)

[10.1001/jama.274.24.1942](https://doi.org/10.1001/jama.274.24.1942)

document version

Publisher's PDF, also known as Version of record

[Link to publication in VU Research Portal](#)

citation for published version (APA)

Assendelft, W. J. J., Koes, B. W., Knipschild, P. G., & Bouter, L. M. (1995). The Relationship Between Methodological Quality and Conclusions in Reviews of Spinal Manipulation. *JAMA*, 274(24), 1942-1948. <https://doi.org/10.1001/jama.274.24.1942>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:

vuresearchportal.ub@vu.nl

The Relationship Between Methodological Quality and Conclusions in Reviews of Spinal Manipulation

Willem J. J. Assendelft, MD; Bart W. Koes, PhD; Paul G. Knipschild, PhD; Lex M. Bouter, PhD

Objective.—To study the relationship between the methodological quality and other characteristics of reviews of spinal manipulation for low back pain on the one hand and the reviewers' conclusions on the effectiveness of manipulation on the other hand.

Data Sources.—Reviews identified by MEDLINE search, citation tracking, library search, and correspondence with experts.

Study Selection.—English- or Dutch-language reviews published up to 1993 dealing with spinal manipulation for low back pain that include at least two randomized clinical trials (RCTs).

Data Extraction.—Methodological quality was assessed using a standardized criteria list applied independently by two assessors (range, 0% to 100%). Other extracted characteristics were the comprehensiveness of the search, selective citation of studies, language, inclusion of non-RCTs, type of publication, reviewers' professional backgrounds, and publication in a spinal manipulation journal or book. The reviewers' conclusions were classified as negative, neutral, or positive.

Data Synthesis.—A total of 51 reviews were assessed, 17 of which were neutral and 34 positive. The methodological quality was low, with a median score of 23%. Nine of the 10 methodologically best reviews were positive. Other factors associated with a positive reviewers' conclusion were review of spinal manipulation only, inclusion of a spinal manipulator in the review team, and a comprehensive literature search.

Conclusions.—The majority of the reviews concluded that spinal manipulation is an effective treatment for low back pain. Although, in particular, the reviews with a relatively high methodological quality had a positive conclusion, strong conclusions were precluded by the overall low quality of the reviews. More empirical research on the review methods applied to other therapies in other professional fields is needed to further explore our findings about the factors related to a positive reviewers' conclusion.

(*JAMA*. 1995;274:1942-1948)

IN RECENT years, the amount of published medical scientific information has grown enormously. For practitioners

and health policymakers, it is almost impossible to keep up with the rapidly increasing number of articles dealing with original research,¹ and thus they have to rely increasingly on review articles and other summaries of medical information.²

The value of a review greatly depends on the quality of the review process itself.³ A number of recent articles discuss the obligatory components of an up-to-date review, while some methodologists also present checklists to assess the ad-

herence of reviewers to these new standards for review articles.⁴⁻¹² During the last decade, there has been a refinement of the scientific guidelines for various aspects of reviewing the literature: optimization of the search strategy,¹³ an explicit assessment of the methodological quality of the studies included,^{5-7,14} and quantitative pooling of the outcomes of different studies.¹⁵ In addition, empirical studies have shown the influence of selective citing on review outcomes (reference bias)¹⁶ and the influence of language in the review process.¹⁷

We wondered whether there might be a relationship between various characteristics of a review article and its final conclusions. During our previous work on the effectiveness of spinal manipulation, we published a review of the randomized clinical trials (RCTs) on spinal manipulation for low back pain that were available at that time.¹⁸ As a part of our search strategy for RCTs we collected and screened review articles on this topic, and we were impressed by the vast number of review articles available.

See also pp 1935 and 1962.

Historically, spinal manipulation has been associated with chiropractic and osteopathy.¹⁹ Political controversies between chiropractors and related professionals and the medical establishment possibly played a role in the scientific assessment of the effectiveness of spinal manipulation.²⁰ There are many different approaches to spinal manipulation, and a great variety of control treatments are used in RCTs.²¹ For spinal manipulation, heterogeneity in the reviewers' opinions can be expected, thus making it a good subject for investiga-

From the Institute for Research in Extramural Medicine, Vrije Universiteit, Amsterdam (Drs Assendelft, Koes, and Bouter); Department of Epidemiology, University of Limburg, Maastricht (Dr Knipschild); and Department of Epidemiology and Biostatistics, Vrije Universiteit, Amsterdam (Dr Bouter), the Netherlands.

Reprint requests to Institute for Research in Extramural Medicine, Vrije Universiteit Amsterdam, Van Der Boerhorststraat 7, 1081 BT Amsterdam, the Netherlands (Dr Assendelft).

Table 1.—Criteria List for Assessment of Methodological Quality of Review Articles, With Weights Applied*

Study Selection	(30)
A. Description of inclusion and exclusion criteria	10
B. Search strategy	10
C. Emphasis on randomized clinical trials (RCTs)	10
Methodological Quality Assessment	(20)
D. Assessment of the validity RCTs	12
E. Number of reviewers	4
F. Blinding of reviewer(s)	2
G. Agreement of reviewer(s)	2
Intervention	(15)
H. Description of manipulative intervention(s)	8
I. Description of control intervention(s)	7
Data Presentation	(20)
J. Outcome presentation	14
K. Statistical pooling	3
L. Discussion power of negative RCTs	3
Evaluation	(15)
M. Overall conclusion	5
N. Discussion of heterogeneity of RCTs and outcomes	10

*The detailed criteria, with weights attached, are as follows:

