The Effectiveness of Chiropractic for Treatment of Low Back Pain: An Update and Attempt at Statistical Pooling

Willem J. J. Assendelft, M.D., Ph.D.*, Bart W. Koes, Ph.D.*, Geert J. M. G. van der Heijden, M.Sc.† and Lex M. Bouter, Ph.D.§

ABSTRACT

Objective: To determine the effectiveness of chiropractic treatment for patients with low back pain by means of a systematic review of the literature.

Data Sources: Randomized clinical trials (RCTs) on chiropractic were identified with a Medline and Embase search (1966-1995), by citation tracking, and by hand searching of the relevant chiropractic reference systems (CRAC and Index to Chiropractic Literature).

Study selection: All RCTs on low back pain that involved chiropractors as therapists.

Data Extraction: Methodological quality was assessed independently by two reviewers on 14 items covering internal validity, informativeness and study size. Data were extracted on: patients (initial referral, duration of complaints, radiation of pain); outcomes (four different types); and timing of follow-up (short-term, intermediate and long-term). Statistical pooling was intended, according to a preset analysis plan, to include subgroup analysis.

Data synthesis: Eight RCTs were identified. All RCTs had serious flaws in their design, execution and reporting. Because of the great variety of outcome measures and follow-up timing, there was insufficient data to enable statistical pooling of the RCTs. A narrative review, however, did not provide convincing evidence for the effectiveness of chiropractic for acute or chronic low back pain.

Conclusions: There is certainly a need for correctly executed trials. In future research on the effectiveness of chiropractic, guidelines for uniform execution and reporting of RCTs should first be established to enable subsequent statistical pooling in systematic reviews of chiropractic trials. (J Manipulative Physiol Ther 1996; 19:499–507).

Key Indexing Terms: Chiropractic; Low Back Pain; Effectiveness; Publication Bias; Meta-analysis; Systematic Review

INTRODUCTION

In 1992, we published a review on the effectiveness of chiropractic in the treatment of low back pain, in which we summarized the results of five randomized clinical trials (RCTs) available at that time (1). The methodological quality of these RCTs was assessed on the basis of a criteria list. A weight was applied to each criterion, thereby providing a methodological quality score for each RCT. The percentage scores for the 5 RCTs ranged from 20 to 48 of a maximum of 100 points. We refrained from statistical pooling because of the low methodological scores and the clinical heterogeneity of the studies. Our conclusion was that, although the limited number of chiropractic RCTs and their poor general methodological quality precluded the drawing of strong conclusions, chiropractic seemed to be an effective treatment for back pain. However, more studies of higher methodological quality were clearly still needed (1).

During the same period, two other reviews [Shekelle et al. (2) and Anderson et al. (3)] that included a statistical pooling were published. In a recent overview of reviews on spinal manipulation, both reviews ranked among the methodological best (4) and have undoubtedly contributed to the favorable recommendations made in recently issued clinical guidelines on the role of spinal manipulation in the treatment of acute low back pain (5, 6). These reviews, however, evaluated the effectiveness of spinal manipulation in general, not chiropractic manipulation specifically. In our previous review, we argued that conclusions on the effectiveness of chiropractic should be based on chiropractic studies only (1). This standpoint is supported by the recent fierce debate on the Manga Report, commissioned by the Ontario Ministry of Health, which reviewed the effectiveness and cost-effectiveness of chiropractic management of low back pain (7, 8). In a reaction to this report, professional organizations of nonchiropractic spinal manipulators argued that the favorable conclusions on chiropractic published in this report were, in fact, based on RCTs on spinal

* Senior Investigator, Institute for Research in Extramural Medicine, Vrije Universiteit, Amsterdam, The Netherlands.
† Research Fellow, Department of Epidemiology, University of Limburg, Maastricht, The Netherlands.
‡ Director and Professor of Epidemiology, Department of Epidemiology and Biostatistics, Vrije Universiteit, Amsterdam, The Netherlands.
Send reprint requests to: Willem J. J. Assendelft, Institute for Research in Extramural Medicine, Vrije Universiteit, Amsterdam, Van der Boechorststraat 7, 1081 BT Amsterdam, The Netherlands.
Paper submitted February 20, 1996; in revised form March 25, 1996
manipulation in the broadest sense, mainly involving nonchiropractors as therapists (9-11). In fact, in the review of Shekelle et al., the subset of RCTs involving patients with acute low back pain, which generated the favorable conclusions on spinal manipulation, did not contain one single chiropractic RCT (2). On the other hand, neither the review by Shekelle et al. (2) nor the review by Anderson et al. (3) included the extensive Meade study, which compared chiropractic and hospital treatment. In this pragmatic trial, the patients in the hospital treatment group, which was the comparison group for chiropractic, were also treated with manipulation: 72% received Maitland manipulation or mobilization and 12% Cyriax manipulation (12).

