0008-3194/2004/74–88/$2.00/©JCCA 2004

Report

A review of the literature pertaining to the efficacy, safety, educational requirements, uses and usage of mechanical adjusting devices Part 1 of 2

Shane H Taylor, DC* Chairman
Nicole D Arnold, BSc, DC**
Lesley Biggs, PhD†
Christopher J Colloca, BS, DC††
Dale R Mierau, DC, FCCS, MSc**
Bruce P Symons, BSc, MSc, DC¶
John J Triano, DC, PhD, FCCS(C)(H)¶¶

Over the past decade, mechanical adjusting devices (MADs) were a major source of debate within the Chiropractors' Association of Saskatchewan (CAS). Since Saskatchewan was the only jurisdiction in North America to prohibit the use of MADs, the CAS established a committee in 2001 to review the literature on MADs. The committee evaluated the literature on the efficacy, safety, and uses of moving stylus instruments within chiropractic practice, and the educational requirements for chiropractic practice. Following the rating criteria for the evaluation of evidence, as outlined in the Clinical Guidelines for Chiropractic Practice in Canada *(1994), the committee reviewed 55 articles – all of which pertained to the Activator. Of the 55 articles, 13 were eliminated from the final study. Of the 42 remaining articles, 6 were rated as class 1 evidence; 11 were rated as class 2 evidence and 25 were rated as class 3 evidence.*

In this article – the first in a series of two – the background and the methods utilized by the MAD committee's activities are described, as well as the results for the review of the literature on efficacy. Of the 21 articles related to efficacy, five were identified as Class 1

Au cours de la dernière décennie, les appareils à mise au point mécanique (MAD) ont été une source majeure de débat au sein de l'Association des chiropraticiens de Saskatchewan (CAS). Comme la Saskatchewan était la seule juridiction nord-américaine à interdire l'utilisation des appareils à mise au point mécanique, l'Association a mis sur pied, en 2001, un comité chargé de revoir la documentation de ces appareils. Ce comité a évalué la documentation selon l'efficacité, la sécurité et l'utilisation d'instruments palpeurs mobiles dans la chiropractie et les exigences académiques de la pratique chiropratique. Suivant les critères d'évaluation lors de l'appréciation des preuves, tel que décrits dans les Directives cliniques des pratiques chiropratiques du Canada (1994), le comité a révisé 55 articles, tous en relation avec le Activator. Sur les 55 articles, 13 ont été éliminés de l'étude finale. Sur les 42 articles restants, 6 ont été classés dans les éléments de preuve de classe 1; 11 dans les éléments de preuve de classe 2; et 25 dans les éléments de classe 3.

Dans cet article, premier d'une série de deux, le contexte et les méthodes utilisées lors des activités du comité sur les appareils à mise au point mécanique ont

* Private practice of chiropractic, Regina, SK, Canada.
** Private practice of chiropractic, Saskatoon, SK, Canada.
† Department of Women's and Gender Studies, University of Saskatchewan, Saskatoon, SK, Canada.
†† Private practice of chiropractic, Phoenix, AZ, USA.
¶ Private practice of chiropractic and Human Performance Lab, Faculty of Kinesiology, University of Calgary, Calgary, AB, Canada.
¶¶ Chiropractic Division, Texas Back Institute, Plano, TX and University of Texas, Arlington Biomedical Engineering Program, USA.
Thank you to the Chiropractors' Association of Saskatchewan for funding of this project.
Submit reprint requests to the Chiropractors' Association of Saskatchewan, 3420A Hill Avenue, Regina, SK, Canada S4S 0W9.

evidence; 4 were identified as Class 2 evidence; and 12 were identified as Class 3. Overall, the committee reached consensus that the MAD procedures using the Activator were as effective as manual (HVLA) procedures in producing clinical benefit and biological change. A minority report was also written, arguing that there was not enough evidence to support or refute the efficacy of MADs.
(JCCA 2004; 48(1):74–88)

KEY WORDS: Activator, mechanical adjusting device.

été décrits, de même que les résultats de la révision de la documentation sur l'efficacité. Sur les 21 articles liés à l'efficacité, cinq ont été classés dans les éléments de preuve de classe 1, 4 dans les éléments de preuve de classe 2 et 12 dans les éléments de preuve de classe 3. Pour l'ensemble, le comité en est arrivé à un consensus : les méthodes des appareils à mise au point mécanique utilisant le Activator étaient aussi efficaces que les méthodes manuelles (HVLA) pour produire des avantages cliniques et des changements biologiques. Un rapport minoritaire a aussi été rédigé, expliquant qu'il n'y avait pas assez de preuves pour appuyer ou réfuter l'efficacité des appareils à mise au point mécanique.
(JACC 2004; 48(1):74–88)

MOTS CLÉS : Activator, appareils à mise au point mécanique.

Introduction

The use of Mechanical Adjusting Devices (MAD) in Saskatchewan has been debated for the past decade. The use of MADs in Saskatchewan is currently not sanctioned by the CAS as part of chiropractic scope of practice and contrary to Regulatory Bylaw 19(1)C which states that : **"no member shall use a machine or mechanical device as a substitute method of adjustment by hand of any one or more of the several articulations of the human body."** This has led to several motions and votes to change the bylaw to allow its use within the scope of practice in Saskatchewan. Saskatchewan is the only jurisdiction in North America which prohibits the use of mechanical adjusting devices. The membership has repeatedly voted against its use within the province, which has led to this review of the literature.

On June 2, 2001 at the Annual General Meeting of the Chiropractors' Association of Saskatchewan (CAS), a motion was passed to strike a committee to review the literature on the Activator and other mechanical adjusting devices. The motion read as follows:

> THAT, a separate committee be created within the Modes of Care committee to review the literature pertaining to the use of the Activator and other similar instruments. This committee shall provide a report to the CAS membership at least one month prior to the Fall General Meeting of 2001. A reasonably held minority opinion will also be allowed for in this report. The committee shall be comprised of 3 members of the Board's choosing and 3 members acceptable to the 42 members' who called for the special meeting in April, 2001. At least two of the members are to be female. No member of the Board since 1990 shall sit on this committee, nor shall any member directly involved in the lawsuit from either side.