- A. Description of study setting(s) included (ie, industry, general practice, hospital), intervention type(s) included (manipulation, mobilization, or both; specific techniques or professions [chiropractors, osteopaths, Cyriax technique]), outcome type(s) included (pain, global assessment, mobility, functional status, time until recovery, medical consumption), years covered, language(s) covered (2 points each).
- B. Established bibliographical database included (eg, *Index Medicus*, EMBASE) (5 points). Additional efforts to locate nonindexed RCTs (eg, citation tracking, correspondence with experts, manual search of nonindexed journals) (5 points).
- C. Randomized clinical trials only, or results of RCTs discussed separately from other study designs (10 points).
- D. Assessment per RCT included that is explicit (reproducible by readers of the review) regarding the similarity of treatment groups, similarity of treatment characteristics, adequacy in treatment of missing values (dropouts, loss to follow-up), success regarding blinding of outcome assessment, relevance of outcome measures, and adequacy of statistical analysis (2 points each).
- E. At least two independent reviewers (4 points).
- F. Reviewer(s) blinded for at least the outcomes of the RCTs (2 points).
- G. Agreement between reviewers reported (quantitative) and acceptable. In the event of reviewer, use of an assessment list with established reliability (2 points).
- H. Description of manipulative intervention per RCT: manipulation technique (3 points), profession or training manipulator (1 point), treatment frequency or number of treatments (2 points), and duration of treatment period (2 points).
- I. Description of control treatment per RCT: type (3 points), treatment frequency or number of treatments (2 points), duration of treatment period (2 points).
- J. Main outcome(s) (at least one of the following: pain, global assessment, mobility, functional status, medical consumption) presented separately per RCT per group (5 points). Presentation of the mean difference or ratio of outcome(s) between intervention group(s) and control group(s) (3 points) with confidence interval or SD per RCT (3 points). Graphic presentation of the most important outcome(s) (indicating outliers and distribution) per RCT (3 points). (That is, the most important outcome[s] that are rational in relation to the research question of the review. If not explicitly stated, at least pain, recovery, or functional status.)
- K. Statistical pooling of the most important outcome(s) or discussion of the reason why pooling is not indicated or warranted or pooling of the subset considered to be valid and similar enough (3 points).
- L. Elaboration on the power of negative RCTs: calculation of the power of each negative RCT (3 points) or narrative elaboration on the power of each negative RCT (2 points) or overall narrative elaboration on the power of the negative RCTs (1 point).
- M. Overall conclusion on the aggregated level of available RCTs on the effectiveness of manipulation presented (5 points).
- N. Identification of relevant subgroups (eg, study setting, disease classification) with explicit motivation (4 points). Discussion of the variety of treatment modalities in the intervention groups (eg, mobilization, manipulation; chiropractic, osteopathic, Cyriax technique) (2 points). Discussion of the variety of treatment modalities in the control groups (placebo, existing modality) (2 points). Discussion of the relationship between the methodological quality of the RCTs and outcome (2 points).

tion of which characteristics of reviews are associated with its final conclusions.

Our study question is, what is the relationship between review characteristics and the final conclusion of the reviewers regarding the effectiveness of spinal manipulation for low-back pain?

METHODS

Selection of Review Articles

A review article was included in our study if (1) the article dealt with the effectiveness of manipulation or mobilization for low back pain (not necessarily exclusively); (2) at least two randomized clinical trials were included in the assessment of effectiveness; (3) the article was not the original publication of an RCT itself; (4) it was published as a book chapter, report, or journal article; (5) the year of publication was up to 1993; and (6) the language was English or Dutch.

Relevant articles were identified by using the key words back-pain therapy, musculoskeletal diseases, joint diseases, manipulation-orthopedic, osteopathic medicine, chiropractic, evaluation studies, outcome, and process assessment for a computer search of *Index Medicus* (MEDLINE) for the period 1967 through 1993; by article citation tracking and library search; and through correspondence with experts in the field.

To check for the completeness of our search strategy, all available titles of RCTs up to 1991¹⁸ were selected in the *Science Citation Index* for 1970 through 1993. All articles citing these RCTs were scanned on title and type of publication in the *Science Citation Index* and subsequently in *Index Medicus* (MEDLINE).

Assessment of Methodological Quality

For the assessment of the methodological quality of the review articles, we used a list of standardized criteria based on previous work of Oxman and Guyatt,⁵ Light and Pillemer,⁷ and Mulrow.⁶ Further operationalization for the specific topic of spinal manipulation was based on our previous experience in this field.¹⁸ The list contained specified criteria covering the following domains: study selection, methodological quality assessment, description of the intervention, data presentation, and evaluation (Table 1).

A specific, arbitrary weight was applied to each criterion, as described in detail in the footnote to Table 1. The scoring was done by two nonblinded assessors independently. (By "assessors" we mean the first two authors of this article, W.J.J.A. and B.W.K., who ap-

plied the methodological criteria and performed the data extraction; by "reviewers" we mean authors who have written a review based on RCTs.) The final score was determined in a consensus meeting. If disagreements were not resolved during this meeting, a third blinded reviewer (L.M.B.) was consulted to make the final judgment. The total score thus determined is further referred to as the "review score."

Data Extraction

Characteristics of the review that were considered to be relevant were classified into two categories, methodological and subject related.

Methodological characteristics included the comprehensiveness of the literature search, selective citing of the available positive and negative RCTs, language, inclusion of non-RCTs, type of publication (ie, journal article or book chapter), and being indexed in *Index Medicus*. The previously mentioned review score was also considered to be a methodological characteristic. Since there is no registry of peer-reviewed journals, being indexed in *Index Medicus* was included as a substitute indicator for peer review.

Subject-related characteristics included publication in a specific spinal manipulation journal or book and the professional backgrounds of the reviewers (spinal manipulator or not). These characteristics were included to investigate whether a professional link between the reviewer or the publication medium and spinal manipulation influenced the reviewers' final conclusions.

For the determination of the comprehensiveness of the literature search of reviews, we assumed an RCT to be available for the reviewer(s) if it was published more than 2 years before the review.

The dependent variable (ie, the final conclusion of the reviewer[s]) was expressed as "negative," "neutral/no conclusion drawn," or "positive." The conclusion was considered to be negative if the reviewer(s) stated that the reference treatments were more effective than manipulation, and the conclusion was considered to be positive if the reviewer concluded that spinal manipulation was more effective than the reference treatments. The review was rated neutral if no conclusion was drawn or if the reviewers were neutral in their final conclusion. Reviews with the positive conclusion restricted to subgroups (eg, [sub]acute complaints) or to the duration of the effect (eg, short-term effects) were rated as positive. For these reviews, the nature of the restriction is reported separately.