Since the publication of our previous review (1), three additional chiropractic RCTs [Triano et al. (13), Pope et al. (14, 15), Herzog et al. (16)] and the results from the extended follow-up of the Meade study (17) have been published. Therefore, an update of our previous review seems to be warranted.

**Changes in Review Methodology**

In recent years, several new insights into review methodology have developed. In 1992 we refrained from statistical pooling because of the low methodological quality of the available RCTs (1). However, not pooling is only one of the possible ways of dealing with variation of methodological quality. Detsky et al. provide several other options, which focus on incorporating methodological quality if statistical pooling of RCTs is performed (18).

In the description of the methodological checklist (19) we used in our previous review (1), we did not actually define "methodological quality." Our list, like many other methodological quality assessment scales (20), contained items referring to different domains of the publication of an RCT. Specifically, our checklist contained items on the internal validity, the informativeness of reporting and the size of the RCT at issue. The internal validity items (e.g., quality of randomization, completeness of follow-up, blinding of outcome assessment) concern potential bias. Informativeness items (e.g., prognostic information on the patients included, type of control treatment, timing of follow-up) are subject-specific and refer to the information needed to judge the generalizability of the studies, and also provide information essential for performing subgroup analyses (e.g., patients with acute vs. chronic complaints, placebo-controlled vs. pragmatic studies). The item concerning the size of the RCT relates to the precision of the quantitative estimate and the "power" to detect relevant differences between the intervention at issue (in this case chiropractic) and control treatments. Statistical pooling is meant, in essence, to combine studies to increase the power to detect (meaningful) differences between the intervention and control treatments (21). Therefore, the precision item is only relevant if statistical pooling is not performed. In the latter case, the quantitative synthesis is often restricted to "vote counting," which consists of a comparison of positive and negative studies (22). At present, all available methodological checklists are still based on general principles of intervention research, not on empirical research (20). Only very recently has empirical research on the relative importance (weight) of validity items been published. Schulz et al. found concealment of treatment allocation and blinding of outcome measurement to be related to study results, indicating that these items are sources of bias (23). Generation of allocation and completeness of follow-up did not seem to be sources of bias in this study. Thus, based on empirical research, the relative importance of each of the items on our checklist that were based on "generally accepted principles of intervention research" (19) remains uncertain. Applying weights to items, as we did in the previous study, is therefore not yet substantiated by empirical research (20).

An update of our previous review of RCTs on the effectiveness of chiropractic for low back pain seems to be warranted, because a number of new RCTs have been published and because review methodology is rapidly evolving. Compared with our 1992 review (1), several aspects of the methodology should be revised:

- Statistical pooling should be attempted, and methodological quality can be incorporated in this pooling, even if the quality is relatively low.
- The criteria for the methodological assessment should be divided into validity items, informativeness items and precision items. The validity items provide the components of the methodological quality that are directly related to the statistical pooling. The informativeness items will be used in subgroup analyses and are important for judging the generalizability of the results. The precision item (size of the study) is only of relevance if pooling is not established.
- No weights will be applied to the methodological items.
- The items on randomization and completeness of follow-up should be revised to make these comparable to those of Schulz et al. (23).

**METHODS**

**Selection of Studies**

A MEDLINE literature search (keywords: backache, musculoskeletal diseases, joint diseases, manipulation, osteopathy, chiropractic, evaluation studies, outcome and process assessment) and an Embase search (keywords: manipulative-medicine, chiropractic) were carried out for the period January 1966–June 1995. In addition, the references given in relevant publications were further examined. Abstracts and unpublished studies were not selected. The following criteria were applied:

1. The (experimental) treatment regimen included chiropractic for treatment of low back pain, consisting of manipulation or mobilization of the spine. Additional interventions were allowed.
2. The follow-up of patients was clinically relevant (> 1 day).
3. Only truly randomized studies were included (quasirandom procedures, like alternate allocation or selecting birth dates, were excluded).
4. The article was written in English.
Assessment of the Methodological Quality of the Studies

All trials were scored according to the criteria listed in Table 1. The list is based on the criteria of Koes et al. (19), which we also used in our previous review (1). Some items were revised slightly to make these comparable with those of Schulz et al. (23) and weights were no longer applied (see Appendix A). For each item, we determined whether it concerned (internal) validity (V), informativeness (I) or precision (P). All studies were assessed by two reviewers (WJJA, GJMvdH) independently of one another. In a subsequent meeting, the reviewers tried to reach a consensus on each of the criteria upon which they disagreed. Where disagreement persisted, a third reviewer (LMB) made the decision.

Data Extraction

**Patient characteristics.** For the analysis of patient characteristics, the following categories were detailed: (a) duration: acute (≤ 3 wk), subacute (3 wk < 13 wk), chronic (≥ 13 wk); (b) radiating complaints: pain radiating below the knee (indicative for nerve root involvement) or not radiating below the knee; (c) location of initial referral: a chiropractic practice or a medical setting [which seems, according to the results of Meade et al. (12), to be an important prognostic factor].

