REVIEWS OF THE LITERATURE

Efficacy of Spinal Manipulation for Chronic Headache: A Systematic Review

Gert Bronfort, DC, PhD,[a] Willem J.J. Assendelft, MD, PhD,[b] Roni Evans, DC,[a] Mitchell Haas, DC,[c] and Lex Bouter, PhD[d]

ABSTRACT

Background: Chronic headache is a prevalent condition with substantial socioeconomic impact. Complementary or alternative therapies are increasingly being used by patients to treat headache pain, and spinal manipulative therapy (SMT) is among the most common of these.

Objective: To assess the efficacy/effectiveness of SMT for chronic headache through a systematic review of randomized clinical trials.

Study Selection: Randomized clinical trials on chronic headache (tension, migraine and cervicogenic) were included in the review if they compared SMT with other interventions or placebo. The trials had to have at least 1 patient-rated outcome measure such as pain severity, frequency, duration, improvement, use of analgesics, disability, or quality of life. Studies were identified through a comprehensive search of MEDLINE (1966-1998) and EMBASE (1974-1998). Additionally, all available data from the Cumulative Index of Nursing and Allied Health Literature, the Chiropractic Research Archives Collection, and the Manual, Alternative, and Natural Therapies Information System were used, as well as material gathered through the citation tracking, and hand searching of non-indexed chiropractic, osteopathic, and manual medicine journals.

Data Extraction: Information about outcome measures, interventions and effect sizes was used to evaluate treatment efficacy. Levels of evidence were determined by a classification system incorporating study validity and statistical significance of study results. Two authors independently extracted data and performed methodological scoring of selected trials.
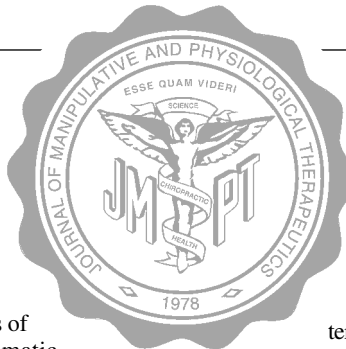
Data Synthesis: Nine trials involving 683 patients with chronic headache were included. The methodological quality (validity) scores ranged from 21 to 87 (100-point scale). The trials were too heterogeneous in terms of patient clinical characteristic, control groups, and outcome measures to warrant statistical pooling. Based on predefined criteria, there is moderate evidence that SMT has short-term efficacy similar to amitriptyline in the prophylactic treatment of chronic tension-type headache and migraine. SMT does not appear to improve outcomes when added to soft-tissue massage for episodic tension-type headache. There is moderate evidence that SMT is more efficacious than massage for cervicogenic headache. Sensitivity analyses showed that the results and the overall study conclusions remained the same even when substantial changes in the prespecified assumptions/rules regarding the evidence determination were applied.

Conclusions: SMT appears to have a better effect than massage for cervicogenic headache. It also appears that SMT has an effect comparable to commonly used first-line prophylactic prescription medications for tension-type headache and migraine headache. This conclusion rests upon a few trials of adequate methodological quality. Before any firm conclusions can be drawn, further testing should be done in rigorously designed, executed, and analyzed trials with follow-up periods of sufficient length. (J Manipulative Physiol Ther 2001;24: 457-66)

Key Indexing Terms: Headache; Orthopedic Manipulation; Chiropractic Manipulation; Osteopathy; Systematic Review

[a]Department of Research, Wolfe-Harris Center for Clinical Studies, Northwestern Health Sciences University, Bloomington, Minn.

Department of General Practice, Academic Medisch Centrum, The Netherlands.

[c]Center for Outcomes Studies, Western States Chiropractic College, Portland, Ore.

[d]Institute for Research in Extramural Medicine and Department of Epidemiology and Biostatistics, Vrije Universiteit, The Netherlands.

INTRODUCTION

Headaches vary widely in terms of severity, frequency, and disability—from rare episodes of minor discomfort to daily, incapacitating headaches. By far the most common type of headache is tension-type headache, with a 1-year prevalence ranging from 40 to 80%.[1,2] The cost, in terms of work loss and decreased quality of life, is high among those who experience tension-type headache, with 10% reporting lost workdays and almost half reporting decreased effectiveness at work, home, or school.[2]

Although less common than tension-type headaches, migraines affect a sizable proportion of individuals and tend to be more severe and disabling.[1,3] Recent population-based studies which use International Headache Society (IHS) criteria estimate that 10 to 12% of adults have experienced a migraine in the previous year,[1,3] with 2 to 4 times more women affected than men. In one study, almost all mi-

458 Journal of Manipulative and Physiological Therapeutics
Volume 24 • Number 7 • September 2001
Headache and Spinal Manipulation • *Bronfort et al*

graineurs reported that headache pain affected their ability to work and perform daily activities, and 43% missed work because of their headaches.[1]

The financial cost of headaches is great, with billions of dollars spent annually for lost productivity and treatment.[4] Although persons affected with headaches are commonly treated by traditional medical practitioners, they are increasingly turning to non-medical or alternative therapies for relief. A recent study by Eisenberg et al[5] reported that one of the most common alternative practitioners sought out for the treatment of headaches was the chiropractor—the most common provider of spinal manipulation in the United States. For the purpose of this study, spinal manipulative therapy (SMT) is defined as the application of high-velocity, low-amplitude manual thrusts to the spinal joints slightly beyond the passive range of joint motion.[6]

A systematic review of cervical SMT for neck pain and headache published in 1996 concluded that SMT might be beneficial for tension-type headaches, but further studies of higher methodological quality were needed in order to reach firmer conclusions.[7] Further studies have been performed since that time. The purpose of this article is to assess the clinical efficacy of cervical SMT for chronic headache based on the results of the existing randomized clinical trials (RCTs).

METHODS

Study Selection

Randomized clinical trials on chronic headache were included in this review if they compared SMT with a placebo or other interventions, and if they had at least 10 subjects in the SMT arm of the trial. The trials also had to have at least one patient-rated outcome measure such as headache pain severity, frequency, duration, improvement, analgesic use, disability, or quality of life. "Chronic headache" included, but was not limited to, tension-type, cervicogenic, and migraine headaches classified according to the International Headache Society (IHS) criteria (some studies were anticipated to predate or not adhere to the 1988 IHS classification system). Studies were identified by a comprehensive search of MEDLINE (1966-11/1998) and EMBASE (1974-11/1998). The primary MeSH headings and keywords were: headache, manipulation/orthopedic, randomized controlled trials, comparative study, review literature, chiropractic, and osteopathy medicine. Studies were further identified through the Cumulative Index of Nursing and Allied Health Literature, the Chiropractic Research Archives Collection, and the Manual, Alternative, and Natural Therapies Information System (MANTIS), and by citation tracking and hand searching of the non-indexed chiropractic, osteopathic, and manual medicine journals. References found in relevant publications were also examined. Abstracts from proceedings and unpublished studies were not included.