The two assessors (W.J.J.A. and B.W.K.) were not blinded for the review

Table 2.—Methodological Review Scores and Reviewers' Conclusions

Source, y	Methodological Scoring*															Conclusion†	Restriction‡	
	A 10	B 10	C 10	D 12	E 4	F 2	G 2	H 8	I 7	J 14	K 3	L 3	M 5	N 10	Total 100		(Sub)acute	Short- Term
Shekelle et al, ²¹ 1992	6	10	10	12	4	...	...	3	3	14	3	...	5	6	76	Positive	Yes	Yes
Koes et al, ¹⁸ 1991	4	10	10	12	4	2	2	4	3	5	3	2	5	6	72	Neutral	...	...
Assendelft et al, ³³ 1992	4	10	10	12	4	2	2	4	3	5	...	...	5	...	61	Positive	No	Yes
Anderson et al, ³² 1992	6	10	10	...	...	...	...	3	3	...	3	...	5	6	46	Positive	No	No
Conlon, ³⁹ 1992	2	10	10	4	...	...	...	1	3	...	3	...	5	8	46	Positive	No	Yes
Shekelle et al, ⁷⁶ 1991	4	10	10	...	...	...	...	5	3	...	...	1	5	8	46	Positive	Yes	Yes
DiFabio, ⁴⁵ 1992	2	10	10	...	...	...	...	3	3	...	3	1	5	6	43	Positive	Yes	No
Lynou and Bygren, ⁶⁵ 1987	2	10	10	6	...	...	...	4	3	...	...	...	5	...	40	Positive	No	Yes
Abenhaim and Bergeron, ³¹ 1992	8	5	10	2	...	...	...	3	...	...	...	...	5	4	37	Positive	No	Yes
Ottenbacher and DiFabio, ⁷³ 1985	4	10	10	...	...	...	...	...	...	...	3	...	5	2	34	Positive	No	Yes
Winer, ⁷⁹ 1985	...	...	10	...	...	...	...	7	7	...	...	...	5	4	33	Positive	Yes	Yes
O'Donoghue, ⁷² 1986	...	...	10	...	...	...	...	8	7	...	...	...	5	2	32	Positive	No	No
Aufdemkampe, ³⁴ 1991	2	...	10	4	...	...	...	3	3	...	...	...	5	4	31	Positive	Yes	No
Bronfort, ³⁷ 1992	...	...	10	4	...	...	...	3	3	5	...	...	5	...	30	Positive	No	Yes
Deyo, ⁴² 1983	2	10	10	...	...	...	...	...	...	...	...	1	5	2	30	Positive	No	Yes
Evans, ⁴⁶ 1985	...	...	10	2	...	...	...	3	3	...	...	...	5	6	29	Positive	Yes	Yes
Jacobs et al, ⁵⁸ 1991	2	10	10	...	...	...	...	3	3	...	...	...	...	...	28	Neutral	...	...
Curtis, ⁴¹ 1988	...	...	10	...	...	...	...	8	...	...	...	1	5	4	28	Positive	Yes	Yes
Brunarski, ³⁸ 1984	2	...	10	4	...	...	...	...	...	...	...	2	5	4	27	Positive	No	No
Haldeman and Phillips, ⁵⁶ 1991	...	...	10	...	...	...	...	2	5	5	...	...	5	...	27	Positive	No	Yes
DiFabio, ⁴⁴ 1986	2	...	10	...	...	...	...	3	3	...	...	...	5	2	25	Positive	No	Yes
Branson and Buerger, ³⁶ 1984	...	...	10	2	...	...	...	...	3	...	...	3	5	2	25	Neutral	...	...
Haldeman, ⁵³ 1980	2	...	10	...	...	...	...	...	3	5	...	...	5	...	25	Neutral	...	...
Moritz, ⁶⁷ 1979	...	...	10	4	...	...	...	3	3	...	...	...	5	...	25	Positive	No	Yes
Spitzer, ⁷⁷ 1987	4	10	10	...	...	...	...	...	...	...	...	...	...	...	24	Neutral	...	...
Manga et al, ⁶⁶ 1993	...	...	10	...	...	...	...	4	3	...	...	...	5	...	22	Positive	No	No
Coulehan, ⁴⁰ 1985	...	...	10	...	...	...	...	...	3	...	...	...	5	4	22	Positive	No	Yes
Haldeman, ⁵⁴ 1983	...	...	10	...	...	...	...	...	3	...	...	...	5	4	22	Positive	Yes	Yes
Haldeman, ⁵⁵ 1986	...	...	10	...	...	...	...	...	...	...	...	...	5	6	21	Positive	Yes	Yes
Greenland and Haldeman, ⁵² 1980	...	...	10	...	...	...	...	...	3	...	...	...	5	2	20	Positive	No	Yes
Lawrence et al, ⁶⁴ 1992	4	5	10	...	...	...	...	...	...	...	...	...	...	...	19	Neutral	...	...
Jayson, ⁵⁹ 1986	...	...	10	...	...	...	...	...	...	...	...	...	5	4	19	Positive	Yes	Yes
Bouter, ³⁵ 1985	...	...	10	...	...	...	...	...	3	...	...	1	5	...	19	Neutral	...	...
Yates, ⁸⁰ 1980	...	...	10	...	...	...	...	...	...	...	...	...	5	4	19	Positive	Yes	Yes
Tan, ⁷⁸ 1992	...	...	10	...	...	...	...	...	3	...	...	...	5	...	18	Neutral	...	...
Deyo, ⁴³ 1991	...	...	10	...	...	...	...	...	3	...	...	...	5	...	18	Positive	No	Yes
Quinet and Hadler, ⁷⁵ 1979	...	...	10	...	...	...	...	...	3	...	...	...	5	...	18	Neutral	...	...
Grahame, ⁵¹ 1980	...	...	10	...	...	...	...	3	3	...	...	...	...	...	16	Neutral	...	...
LaBan and Taylor, ⁶² 1992	...	...	10	...	...	...	...	...	3	...	...	...	...	2	15	Neutral	...	...
Gatterman, ⁴⁸ 1990	...	...	...	...	...	...	...	6	...	...	...	...	5	4	15	Positive	No	No
Lankhorst, ⁶³ 1987	...	...	10	...	...	...	...	...	...	...	...	...	5	...	15	Positive	No	Yes
Gilbert, ⁵⁰ 1986	...	...	10	...	...	...	...	...	...	...	...	...	5	...	15	Neutral	...	...
Kukurin, ⁶¹ 1985	...	...	10	...	...	...	...	...	...	...	...	...	5	...	15	Positive	No	No
Paris, ⁷⁴ 1983	...	...	10	...	...	...	...	...	...	...	...	...	5	...	15	Positive	No	No
Inglis, ⁵⁷ 1979	...	...	10	...	...	...	...	...	...	...	...	...	5	...	15	Positive	No	No
Nachemson, ⁶⁹ 1977	...	...	10	...	...	...	...	...	...	...	...	...	5	...	15	Neutral	...	...
Geiringer et al, ⁴⁹ 1993	...	...	10	...	...	...	...	...	...	...	...	...	...	2	12	Neutral	...	...
Nachemson, ⁷⁰ 1985	...	...	10	...	...	...	...	...	...	...	...	...	...	...	10	Neutral	...	...
Nachemson, ⁷¹ 1985	...	...	10	...	...	...	...	...	...	...	...	...	...	...	10	Positive	No	Yes
Flor and Turk, ⁴⁷ 1984	...	...	...	...	...	...	...	...	...	...	...	...	5	...	5	Neutral	...	...
Nachemson, ⁶⁸ 1980	...	...	...	...	...	...	...	...	...	...	...	...	5	...	5	Neutral	...	...
Maximum score per item, %	12	25	94	11	6	4	4	24	23	5	12	8	85	22	26	...	...	...