**Outcome measurement.** Four types of outcomes were extracted (listed in order of presentation): (a) Success rate: proportion of patients recovered, according to a (forced) dichotomized overall assessment of the clinical state by either the patient or an assessor; (b) Visual analog score (VAS) pain: severity of pain in the individual patient (usually expressed on a VAS); (c) Back pain-specific functional status (e.g., Oswestry back-pain disability scale); (d) Generic functional status (e.g., Sickness Impact Profile).

Physiological variables, such as spinal flexibility and degrees of straight leg raising, were not used in the assessment of results, because in low back pain research, this kind of outcome correlates poorly with the clinical status of the patient (24).

**Timing of follow-up measurements.** To assess the duration of effects, the timing of follow-up measurements were divided into: (a) short-term (≤ 3 wk after randomization); (b) intermediate-term (3 wk < 3 months) and (c) long-term (≥ 3 months) (2).

Data Analysis

This section describes the intended analysis plan. In Results, we explain that the RCTs did not include sufficient data to enable a statistical pooling of the results of the RCTs; we have retained this section for didactic reasons and to support the recommendations in Discussion.

We intended to analyze the data according to a preset plan. A study was judged to be positive if the authors concluded (in their abstract or conclusions) that chiropractic was more effective than the control treatment. This usually meant that the difference in effect for the primary outcome measure was statistically significant at the conventional 5% level. In a negative study, the authors generally reported no differences between the study treatments or more favorable results for the control treatment. Effects for the binary data (success rate) (see Types of Outcome) were to be expressed as pooled odds ratios, using either a fixed or a random effects model (25). Confidence levels were to be set at 95%. The continuous data (VAS pain, back pain-specific functional status, generic functional status) were to be expressed in an effect size (Cohen’s D) (26).

Separate stratified analyses, according to the preset analysis plan were: (a) placebo-controlled trials vs. pragmatic trials, (b) low quality trials vs. high quality trials, (c) acute vs. subacute vs. chronic complaints, (d) initial referral to a chiropractic practice vs. initial referral to a medical setting and (e) short-term vs. intermediate vs. long-term follow-up. For analyses not related to follow-up timing (a, b, c and d) in trials presenting various follow-up moments, we chose the follow-up moment considered most important according to the authors of the article at issue.

### RESULTS

#### Trials Included

A total of 8 chiropractic RCTs on the effectiveness of chiropractic treatment for low back pain were identified (12-17, 27-30). Two RCTs with a follow-up time of < 1 day were excluded (31, 32).

All RCTs included patients with subacute and chronic complaints (Table 2). None of these publications enabled a further

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Domain*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study population</td>
<td>A Homogeneity</td>
</tr>
<tr>
<td>Base comparability of groups</td>
<td>B Baseline comparability of groups</td>
</tr>
<tr>
<td>Concealment of randomization</td>
<td>C1 Concealment of randomization</td>
</tr>
<tr>
<td>Generation of allocation sequence</td>
<td>C2 Generation of allocation sequence</td>
</tr>
<tr>
<td>Reason for loss to follow-up described for each study group separately</td>
<td>D Reason for loss to follow-up described for each study group separately</td>
</tr>
<tr>
<td>&lt;20% loss to follow-up (for follow up ≥3 months)</td>
<td>E &lt;20% loss to follow-up (for follow up ≥3 months)</td>
</tr>
<tr>
<td>&lt;10% loss to follow-up (for follow up &lt;3 months)</td>
<td>F &lt;10% loss to follow-up (for follow up &lt;3 months)</td>
</tr>
<tr>
<td>Number of subjects in the smallest group</td>
<td>G Number of subjects in the smallest group</td>
</tr>
<tr>
<td>Interventions</td>
<td>H Interventions protocolized and described</td>
</tr>
<tr>
<td>Pragmatic study</td>
<td>I Pragmatic study</td>
</tr>
<tr>
<td>Co-interventions avoided or equal</td>
<td>J Co-interventions avoided or equal</td>
</tr>
<tr>
<td>Placebo-controlled</td>
<td>K Placebo-controlled</td>
</tr>
<tr>
<td>Mentioning good qualification of manipulative therapist</td>
<td>L Mentioning good qualification of manipulative therapist</td>
</tr>
<tr>
<td>Effect</td>
<td>M Effect size</td>
</tr>
<tr>
<td>Patients blinded</td>
<td>N Intention-to-treat analysis</td>
</tr>
<tr>
<td>Outcome measures relevant</td>
<td>O Follow-up period adequate</td>
</tr>
<tr>
<td>Blinded outcome assessment</td>
<td>P Data presentation and analysis</td>
</tr>
<tr>
<td>VIntention-to-treat analysis</td>
<td></td>
</tr>
<tr>
<td>Q Frequencies of most important outcomes presented for each treatment group</td>
<td>P Precision of quantitative estimate of individual study</td>
</tr>
</tbody>
</table>

Adapted from Koes et al. (19). Note: no weights applied.