Data Extraction

A best-evidence synthesis method incorporating explicit information about outcome measures, interventions, and effect sizes was used to evaluate treatment efficacy. Two of us (GB and WJJA) independently extracted and recorded relevant data from each article. All original data on outcomes were standardized into percentage-point scores whenever possible. Contrary to meta-analysis, studies for which effect sizes cannot be computed are retained as primary evidence in a best-evidence synthesis.[8-11] Statistical pooling of effect sizes was considered to be an adjunct to the systematic review and not the primary goal.

Effect Size Computations

The effect size (ES) differences between the SMT and comparison groups were calculated at the end of the treatment intervention phase and at the main post-treatment follow-up and were adjusted for baseline differences in main outcome measures. Effect sizes were computed as described by Glass et al[12] and Cohen[13] (difference in treatment and control group means divided by the pooled standard deviation). In the absence of these statistics, effect sizes were calculated from T-scores, F-values, and confidence intervals, provided sample sizes were given.[12,14] Effect sizes for differences in proportions were estimated by using probit transformation.[14] Correction for ES estimate bias associated with small sample sizes (n < 50) was accomplished by using the method described by Hedges and Olkin.[15] If confidence intervals could not be directly calculated for effect sizes, they were estimated by using p-values and sample sizes.

Assessment of Methodological Quality of RCTs

A critical evaluation list of 20 methodological items and their operational definitions was used to assess methodological quality. This list represents a modification of previously used instruments.[16,17] Fourteen of the items addressed validity issues, yielding a validity score. An additional 6 items concerned descriptive information. For example, we awarded points for the following: a statistical analysis that included a power calculation based on a predetermined clinically important difference between treatment and control groups; an adjustment made for baseline differences between treatment and control groups (eg, analysis of covariance); an adjustment of significance levels to account for multiple comparisons, primary outcomes, and follow-up time; and an appropriate analysis of dropouts, compliance, and missing data. The study conclusions also had to be supported by the design and data analyses. A detailed list of the individual items and operational definitions is described in the Appendix. The methodological scoring of the RCTs was performed by 2 reviewers independently (WJJA and GB). Differences in scores were resolved through consensus by the 2 reviewers. The validity scores of the individual RCTs were used as part of the evidence determination. Two of us (GB and RE) performed a supplementary methodological scoring of the studies by using the short checklist developed by Jadad et al.[18]

Assessment of the Level of Evidence of Efficacy

The criteria for determining the level of evidence of efficacy has been adapted from the Agency for Health Care

Journal of Manipulative and Physiological Therapeutics   459
Volume 24 • Number 7 • September 2001
Headache and Spinal Manipulation • *Bronfort et al*

Table 1. *Definition of levels of evidence modified from the U.S. and British low back pain clinical guidelines[19,71]*

Level of evidence of efficacy or inefficacy	Number of RCTS with validity score ≥50	Number of RCTs with validity score =21-49	Number of RCTs with validity score ≤20	Statistically significant results
A. Strong	≥2	—	—	Yes
B. Moderate	1	—	—	Yes
C. Limited				
either I	—	≥1	—	Yes
or II	—	—	≥2	No
D. Inconclusive	Minimal standards for classification as limited evidence were not met **or** the evidence from eligible RCTs was conflicting			

Policy and Research panel that evaluated the efficacy of various treatments for acute low back pain.[19] Our system evaluates the evidence taking into account: (1) the type of comparison intervention (established efficacious treatment, commonly used therapy, or placebo); (2) methodological quality (validity scores); (3) the number of studies; and (4) statistical significance of study findings. Four categories were used to describe evidence levels: strong, moderate, limited, and inconclusive. All eligible RCTs were considered regardless of their results. Statistical pooling of 2 or more trials was considered if they were homogeneous in terms of headache type, subjects, treatments, outcomes, and follow-up time.

For determination of the outcome of each RCT, we prioritized patient-rated pain severity, frequency, and duration, unless otherwise specified.

The assessment of efficacy depended on the type of comparison intervention. If the study showed that SMT had at least a similar magnitude of effect compared with an established efficacious treatment or was superior to a placebo or a commonly used therapy, it was considered to be evidence of efficacy. If the study showed that SMT was inferior to an established efficacious treatment, commonly used therapy or placebo, or showed an effect similar to a placebo intervention, it was considered evidence of inefficacy. We prespecified that SMT was considered superior/inferior to a comparison therapy or placebo if the ES was equal to ± 0.5.

Methodological quality and statistical significance were then considered to determine the evidence level, as defined in Table 1.

RESULTS

Our literature search identified 22 original studies that assessed the effect of SMT in the treatment of headache. We excluded 13 papers because one was a case study[20] and 12 were prospective or retrospective clinical series without comparison groups.[21-32] The reports of 9 RCTs involving 683 patients were retained in our review. The main features of these trials are summarized in Table 2.

A total of 386 patients received spinal manipulation. Age inclusion criteria ranged from 15 to 70 years of age. The number of SMT treatments ranged from 1 to 12 (average of 6) over a period of 1 day to 8 weeks (average of 4 weeks). In 3 studies, SMT was combined with other therapies (massage,[33] azapropazone,[34] and deep friction massage with

placebo).[35] In 5 of the studies, SMT was performed by chiropractors[35-39]; in 3 studies, by medical doctors[34,40,41]; in 1 study, by medical doctors or physical therapists;[38] and in another study, by osteopaths.[33]

Comparison groups included amitriptyline,[36,39] deep friction with placebo,[35,37] mobilization,[38,41] palpation and rest,[33] cold packs,[40] azapropazone,[34] and waiting list.[41] None of the studies compared spinal manipulation directly with a sham or placebo spinal manipulation procedure. Outcome measures varied greatly across studies. The main outcomes abstracted from the 9 trials were pain intensity and frequency of headaches, medication use, and general health status.

The methodological quality (validity) scores of the trials ranged from 21 to 87 on a 100-point scale. Detailed results of the methodological quality assessment of the trials are noted in Table 3. The 2 methodological quality assessors initially agreed on 74% of the 20 quality items for the 9 RCTs; all disagreement was resolved through joint review by the assessors.

The ES differences including 95% confidence intervals for the 9 trials are depicted in Figure 1.