*See the footnote to Table 1 for an explanation of the methodological scoring criteria.

†Positive conclusion indicates spinal manipulation was better than control treatments; and neutral, spinal manipulation was equal to control treatments or no conclusion drawn. Ellipses indicate not applicable.

‡(Sub)acute restriction indicates restriction of reported positive effects to patients with (sub)acute complaints (mainly <6 weeks); and short-term, restriction of duration of reported positive effects of spinal manipulation to a short duration (the exact period was mostly not specified).

characteristics when they determined the value of the independent variable (the final conclusion). Therefore, the final conclusion per review was recorded

both by the two assessors (W.J.J.A. and B.W.K., pair A) and also by two other trained clinical epidemiologists (A.F. de Winter, MSc, and D.A.W.M. van der

Windt, MSc, pair B). Since the opinion of pair B had the least risk of bias, as they were not authors of a review themselves and had no knowledge of the other

The relation between the methodological review score and the reviewers' conclusion.

items recorded (including the methodological criteria), their recordings were used in the final analysis.

Data Entry and Analysis

All data were entered and analyzed in SPSS/PC. Data were initially analyzed in frequency tables and by cross tabulation. To calculate crude and adjusted odds ratios (ORs), the ordinal and continuous independent variables (review characteristics) were all dichotomized. For continuous data the median was taken as the cutoff point. Univariate analysis was reported as ORs with corresponding 95% confidence intervals (CIs).

All characteristics with a statistically significant univariate OR and the methodological review score, which we considered to be essential a priori, were entered in a logistic regression model.

RESULTS

Review Identification

Initially 60 review articles were identified. Nine did not fulfill all the inclusion criteria: one was written in German,²² one dealt with methodological issues only,²³ and seven did not include two or more RCTs.²⁴⁻³⁰ The number of reviews on the effectiveness of spinal manipulation for low back pain in English or Dutch we thus identified (51) (Table 2)^{18,21,31-50} by far exceeded the number of original RCTs (28).¹⁸

The Reviewers' Conclusions

The agreement between the two independent pairs of assessors regarding the final conclusions of the reviewers was high: 92%. None of the reviewers stated that spinal manipulation was less

effective than reference treatments, 17 were neutral or offered no conclusion, and 34 were positive (Table 2). Of the 34 positive reviews, 15 (44%) restricted the positive conclusion to short-term results for manipulation and two (6%) restricted the conclusion to (sub)acute back pain, while nine (26%) restricted conclusions to both (sub)acute back pain and short-term results and eight positive reviews (24%) mentioned no restriction in their conclusions. Most reviewers did not define precisely what was meant by the terms (sub)acute and short-term.

Review Score

The two nonblinded reviewers initially agreed on 94% of all subitems scored (total, $51 \times 35 = 1785$ subitems). Nearly all the disagreements were due to reading errors or to a difference in interpretation of the methodological criteria. Since all the disagreements were resolved during the consensus meeting, the third reviewer did not have to be consulted.

The results of the assessment of the methodological quality of the review articles are presented in Table 2. The median score was 23 (range, 5 to 76), indicating the overall poor methodological quality of the reviews. Only three reviews had a score higher than 50 points of the 100 points possible. Recently published reviews tended to have higher scores.

The Figure presents the relationship between the review score and the reviewers' conclusions. Nine of the 10 highest-ranking (methodologically best) reviews had a positive reviewers' conclusion (see also Table 2).

Review Characteristics

The prevalence of the selected characteristics in the review articles included is presented in Table 3. In 16 (31%) of the 51 reviews, treatments other than spinal manipulation were also included. In 18 (40%) of the 45 reviews presenting data on this characteristic, a spinal manipulator was (co)author of the review. One third of the reviews were published in journals or books dealing exclusively with spinal manipulation. A minority of the reviews (22%) also included non-RCTs. Four reviews were in Dutch. Only a few reviews included statistical pooling. Half of the reviews included fewer than 60% of the RCTs available at the time of writing. Remarkably, the mean percentage of available RCTs included did not differ between the 12 reviews that reported an extensive search (use of a database as well as additional efforts [10 points on item B]) and the 39 reviews that had no description of such a comprehensive search (59% and 60%, respectively) (data not given in Table 3).