* Domain of item: V, validity of study; I, information necessary for analysis of sensitivity and to assess generalizability; P, precision of quantitative estimate of individual study.
Table 2. Randomized clinical trials on the effectiveness of chiropractic treatment of back pain: patient characteristics, interventions and results

<table>
<thead>
<tr>
<th>First author (ref)</th>
<th>No. pts.</th>
<th>Indication</th>
<th>Chiropractic treatment (no. patients)</th>
<th>Control treatment [type] (no. patients)</th>
<th>Results</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meade (12, 17)</td>
<td>741</td>
<td>Patients with acute (&lt;1 month) and chronic (&gt;1 month) low back pain with or without radiation; nerve root not affected</td>
<td>Standardized chiropractic technique (384)</td>
<td>Physiotherapy including Maitland mobilization/manipulation, Cytaphrac manipulation, traction and exercises [pragmatic] (357)</td>
<td>Difference in change Oswestry questionnaire (mean score chi − mean score control) after 6 wk: 6 months: 1 yr, 2 yr and 3 yr: 1.69, 3.31, 2.04, 3.02, 3.18. Chiropractic treatment more effective on Oswestry scale, especially for patients with chronic and severe pain; difference at 6 wk and 1 yr not significant</td>
<td>Positive</td>
</tr>
<tr>
<td>Pope (14, 15)</td>
<td>164</td>
<td>Low back pain between 3 wk and 6 months, no sciatica</td>
<td>Short lever, high velocity, low-amplitude thrust technique (70) i. Soft tissue massage [pragmatic] (37) ii. Transcutaneous muscle stimulation [pragmatic] (28) iii. Cortisone [pragmatic] (29)</td>
<td></td>
<td>Change VAS pain after 3 wk: chi 24; i: (i) 17.2, (ii) 9.6, (iii) 15.9; None of the changes in outcomes were significantly different among any of the groups</td>
<td>Negative</td>
</tr>
<tr>
<td>Triano (13)</td>
<td>209</td>
<td>Pain in lumbar or sacroiliac region; extremity pain permitted; pain &gt;50 days or more than 6 months pain in preceding 12 months</td>
<td>High velocity, low amplitude manipulation (39)</td>
<td>i. High velocity, low force mimic manipulation [placebo] (40) ii. Back education program [pragmatic] (38)</td>
<td>Mean pain score VAS at baseline, after 2 and 4 wk: chi 38.4, 13.9, 13.3; co 35.6, 19.6, 15.1. Mean score on Oswestry questionnaire at baseline, after 2 and 4 wk: chi 17.5, 9.5, 10.6; co 20.2, 12.3, 11.4. Significant difference in Oswestry score at 2 wk in favor of chiropractic; nonstatistical advantage of chiropractic at 2 wk on VAS pain; no differences on VAS and Oswestry at 4 wk</td>
<td>Positive</td>
</tr>
<tr>
<td>Waagen (27)</td>
<td>29</td>
<td>Chronic low back pain (&gt;1 wk); intensity mild to moderate</td>
<td>Chiropractic full spine adjustments (11)</td>
<td>Sham adjustment using minimal force and paraspinal soft tissue massage [placebo] (18)</td>
<td></td>
<td>Improvement on VAS pain (10 cm) after 2 wks: chi 2.3; co 0.6. Significant greater improvement on “objective index” and VAS. Results considered preliminary because of small sample size</td>
</tr>
<tr>
<td>Postacchini (28)</td>
<td>398</td>
<td>5 different groups: IA = acute (&lt;4 wk)/no radiation; IB = chronic (&gt;8 wk)/no radiation; IC = chronic with acute episode; IA = acute with radiation; IB = chronic with radiation</td>
<td>Standardized chiropractic technique (87) i. Placebo (antiedema gel) [placebo] (73) ii. NSAID (diclofenac) [pragmatic] (81) iii. Physiotherapy (massage, infrared or short wave) [pragmatic] (78) iv. Bed rest (6–8 days) [pragmatic] (29) v. Low back school [pragmatic] (50)</td>
<td></td>
<td>Mean improvement on combined pain, disability, and spinal mobility score (range of score 5 to 32) after 3 wk, 2 and 6 months for group IA (acute, no radiation): chi 7.5, 9.7, 12.7; co: (i) 1.8, 7.3, 11.0; (ii) 5.0, 8.4, 10.2; (iii) 3.0, 10.7, 14.0; (iv) 5.4, 7.5, 7.3. Best results with chiropractic treatment at 3 wk in acute patients (IA and IA); at longer term and for other indications (chronic and chronic with acute episode; IB, IC and IIB) no superiority over other treatments</td>
<td>Positive for subgroup (acute patients only)</td>
</tr>
<tr>
<td>Herzog (16)</td>
<td>37</td>
<td>Chronic sacroiliac problems</td>
<td>Spinal manipulative therapy (16)</td>
<td>Back school including exercises and instructions [pragmatic] (13)</td>
<td></td>
<td>VAS pain 4 wk: chi 1.85 co 0.77; Oswestry questionnaire chi 13.8; co 8.1. Greater pain reduction and change on Oswestry scale for back school therapy; no difference on sacroiliac joint mobility; chiropractic more effective in correcting gait asymmetry (no adequate statistical testing of results)</td>
</tr>
<tr>
<td>Bronfort (29)</td>
<td>21</td>
<td>Low back pain of various duration with or without radiation</td>
<td>Chiropractic procedures: low amplitude/high velocity (11)</td>
<td>Medical treatment including analgesics, injections, bed rest and/or physiotherapy [pragmatic] (10)</td>
<td>No of patients (%) improved after 1, 3 and 6 months: chi 60%, 70%, 80%; co 44%, 66%, 44%. Patient’s assessment of improvement more positive in chiropractic patients; other outcomes no differences</td>
<td>No conclusions because of small sample size</td>
</tr>
<tr>
<td>Rupert (30)</td>
<td>145</td>
<td>Low back or leg pain and/or restriction in lumbar range of motion; acute (&lt;30 days) and chronic (&gt;30 days)</td>
<td>Specific short lever technique (–)</td>
<td>i. Sham manipulation consisting of touching and palpatting [placebo] ii. Drugs and bed rest [pragmatic] (–)</td>
<td>Percentage change on VAS pain at 3 wk: chi 45%; co: (i) 20%, (ii) 40%. Greater pain reduction in chiropractic patients compared with other two treatments; more immediate relief for chronic patients and patients under age 40; no reporting of results of footprint-to-floor distance and straight leg raising</td>
<td>Positive</td>
</tr>
</tbody>
</table>