Description of the Individual Trials

Tension-Type Headache. The trial by Boline et al[36] included a mix of patients with chronic headache and patients with episodic tension-type headache. The primary research goal was to assess the sustained treatment effect (4 weeks after treatment) of 6 weeks of SMT when compared with 6 weeks of amitriptyline (an efficacious prescription medication). At the 4-week post-treatment follow-up, the results showed an advantage for SMT in headache pain, use of non-prescription analgesics, and general health status ($P < .05$). Conversely, at the end of the 6-week treatment period the amitriptyline group fared slightly better than the SMT group in terms of headache pain ($P = .05$). Amitriptyline patients reported more side effects than those receiving SMT.

The trial by Bove and Nilsson[35] assessed whether the addition of SMT to soft tissue massage would improve outcomes for episodic tension-type headache. There were 2 treatment groups: deep friction massage with SMT and deep friction massage with placebo laser treatment. All participants received 8 treatments over a 4-week period. Outcomes were assessed at the end of treatment and 3 months after treatment. Both groups improved at similar rates during the

460 Journal of Manipulative and Physiological Therapeutics
Volume 24 • Number 7 • September 2001
Headache and Spinal Manipulation • *Bronfort et al*

Table 2. *Randomized controlled trials of spinal manipulation for headaches (arranged in order of methodological quality [validity score])*

First author, reference, year, validity score	Headache type	Study groups (n)	Results of abstracted data
Nelson[39] 1998 (VS = 87)	Chronic migraine	G1: SMT-DC (77) G2: SMT-DC+ Amitriptyline (70) G3: Amitriptyline (71)	Small group difference in headache pain index (pain × frequency) and medication use at the end of 2 months of treatment (NS) G1 had greater reduction in headache pain index at 1-month follow-up (NS)
Boline[36] 1995 (VS = 75)	Chronic and episodic tension-type	G1: SMT-DC (75) G2: Amitriptyline (75)	G2 had greater pain reduction after 6 weeks of tx (SS) G1 had greater reduction in pain, frequency, and medication use 4 weeks after tx (SS)
Nilsson[37] 1997 (VS = 67)	Chronic cervicogenic	G1: SMT-DC (28) G2: Deep friction massage + placebo laser (25)	G1 had greater reduction in headache intensity and hours of headache per day at the end of 3 weeks of treatment (SS)
Parker[38,42] 1978 and 1980 (Re-analyzed by New Zealand Gov't Commission) (VS = 67)	Chronic migraine	G1: SMT-DC (30) G2: SMT-PT/MD (27) G3: Cervical mobilization-PT/MD (28)	G1 had greater pain reduction than G3 after 2 months of tx (NS)
Bove[35] 1998 (VS = 56)	Episodic tension-type	G1: SMT-DC + friction massage (38) G2: friction massage + placebo laser (37)	Relatively small group difference in headache pain severity, duration, and medication use at the end of 2 months of treatment and at 1- and 3-month follow-ups (NS)
Hoyt[33] 1979 (VS = 45)	Chronic muscle-tension	G1: Massage + SMT-DO (10) G2: Palpation (6) G3: Rest 10 min (6)	Immediately after one tx G1 had much greater pain reduction than G2 and G3 (SS)
Jensen[40] 1990 (VS = 34)	Post-traumatic	G1: SMT-MD (11) G2: Cold Packs (12)	G1 had much greater pain reduction after 3 weeks of tx (SS)
Howe[34] 1983 (VS = 32)	Neck pain related	G1: NSAID (Azapropazone) + SMT-MD (14) G2: NSAID (13)	After one and only tx G1 had higher % of patients showing improvement immediately (NS) No group difference 3 weeks after tx (NS)
Bitterli[41] 1977 (VS = 21)	Cervicogenic-like	G1: SMT-MD (10) G2: Mobilization-MD (10) G3: 3 week waiting list, then SMT-MD (10)	G1 had greater pain reduction than G2, and much higher than G3 after 3 weeks of tx (NS)

VS, Validity score; *MD*, medical doctor; *DO*, osteopathic doctor; *PT*, physiotherapist; *DC*, chiropractor; *G1*, group 1; *G2*, group 2; *G3,* group 3; *tx,* treatment; *SMT*, spinal manipulative therapy; *MOB*, spinal mobilization; *NSAID, nonsteroidal anti-inflammatory drug; SS*, statistically significant ($P \leq$ .05); *NS*, not statistically significant.

Table 3. *Methodological quality scores of randomized clinical trials evaluating spinal manipulation for chronic headache.*

First author, reference, year	Validity items*														Validity % score	Study information items*					
	B	C	D	E	F	G	J	L	M	N	P	Q	R	S		A	H	I	K	O	T
Nelson (39) 1998	+	+	+	na	na	p	–	+	+	+	+	+	+	+	87	+	p	+	+	+	+
Boline (36) 1995	+	p	+	na	na	p	–	+	+	+	+	+	+	+	75	+	p	+	+	+	+
Nilsson (37) 1997	p	p	+	na	na	p	+	+	+	–	+	p	+	–	67	+	–	p	p	p	p
Parker (38,42) 1978+80	p	–	+	na	na	p	+	+	+	–	p	p	+	+	67	+	+	p	–	–	p
Bove (35) 1998	+	p	+	na	na	p	p	+	p	p	p	p	p	–	56	+	+	p	p	+	p
Hoyt (33) 1979	p	–	p	p	na	p	p	+	p	–	na	na	na	na	45	–	–	p	–	–	+
Jensen (40) 1990	p	p	p	na	na	p	–	–	p	–	p	–	+	–	34	p	p	p	–	+	–
Howe (34) 1983	p	p	p	na	na	p	–	–	p	–	–	–	+	na	32	p	–	p	–	+	–
Bitterli (41) 1977	–	–	+	–	na	p	–	p	–	–	p	–	–	na	21	p	+	p	–	+	p

+, Yes; –, no; *p*, unclear/partly; *na*, not applicable.