This exclusion clause was included in the above motion because a group of practitioners sued the CAS when the Activator became an issue. Some members of the Board were named individually as defendants and others were involved as witnesses. It was felt that anyone involved in the lawsuit should be excluded from the MAD committee.

In the establishment of the committee, it became clear that there would be a significant cost of conducting a review of the literature, and therefore, the CAS allocated a budget for the MAD committee. In addition, a chairperson, who was named separately from the committee members, was responsible for the administration of the commitee. The chairperson would not hold voting privileges and served as a facilitator.

During the November of 2001 fall general meeting of the CAS, debate arose whether the committee should proceed or not and the motion was amended to read the following:

> THAT, the Activator committee evaluate the literature with the intent that they will report back to the membership with a recommendation on the efficacy, safety, uses and educational requirements/standards of mechanical adjusting devices for consideration for use in Saskatchewan and this to be accomplished, if possible, by May 1, 2002.

The motion was intended to direct the committee to evaluate the scientific literature concerning mechanical adjusting devices and provide a time line for filing the report. The motion was passed and the chairperson was given the budget and charged to proceed with the committee.

The committee consisted of the following:

Dr. Shane Taylor, Chair	
Dr. John Triano	Dr. Dale Mierau
Dr. Lesley Biggs	Dr. Christopher Colloca
Dr. Nicole Arnold	Dr. Bruce Symons

The committee held several conference calls to determine the process they would follow and set the guidelines which they felt would best suit the needs of the members of the CAS, as well the needs of the committee members who were spread throughout North America.

Methods

Due to the fact that the members of this committee live throughout North America, teleconference and e-mail were considered the easiest and least expensive way to carry out the mandate of the committee.

It was agreed by the committee members that a conflict of interest statement which was signed by each committee member, would become part of the final report. (See Appendix A) One member raised a potential conflict of interest which was discussed within the group. It was agreed that this member should remain part of the process due to his/her expertise in this area.

At the outset, the committee decided that there needed to be agreement on exactly what the questions should be asked. After reviewing the CAS motion, the committee decided that the following questions needed to be answered:

> **What is the evidence in the literature on efficacy, safety, and uses of moving stylus instruments within chiropractic practice?**
>
> **If evidence exists, what are the educational requirements for moving stylus instruments within chiropractic practice?**

The committee agreed that each member would submit keywords for a literature search by February 8, 2002. The literature search was performed at the Canadian Memorial Chiropractic College utilizing the following keywords and searching MEDLINE, MANTIS, and CINAHL and the INDEX TO CHIROPRACTIC LITERATURE:

Literature Search for Mechanical Adjusting Devices Committee
March 2002

Index to Chiropractic Literature and Mantis

Subject terms searched:
Activator method
Chiropractic/adverse effects
Chiropractic/instrumentation
Chiropractic/methods
Diagnosis/instrumentation
Pettibon method

Keywords searched:
Activator
Instrumentation
Instruments
Stylus
Mechanical adjusting device(s)

CINAHL

Headings searched:
Chiropractic
Chiropractic assessment
Chiropractic manipulation

Text words searched:
Chiropract*
Activator*
Spinal manipulation
Mechanical adjusting device*
Integrator
Stylus

MEDLINE

MeSH terms used:
Electromyography
Manipulation, Spinal
Chiropractic/instrumentation

Text words searched:
Same as for the other indexes

The MAD Committee, via the CAS, invited the membership of the CAS to submit information for the committee's consideration by February 28, 2002. All information was to be sent directly to the chairperson. The committee decided not to accept newspaper articles, magazine articles, conference proceedings and journal articles that were not peer-reviewed. Dr. Colloca had an initial reference list[a] which he submitted to the group to use as a cross reference for material gathered. All new material was added to Dr. Colloca's list.

In total there were 55 pieces of evidence that were accepted for review. The committee decided that the chairperson would separate the literature into six equal parts and distribute 1/6th to each members of the committee. Each member classified each piece of evidence according to an agreed data extraction template (See Appendix B). The assessment of the evidence followed a procedure rating which answered to one of three statements: 1) That the evidence supports safety, uses, efficacy and educational requirements. 2) That evidence does NOT support safety, uses, efficacy and education requirements. 3) That there is NO evidence to support safety, uses, efficacy and educational requirements. The quality of the evidence was determined according to the rating guidelines in the Glenerin proceedings found in *Clinical Guidelines for Chiropractic Practice* in Canada 1994. The guidelines were amended (as identified in bold) to reflect the mandate of the committee

> Class 1: Evidence provided by one or more well-designed controlled clinical trials; or well-designed experimental studies that address reliability, validity,
>
> positive predictive value, discriminability, sensitivity, **efficacy or safety**.
>
> Class 2: Evidence provided by one or more well-designed uncontrolled, observational clinical studies, such as case-control, cohort studies, etc; or clinically relevant basic science studies that address reliability, validity, positive predictive value, discriminability, sensitivity, specificity, **efficacy or safety**; and published in refereed journals.
>
> Class 3: Evidence provided by expert opinion, descriptive studies or case reports **on the topics of safety or efficacy.**

Once completed, each member submitted their review to the chairperson who copied and sent out all the information to each member who, in turn, reviewed all pieces of evidence and determined whether they agreed or disagreed with the classification.

The committee then held a conference call to discuss and resolve points of dispute. There were nineteen articles disputed. Some disputes were minor word changes and others changed the class of evidence. The final classification for all nineteen were unanimously agreed upon and changed accordingly. The chairperson then took all of the information and formed evidence tables under the headings of SAFETY, USES OR USAGE, EFFICACY and EDUCATIONAL REQUIREMENTS. Utilizing these evidence tables, the members of the committee were asked to write an essay on the four topics indicating whether they thought there was enough evidence, not enough evidence or no evidence to support safety, uses or usage, efficacy and educational requirements.

a Dr. Colloca's original reference list included conference proceedings which were eliminated from the list. In addition, two letters to the editor were included on the reference list which unfortunately were included as part of the package that was sent out to the members. Upon discussion, the committee agreed that these letters would be included in the reference list but were not included as part of the study.