a Ranked in decreasing order of number of validity items scored positive (Table 3)
b Total number of patients directly randomized
c Number of patients for each group directly after randomization, if not provided by original author(s): at main follow-up moment
d Quantitative data on main outcome measure: chi, chiropractic; co, control group
e Narrative summary according to original authors
f Numbers not provided
subdivision of subacute and chronic patients. Therefore, both categories will be summarized as “chronic” (>3 wk). Four studies also included acute patients (≤3 wk), of which 3 studies provided separate data for patients with acute complaints and those with chronic complaints (see also Table 4). The control treatments covered a great variety of realistic or (intended) placebo interventions. Three RCTs were pragmatic, one placebo-controlled and three studies included both a pragmatic and a placebo comparison (Table 2).

Methodological Assessment

Table 3 presents the results of the methodological assessment. Initially, the two reviewers agreed on 81% of the instances in which a criterion was applied. All disagreements were solved in a consensus meeting. The emphasis was placed on the validity items because insufficient attention to them could result in bias. Only three RCTs had an adequate randomization procedure (items C1 and C2). Loss to follow-up (item E) is also a highly prevalent shortcoming. Most studies failed to include blind or naive patients (L), often resulting in unblinded outcome measurements (N). All but two studies failed to include an intention-to-treat analysis, and none of them included a worst-case analysis to correct for missing data (item P). Most studies were too small to establish an adequate precision (item F). In addition, for all RCTs, an adequate reporting on many of the informativeness items was lacking.

Completeness of Report

Table 4 provides an overview of the availability of the data required for the analysis plan described in the Methods paragraph. The four lines at the top of Table 4 indicate that the studies, with a few exceptions, report sufficient details on the location of initial referral, duration of complaints, radiation (nerve root involvement) and type of comparison to enable

### Table 3. Methodological quality of randomized trials on the effectiveness of chiropractic treatment of back pain, ranked in decreasing order of number of validity items scored positive. See Table 1 and Appendix A for explanation of the criteria

<table>
<thead>
<tr>
<th>First author (reference)</th>
<th>Validity items</th>
<th>Precision</th>
<th>Informativeness items</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B C1 C2 E I L N P F A D G H J K M O Q</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Meade (12, 17)</td>
<td>+ + + + - - - - -</td>
<td>+ + + + + + + + + +</td>
<td>+ + + + + + + + + +</td>
</tr>
<tr>
<td>Pope (14, 15)</td>
<td>+ + + + - - - - -</td>
<td>+ + + + + + + + + +</td>
<td>+ + + + + + + + + +</td>
</tr>
<tr>
<td>Triano (13)</td>
<td>± ± ± + - - - - -</td>
<td>± ± ± ± ± ± ± ± ± ±</td>
<td>± ± ± ± ± ± ± ± ± ±</td>
</tr>
<tr>
<td>Waagen (27)</td>
<td>± ± ± + - - - - -</td>
<td>± ± ± ± ± ± ± ± ± ±</td>
<td>± ± ± ± ± ± ± ± ± ±</td>
</tr>
<tr>
<td>Postacchini (28)</td>
<td>± ± ± + - - - - -</td>
<td>± ± ± ± ± ± ± ± ± ±</td>
<td>± ± ± ± ± ± ± ± ± ±</td>
</tr>
<tr>
<td>Herzog (16)</td>
<td>± ± ± + - - - - -</td>
<td>± ± ± ± ± ± ± ± ± ±</td>
<td>± ± ± ± ± ± ± ± ± ±</td>
</tr>
<tr>
<td>Bronfort (29)</td>
<td>± ± ± + - - - - -</td>
<td>± ± ± ± ± ± ± ± ± ±</td>
<td>± ± ± ± ± ± ± ± ± ±</td>
</tr>
<tr>
<td>Rupert (30)</td>
<td>± ± ± + - - - - -</td>
<td>± ± ± ± ± ± ± ± ± ±</td>
<td>± ± ± ± ± ± ± ± ± ±</td>
</tr>
</tbody>
</table>