*The critical evaluation list contains 20 items (A-T) of which 14 (B-G, J, L-N, P-S) have been classified as (internal) validity items and six (A, H, I, K, O and T) as information items. The Appendix contains a description of each item as worded in the list (italicized), accompanied by operationalization where needed.
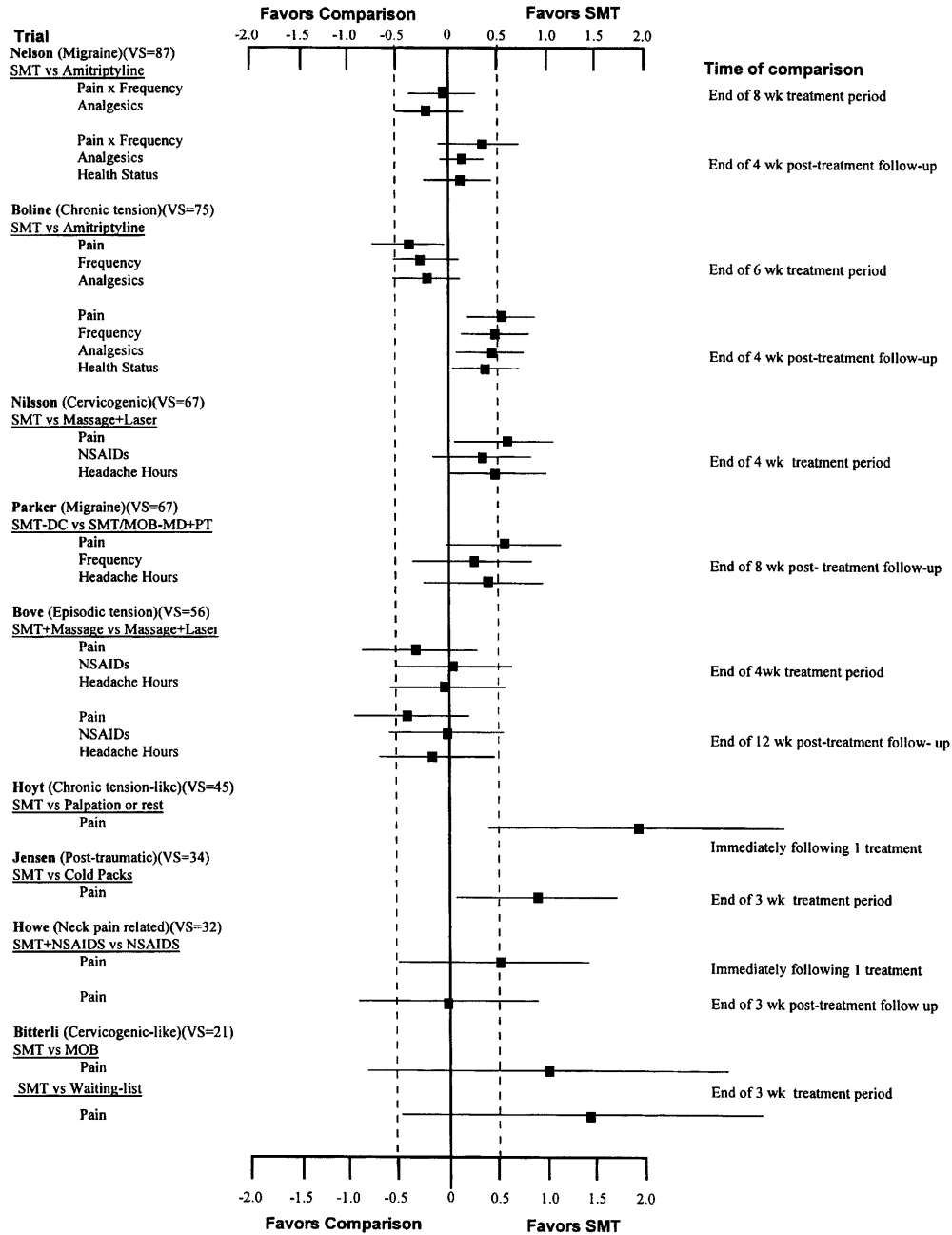
Journal of Manipulative and Physiological Therapeutics   461
Volume 24 • Number 7 • September 2001
Headache and Spinal Manipulation • *Bronfort et al*

Fig 1. *Effect size (ES) differences for randomized trials on the role of spinal manipulation (SMT) for chronic headache.* Line with box, *ES difference with 95% confidence interval; 0.2, small ES difference; 0.5, medium ES difference; 0.8, large ES difference; VS, validity score; DC, chiropractor; MD, medical doctor; PT, physical therapist; MOB, spinal mobilization.*

treatment and follow-up period. There were no important differences between the groups in either daily headache hours, pain intensity per episode, or daily analgesic use. Headache pain intensity per episode showed no important change during the trial in either of the groups.

Migraine. In the RCT by Parker et al,[38,42] chiropractic SMT showed an advantage in pain intensity, disability score, duration, and frequency of attacks after 8 weeks of treatment when compared with SMT and mobilization delivered by medical physicians and physical therapists. There was

also a slight advantage in headache frequency in patients receiving chiropractic SMT at 20 months' follow-up. However, there was uncertainty regarding the appropriateness of the original statistical analyses and subsequent reanalyses.[42]

In a recently published trial on chronic migraine by Nelson et al,[39] patients were randomized into 3 groups: SMT, amitriptyline, or a combination of the 2 therapies. Patients were treated for 8 weeks and were evaluated at the end of treatment and 4 weeks after treatment. In the follow-

462   Journal of Manipulative and Physiological Therapeutics
Volume 24 • Number 7 • September 2001
Headache and Spinal Manipulation • *Bronfort et al*

up period, there was a reduction in headache index score for the SMT group compared with the other 2 groups, which was of borderline statistical significance. There was no advantage to combining amitriptyline and SMT. During the treatment phase, the SMT group experienced a clinical effect of similar magnitude to the amitriptyline group, but reported fewer side effects.

Cervicogenic Headache. The trial by Nilsson et al[37] on cervicogenic headache compared 3 weeks of SMT with 3 weeks of deep friction massage with placebo laser therapy with no follow-up period. The study was performed in 2 stages. The first stage, reported in 1995,[43] was unable to detect a clinically important difference between the 2 treatment arms. The trial was continued after a pause in recruitment; an additional 14 patients added to the original sample, resulting in a total sample of 53 patients. The results of the extended trial showed a decrease of 69% in headache hours in the SMT group compared with a 47% decrease in the massage group ($P < .05$). Patients receiving SMT also reported approximately twice as much reduction in headache intensity per episode than the massage group ($P < .05$).

Other Types of Chronic Headache. The remainder of the trials either predated the IHS classification system or did not adequately adhere to it. They were also of lower methodological quality. Jensen et al[40] compared 2 sessions of SMT with 2 sessions of cold packs for patients with post-traumatic chronic headaches. The cold pack group did not improve over the 8-week study duration, and the advantage of SMT was evident at both 3 ($P = .03$) and 8 weeks ($P > .05$). The trial by Howe et al[34] on headache-related neck pain showed that the addition of 1 SMT session to nonsteroidal anti-inflammatory drug therapy compared with nonsteroidal anti-inflammatory drug therapy alone was superior immediately following treatment, but this difference was lost at 3 weeks after treatment ($P > .05$). Bitterly et al[41] showed a clinically important advantage for SMT compared with mobilization and waiting list control ($P > .05$) after 3 weeks of treatment for cervicogenic-like headache. The trial by Hoyt et al[33] showed a large advantage for SMT over the 2 controls (palpation and rest) for muscle-tension headache ($P = .0001$). However, the investigators tested the effect of only 1 SMT session immediately following treatment; no follow-up was included.