The committee decided beforehand how the vote would be interpreted in order to determine if the committee had reached consensus. If there was a vote of 4:2, then the committee had reached consensus. The two members who disagreed would be invited to submit a minority report to attach to the final report. If there was a tie (a tie meaning either a 3:3 or 2:2:2 split), then the committee had not reached consensus. If there was a 3:2:1 vote split after discussions, then three would be considered majority and a consensus had been reached. The others who voted 2:1 would be invited to give a minority report to attach to the final report.

The chairperson distributed the essays to each member for their review prior to the final conference call, at which time the committee voted on the questions regarding the safety, efficacy and uses of MAD. Before the voting took place, each question was discussed, giving each member an opportunity to provide a rationale for their point of view. Following voting, the members gave final instructions and guidance to the chairperson, who compiled the final report which was then submitted to committee members for their comments on style but not content.

Although the committee was asked to review all mechanical adjusting devices, only research about the Activator instrument was found utilizing the inclusion criteria established by the committee. Since no other mechanical adjusting device material was found , Activator Methods was the only device that the committee reviewed.

"The Activator Adjusting Instrument (Activator Methods, Inc., Phoenix, Ariz.) is a low-force, moving stylus-type of mechanical instrument. The AAI is powered by the fixed potential energy of a spring that propels a 16-g hammer into a 30–g stylus. The spring is compressed manually by squeezing a sliding handle located on the shank of the instrument, and at a predetermined point is activated, propelling the hammer into the stylus. An 80-durometer rubber tip is attached to the end of the stylus and reduces the impulse force shock delivered to the spine slightly when the instrument is activated."[26] AMCT. St. Louis: Mosby, 1997, p 447.

Statistics

Following the conference call discussing the classification of articles, the following statistics were extrapolated.

1) 55 articles were accepted for review.
2) 13 of them did not relate at all to the questions to be answered and were therefore not included in the evidence tables.
3) The 42 remaining articles were rated as follows:
   a. 6 were rated as class 1 evidence
   b. 11 were rated as class 2 evidence
   c. 25 were rated as class 3 evidence
4) 30 of the articles were related to usage or uses.
5) 20 of the articles related to efficacy.
6) 16 of the articles related to safety.
7) 5 of the articles related to educational standards.

Evidence tables[b] were created for each subcategory of efficacy, safety, usage or uses and educational standards.

The remainder of this article provides a review of the literature on the efficacy of MAD. The issues of safety, use and usage, and educational requirements will be discussed separately in a subsequent article appearing the JCCA.

Results

Summary of the literature on efficacy

Of the 21 studies examining efficacy either implicitly or explicitly, 4 were RCT and 1 was a cohort (Class 1 Evidence), 2 were experimental (Class 2 Evidence), 1 was a clinical trial (Class 2 evidence), 11 were case studies, 1 was a case series and 1 was a review of the literature.

Class 1 Evidence

Of the RCT studies, Wood, Colloca, Mathews (2001) compared standard Diversified Technique to Mechanical Force Manually Assisted (MFMA) manipulation in the treatment of cervical spine dysfunction in a sample of 30 patients.[1] These authors found no statistical differences between the two groups; both groups showed significant improvement after the treatment phase and at a one month follow-up. Cervical range of motion (ROM) showed statistically significant changes for both groups during the treatment phase, but the differences between groups was not significant at the end of treatment or one month following.

b The evidence tables can be found on the JCCA website; the references for efficacy, safety, use and usage, and educational requirements will be presented separately in the article reviewing these issues.

Keller and Colloca (2000) demonstrated that maximal voluntary contraction of the lumbar paraspinal musculature was increased according to electromylography (EMG) measurements following Activator adjusting.[2] It is uncertain if the patients were randomized into treatment, sham and control groups. But because of the nature of treatment, it would have been obvious to the control group that they were in fact the control. This article is an interesting demonstration of the various factors mechanical adjusting devices and potentially manual adjusting can effect.

In a pilot study (n = 14) of patients with unilateral neck pain, Yurkiw and Mior (1996) found no statistically significant differences on left and right lateral flexion scores, and VAS scores differences for patients receiving MAD and SMT treatments.[3] Although the trend was toward clinical improvement for both treatments, it was not statistically significant. The clinical significance and clinical relevance of the results of both of these studies are limited by the small sample sizes of subjects participating in the research protocol making them both prone to Type II error. A lack of the ability to blind the experimenters could be a source of experimenter bias. Similarly, Gemmell and Jacobson (1995) found in sample of 30 patients no statistical differences between Meric and Activator adjustments to reduce acute low back pain.[4]

Yates et al., (1988) conducted a study (n = 21) of patients with elevated blood pressure who were randomly assigned to one of three conditions, active treatment (which received a chiropractic adjustment delivered by an Activator); a placebo group (which received a sham adjustment delivered by an Activator delivered in the off position); a control group (which received no treatment).[5] The study found statistically significant differences between the Active Treatment Condition Group, and the placebo group and control group. Lower blood pressure readings were documented for the active treatment group. The study also reported lowered states of anxiety for the active treatment group and control groups but an elevated anxiety score for the placebo group. This study was prone to the placebo effect; however it is unclear whether or not the lower blood pressures and lowered states of anxiety were statistically significant.

The number of patients/subjects included in the studies of the effect of the activator instrument on musculoskeletal conditions:

	N	number treated with activator
Low Back Pain		
8. Keller and Colloca	40	20
31. Gemmel and Jacobson	30	15
	70	35
Cervical spine pain		
6. Wood Colloca Mathews	30	15
28. Yurkiw and Mior	14	7
	44	22
Total	114	57

When the data from the four Class I clinical studies was pooled, the total number of subjects treated with the activator instrument was 57; 35 subjects experienced cervical spine pain and 22 experienced low back pain.

Class 2 Evidence

Basic science research comprises two of the three studies rated as Class II Evidence. Symons et al., (2000) demonstrated physiologic responses associated with Instrument delivered spinal adjustments.[6] In a sample of 9 patients and 83 observations reported, Symons et al reported that thrusts delivered by an Activator instrument to the entire spine elicited an 68% positive response rate overall. However, positive responses varied across the spine ranging from 94% for sacroiliac SMT thrusts to 50% for cervical thrusts. They concluded that a reflex response elicited by treatment with an Activator instrument is quantitatively and qualitatively different than the response elicited by a manual treatment and that the physiologic and clinical relevance of the reflex response they observed remains unknown.