The results of the univariate analysis are presented in Table 3. If a review dealt with spinal manipulation only ($n=35$), it had a positive reviewers' conclusion much more often. Of these 35 reviews, 29 (82.9%) had a positive conclusion. Of the 16 reviews not dealing with spinal manipulation exclusively, only five (31.3%) had a positive conclusion. The resulting OR was 10.6 (95% CI, 2.68 to 42.1). The presence of a spinal manipulator in the review team also provided a strong association with a positive conclusion. This review characteristic was unknown for seven reviews. A third characteristic, a comprehensive literature search, also had a statistically significant OR. Also positively but not significantly related were publication in a spinal manipulation journal or book, an explicit validity assessment in the review, and a high review score. There were no clear associations with the reviewers' conclusion for selective citing of positive RCTs, language, inclusion of non-RCTs, publication in a journal (as opposed to publication in a book or report), or being indexed in *Index Medicus*.

In the logistic regression analysis we included the characteristics "review of manipulation only," "spinal manipulators in the review team," and "completeness of search," which were the characteristics with a statistically significant crude OR, and "methodological review score." Statistical analysis revealed a strong association between the characteristics spinal manipulator in review team and review of spinal manipulation only. Therefore, in the logistic regression analysis we present the data for two models, alternately including these colinear variables. In these two models, only "review of manipulation only" still had a statistically significant OR.

COMMENT

Potential Biases

As in other kinds of empirical research, reviewing the literature is also prone to several types of bias. Our first consideration is reference bias,¹⁶ a form of selection bias. In general, review articles are not systematically cited by others, as is the case with RCTs in review articles. Therefore, citation tracking is not a guarantee for completeness. In fact, we expect that there are more reviews existing on this subject than we have actually identified. Some reviews were obtained through "correspondence with experts," which is an accepted policy in literature search.¹¹ In most cases these experts were spinal manipulators themselves, which suggests that our sample of reviews may have a bias toward more positive conclusions. Being indexed in *Index Medicus* was recorded as a separate review char-

Table 3.—Relationship Between the Review Characteristics and the Reviewers' Conclusions

Characteristic	Total No. of Reviews	No. (%) of Reviewers' Conclusions Positive	No. (%) of Reviewers' Conclusions Negative or No Conclusion	Crude Odds Ratio (95% Confidence Interval)	Adjusted Odds Ratio Model 1* (95% Confidence Interval)	Adjusted Odds Ratio Model 2† (95% Confidence Interval)
Review manipulation only						
Yes	35	29 (82.9)	6 (17.1)	10.6 (2.68-42.1)	6.40 (1.44-28.4)	...
No	16	5 (31.3)	11 (68.8)			
Spinal manipulator in review team						
Yes	18	16 (88.9)	2 (11.1)	7.42 (1.42-38.8)	...	4.63 (0.80-26.9)
No	27	14 (51.9)	13 (48.1)			
Unknown	6					
Available randomized clinical trials (RCTs) included, %‡						
≥60	25	21 (84.0)	4 (16.0)	5.25 (1.40-19.6)	2.69 (0.61-11.9)	3.69 (0.85-16.1)
<60	26	13 (50.0)	13 (50.0)			
Methodological review score§						
≥23	25	20 (80.0)	5 (20.0)	3.42 (0.99-11.9)	1.90 (0.45-8.11)	2.01 (0.44-9.25)
<23	26	14 (53.8)	12 (46.2)			
Explicit validity assessment						
Yes	10	9 (90.0)	1 (10.0)	5.76 (0.66-49.9)	...	...
No	41	25 (61.0)	16 (39.0)			
Statistical pooling						
Yes	3	3 (100.0)	0 (0.0)	...	...	...
No	48	31 (64.6)	17 (35.4)			
Spinal manipulation journal or book						
Yes	17	14 (82.4)	3 (17.6)	3.27 (0.78-13.5)	...	...
No	34	20 (58.8)	14 (41.2)			
Inclusion of nonrandomized clinical trials (RCTs) in review						
Yes	11	9 (81.8)	2 (18.2)	2.70 (0.51-14.2)	...	...
No	40	25 (62.5)	15 (37.5)			
Language						
English	47	32 (68.1)	15 (31.9)	2.13 (0.27-16.6)	...	...
Dutch	4	2 (50.0)	2 (50.0)			
Positive-negative RCTs included						
% Positive RCTs >% negative RCTs	20	14 (70.0)	6 (30.0)	1.28 (0.38-4.28)	...	...
% Negative RCTs >% positive RCTs	31	20 (64.5)	11 (35.5)			
Indexed in <i>Index Medicus</i>						
Yes	26	17 (65.4)	9 (34.6)	0.89 (0.28-2.85)	...	...
No	25	17 (68.0)	8 (32.0)			
Publication in a journal						
Yes	32	20 (62.5)	12 (37.5)	0.60 (0.17-2.07)	...	...
No	19	14 (73.7)	5 (26.3)			

*Model 1 includes all characteristics with a statistically significant crude odds ratio and review score. "Spinal manipulators in review team" was excluded because of colinearity with "review of manipulation only." N=51. Ellipses indicate that this variable is not in the model.

†Model 2 includes all characteristics with a statistically significant crude odds ratio and review score. "Review of manipulation only" was excluded because of colinearity with "spinal manipulators in review team." N=45.

‡Dichotomized at median value.

acteristic. Table 3 shows that there was no difference in conclusion between the indexed and nonindexed reviews. Therefore, we conclude that the reviews obtained through our additional search efforts did not lead to substantial bias. We did not include unpublished reviews, since unpublished reviews do not have any impact on the general opinion. As a consequence, our conclusions apply to published reviews only.

The nonblinded recording of review characteristics and the reviewers' conclusions implies the possibility of information bias.⁸¹ However, we have presented our findings in a reproducible manner (Table 2). In this way, the critical reader is able to check our scores of the different items and can even execute a recalculation with application of other weights. Independent scoring by the two

assessors resulted in a high initial agreement (94%), and a subsequent consensus meeting provided correction of the eventual individual biases. The dependent variable (the reviewers' conclusion) was even recorded by two pairs of assessors independently (also with an excellent agreement between the pairs of 92%).