Evidence of Efficacy. Statistical pooling was not justified due to study differences in chronic headache type, main outcome measures, baseline clinical characteristics, assessment time points, and number and type of SMT interventions. Assessing the preponderance of the evidence based on our predefined criteria, and factoring in the magnitude and direction of effect size differences, we conclude that there is moderate evidence that SMT has short-term efficacy comparable with amitriptyline in the prophylactic treatment of chronic tension-type headache and migraine. SMT does not appear to improve outcomes when added to soft-tissue massage for episodic tension-type headache. There is moderate evidence that SMT is more efficacious than massage for cervicogenic headache.

Several sensitivity analyses were performed to test the robustness of the assumptions behind the weighting of the evidence by using headache pain as the main outcome measure. An increase of 33% or any amount of decrease of the prespecified cut-off point for adequate methodological quality (validity) did not change the weight of the evidence or the overall conclusions. Neither a decrease of the prespecified cut-point for what constituted a minimal clinically important effect size difference from 0.5 to 0.1 nor an increase to 0.6 changed the weight of the evidence or the overall conclusions. The effect size cut-point had to be set to 0.7 or above (corresponding to a near large effect size difference) to invalidate the conclusions regarding the effect of SMT for cervicogenic headache.

Side Effects. In the studies comparing SMT with amitriptyline,[36,39] more than half the patients taking amitriptyline reported side effects such as drowsiness, dry mouth, and weight gain, and approximately 10% were withdrawn from the studies due to drug intolerance. In comparison, only 5% of the patients receiving SMT reported side effects, the most frequent being muscle soreness and neck stiffness. These effects are common and considered normal reactions to spinal manipulation.[44] No serious complications (ie, vertebrobasilar accidents) were reported in any of the studies included in this review. The risk of serious complications from SMT is considered low. Estimates of vertebrobasilar accidents range from 1 per 20,000 patients to 1 per 1 million cervical manipulations;[45] however, large prospective studies are needed to provide more reliable estimates.

DISCUSSION

A previous systematic review assessing the effect of SMT on chronic headaches has suggested that SMT may be a worthwhile therapy for tension-type headache[7]. The findings of our review, which includes 3 additional relatively high-quality RCTs, provide a basis for considering SMT in the therapeutic management of migraine, chronic tension-type and cervicogenic headaches. Although migraine, cervicogenic headache and tension-type headache generally are considered to be separate conditions, there is some support in the literature for the notion that they represent a continuum with several common underlying mechanisms, including cervical spine dysfunction.[46,47] One possible explanation of the apparent effect of SMT in chronic headache comes from the results of several studies that have demonstrated that headache can be induced experimentally by noxiously stimulating tissues, including joint capsules, ligaments, and paraspinal muscles, enervated by the cervical nerve roots (C1-C3).[48] Headache pain caused by such stimulation may be possible because of the common neurological pathways shared by the trigeminal nucleus and the C1-C3 nerves.[48]

Methodological Limitations

Different methodologies have been advocated for the systematic review of studies addressing therapeutic efficacy.[15,18,49-52] Given the nature of RCTs available for this review, we chose to evaluate the strength of the evidence

Journal of Manipulative and Physiological Therapeutics   463
Volume 24 • Number 7 • September 2001
Headache and Spinal Manipulation • *Bronfort et al*

based on the best-evidence synthesis method rather than a formal meta-analysis.[9,53] A number of meta-analytical methods have been advocated for combining results of RCTs.[15,54] It is recognized by international experts that one of the most important limitations of published meta-analyses is inadequate control for clinical heterogeneity among synthesized studies.[8,55,56] There is currently little consensus on decision rules regarding statistical pooling of study results.[57] The clinical heterogeneity of the trials, in terms of headache type, patient characteristics, interventions, comparison therapies, and outcome measure prevented statistical pooling in this review.

A possible limitation of the current review is publication bias, of which there are several potential sources.[58] No effort was made to identify unpublished research,[59] which is more likely to have negative outcomes.[60] However, it is recognized that attempts to retrieve unpublished trial data may also bias studies.[60] The search strategy may have missed important studies not currently indexed, but by including citation tracking of non-indexed journals it is unlikely that many were overlooked. Optimally, reviews should include all trials regardless of language.[61-63] However, this review was initially restricted to the languages we spoke: English, German, French, Dutch, and the Scandinavian languages. Although an attempt was made to identify trials in other languages, this approach was not fully systematic; the possibility that some relevant trials may have been overlooked must be acknowledged.

The evidence for efficacy or inefficacy rests primarily on the results of a small number of RCTs of acceptable methodological quality. A few additional high-quality RCTs in the future could easily change the conclusions of our review.[62,64] Little research has been done to determine what constitutes a minimal clinically-important difference in headache outcomes. The chosen cut-point of a medium effect-size (0.5) difference to determine inferiority/superiority of an intervention is somewhat arbitrary but similar to other reported estimates.[65,66] Also, sensitivity analyses showed that the results and the overall study conclusions remained the same even when substantial changes in the pre-specified assumptions/rules regarding the evidence determination were applied.

The reliability with which different reviewers use similar methodological scoring systems is a source of uncertainty.[67] Conclusions regarding the weight of evidence are largely dependent on the exact definition of the evidence classification system used.[64] An additional methodological assessment of the studies included in this review was performed by using a 5-point scoring system developed by Jadad et al.[18] This scale addresses 3 areas—randomization, double blinding, and description of dropouts—which, if not addressed adequately, may be important sources of bias. Studies that scored highly with our system also scored relatively high with the Jadad scale (correlation coefficient of .62). It is important to note that none of the studies could achieve higher than a 3-point score with the Jadad scale because none of them were double-blinded.

Another possible limitation of this review is that we who performed the methodological scoring were not blinded to the authors and results of the individual RCTs because of our familiarity with the SMT literature. Some maintain that blinding yields significantly lower methodological scores,[18] whereas others contend that it does not make a difference.[68] Berlin et al[69] have demonstrated that the overall results of meta-analyses are uninfluenced by blinding.

Limitations of the Individual Trials

Most of the headache trials, including those of acceptable quality, have substantial methodological limitations. In the trials by Boline et al[36] and Nelson et al,[39] withdrawal of amitriptyline at the end of treatment is inconsistent with normal clinical practice. The return of these patients to near baseline values could be largely due to a medication rebound effect, making the apparent advantage of the SMT group less impressive. Longer periods of observation after treatment are necessary to adequately judge the value of SMT as a potential first line of therapy for tension-type headache.