Herzog, Kawchuk and Conway (1993) attempted to quantify the pre-load and peak forces associated with moving stylus instruments and spinal manipulative therapy (SMT).[8] They reported no significant correlation between preload and ΔF forces for treatments using the Activator instrument in contrast to the four of the five manual techniques. A statistically significant correlation between preload and ΔF was found for the manual techniques.

A moving stylus device has been found to be effectively used in a research setting "detuned" as a placebo in a study by Hawk et al. (1999).[7] In a comparison of flexion-distraction table technique with the AAI set on 0 used to perform

a sham adjustment, they found that VAS and GWBS scores improved with both treatments; a somewhat greater improvement occurred in most cases with the active treatment. This study was also subject to the placebo effect. That same cohort study (n = 18) indicates that the role of placebos needs to be examined more thoroughly.

Class 3 Evidence

In a descriptive case series study of 10 patients suffering whiplash, Osterbauer et al., (1992) reported a statistically significant decrease in overall mean pain scores and increased range of motion after treatment.[18] In case series study of 10 patients with low back pain, Osterbauer et al. (1993) found a statistically significant difference in VAS scores and Oswestry Index scores after receiving a MFMA SMT.[17] The majority, but not all patients, reported a decline in back pain and increased function; these improvements remained stable at a one year follow-up.

Improved clinical outcomes were reported in case reports of patients with post-surgical neck syndrome[9], coccygodynia[10], lumbar disc herniation[11], frozen shoulder[13], frozen shoulder with metastatic carcinoma[14], plantar fascitis[15], torn medial meniscus[16], two cases of Bell's palsy[19], otitis media[20], and sciatic neuropathy and lumbar disc herniation[21]. Of these studies, 5 provided opinion that MFMA SMT may provide an alternative when there are contraindications to using manual SMT.[10,11,13–15]

Conclusion

After reviewing the literature and after much debate the committee reached consensus (4 to 2) that, while all of these studies are flawed to varying degrees and the literature is generally weak, the evidence in the literature supports the statement that MAD procedures using Activator are as effective as manual HVLA in producing clinical benefit and biological change. More research, particularly a larger scale randomized controlled trial, would be helpful in determining efficacy to a further degree.

MAD Minority Report
Submitted by Dale Mierau DC, MSc, FCCSC and Lesley Biggs, PhD
October 4, 2002

Introduction

As the Report indicates "the committee reached consensus that while all of these studies are flawed to varying degrees and the literature is generally weak in strength, the evidence in the literature supports the statement that MAD procedures using Activator are as effective as manual HVLA in producing clinical benefit and biological change."

We agree with the conclusion that "the studies are flawed to varying degrees" and that "the literature is generally weak in strength." Where we disagree is over the statement that "the evidence in the literature supports the statement that MAD procedures using Activator are as effective as manual HVLA in producing clinical benefit and biological change." Based on our reading of the literature, we believe that the findings of studies classified as Class I do not indicate MFMA as more or less efficacious than other SMT techniques. In total, only 56 subjects over 4 studies were treated with the Activator instrument in studies classified as Class 1. Of those 56, 10 had subacute neck pain (7 in Yurkiw and Mior, 1996; 3 in Wood et al., 2001), 12 had chronic neck pain (Wood et al., 2001) 14 had acute low back pain (Gemmell and Jackson, 1995), and the duration of the LBP was unknown for 20 (Keller et al., 2000).

The tally of subjects treated with the Activator instrument across the 4 studies was:

Neck Pain	
Acute neck pain	0
Subacute neck pain	10
Chronic neck pain	12
Total	22
Low Back Pain (LBP)	
Acute LBP	14
Subacute LBP	0
Chronic LBP	0
Unknown LBP	20
Total	34
Total spine pain	56
Other	
Blood pressure and anxiety	21
TOTAL	77

The review of the literature did not reveal any studies

of the efficacy of MAD for patients with acute neck pain, subacute LBP or chronic LBP.

For studies with outcome measures that can be directly related to patient centered outcomes such as pain and function (omitting the study that used sEMG as the only outcome measure), the review of the literature documented results for 24 patients with neck pain (10 patients with subacute neck pain, 14 patients with chronic neck pain) and 14 patients with acute LBP were treated with MAD.

Smaller studies are, on average, conducted with less methodological rigor than larger studies. Trials of lower quality tend to show larger treatment effects (Schultz 1995, Moher 1998)

In our view, there is not enough evidence in the literature at this time to draw a definitive conclusion that MAD are more or less efficacious. Moreover, the limitations of the studies outlined by the authors themselves (see below) should not be taken lightly, and deserve consideration. In each of the 4 clinical studies, the authors present their findings as preliminary (i.e. as pilot studies) and call for a full-scale randomized controlled trial in order to verify their findings.

In the following sections, we present a more detailed analysis in support of our argument. We examine the types and strengths of evidence and its relationship to randomization; the relationship between statistical sources of error, statistical and clinical significance, effect size and sample size; measurement error; outcome measures; and a summary of the investigators' comments on the relevance of their work to clinical practice. Where applicable, we have included definitions of technical terms.

A. Types and Strength of Evidence

Strength of evidence	Method of study
Strongest	Randomized controlled trial
	Controlled clinical trial
	Comparative clinical trial
	Cohort study
	Case control study
	Case series
Weakest	Case report

A **randomized controlled trial** is without any doubt the best way to address questions of therapeutic efficacy. A random numbers table or some other mathematical method of randomization is used. Two or more groups are chosen at random, one receives the treatment and one does not. The only study in the literature reviewed that could be considered a full-scale randomized controlled trial was by Yates et al (1988).

Controlled clinical trials study at least one treatment and one control treatment with concurrent enrolment and follow-up of the test and control treated groups. The treatments to be administered are selected by a pseudo-random process, (e.g. a coin toss, odd-even numbers, medical record number). The Keller and Colloca (2000) study would likely have fallen into this category had they described their method of randomization of subjects.

A **comparative clinical trial** differs from a randomized control trial and a controlled clinical trial because it does not include an untreated control group. It is designed to compare two treatments by randomly allocating subjects to treatment groups. It is useful tool to compare treatments for a condition but it has limited strength because of the lack of an untreated control group. The trials published by Yurkiw and Mior (1996), Gemmel and Jacobson and Wood et al (1995) fall into this category.