+, item completely addressed; ±, item partially addressed; −, item not addressed or not described

### Table 4. Study characteristics of randomized clinical trials on the effectiveness of chiropractic of back pain

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient characteristics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acute complaints</td>
<td>Ac + Chr</td>
<td>Chr</td>
<td>Chr</td>
<td>Chr</td>
<td>Ac + Chr</td>
<td>Chr</td>
<td>Ac + Chr</td>
<td>Ac + Chr</td>
</tr>
<tr>
<td>Nonradiating</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR + R</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Location initial referral</td>
<td>M + C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
</tr>
<tr>
<td>Intervention</td>
<td>Pr</td>
<td>Pr</td>
<td>Pr + Pl</td>
<td>Pr</td>
<td>Pr + Pl</td>
<td>Pr</td>
<td>Pr + Pl</td>
<td>Pr</td>
</tr>
<tr>
<td>Type of measurement</td>
<td>S + I + L</td>
<td>(S)'</td>
<td>S + I</td>
<td>(S)'</td>
<td>(S + I + L)'</td>
<td>I</td>
<td>(S)'</td>
<td>I</td>
</tr>
<tr>
<td>Success rate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Back pain specific functional status</td>
<td>I + L</td>
<td>(S)'</td>
<td>S + I</td>
<td>(S)'</td>
<td>(S + I + L)'</td>
<td>I</td>
<td>(S)'</td>
<td>I</td>
</tr>
<tr>
<td>Generic functional status</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Ac, acute complaints (≤3 wk); Chr, chronic complaints (>3 wk); Ac + Chr, acute and chronic reported separately; Ac/Cr, acute and chronic reported combined

R, radiating complaints—nerve root affected; NR, nonradiating complaints; NR + R, nonradiating and radiating reported separately

C, chiropractic clinic; M, medical practice/hospital; NS, not specified

Pr, pragmatic; Pl, placebo; Pr + Pl, pragmatic and placebo reported separately

S, short term follow up (<3 wk); L, intermediate term follow up (3 wk ≤3 months); L, long term follow up (≥3 months)

Items in brackets: outcome reported, but information not suitable for statistical pooling

Nonvalidated index of pain, functional status and physical examination

Only data from a sample of the trial (85 of the total of 164 patients)

No standard deviation reported

No patients per group and no standard deviation reported
subgroup analyses to be made. However, the four lines at the bottom of Table 4, dealing with the outcome measures and follow-up timing, show that the different studies contain insufficient data to enable a sensible statistical pooling of the results. A success rate was reported in one study only. Usable VAS scores for pain were only presented four times in three RCTs: two with short-term and two with intermediate-term follow-up. Back pain-specific functional status [in all cases using the Oswestry scale (33)] was reported for all patients in three trials and only some of the patients in one trial. General functional status was not reported in any of the RCTs.

Outcomes of Studies
Because there is insufficient data for the statistical pooling mentioned in the analysis plan (see Methods), we are only able to focus on the quantitative data and conclusions based on the main outcome measures on a study-to-study (“vote-counting”) basis. These data are presented in Table 2. In this table, the results of the Meade study (12) are updated according to the recently published report on the extended follow-up (17). Compared with the first publication on this study, when the long-term follow-up had not yet been completed (12), the results of the 2- and 3-yr follow-up now show remarkably fewer benefits from chiropractic treatment than previously reported.

Results for Acute Patients
For the acute patients (complaints ≤ 3 wk), all three trials [Postacchini (28), Meade (12, 17), Rupert (30)] report benefits from chiropractic. However, the trials by Rupert et al. (30) only report on a follow-up of 3 wk. Postacchini et al. (28) report positive results for patients with acute complaints at 3 wk follow-up, but not after 2 and 6 months. In the first publication by Meade et al. (12), when the follow-up was not yet complete, the figures for acute patients at 6 wk and 6 months did not reach statistical significance (the number of patients available for follow-up was 276 and 254, respectively). In the subsequent article on the extended follow-up, no detailed figures or statistical analysis of acute complaints are provided (17). In this recent article, however, the authors state that acute patients tended to derive most benefit from chiropractic.