The fact that the authors of this article had a hand in writing three of the reviews^{18,33,35} introduces another form of information bias, which we might call reviewer bias. We were aware of this and considered exclusion of our own reviews. However, this would have had no significant influence on the conclusions of this article.

Criteria and Weights

The choice of items and the weighing of the items in a methodological quality

assessment are prone to subjective preferences. However, even in the more extensively researched field of methodological assessment of RCTs, a published validated list is still missing.⁸² Most lists used in the assessment of RCTs use an additive model,⁸² as we did. Therefore, in the absence of empirical evidence to the contrary, we think that use of a list based on an additive model is still justified. The choice of the items for our list closely matches similar recent publications on quality of reviews,^{5-7,12} indicating at least concurrent construct validity. However, the process of weighing items provides more variation and cannot easily be compared among publications. Therefore, we also performed an alternative (sensitivity) analysis, applying equal weights to all items. In this procedure, the ranking of the better re-

views was almost identical (data not shown).

Quality of the Reviews Included

The methodological quality of reviews can be well assessed with the type of criteria list we used. The agreement reached among the assessors is acceptable, and the list is easy to use and robust for changes in the weighing of items. However, a list of this sort mainly assesses the description of the review procedures as such, not the quality of the performance of the review procedure. For item B (comprehensive search strategy), there was a clear indication that this was the case. This disadvantage is similar for the lists used to evaluate the quality of RCTs.⁸² The low review score for most reviews, indicating low methodological quality, was striking. This finding is consistent with other critical assessments of reviews in the medical field.^{6,10,12,83} The great majority of reviews lacked essential components (Table 2). This low review quality clearly indicates the need for the development of standardized high quality in the conducting and reporting of meta-analyses, similar to the recent efforts to improve the reporting of RCTs.⁸⁴

The Relationship Between Review Characteristics and the Overall Conclusion of the Review

This study shows that a review of spinal manipulation is more likely to have a positive outcome if the review deals with spinal manipulation only, if a spinal manipulator was among the authors, and if a comprehensive search was described. Some characteristics were unlikely to produce significant results in our sample because of the low prevalence of the characteristic (eg, pooling, explicit quality assessment, inclusion of non-RCTs, and language).

It is difficult to generalize our remarkable findings to other fields in which reviews are written, because spinal manipulation and back pain combine certain specific methodological features.^{85,86} Back pain is not measured with "hard" outcomes,⁸⁷ and in the RCTs numerous different outcome measures are applied. No uniform classification into prognostically homogeneous subgroups is available. Most importantly, the RCTs on spinal manipulation are, on the average, of a poor methodological quality.^{18,21} These aspects leave relatively much room for subjective interpretation of the results of the relevant effectiveness studies. For spinal manipulators as reviewers there is an increased chance of resulting bias, because there is a tight link of the profession to a single type of treatment. For example, cardiologists have

several options to prevent reinfarction and therefore will not experience a professional threat if one of the available treatments proves to be ineffective. For spinal manipulators, however, proof of the effectiveness of their main therapy is of utmost importance for the validation of their professional identity. It is not clear at all whether this professional relationship with the treatment indeed caused bias in reviewing. It can be equally true that reviewers who were not spinal manipulators were biased against the therapy.

Because of the above-mentioned considerations, more empirical research on the review methods applied to other therapies in other professional fields is needed to further explore our findings.