For the purposes of this review, randomized controlled trials, controlled clinical trials and comparative clinical trials were assigned to the category of Class 1 evidence.

The Importance of Randomization

There are two important reasons for the use of caution when interpreting the results of non-randomized studies.

1. Randomization is the only way to control for unknown or unmeasured confounders. Non-randomized studies tend to overestimate the effects of health care intervention or treatment (Sacks et al., 1982; Chalmers et al., 1983; Schulz et al., 1995).
2. The inclusion of studies other than controlled trials in a review can increase the risk that the result of the review is influenced by publication bias (Dickersin and Min, 1993) or selection bias (Kunz et al., 1998).

It is appropriate to conduct a review of non-randomized studies of the effects of an intervention if the effects of the condition are so uniform or dramatic that it is unnecessary or unethical to wait for an RCT. The only method to establish confidence that a treatment if effica-

cious, without a randomized controlled trial, is in the circumstance that the treated condition is followed by, or results in, death. (Sackett et al., 1985).

In the absence of evidence from randomized controlled trials, it is incorrect to simply default to the best available evidence, such as weak cohort studies or case series (Sackett et al, 1985)[c].

B. The relationship between statistical sources of error, statistical and clinical significance, effect size and sample size

Conclusions from a study should not be considered as evidence of efficacy unless there are clear statistical and clinical differences between groups. The lack of a statistical difference in outcome measures between two or more groups (treated with different interventions in a small, uncontrolled study) does not allow one to reach a conclusion that the effect(s) of the interventions are equivalent. What follows is an explanation and rationale for this statement.

Definitions

Clinical significance refers to the practical importance of a reported difference in clinical outcomes between treated and control patients. It is usually expressed in terms of the size of the treatment effect.

Statistical significance is used to identify whether or not the results and conclusions drawn by the authors are likely to be true (regardless of the clinical importance or significance).

The two potential sources for statistical error are summarized in Table I (appendix A) (Sackett el. al., 1985: 183):

- a false positive result (cell x) is called a type I error. A type I error is easy to spot because the reported P value (or α as it is called before the study begins i.e. $p < .05$) is greater than 0.05 or 1 chance in 20. The smaller the reported p value, the more confident one can be that MAD is better than MM.
- an erroneous false negative conclusion is called a type II error (cell y). The size of the risk to arrive at this erroneous conclusion is called B.

By convention, investigators usually accept a 5% risk of drawing a false-positive type I error (a error of 0.05) and accept a 20% risk of concluding that the outcomes of the compared treatments do not produce different clinical outcomes when they really do (a false negative type II error of 0.20). If B is 0.20 then the power of the study is 1-B (80%).

The probability of arriving at a true positive conclusion when one is correct in doing so is called the power of the study.

The accepted settings for statistical risk are:

- the false positive risk (α) at 0.05
- the false negative risk (B) at 0.20.

To the risks of statistical error are added:

- the expected rate of outcome events for patients assigned to treatment and/or control groups.
- the degree of difference in outcomes that are considered clinically significant between treated and non treated groups or between two or more treated groups.
- the number of subjects in each group.

If one can estimate, 4 of the 5 variables, one can calculate the 5th. Pilot studies are done to allow investigators to more accurately predict or estimate some of the error variables. A larger sample size can counteract underestimation of the other sources of error including α and B.

Relevance

Three of the 5 studies categorized as Class 1 did not re-

c An example of the misuse of poor evidence is illustrated by an example of 'historical comparison' (not to be confused with 'historical controls'). Clinicians sometimes judge the efficacy of a modern treatment by comparing an experience with a new treatment to a former experience with older methods. An example cited by Sackett et al. is the rise and fall of the 'gastric freeze'. Sackett et al. 1985; p. 176–77). After reporting their results, the inventors and purveyors of a procedure called the 'gastric freeze' were pleased that their surgical service was inundated with patients who all reportedly did well after the procedure. Years later, a randomized controlled trial demonstrated that patients who underwent the 'gastric freeze' had more complications and did no better (and sometimes worse) than patients who underwent a sham treatment (Ruffin et al., 1969). The documentation of this experience speaks volumes about the confidence in new methods by patients and surgeons and highlights the importance of being careful about substituting a new untested treatment for a tested conventional one without compelling evidence from well-designed randomized controlled studies.

port a statistically significant difference between the MAD and MM groups. If the difference between MAD and MM is not statistically significant, were the trials large enough to show a clinically important difference if it was really there? This decision can be easily made with tables that provide sample size guidelines based on predetermined statistical criteria. (Sackett 1985:186–87).

Yurkiw and Mior (1996) estimated that a sample size of 150 subjects was required in each treatment group to adequately study the relative effects of MAD and MM on subacute neck pain. Gemmel and Jacobsen (1995) estimated that 1200 subjects in each treatment group were required to study the relative effects on acute low back pain. These sample sizes are far larger than the number of subjects used in the Yurkiw and Mior (1996), and Gemmel and Jacobsen (1995) studies.

This review of the literature did not identify published sample size estimates for the future investigation of acute neck pain, chronic neck pain, subacute low back pain and chronic low back pain with the MAD device. Clinical investigation of the treatment of these conditions with MAD still requires more investigation to estimate sample size.

C. Measurement Error

Wood et al., 2001: 264)) and Yurkiw and Mior (1996: 161) alerted readers to the precision, or lack of it, in the cervical range of motion measurement device (CROM). The CROM scale is in 2 degree increments. One could safely assume at least a +/- 2 degrees (4 degree total margin of error), although the margin of error could be greater due to error inherent in examiner testing as described by Yurkiw and Mior (1996).