Results for Chronic Patients
For the studies that included chronic patients (follow-up > 3 wk) Waagen et al. (27) and Rupert (30) report positive overall findings, but do not perform a separate statistical test of the results for this subgroup. Meade et al. also report favorable results for chronic patients. However, also for this subgroup, the results were not statistically significant in the first publication (12) (number of patients available at 6 wk, 390; at 6 months, 353) and were not reported in detail or tested statistically in the recent publication (17). In the study by Triano et al. (13), the advantage of chiropractic was only clear on the VAS scale at 2 wk, not on the Oswestry scale. At 4 wk (2 wk after completion of the intervention) no significant difference could be detected. In the studies by Postacchini et al. (28) and Herzog et al. (16), no clear advantage of chiropractic over control treatments could be detected for chronic patients.

DISCUSSION
This aim of this review was to update our previous review (1) in two ways: to incorporate new insights on research methodology, including statistical pooling, and to provide an update that included the newly issued reports on RCTs.

Register of Trials
In our review published 4 yr ago (1), we described five chiropractic RCTs on low back pain that were in the preparation, execution or reporting phase [Triano et al. (13), Pope et al. (14), Bronfort et al. (29), Waagen et al. (27; second trial) and Jochemsens (35)]. The same trials were also cited in an issue of the Chiropractic Report (34). Two trials (Triano et al. and Pope et al.) are included in this update. A third trial (Bronfort et al.) has recently been submitted for publication (G. Bronfort, personal communication). The two other trials [Waagen et al. (second trial) and Jochemsen] , although completed some years ago (34, 35), have yet to be published. The reasons for this can vary. However, all randomized clinical trials should be published to prevent publication bias (36). Because chiropractic research is still rapidly expanding, a register of trials (36) should be established to monitor the conduct and publication of RCTs in this field.

Methodological Quality and Internal Validity
The separation of the items of the methodological assessment into three domains indicated, perhaps even more than in our 1992 review, the poor reporting or execution of essential prerequisites for the (internal) validity of chiropractic RCTs. An adequate randomization procedure and a blinded outcome assessment, which seem to be components that are related to internal validity (23), were only found in a few studies. Loss to follow-up is another important validity item. The loss to follow-up was especially prominent in two trials conducted at chiropractic colleges. In the study by Waagen et al. (27), 10 out of 29 subjects (34%) were not reported on at the 2-wk follow-up. Triano et al. (13) had complete data at 4 wk for only 117 of the 209 subjects, indicating a 44% loss to follow-up. Reasons for noncompliance with treatment or follow-up measurements can be strongly related to treatment results; therefore, the conclusions of these studies should be interpreted with the greatest caution. The statistical analysis was inadequate in all studies, and is certainly in need of improvement (37).

Attempted Statistical Pooling
We were not able to perform the planned statistical pooling according to our preset analysis plan. Table 4 clearly shows that most outcome measures, in combination with the various follow-up moments, were simply not covered by enough RCTs to perform a sensible pooling. It is regrettable that this was partially caused by such examples of insufficient reporting as: no explicit description of group size or standard deviation, reporting on only some of the patients or unusable information.
from inseparable composite indexes of pain, disability and physical measures. However, the main obstacle was the substantial heterogeneity in outcome measures and follow-up timing. We considered pooling the VAS pain scores with the back pain-specific (Oswestry) scores to increase the number of combinable studies. However, the discrepancy between the results of the VAS scores for pain and the Oswestry scores in all three trials reporting on both outcome measures (13-16) (data not shown) clearly indicates that these outcome measures do cover different domains and should consequently be analyzed separately.

**Effectiveness of Chiropractic**

Because pooling of the study results was not possible, we decided to summarize the outcomes of the various studies in a narrative way. Acute and chronic low back pain are distinct prognostic categories and were therefore reported separately.

All three studies that included acute patients reported positive overall results. For two studies, however, this conclusion is not wholly supported by the report on these studies. Rupert did not analyze statistically the results for the subgroup of acute patients (30). In the extensive Meade study, the results were in favor of chiropractic: the subgroup of acute patients, however, despite the large number of patients, did not reach statistical significance (12). If results of these studies are regarded as convincingly positive (which is not our opinion), the timing of the maximal effect remains uncertain: Rupert (30) and Postacchini et al. (28) report short-term effects only, whereas the Meade study indicates that the (nonsignificant) effects only appear at 6 months, not at 6 wk (12).

For chronic low back pain, the effectiveness is inconsistently reported. The positive studies by Rupert (30) and Waagen et al. (27) did not statistically analyze the results for chronic patients separately. Triano et al. (13) report positive outcomes for only one of the two main outcome measures, and then only directly after the intervention period. Two weeks later, there was no longer any apparent advantage. Finally, although there were 390 patients in the subgroup of chronic patients in the Meade study, at the 6-wk and 6-month follow-up, the results did not reach a level of statistical significance (12). For this indication, the studies by Pope et al. (14, 15), Postacchini et al. (28) and Herzog et al. (16) can be regarded as negative studies.