In the trial by Nelson et al,[39] it appears that SMT has a magnitude of effect similar to the commonly used prophylactic medication amitriptyline. However, the trial was not designed to assess equivalence and did not have sufficient power to do so. Thus, whether the 2 therapies are equivalent is still unknown. Another concern regarding this study is the substantial loss of patients to follow up (28%). Although the study investigators performed missing data analyses, these can never fully compensate for the loss of data.

The authors of the trials by Bove and Nilsson[35] conclude that, as an isolated intervention, SMT does not have a positive effect on episodic tension-type headache. However, by its design the Bove and Nilsson trial did not assess the isolated effect of SMT; rather it looked at the combined effect of SMT with soft tissue massage. Whether there is an interaction that results from combining SMT with soft tissue massage is unknown. A more appropriate conclusion would have been that SMT, when combined with soft tissue massage, is no better than soft tissue therapy alone for episodic tension-type headache. This conclusion neither supports nor refutes the efficacy of SMT as a separate therapy.

In the trial by Parker et al,[38,42] there is no description of the dropouts, increasing the likelihood of bias. The extended trial by Nilsson et al[37] on cervicogenic headache is somewhat unorthodox in that the decision to recruit more patients was made after the original analyses of the data. No pre-specifications were made regarding separate analyses of the data, and one must be concerned about the possibility of a Type I error.

The results of the remainder of the trials, which were of lower methodological quality, all tend to suggest that SMT was better than the comparison therapies. This is consistent with studies in other fields that have shown that those of lower methodological quality tend to have positive outcomes.[52,64,70] Thus, one must interpret the results of these trials with caution.

464 Journal of Manipulative and Physiological Therapeutics
Volume 24 • Number 7 • September 2001
Headache and Spinal Manipulation • *Bronfort et al*

None of the studies reviewed evaluated the cost-effectiveness of SMT for chronic headaches. Trials are needed to establish SMT's relative cost-effectiveness to other commonly used therapies, and are particularly needed to address the potential for long-term effects. Finally, caution should be exercised when extrapolating from studies of SMT, because there is substantial diversity in terms of training and technique among providers.

CONCLUSION

SMT appears to have a better effect than massage for cervicogenic headache. It also appears that SMT has an effect comparable with commonly used first-line prophylactic prescription medications for tension-type headache and migraine headache. This conclusion rests on a few trials of adequate methodological quality. Before any firm conclusions can be drawn, further testing should be done in rigorously designed, executed, and analyzed trials with follow-up periods of sufficient length.

REFERENCES

1. Rasmussen BK. Epidemiology of headache. Cephalalgia 1995; 15:45-68.
2. Schwartz BS, Stewart WF, Simon D, Lipton RB. Epidemiology of tension-type headache. JAMA 1998;279:381-3.
3. Stewart WF, Lipton RB, Celentano DD, Reed ML. Prevalence of migraine headache in the United States. Relation to age, income, race, and other sociodemographic factors. JAMA 1992;267:64-9.
4. Business and Health Special Report: Controlling headache costs. Montvale (NJ): Medical Econo Pub; 1992.
5. Eisenberg DM, Davis RB, Ettner SL, Appel S, Wilkey S, Van Rompay M. Trends in alternative medicine use in the United States, 1990-1997: results of a follow-up national survey. JAMA 1998;280:1569-75.
6. Haldeman S, Phillips RB. Spinal manipulative therapy in the management of low back pain. In: Frymoyer JW, Ducker TB, Hadler NM, Kostuik JP, Weinstein JN, Whitecloud TS, editors. The adult spine: principles and practice. New York: Raven Press; 1991:1581-605.
7. Hurwitz EL, Aker PD, Adams AH, Meeker WC, Shekelle PG. Manipulation and mobilization of the cervical spine. A systematic review of literature. Spine 1996;21:1746-60.
8. Slavin RE. Best-evidence synthesis: an intelligent alternative to meta-analysis. J Clin Epidemiol 1995;48:9-18.
9. Slavin RE. Best-evidence synthesis: an alternative to meta-analytic and traditional reviews. Educ Res 1986;15:5-11.
10. Spitzer WO, Lawrence V, Dales R, Hill G, Archer MC, Clark P, et al. Links between passive smoking and disease: a best-evidence synthesis. A report of the Working Group on Passive Smoking. Clin Invest Med 1990;13:17-46.
11. Report of the Quebec Task Force. Scientific approach to the assessment and management of activity-related spinal disorders. A monograph for clinicians. Spine 1987;12:S1-59.
12. Glass GV, McGaw B, Smith ML. Meta-analysis in social research. Beverly Hills (CA): Sage Publications; 1981. p. 10-55.
13. Cohen J. Statistical power analysis for the behavioral sciences. Hillsdale (NJ): Lawrence Erlbaum Associates; 1988. p. 8-14.
14. Friedman H. Magnitude of experimental effect and a table for its estimation. Psychol Bull 1968;70:245-51.
15. Hedges LV, Olkin I. Statistical methods for meta-analysis. Orlando (FL): Academic Press: 1985. p. 2-46,286-306.
16. Koes BW, Assendelft WJ, van der Heijden GJ, Bouter LM, Knipschild PG. Spinal manipulation and mobilization for back and neck pain: a blinded review. BMJ 1991;303:1298-303.
17. Assendelft WJ, Koes BW, Geert JMG, van der Heijden MS, Bouter LM. The effectiveness of chiropractic for treatment of low back pain: an update and attempt at statistical pooling. J Manipulative Physiol Ther 1996;19:499-507.
18. Jadad AR, Moore RA, Carroll D. Assessing the quality of reports of randomized clinical trials: is blinding necessary? Controlled Clin Trials 1996;17:1-12.
19. Bigos S, Bowyer O, Braen G, Brown K, Deyo R, Haldeman S. Acute low back problems in adults. Clinical Practice Guideline Number 14. Rockville: U.S. Department of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research; 1994.
20. Stude DE, Sweere JJ. A holistic approach to severe headache symptoms in a patient unresponsive to regional manual therapy. J Manipulative Physiol Ther 1996;19:202-7.
21. Lewit K. Ligament pain and anteflexion headache. Eur Neurol 1971;5:365-78.
22. Vernon H. Chiropractic manipulative therapy in the treatment of headaches: a retrospective and prospective study. J Manipulative Physiol Ther 1982;5:109-12.
23. Turk Z, Ratkolb O. Mobilization of the cervical spine in chronic headaches. Manual Med 1987;3:15-7.
24. Wight JS. Migraine: a statistical analysis of chiropractic treatment. ACA J Chiro 1978;15:28-32.
25. Droz JM, Crot F. Occipital headaches. Ann Swiss Chiro Assoc 1985;8:127-35.
26. Mennell JM. The validation of the diagnosis "joint dysfunction" in the synovial joints of the cervical spine. J Manipulative Physiol Ther 1990;13:7-12.
27. Jirout J. Comments regarding the diagnosis and treatment of dysfunctions in the C2-C3 segment. Manual Med 1985;2:16-7.
28. Schoensee SK, Jensen G, Nicholson G, Gossman M, Katholi C. The effect of mobilization on cervical headaches. J Orthop Sports Phys Ther 1995;21:184-96.
29. Mootz RD, Dhami MSI, Hess JA, Cook RD, Schorr DB. Chiropractic treatment of chronic episodic tension-type headache in male subjects: a case series analysis. J Can Chiro Assoc 1994;38:152-9.
30. Rose P. Statistical analysis of chiropractic patients. Bull Eur Chiro J 1978;12:21-2.
31. Whittingham W, Ellis WB, Molyneux TP. The effect of manipulation (toggle recoil technique) for headaches with upper cervical joint dysfunction: a pilot study. J Manipulative Physiol Ther 1994;17:369-75.
32. Stodolny J, Chmielewski H. Manual therapy in the treatment of patients with cervical migraine. J Manual Med 1989;4:49-51.
33. Hoyt WH, Shaffer F, Bard DA, Benesler JS, Blankenhorn GD, Gray JH, et al. Osteopathic manipulation in the treatment of muscle-contraction headache. J Am Osteopath Assoc 1979; 78:322-5.
34. Howe DH, Newcombe RG, Wade MT. Manipulation of the cervical spine—a pilot study. J R Coll Gen Pract 1983;33:574-9.
35. Bove G, Nilsson N. Spinal manipulation in the treatment of episodic tension-type headache: a randomized controlled trial. JAMA 1998;280:1576-9.
36. Boline PD, Kassak K, Bronfort G, Nelson C, Anderson AV. Spinal manipulation vs. amitriptyline for the treatment of chronic tension-type headaches: a randomized clinical trial. J Manipulative Physiol Ther 1995;18:148-54.
37. Nilsson N, Christensen HW, Hartvigsen J. The effect of spinal manipulation in the treatment of cervicogenic headache. J Manipulative Physiol Ther 1997;20:326-30.
38. Parker GB, Tupling H, Pryor DS. A controlled trial of cervical manipulation of migraine. Aust N Z J Med 1978;8:589-93.
39. Nelson CF, Bronfort G, Evans R, Boline P, Goldsmith C, Anderson AV. The efficacy of spinal manipulation, amitriptyline and the combination of both therapies for the prophylaxis of migraine headaches. J Manipulative Physiol Ther 1998;21:511-9.