D. Outcome Measures

Outcomes reported in clinical trials of efficacy should measure changes that are important to patients. At the Mercy Center Conference, in his discussion of selecting outcome measures when evaluating clinical interventions, Paul Shekelle (1993) emphasized that every effort should be made to base decisions regarding efficacy on 'scientific demonstrations of benefit to patients.' Further, he argued that 'benefit to patients' means outcomes that matter to patients such as 'relief of pain or ability to resume usual activities.' 'Benefit to patients' does not mean improvement in the results of diagnostic tests such as EMG or x-rays (Shekelle, 1993). One investigation, included as Class I evidence of efficacy, studied a treated group, a sham treated group and a control group (Keller and Colloca, 2000). The outcome measure used in the study (sEMG), while it may be of interest to those who study spinal pain, has no application to the treatment, or the outcome of a treatment for spinal pain. The results of this study do not support the efficacy of the Activator instrument because the outcome measure has no relevance to patients.

E. Sources of Error in Uncontrolled Studies, including Comparative Clinical Trials

Definitions

Placebo effect refers to the psychological or psycho physiological effects of a placebo; i.e. a patient's need or tendency to report a treatment effect. A patient's perception of the effects of a treatment can have an effect on the subject (placebo effect) that may be as profound or measurable as the treatment itself (Turner et al., 1994).

Hawthorne Effect refers to the tendency for subjects who are being watched or studied to perform in an unusual manner. Usually subjects perform at a higher level, or a level that the subject perceives to be better, when being watched or studied. The intangible effect of participating in a study (Hawthorne effect) can affect the results of such participation.

Relevance

Four studies compared MAD (in one case MAD set to 0 to provide a placebo treatment) to another intervention for spinal pain. In all four studies, there was documented improvement for the entire subject sample without a significant difference between the groups. (Yurkiw and Mior, 1996; Gemmell and Jacobson, 1995; Wood et al., 2001; Hawk et al; 1999.)

The study with the placebo treatment was a two-period crossover study of a sham adjustment with an Activator instrument (set to 0) and a flexion distraction technique. Both interventions demonstrated the same phenomenon; that is, improvement in outcome measures in both treated groups without a significant difference in outcomes between the groups (Hawk et al., 1999). The authors attrib-

uted the improvement in outcome measures of both groups without a significant difference between the groups to the placebo effect.

Three comparative clinical studies documented a very slight, but statistically significant improvement in outcome measures for both interventions, but no difference in outcome measures for the interventions individually or compared to one another (Yurkiw and Mior, 1996; Gemmell and Jacobson, 1995; Wood et al., 2001). Yurkiw and Mior (1996) mentioned the possibility of a placebo effect to explain the improvement in outcomes for both groups. Keller and Colloca (2000) couldn't attribute the positive changes in MVC lumbar sEMG changes directly to the Activator treatment. Wood et al. (2001) concluded that it was necessary to include an untreated control to understand the true clinical effects of the manipulative procedures. Gemmell and Jacobson (1995) discussed the possibility that subjects, knowing that they were involved in a research project, may have been biased toward reporting an expected reduction of pain. The design and results of these studies does not allow one to reach a conclusion about the efficacy of the treatments.

F. Summary of the Investigators' Comments on the Relevance of their Work to Clinical Practice

The review performed by the MAD committee was conducted to assess the present state of the literature on the efficacy of the Activator instrument. All clinical studies to date are described as inadequate to support a conclusion of efficacy by the authors who conducted the studies. We agree with these conclusions for the reasons given above. All the authors state that studies with more subjects, and an untreated control group, are required to draw conclusions regarding the relative efficacy of MAD in a clinical setting. To illustrate this point we submit the following:

1. Keller and Colloca (2000) reported a statistical difference between MAD treatment, sham treatment and a control group using sEMG as the outcome measure. The study was not randomized, the investigators were not blinded and the nature of the LBP was not documented (i.e. acute, subacute or chronic). The clinical significance and relevance of this finding should be carefully assessed. The value of the outcome measure used (sEMG) with respect to 'benefit to patients' is not known. To quote the authors:
   - 'a larger group size comparison is necessary to substantiate this finding.'
   - 'the positive changes in MVC lumbar sEMG output cannot be directly attributed to the SMT treatment alone ...'

2. Wood et al. (2001) stated in their conclusion: ' ... these results would have supported the use of MFMA (MAD) and HVLA (MM) manipulation for cervical spine dysfunction only if a control group had been studied with the investigation.'

3. Yurkiw and Mior (1996) identified their results as not statistically significant and recommended modifications for future study including a larger sample size. Their sample size estimate was for at least 150 subjects in each group.

4. Gemmell and Jacobsen (1995) documented no difference in relative effectiveness between MAD and MM. They implicated a Type II error and recommend that 'judgment be suspended until the study can be repeated by other researchers.'

5. Yates et al. (1988) authored a study that was sound in design and recommendations for future study. It is a shame that the work was not carried forward.

G. Conclusion

Conclusions regarding the strength of published evidence for efficacy of a treatment can be subjected to the following 5 considerations (Hill, 1971):

1. How good is the quality of the reviewed trials?
   The committee agreed unanimously that the quality of the reviewed trials was weak.
2. How large and significant are the observed effects?
   The observed effects were not large. Rather, relative to the hypotheses, they were insignificant.
3. How consistent are the effects across trials?
   The lack of a significant difference between compared treatments was consistent across trials.
4. Is there a clear dose-response relationship?
   A dose-response relationship was not discussed in any of the reviewed trials.
5. Is there indirect evidence that supports the inference?

Indirect evidence (i.e. both interventions improved scores on outcome measures) was present and it was discussed, but the indirect evidence (improvement in outcome measures for both treatments (including a detuned Activator) was not specific to either intervention.

6. Have other plausible competing explanations of the observed effects (e.g. bias or co-intervention) been ruled out?

Plausible competing explanations for the observed effects are presented above.

We submit that the deliberations of this committee overestimated the validity and clinical relevance of the literature reviewed about the Activator instrument. We feel compelled to put forward the position, that a larger, robust randomized controlled trial is necessary to support the conclusion of a level of efficacy equivalent to that of manual manipulation/adjustment for the treatment of spinal pain/dysfunction.

References: Majority Report

1 Wood TG, Colloca CJ, Matthews R. A pilot randomized clinical trial on the relative effect of instrumental versus manual thrust manipulation in the treatment of cervicalspine dysfunction. J Manipulative & Physiological Therapeutics 2001; 24(4):260–271.

2 Keller TS, Colloca CJ. Mechanical force spinal manipulation increases trunk muscle strength assessed by electromyography: a comparative clinical trial. J Manipulative & Physiological Therapeutics 2000: 23(9):585–595.