References

1. Sackett DL, Haynes RB, Guyatt GH, Tugwell P. *Clinical Epidemiology: A Basic Science for Clinical Medicine*. 2nd ed. Boston, Mass: Little Brown & Co Inc; 1991:359-378.
2. Guyatt GH, Rennie D. Users' guides to the medical literature. *JAMA*. 1993;270:2096-2097.
3. Haynes RB. Clinical review articles—should be as scientific as the articles they review. *BMJ*. 1992;304:330-331.
4. Milne R, Chambers L. Assessing the scientific quality of review articles. *J Epidemiol Community Health*. 1993;47:169-170.
5. Oxman AD, Guyatt GH. Guidelines for reading literature reviews. *Can Med Assoc J*. 1988;138:697-703.
6. Mulrow CD. The medical review article: state of the science. *Ann Intern Med*. 1987;106:485-488.
7. Light RJ, Pillemer DB. *Summing up the evidence: the science of reviewing research*. Cambridge, Mass: Harvard University Press; 1984:160-173.
8. O'Rourke K, Detsky AS. Meta-analysis in medical research: strong encouragement for higher quality in individual research efforts. *J Clin Epidemiol*. 1989;42:1021A.
9. Goldschmidt PG. Information synthesis: a practical guide. *Health Serv Res*. 1986;21:215-237.
10. Williamson JW, Goldschmidt PG, Colton T. The quality of medical literature: an analysis of validation assessments. In: Bailar JC, Mosteller F, eds. *Medical Uses of Statistics*. Waltham, Mass: NEJM Books; 1986:370-391.
11. Oxman AD. Checklists for review articles. *BMJ*. 1994;309:648-651.
12. Sacks HS, Berrier J, Reitman D, Ancona-Berk VA, Chalmers TC. Meta-analyses of randomized controlled trials. *N Engl J Med*. 1987;316:450-455.
13. Dickersin K, Scherer R, Lefebvre C. Identifying relevant studies for systematic reviews. *BMJ*. 1994;309:1286-1291.
14. Oxman A, ed. *Preparing and Maintaining Systematic Reviews: The Cochrane Collaboration Tool Kit*. Oslo, Norway: National Institute of Public Health; 1994. Available by Internet from the Cochrane Collaboration Centres.
15. Dickersin K, Berlin JA. Meta-analysis: state-of-the-science. *Epidemiol Rev*. 1992;14:154-176.
16. Gotzsche PC. Reference bias in reports of drug trials. *BMJ*. 1987;295:654-656.
17. Gregoire G, Derderian F, LeLorier J. Selecting the language of the publications included in a meta-analysis: is there a Tower of Babel bias? *J Clin Epidemiol*. 1995;48:159-163.
18. Koes BW, Assendelft WJJ, van der Heijden GJMG, Bouter LM, Knipschild PG. Spinal manipulation and mobilization for back and neck pain: a blinded review. *BMJ*. 1991;303:1298-1303.
19. Wardwell WI. *Chiropractic: History and Evolution of a Profession*. St Louis, Mo: Mosby-Year Book; 1992.
20. Chapman-Smith D. The Wilk case. *J Manipulative Physiol Ther*. 1989;12:142-146.
21. Shekelle PG, Adams AH, Chassin MR, Hurwitz EL, Brook RH. Spinal manipulation for back pain. *Ann Intern Med*. 1992;117:590-598.
22. Zinn WM. Physiotherapie degenerativer Wirbelsäulenerkrankungen. *Ther Umsch*. 1981;38:642-650.
23. Leboeuf C. A review of data reports published in the *Journal of Manipulative and Physiological Therapeutics* from 1986 to 1988. *J Manipulative Physiol Ther*. 1990;13:89-95.
24. Lee CK. Office management of low back pain. *Orthop Clin*. 1988;19:797-804.
25. MacDonald MJ, Morton L. Chiropractic evaluation study—reprint of the Task III report of relevant literature. *J Chiropractic Res*. Autumn 1988:2-15.
26. Nachemson A. Physiotherapy for low back pain patients—a critical look. *Scand J Rehabil Med*. 1969;85-90.
27. Nachemson AL. Newest knowledge of low back pain—a critical look. *Clin Orthop*. 1992;279:8-20.
28. Pedersen P. The identity of chiropractic practice with special reference to Western Europe: a literature review. *Eur J Chiropractic*. 1990;38:41-55.
29. Nachemson A. Physiotherapy for back pain patients. *Scand J Rehabil Med*. 1969;1:85-90.
30. Frymoyer JW. Back pain and sciatica. *N Engl J Med*. 1988;318:291-300.
31. Abenhaim L, Bergeron AM. Twenty years of randomized clinical trials of spinal manipulative therapy for back pain: a review. *Clin Invest Med*. 1992;15:527-535.
32. Anderson R, Meeker WC, Wirrick BE, Mootz RD, Kirk DH, Adams A. A meta-analysis of clinical trials of spinal manipulation. *J Manipulative Physiol Ther*. 1992;15:181-194.
33. Assendelft WJJ, Koes BW, van der Heijden GJMG, Bouter LM. The efficacy of chiropractic for back pain—blinded review of the relevant randomized clinical trials. *J Manipulative Physiol Ther*. 1992;15:487-494.
34. Aufdenkampe G. Effektonderzoek naar manuele therapie van de wervelkolom. In: Winkel D, ed. *Orthopedische Geneeskunde en manuele Therapie, Deel 4a: Wervelkolom—Diagnostiek en Therapie*. Houten/Zaventem, the Netherlands: Bohn/Stafleu/Van Lochem; 1991:115-131.
35. Bouter LM. Het effect van manuele therapie bij lage rugpijn. *Medisch Contact*. 1985;40:357-360.
36. Branson MH, Buerger AA. Randomized clinical trials in the validation of cervical and lumbar manipulation. In: Greenman PE, ed. *Concepts and Mechanisms of Neuromuscular Functions*. New York, NY: Springer; 1984:90-105.
37. Bronfort G. Effectiveness of spinal manipulation and adjustments. In: Haldeman S, ed. *Principles and Practice of Chiropractic*. 2nd ed. New York, NY: Appleton-Century-Crofts; 1992:415-441.
38. Brunarski DJ. Clinical trials of spinal manipulation: a critical appraisal and review of the literature. *J Manipulative Physiol Ther*. 1984;7:243-249.
39. Conlon J. *Chiropractic Treatment of Neck and Back Disorders: A Review of Selected Studies*. Ottawa, Ontario: Canadian Coordinating Office for Health Technology Assessment; 1992.
40. Coulehan JL. Adjustments, the hands and healing. *Cult Med Psychiatry*. 1985;9:353-382.
41. Curtis P. Spinal manipulation: does it work? *Occup Med*. 1988;3:31-44.
42. Deyo RA. Conservative therapy for low back pain: distinguishing useful from useless therapy. *JAMA*. 1983;250:1057-1062.
43. Deyo RA. Historic perspective on conservative treatments for acute back problems. In: Mayer TG, Mooney V, Gatchel RJ, eds. *Contemporary Conservative Care for Painful Spinal Disorders*. Philadelphia, Pa: Lea & Febiger; 1991:169-180.
44. DiFabio RP. Clinical assessment of manipulation and mobilization of the lumbar spine. *Phys Ther*. 1986;66:51-54.
45. DiFabio RP. Efficacy of manual therapy. *Phys Ther*. 1992;72:853-864.