Journal of Manipulative and Physiological Therapeutics 465
Volume 24 • Number 7 • September 2001
Headache and Spinal Manipulation • *Bronfort et al*

40. Jensen OK, Nielsen FF, Vosmar L. An open study comparing manual therapy with the use of cold packs in the treatment of post-traumatic headache. Cephalalgia 1990;10:241-50.

41. Bitterli J, Graf R, Robert F, Adler R, Mumenthaler M. Zur Objektivierung der manualtherapeutischen Beeinflußßarkeit des spondylogenen Kopfschmerzes. Nervenarzt 1977;48:259-62.

42. Parker GB, Pryor DS, Tupling H. Why does migraine improve during a clinical trial? Further results from a trial of cervical manipulation for migraine. Aust N Z J Med 1980;10:192-8.

43. Nilsson N. A randomized controlled trial of the effect of spinal manipulation in the treatment of cervicogenic headache. J Manipulative Physiol Ther 1995;18:435-40.

44. Leboeuf-Yde C, Hennius B, Rudberg B, Leufvenmark P, Thunman M. Side effects of chiropractic treatment: a prospective study. J Manipulative Physiol Ther 1997;20:511-5.

45. Assendelft WJ, Bouter LM, Knipschild PG. Complications of spinal manipulation: a comprehensive review of the literature. J Fam Pract 1996;42:475-80.

46. Nelson CF. The tension headache, migraine headache continuum: a hypothesis. J Manipulative Physiol Ther 1994;17:156-67.

47. Featherstone HJ. Migraine and muscle contraction headaches: a continuum. Headache 1985;25:194-8.

48. Bogduk N. Headaches and the cervical spine. Cephalalgia 1984;4:7-8.

49. Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. Med Care 1989;27:S178-89.

50. Sackett DL. Rules of evidence and clinical recommendations on the use of antithrombotic agents. Chest 1989;95:2S-4S.

51. Chalmers TC, Smith H Jr, Blackburn B, Silverman B, Schroeder B, Reitman D, et al. A method for assessing the quality of a randomized control trial. Controlled Clin Trials 1981;2:31-49.

52. Moher D, Jadad A, Tugwell P. Assessing the quality of randomized controlled trials: current issues and future directions. Int J Technol Assess Health Care 1996;12:195-208.

53. Spitzer WO. Meta-meta-analysis: unanswered questions about aggregating data. J Clin Epidemiol 1991;44:103-7.

54. Eddy DM, Hasselblad V, Shachter RD. The statistical synthesis of evidence: meta-analysis by the confidence profile method. Orlando (FL): Academic Press: 1992. p. 35-108,351-66.

55. Victor N. "The challenge of meta-analysis": discussion. Indications and contra-indications for meta-analysis. J Clin Epidemiol 1995;48:5-8.

56. Cook DJ, Sackett DL, Spitzer WO. Methodologic guidelines for systematic reviews of randomized control trials in health care from the Potsdam Consultation on Meta-Analysis. J Clin Epidemiol 1995;48:167-71.

57. Feinstein AR. Meta-analysis: statistical alchemy for the 21st century. J Clin Epidemiol 1995;48:71-9.

58. Dickersin K. The existence of publication bias and risk factors for its occurrence. JAMA 1990;263:1385-9.

59. Cook DJ, Guyatt GH, Ryan G, Clifton J, Buckingham L, Willan A, et al. Should unpublished data be included in meta-analyses? Current convictions and controversies. JAMA 1993; 269:2749-53.

60. Rosenthal R. The "file drawer problem" and tolerance for null results. Psychol Bull 1979;86:638-41.

61. Dickersin K, Chan S, Chalmers TC, Sacks HS, Smith H Jr. Publication bias and clinical trials. Controlled Clin Trials 1987; 8:343-53.

62. Moher D, Fortin P, Jadad AR, Juni P, Klassen T, Le Lorier J, et al. Completeness of reporting of trials published in languages other than English: implications for conduct and reporting of systematic reviews. Lancet 1996;347:363-6.

63. Gregoire G, Derderian F, Le Lorier J. Selecting the language of the publications included in a meta-analysis: is there a Tower of Babel bias? J Clin Epidemiol 1995;48:159-63.

64. 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. Controlled Clin Trials 1995;16:62-73.

65. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. Controlled Clin Trials 1989;10:407-15.

66. Juniper EF. Quality of life questionnaires: does statistically significant = clinically important? J Allergy Clin Immunol 1998;102:16-7.

67. Oxman AD, Guyatt GH, Singer J, Goldsmith CH, Hutchinson BG, Milner RA, et al. Agreement among reviewers of review articles. J Clin Epidemiol 1991;44:91-8.

68. Verhagen AP, de Vet HC, de Bie RA, Kessels AG, Boers M, Knipschild PG. Balneotherapy and quality assessment: interobserver reliability of the Maastricht criteria list and the need for blinded quality assessment. J Clin Epidemiol 1998;51:335-41.

69. Berlin JA, Miles GC, Cirigliano MD, Goldman DR, Horowitz DA. Does blinding of readers affect the results of meta-analyses? Results of a randomized trial. Online J Curr Clin Trials 1997;29:

70. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA 1995;273:408-12.

71. Waddell G, Feder G, McIntosh A, Hutchinson A. Clinical guidelines for the management of acute low back pain. London: Royal College of General Practitioners, 1996:1-35.

APPENDIX

Evaluation list for scoring: descriptions

Scoring: A YES score (+) is only used when all described individual item criteria are met. A NO score (-) is only used when it is clear from the article that none of the described individual item criteria are met. UNCLEAR/PARTLY (p) is used when the documentation or description is insufficient to answer yes or no to whether any or all of the described individual item criteria are met. The validity score (VS) is the percentage score of the applicable validity items (maximum of 14). (+) = 1, (p) = ½, and (-) = 0.

A. Are the inclusion and exclusion criteria clearly defined? They must be stated explicitly. If a more detailed description was needed, or only inclusion or exclusion criteria were clearly defined, the score is UNCLEAR/PARTLY.

B. Is it established that the groups are comparable at baseline? If different, are appropriate adjustments made during the statistical analysis? Comparability should be present especially for main outcomes, but also for important clinical and demographic variables, such as age, gender, duration and severity of condition, and known prognostic indicators.

C. Is the randomization procedure adequately described and appropriate? If it was only noted that randomization was used, the score is NO. To receive a YES score, the randomization process must be described (ie, randomly generated list, opaque envelopes), the method used (simple, block, stratification, minimization) must be appropriate, and the concealment of randomization must be described explicitly. If only one or two of these criteria are met, a score of UNCLEAR/PARTLY is the highest possible.

D. Is it established that at least one main outcome measure was relevant to the condition under study, and were the reliability and validity documented? This must be explicitly

466 Journal of Manipulative and Physiological Therapeutics
Volume 24 • Number 7 • September 2001
Headache and Spinal Manipulation • *Bronfort et al*

established by investigation, appropriately referenced, or generally accepted (eg, VAS scales, Oswestry, or Roland-Morris disability scales). If all of the above conditions are not met the score is NO.

E. Are patients blinded to the degree possible, and did the blinding procedure work? This may not apply to study (na) (eg, a comparison of a drug and physical therapy) and is therefore not included in % scores. If the presence of either "optimal blinding" or "effectiveness of blinding" is not documented, a score of UNCLEAR/PARTLY is the highest attainable. If at least one study involves a "blindable" intervention, then the effectiveness of the blinding must be documented; otherwise a score of UNCLEAR/PARTLY is the highest attainable.

F. Is it established that treatment providers were blinded to the degree possible, and did the blinding procedure work? This may not apply to study (na) and is therefore not included in % scores.

G. Is it established that assessment of the primary outcomes was unbiased? If assessment of outcomes could be blinded, was it done? Was the effectiveness of blinding documented? Was there documentation that patients were not influenced by providers or investigators on how they scored their own outcomes?

H. Is the postintervention follow-up period adequate and consistent with the nature of the condition under study? This may not apply to study (na) (eg, crossover designs) and is therefore not included in % scores. This minimum follow-up period is 1 month for acute conditions and 3 months for chronic conditions in order to receive a YES score. A minimum of 2 weeks for acute conditions and 1 month for chronic conditions must be met for an UNCLEAR/PARTLY score.

I. Are the interventions described adequately? Did all interventions follow a defined protocol? Is it possible from the description in the article or reference to prescribe or apply the same treatment in a clinical setting? If not, YES is not an appropriate score.

J. Were differences in attention bias between groups controlled for and explicitly described? Were time, provider enthusiasm, and number of intervention sessions equivalent among study groups?

K. Is comparison made to existing efficacious or commonly practiced treatment option(s)? If a placebo controlled study, has a comparison to existing efficacious standard therapy been made previously?

L. Is the primary study objective (hypothesis) clearly defined in terms of group contrasts, outcomes, and time points a priori? (Many studies present biased posthoc conclusions.)

M. Is the choice of statistical test(s) of the main results appropriate? Is the main analysis consistent with the design and the type of the outcome variables?

N. Was it established at randomization that there was adequate statistical power ($\beta = 0.2$ with $\alpha = 0.05$) to detect an a priori determined clinically important between-group difference of the primary outcome(s) including adjustment for multiple tests and/or outcome measures?

O. Are confidence intervals (CI), or data allowing CI to be calculated, presented?

P. Are all dropouts described for each study group separately and accounted for in the analysis of the main outcomes? Look for analysis of impact of dropouts or worst/best case analysis. Almost all studies with appropriate follow-up periods that evaluated the effects of therapeutic management of a condition will have some attrition (>5%). If no dropouts, this item does not apply to study (eg, studies with one intervention and outcomes collected in same session) and is not included in % scores.

Q. Are all missing data described for each study group separately and accounted for in the analysis of the main outcomes? Look for analysis of impact of missing data. Almost all studies that evaluated the effects of therapeutic management of a condition will have missing data (>5%). If no missing data, this item will not apply to study (na) and is not included in % scores.

R. If indicated, was an intention-to-treat analysis used? In studies with documented full compliance with allocated treatments, and no differential co-intervention between groups, a YES score can apply. In single session studies (eg, studies with one intervention and outcomes collected in same session) this item does not apply (na) and is therefore not included in % scores.

S. Were adjustments made for the number of statistical tests (2 or more) when establishing cut-off point of *P*-level for each test? If applicable (avoidance of increasing risk of Type I errors), was it documented that this was an issue that could have influenced the outcome of the study, and were adjustments made (eg, Bonferonni's or similar type of adjustment)? If indicated adjustment(s) were incapable of changing main result/outcome of study, or if study involved only one test at one point in time, a score of 'na' applies.

T. Are the conclusions directly related to the primary objectives of the study, and are they valid? Were the a priori testable hypotheses tested and prioritized appropriately in the conclusions (see also item L)?