COMPARISON OF MECHANICAL vs. MANUAL MANIPULATION METHODS FOR LOW BACK PAIN

by

Michael J. Schneider, D.C.

State University of New York, Binghamton, NY 1975-1978

Doctor of Chiropractic, Palmer College of Chiropractic, Davenport, Iowa 1982

Submitted to the Graduate Faculty of

School of Health and Rehabilitation Sciences

in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2008

UNIVERSITY OF PITTSBURGH

SCHOOL OF HEALTH AND REHABILITATION SCIENCES

This dissertation was presented

by

Michael J. Schneider, D.C.

It was defended on

November 24, 2008

and approved by

Jennifer Brach, PhD, PT, GCS

James J. Irrgang, PhD, PT, ATC

Katherine Verdolini, PhD

Steven Wisniewski, PhD

Dissertation Advisor: Anthony Delitto, PhD, PT, FAPTA

COMPARISON OF MECHANICAL vs. MANUAL MANIPULATION METHODS FOR LOW BACK PAIN

Michael J. Schneider, Ph.D., D.C.

University of Pittsburgh, 2008

ABSTRACT:

Purpose and Study Design: Prospective cohort study to explore the clinical treatment effect of mechanical vs. manual manipulation for acute low back pain.

Methods: 92 patients with a history of acute low back pain were recruited from three private chiropractic offices. Two of these offices utilized manual lumbar manipulation and one used mechanical instrument manipulation (Activator) as their primary modes of treatment. The chiropractors used a "treatment as usual" protocol with the participants for a maximum of eight visits or four weeks, which ever occurred first. Primary and secondary outcome measures were the differences in pain and Oswestry scores from baseline to four weeks, respectively.

Results: Socio-demographic characteristics of the two cohorts at baseline were not found to show any significant differences between the groups except for age. The Activator cohort had a significantly higher utilization of adjunctive modalities and x-rays, with a mean number of office visits about twice that of the manual manipulation cohort at four weeks. The pain scores decreased in both groups with the manual manipulation group showing a slightly greater amount of pain reduction at four weeks, but this difference did not reach statistical significance after controlling for baseline pain. The manual manipulation group also showed a slightly greater

reduction in Oswestry scores from baseline to four-weeks, but this difference was not statistically significant after adjusting for baseline Oswestry score.

Conclusions: In this observational study of treatment-as-usual there was no significantly greater reduction in pain scores or Oswestry scores between the manipulation and Activator groups at four weeks. There were many differences between the Activator and manual manipulation groups with respect to treatment beliefs and expectations, modality usage, and frequency/duration of care, which are potential sources of confounding in the interpretation of these results. This study provides important pilot data and research issues for the design of a future randomized clinical trial that can control for these issues of confounding variables.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0 INTRODUCTION

1.1 LOW BACK PAIN AND MANIPULATION

Low back pain has become a serious public health issue in the United States and many other Western societies for three basic reasons, the first of which is the large prevalence of low back pain in American society. The general yearly prevalence of low back pain in the U.S. population is estimated at 15-20% and among working-age adults at 50% (1). Back pain is the most common cause of disability for persons under the age of 45 and the second most common reason for office visits to primary care physicians (2). The second reason is the high rate of disability and activity intolerance due to low back pain. A systematic review of population prevalence studies of low back pain revealed the following: point prevalence ranged from 12% to 33%, one-year prevalence ranged from 22% to 65%, and lifetime prevalence ranged from 11% to 84% (3). The third reason is increasing evidence that low back pain is not necessarily a self-limiting disorder. One recent study showed that in patients who experienced low back pain for greater than thirty days, 40% continued to have symptoms for another one to five years (4).

In 1994, the Agency for Health Care Policy and Research (AHCPR) published clinical practice guidelines for the management of acute low back problems in adults based upon a careful review and meta-analysis of the literature on various treatments for low back pain (1). The AHCPR

guidelines were the first time that an official government health agency recommended manipulation as a primary treatment for patients with acute low back pain. During the past ten years, since the publication of these AHCPR guidelines, many additional clinical trials have been published that continue to show significant benefits of spinal manipulation for certain types of low back and neck pain. These trials have been the subject of several systematic reviews and meta-analyses of the manipulation literature (5-12).

1.2 DEFINITION OF MANIPULATION

This intervention called "spinal manipulation" that appears to be so beneficial to acute low back pain sufferers is not a single treatment procedure. Spinal manipulation is an umbrella term that includes a multitude of different procedures that introduce a variety of manual and mechanical forces into the musculoskeletal structures. Manipulation is practiced by a variety of clinicians, especially those in the professions of chiropractic, osteopathy, and physical therapy.

An analysis by the editor of the Journal of Chiropractic Technique in 1993 stated that there were over 100 chiropractic technique systems using a variety of manipulative methods at that time (13). In an attempt to better understand chiropractic practice methods, the National Board of Chiropractic Examiners (NBCE) conducted two national surveys of the chiropractic profession in 1991 and 1998 from which data were obtained from questionnaires as to the most frequently used manipulative techniques in clinical practice (14, 15). These results indicated that chiropractors on average used seven separate manipulative techniques in their practices. The two most commonly-used methods were reported as Diversified Technique, which utilizes manual

high velocity low amplitude (HVLA) manipulations; and Activator Methods which utilizes a hand-held device to deliver a mechanical thrust in lieu of a manual thrust.