46. Evans DP. The design and results of clinical trials of lumbar manipulation: a review. In: Buerger AA, Greenman PE, eds. *Empirical Approaches to the Validation of Spinal Manipulation*. 2nd ed. Springfield, Ill: Charles C Thomas Publisher; 1985:228-238.
47. Flor H, Turk DC. Etiological theories and treatment for chronic back pain, part I: somatic models and interventions. *Pain*. 1984;19:105-121.
48. Gatterman MI. In the patient's interest. In: Gatterman MI, ed. *Chiropractic Management of Spine-Related Disorders*. Baltimore, Md: Williams & Wilkins; 1990:397-404.
49. Geiringer SR, Kincaid CB, Rechtien JR. Traction, manipulation, and massage. In: DeLisa JA, Gans BM, Currie DM, eds. *Rehabilitation Medicine: Principles and Practice*. 2nd ed. Philadelphia, Pa: JB Lippincott; 1993:451-462.
50. Gilbert JR. Management of low-back pain in family practice: a critical review. *Can Fam Physician*. 1986;32:1855-1861.
51. Grahame R. Clinical trials in low back pain. *Clin Rheum Dis*. 1980;6:143-157.
52. Greenland S, Haldeman S. Controlled clinical trials of manipulation: a review and a proposal. *J Occup Med*. 1980;22:670-676.
53. Haldeman S. Spinal manipulative therapy in the management of low back pain. In: Finneson BE, ed. *Low Back Pain*. 2nd ed. Philadelphia, Pa: JB Lippincott; 1980:245-275.
54. Haldeman S. Spinal manipulative therapy: a status report. *Clin Orthop*. 1983;179:62-70.
55. Haldeman S. Spinal manipulative therapy in sports medicine. *Clin Sports Med*. 1986;5:277-293.
56. Haldeman S, Phillips RB. Spinal manipulative therapy in the management of low back pain. In: Frymoyer JW, ed. *The Adult Spine—Principles and Practice*. New York, NY: Raven Press; 1991:1581-1605.
57. Commission of Inquiry (Inglis BD, chairman). *Chiropractic in New Zealand: Report*. Wellington, New Zealand: Government Printer; 1979:200-226.
58. Jacobs JWG, Rasker JJ, van Riel PLCM, Gribnau FWJ, van Putte LBA. Alternatieve geneeswijzen bij reumatische aandoeningen: een literatuuronderzoek. *Ned Tijdschr Geneesk*. 1991;135:317-322.
59. Jayson MIV. A limited role for manipulation. *BMJ*. 1986;293:1454-1455.
60. Jull GA. The management of acute low back pain. In: Grieve GP, ed. *Modern Manual Therapy of the Vertebral Column*. Edinburgh, Scotland: Churchill Livingstone; 1986:740-749.
61. Kukurin GW. Chiropractic and spinal manipulative therapy: a critical review of the literature. *Am Chiropr Assoc J*. 1985;19:41-49.
62. LaBan MM, Taylor RS. Manipulation: an objective analysis of the literature. *Orthop Clin North Am*. 1992;23:451-459.
63. Lankhorst GJ. Manuele therapie. *Ned Tijdschr Geneesk*. 1987;131:898-901.
64. Lawrence VA, Tugwell P, Gafni A, Kosuwan W, Spitzer WO. Low back pain and economics of therapy—the iterative loop approach. *J Clin Epidemiol*. 1992;45:301-311.
65. Lynou N, Bygren LO. Manipulative treatment of back pain. In: Commission on Alternative Medicine (Westin A, chairman), ed. *Legitimation (State Registration) for Certain Chiropractors*. Stockholm, Sweden: Social Dept; 1987:135-187.
66. Manga P, Angus D, Papadopoulos C, Swan W. *The Effectiveness and Cost Effectiveness of Chiropractic Management of Low-Back Pain*. Richmond Hill, Ontario: Kenilworth; 1993:35-57.
67. Moritz U. Evaluation of manipulation and other manual therapy. *Scand J Rehabil Med*. 1979;11:173-179.
68. Nachemson A. A critical look at conservative treatment for low back pain. In: Jayson MJV, ed. *The Lumbar Spine and Back Pain*. Kent, England: Tunbridge Wells; 1980:453-466.
69. Nachemson A. Pathophysiology and treatment of back pain: a critical look at the different types of treatment. In: Buerger AA, Tobis AA, eds. *Approaches to the Validation of Manipulative Therapy*. Springfield, Ill: Charles C Thomas Publisher; 1977:42-57.
70. Nachemson AL. Advances in low back pain. *Clin Orthop*. 1985;200:266-278.
71. Nachemson A. Recent advances in the treatment of low back pain. *Int Orthop*. 1985;9:1-10.
72. O'Donoghue CE. Manipulation trials. In: Grieve GP, ed. *Modern Manual Therapy of the Vertebral Column*. Edinburgh, Scotland: Churchill Livingstone; 1986:849-859.
73. Ottenbacher K, DiFabio RD. Efficacy of spinal manipulation/mobilization therapy: a meta analysis. *Spine*. 1985;10:833-837.
74. Paris SV. Spinal manipulative therapy. *Clin Orthop*. 1983;179:55-61.
75. Quinet RJ, Hadler NM. Diagnosis and treatment of backache. *Semin Arthritis Rheum*. 1979;8:261-287.
76. Shekelle PG, Adams AH, Chassin MR, Hurwitz EL, Phillips RB, Brook RH. *The Appropriateness of Spinal Manipulation for Low-Back Pain*. Santa Monica, Calif: Rand Corp; 1991.
77. Quebec Task Force on Spinal Disorders (Spitzer WO, chairman). Report of the Quebec Task Force on Spinal Disorders. *Spine*. 1987;12(suppl):9-55.
78. Tan JC. Therapy in the management of common low back pain. *Baillieres Clin Rheumatol*. 1992;6:629-655.
79. Winer CER. A survey of controlled clinical trials of spinal manipulation. In: Glasgow EF, Twomey LT, Kleynhans AM, Ideczak RM, eds. *Aspects of Manipulative Therapy*. 2nd ed. Edinburgh, Scotland: Churchill Livingstone; 1985:97-108.
80. Yates AAH. Treatment of back pain. In: Jayson MJV, ed. *The Lumbar Spine and Back Pain*. Kent, England: Tunbridge Wells; 1980:437-452.
81. Bouter LM, ter Riet G. Meta-analysis for physiotherapists—on the importance of standardization and blinding in the study of literature. In: *Proceedings of the Third International Physiotherapy Congress*. Sydney, Australia: Link Printing; 1990:438-441.
82. Moher D, Jadad AR, Nichol G, Penman M, Tugwell P, Walsh S. Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. *Control Clin Trials*. 1995;16:62-73.
83. Silagy CA. An analysis of review articles published in primary care journals. *Fam Pract*. 1993;10:337-341.
84. Standards of Reporting Trials Group. A proposal for structured reporting of randomized controlled trials. *JAMA*. 1994;272:1926-1931.
85. Hoffman RM, Turner JA, Cherkin DC, Deyo RA, Herron LD. Therapeutic trials for low back pain. *Spine*. 1994;19(suppl 18):2068-2075.
86. Bloch R. Methodology in back pain trials. *Spine*. 1987;12:430-432.
87. Deyo RA. Measuring the functional status of patients with low back pain. *Arch Phys Rehabil*. 1988;69:1044-1053.