3 Yurkiw D, Mior S. Comparison of two chiropractic techniques on pain and lateral flexion in neck pain patients: a pilot study. Chiropractic Technique 1996; 8:155–162.

4 Gemmell HA, Jacobson BH. The Immediate Effect of Activator vs. Meric Adjustment on Acute Low Back Pain: A Randomized Controlled Trial. J Manipulative & Physiological Therapeutics 1995; 18:453–456.

5 Yates RG, Lamping DL, Abram NL, Wright C. Effects of Chiropractic Treatment on Blood Pressure and Anxiety: a Randomized Controlled Trial. J Manipulative & Physiological Therapeutics 1988; 11(6):484-488.

6 Symons BP, Herzog W, Leonard T, Nguyen H. Reflex responses associated with Activator treatment. J Manipulative & Physiological Therapeutics 2000; 23(3):155–159.

7 Hawk C, Azad A, Phongphua C, Long CR. Preliminary study of the effects of a placebo chiropractic treatment with sham adjustments. J Manipulative & Physiological Therapeutics 1999; 22(7):436–443.

8 Herzog W, Kawchuk GN, Conway PJ. Relationship Between Preload and peak forces during Spinal Manipulative Treatments. JNMS 1993; 1(2): 52–58.

9 Polkinghorn BS, Colloca CJ. Chiropractic treatment of postsurgical neck syndrome utilizing mechanical force, manually-assisted short lever spinal adjustments. J Manipulative & Physiological Therapeutics 2001: 24(9):589–595.

10 Polkinghorn BS, Colloca CJ. Chiropractic treatment of coccygodynia via external instrumental adjusting procedures utilizing Activator Methods Chiropractic Technique. J Manipulative & Physiological Therapeutics 1999; 22(6):411–416.

11 Polkinghorn BS, Colloca CJ. Treatment of symptomatic lumbar disc herniation utilizing Activator Methods Chiropractic Technique. J Manipulative & Physiological Therapeutics 1998; 21(3):187–196.

12 Cooperstein R. Activator Methods Chiropractic Technique. Chiropractic Technique 1997; 9(3):108–114.

13 Polkinghorn BS. Chiropractic Treatment of Frozen Shoulder Syndrome (Adhesive Capsulitis) Using Mechanical Force Manually Assisted Short Lever Adjusting Procedures. J Manipulative & Physiological Therapeutics 1995; 18:105–115.

14 Polkinghorn BS. Instrumental Chiropractic Treatment of Frozen Shoulder Associated With Mixed Metastatic Carcinoma. Chiropractic Technique 1995; 7:98–102.

15 Polkinghorn BS. Posterior Calcaneal Subluxation: An Important consideration in Chiropractic Treatment of Plantar Fasciitis. Chiropractic Sports Medicine 1995; 9(2):44–51.

16 Polkinghorn BS. Conservative treatment of torn medial meniscus via mechanical force, manually assisted, short lever chiropractic adjusting procedures. J Manipulative & Physiological Therapeutics 1994; 17:474–484.

17 Osterbauer PJ, DeBoer KF, Widmaier RS, Petermann EA, Fuhr AW. Treatment and biomechanical assessment of patients with chronic sacroiliac joint syndrome. J Manipulative & Physiological Therapeutics 1993; 16(2):82–90.

18 Osterbauer PJ, Derickson KL, Peles JD, DeBoer KF, Fuhr AW, Winters JM. Three-dimensional head kinematics and clinical outcome of patients with neck injury treated with spinal manipulative therapy: A pilot study. J Manipulative & Physiological Therapeutics 1992; 15(8):501–511.

19 Frach JP, Osterbauer PJ. Chiropractic treatment of Bell's Palsy by Activator Instrument adjusting and high voltage electrotherapy: A report of two cases. J Manipulative & Physiological Therapeutics 1992; 15(9):596–598.

20 Phillips NJ. Vertebral subluxation and otitis media: A Case Study. J Chiropractic Research and Clinical Investigation 1992; 8(2):38–39.

21 Richards GL, Thompson JS, Osterbauer PJ, Fuhr AW. Low force chiropractic care of two patients with sciatic neuropathy and lumbar disc herniation. Am J Chiropractic Medicine 1990; 3(1):25–32.

Additional References Included in the Minority Report

Chalmers TC, Cleano P, Sacks HS, Smith H. Bias in treatment assignment in controlled clinical trials. N Engl J Med 1983; 309:1358–1361.

Dickersin K, Min YI. NIH clinical trials and publication bias. Online J Cur Clin Trials 1993; Doc #50.

Hill AB. Principles of Medical Statistics. Lancet, 1971:312–330.

Kunz RA, Oxman AD. The unpredictability paradox; a review of empirical comparisons of randomized and non-randomized trials. BMJ 1998; 317:1185–1190.

Moher D, Pham b, Jones A, Cook DJ, Jadad AR, Moher M, Tugwell P, Klassen TP. Does quality of reports of randomized trials affect estimates of intervention efficacy reported in meta-analyses? Lancet 1998; 352:609–613.

Ruffin JM, Grizzle JE, Hightower NC, et al. A cooperative double-blind evaluation of grastric "freezing" in the treatment of duodenal ulcer. N. Engl. J. Med. 1969; 281:16.

Sackett DL, Haynes BR, Tugwell P. Clinical Epidemiology: A Basic Science for Clinical Medicine. Little Brown and Co. Toronto. 1985.

Sacks HS, Chalmers TC, Smith H. Randomized versus historical controls for clinical trials. Am J Med 1982; 72:233–240.

Shekelle P. The evolution and mechanics of a consensus process. In Haldeman S, Chapman-Smith D, Peterson DM. Guidelines for Chiropractic Quality Assurance and Practice Parameters: Proceedings of the Mercy Center Consensus Conference. Aspen Publishers Inc. Gaithersburg Maryland 1993 pp. xxix.

Schultz KF, Chalmer I, Hayes RJ, Altman DG. Empirical evidence of bias: Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA 1995; 273:408–412.

Turner JA, Deyo R, Loeser JD, Von Korff M, Fordyce WE. The importance of placebo effects in pain treatment and research. JAMA 1994; 271:1609–1614.