**Study Size**

Study size relates to several important methodological features: power, publication bias and clinical relevance. Small study size (item F) was a highly prevalent flaw in the reviewed studies, and may lead to insufficient power to detect clinically relevant differences (38). Consequently, a number of negative studies in our “vote counting” analysis might be negative because the sample size was too small. On the other hand, because small positive studies have an increased likelihood of being published (39), they may have introduced publication bias, as already discussed above. Finally, if large numbers of patients are needed to make a small outcome difference reach statistical significance (like the combined results for acute and chronic patients in the Meade trial) (12, 17), then the clinical significance of that small difference should be questioned.

**Standards of Report**

There is certainly insufficient reporting on trials concerning chiropractic treatment for low back pain. Anderson et al. have already focussed on this in their 1991 review on spinal manipulation (3). There is a great need for researchers in this field to follow consistent patterns of measuring and recording outcome variables and selecting time-intervals for post-treatment evaluations. These aspects should be outlined in standards for research in this area. We recommend that such standards should also include the components outlined in our Methods paragraph. Perhaps the exact type of outcome measurement and a more exact timing of follow-up should be further determined in a consensus meeting of leading chiropractic researchers. In addition, reports on RCTs should conform to the recently issued proposal for structured reporting of RCTs by the Standards of Reporting Trials Group (40). The items in this proposal closely resemble the items of our methodological checklist.

**CONCLUSION**

In this updated review we attempted to pool the results of RCTs that evaluated the effectiveness of chiropractic treatment for low back pain. However, because of the great variety of outcome measures and follow-up timing, there were insufficient data to enable us to perform any pooling of the RCTs. A vote-counting of the available RCTs provided no convincing evidence of the effectiveness of chiropractic for acute or chronic low back pain. All RCTs had serious flaws in their design, execution and reporting. There is certainly a need for further correctly executed trials. In future, research on the effectiveness of chiropractic guidelines for uniform execution and reporting of RCTs should first be established, to enable subsequent statistical pooling in systematic reviews of chiropractic trials. In addition, a register of trials should be established to monitor the conduct and publication of randomized clinical trials in this field.

**ACKNOWLEDGMENTS**

We kindly thank Paul G. Shekelle for sharing his ideas on systematic reviews of spinal manipulation and for his useful comments on the manuscript.

**REFERENCES**


3. Anderson R, Meeker WC, Wirick BE, Mootz RD, Kirk DH,


APPENDIX A

Operationalization of the Criteria from Table I.

Each criterion must be applied independently of the others.

A Description of inclusion and exclusion criteria. Restriction to a homogeneous study population.

B Similarity for: duration of complaints, value of outcome measures, age, recurrence status and radiating complaints.

C1 Randomization procedure guarantees concealment (e.g., centralized provision of allocation by phone, on-site computer system providing allocations in a
locked, unreadable file that can be accessed only after entering the characteristics of an enrolled participant; envelopes numbered, sealed, opaque).

C2 Procedure of sequence generation described and adequate (e.g., reported using random-number table, computer random-number generator, coin-tossing).

D Information from which group and with reason for withdrawal.

E Loss to follow-up: (all randomized patients) − (number of patients at main moment of effect measurement for the main outcome measure) / (all randomized patients) × 100.

F Smallest group immediately after randomization. Rating of item: < 50 subjects in the smallest group, item not addressed; 50 ≤ < 100, item is partially addressed; ≥ 100, item is completely addressed.

G Manipulative treatment protocol developed and described. All control treatments protocol developed and described.

H Comparison with a realistic treatment modality.

I Other physical therapy modalities or medical interventions are avoided in the design of the study (except analgesics, advice on posture or use at home of heat, rest or a routine exercise scheme).

J Comparison with a placebo therapy.

K Mentioning of qualified education and/or experience of the chiropractor(s).

L Placebo-controlled: attempt at blinding, blinding evaluated and fully successful. Pragmatic study: patients fully naive, or time restriction (no manipulative treatment for at least 1 yr), naiveness evaluated and fully successful.

M Use (measured and reported) of: pain, global measure of improvement, functional status (activities of daily living), spinal mobility, medical consumption.

N Each measurement mentioned under point M assessed separately for adequate blinding.

O Two subitems: moment of measurement during or just after treatment and moment of measurement at 6 months or later. Rating of item: no subitem is an item not addressed, one subitem is item partially addressed; both subitems are item completely addressed.

P When loss to follow-up is less than 10%: all randomized patients for most important outcome measures and on the most important moments of effect measurement minus missing values, irrespective of noncompliance and co-interventions. When loss to follow-up > 10%: intention-to-treat as well as an alternative analysis that accounts for missing values.

Q For most important outcome measures and on the most important moments of effect measurement. In the case of (semi)continuous variables: presentation of the mean or median with standard error or percentiles.