Appendix A
Disclosure of Potential Conflict of Interest

For the CAS MAD Committee

Having an interest or affiliation with a corporate organization which is involved in the manufacture of, training for or marketing of mechanical adjusting devices may not prevent a member from participating in the MAD Committee. This relationship or affiliation, however, must be disclosed in advance to MAD Committee and the CAS.

Although it is impossible to list every circumstance which could potentially result in a conflict of interest, the following will serve as a guide to the types of activities or conditions which should be reported:

- **To hold, directly, or indirectly, a position with a commercial company for which payment or financial reward is received.**
- **To render directive, managerial, or consultative services to a company.**
- **To accept substantial entertainment, favours, or rewards from a company.**
- **To have a significant stockholding or other material interests in which performance of the company effects financial gain.**
- **To receive payment for presentations made on behalf of a company, i.e., payment for service on speaker's bureaus, etc.**

Name of Committee: MAD committee membership and deliberations
Date of Activity: 2001–2002
Name of MAD Committee Member:

PLEASE COMPLETE AND SIGN ON THE SIGNATURE LINE BELOW

I certify that I do not have now, nor have ever had, any financial interest in the subject matter to be reviewed by the MAD Committee that was struck by the Chiropractors' Association of Saskatchewan; not do I have now nor have ever had any affiliation with, or involvement in, any organization or entity with a direct financial interest in the subject matter to be reviewed by the MAD Committee.

_____ I do not now, nor have I ever had a conflict of interest

_____ I do now, or have had in the past a financial interest, affiliation with, or involvement in any organization or entity with a direct financial interest in the subject matter to be reviewed by the MAD Committee, the details of which are disclosed below.

If you have now or have you ever had a financial interest in the subject matter to be reviewed by the MAD Committee that was struck by the Chiropractors' Association of Saskatchewan; or any affiliation with, or involvement in, any organization or entity with a direct financial interest in the subject matter to be reviewed by the MAD Committee including financial interest or affiliation with (1) the manufacturer of any products, devices or services to be discussed in your presentation at this activity; or (2) with any of the companies providing commercial support for this activity, please identify your involvement below.

Affiliation or Financial Interest	**Name of Company/Organization**
Grant or Research Support	_____
Employee or Paid Consultant	_____
Speaker's Bureau	_____
Stock/Investment Holder	_____
Other	_____

Signature of Committee Member: **Date:**

Appendix B

AUTHOR(S):
TITLE:

JOURNAL (YEAR/VOLUME/PAGES):

Answer all questions in cell E
Answer yes = Y no = N or does not apply = N/A

DOES THIS STUDY ADDRESS USE OR USAGE?

DOES THIS STUDY ADDRESS EFFICACY?

DOES THIS STUDY ADDRESS SAFETY?

DOES THIS STUDY ADDRESS EDUCATIONAL REQUIREMENTS?

NOTES:

SPECIFIC INFORMATION

STUDY CHARACTERISTICS:
VERIFICATION OF STUDY ELIGIBILITY:
Correct population, interventions, outcome and study design

POPULATION CHARACTERISTICS
1) target population (describe)
2) inclusion criteria
3) exclusion criteria
4) recruitment procedures used
5) characteristics of participants
6) number of participants
7) were interventions and control groups comparable?

METHODOLOGICAL QUALITY OF STUDY
1) DESIGN OF THE STUDY:
A) RCT
B) COHORT STUDY
C) CASE STUDY
D) CASE CONTROL
E) EXPERIMENTAL

INTERVENTIONS
1) TYPE OF INTERVENTION
2) OR DESCRIPTION OF TEST

OUTCOMES, OUTCOME MEASURES
1) WHAT WAS MEASURED AT BASELINE?
2) WHAT WAS MEASURED AFTER THE INTERVENTION?
3) WHO CARRIED OUT THE MEASUREMENT?
4) WHAT WAS THE MEASUREMENT TOOL?
5) WAS/WERE THE TOOL VALIDATED AND HOW?

ANALYSIS
1) STATISTICAL TECHNIQUE USED?
2) DOES TECHNIQUE ADJUST FOR CONFOUNDING?
3) UNIT OF ANALYSIS?
4) ATTRITION RATE?

RESULTS
1) QUANTITATIVE RESULTS?
2) EFFECT OF THE INTERVENTION ON THE OTHER MEDIATING VARIABLES?
3) QUALITATIVE RESULTS?
4) COST OF INTERVENTION?
5) COST EFFECTIVENESS?

CLASS OF EVIDENCE (1, 2, 3):

ARTICLES THAT FALL INTO CLASS 1 SHOULD BE FURTHER ASSESSED USING THE FOLLOWING:
Mark cell under your choice with an X.

WAS THE ASSIGNMENT TO THE TREATMENT REALLY RANDOM?
ADEQUATE PARTIAL INADEQUATE UNKNOWN

WAS THE TREATMENT ALLOCATION CONCEALED?
ADEQUATE INADEQUATE UNKNOWN

WERE THE GROUPS SIMILAR AT BASELINE REGARDING THE PROGNOSTIC FACTORS?
REPORTED UNKNOWN

WERE THE ELIGIBILITY CRITERIA SPECIFIED?
ADEQUATE PARTIAL INADEQUATE UNKNOWN

WERE OUTCOME ASSESSORS BLINDED TO THE TREATMENT ALLOCATION?
ADEQUATE INADEQUATE UNKNOWN

WAS THE CARE PROVIDER BLINDED?
ADEQUATE PARTIAL INADEQUATE UNKNOWN

WAS THE PATIENT BLINDED?
ADEQUATE PARTIAL INADEQUATE UNKNOWN

WERE THE POINT ESTIMATES AND MEASURE OF VARIABILITY PRESENTED FOR THE PRIMARY OUTCOME MEASURE
ADEQUATE PARTIAL INADEQUATE UNKNOWN

DID THE ANALYSIS INCLUDE AN INTENTION TO TREAT ANALYSIS?
ADEQUATE PARTIAL INADEQUATE UNKNOWN

DEALING WITH MISSING VALUES
ADEQUATE PARTIAL INADEQUATE UNKNOWN

LOSS TO FOLLOW UP
ADEQUATE PARTIAL INADEQUATE UNKNOWN