1.3 MANUAL VS. MECHANICAL MANIPULATION

"Diversified technique" is a chiropractic term used to describe several types of high-velocity low-amplitude (HVLA) thrust manipulations performed manually. Most of the literature regarding manipulation for low back pain describes some variation of HVLA manipulation with the patient treated in a side lying position. Traditionally, chiropractors have been trained to use chiefly HVLA procedures for most types of non-complicated acute low back pain, and a review of the literature tends to support this use of the procedure for treatment of low back pain. A typical side posture HVLA manipulation technique is depicted in Figure 1 below:

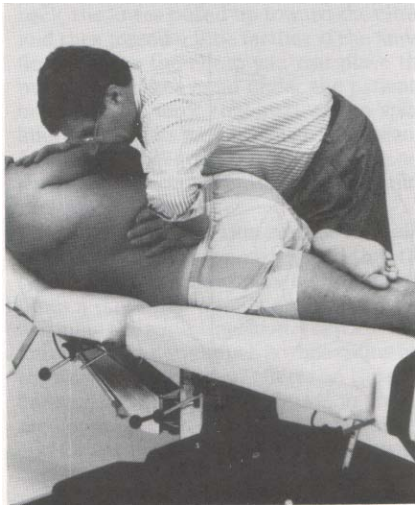

Figure 1: Side posture manual manipulation

However, some chiropractors are reluctant to use HVLA manipulation with certain types of patients, including those with osteoporosis, spinal stenosis, intervertebral disc lesions with radiculopathy, and patients with high fear or anxiety about manual thrust techniques. Mechanical manipulation methods such as the Activator Instrument are more appealing for use in these cases as an alternative to traditional HVLA manipulation, with the belief that the mechanical device may offer a more controlled and safer thrust into the spinal structures (16). Figure 2 below depicts an Activator Instrument manipulation applied to the abdomen of a patient lying supine:

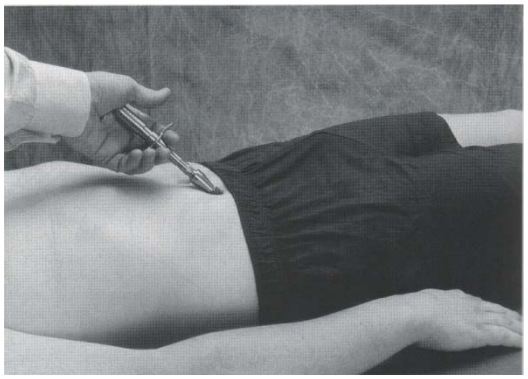

Figure 2: Activator instrument (mechanical manipulation)

Although mechanical devices such as the Activator Instrument are used by a large number of chiropractors with the belief that they are equally effective as HVLA manipulation, only a few small clinical trials comparing these methods for low back pain have been published with equivocal results due to small sample size and other methodological flaws (17, 18). This has created a situation in which an estimated 31,000 chiropractors (16) are using a mechanical instrument to manipulate the spine with little clinical evidence to support or refute their claims of therapeutic effectiveness for the use of such mechanical instruments.

1.4 STATEMENT OF PURPOSE

There is strong evidence for the clinical utilization of manual HVLA manipulation methods for the treatment of acute low back pain, but only weak evidence for mechanical manipulation methods such as Activator (19, 20). Yet Activator is reported as the second most commonly used manipulative method within the chiropractic profession, and is generally considered a routine substitute method for manual HVLA manipulation (14).

This disparity between the extremely high utilization rate of mechanical manipulation methods and the weak literature support for the use of such methods is striking. There is a need to fill this knowledge void by exploring the differences, if any, between manual and mechanical manipulation methods for the treatment of low back pain. Ideally this would involve the implementation of randomized clinical trials (RCT) in which subjects with specific inclusion criteria related to low back pain are randomized into groups receiving either mechanical or manual treatment methods. In the absence of funding for such RCTs, a preliminary step would be a prospective cohort study in which two similar cohorts of low back pain patients are evaluated for their therapeutic responses to care in clinics which specialize in HVLA manipulation and Activator respectively.

1.5 SPECIFIC AIMS AND HYPOTHESES

1.5.1 Specific Aim 1

To perform inter-examiner reliability studies on the prone leg length analysis (Activator) and prone springing palpation for the lumbar facets and sacroiliac joints (manual manipulation).

1.5.2 Hypothesis 1.1

The existing literature reports good reliability for determining the side of leg length inequality but mixed results for determining the magnitude of the inequality and other components of this clinical examination test. The data from this study will be analyzed and compared to the previous results of important studies in the existing literature.

1.5.3 Hypothesis 1.2

The existing literature regarding prone lumbar springing palpation reports good reliability for pain provocation, but less reliability for segmental mobility testing in the absence of pain provocation. The data from this study will be analyzed and compared to the previous results of important studies in the existing literature.

1.5.4 Specific Aim 2

To gather clinical information on two cohorts of patients with sub-acute low back pain who subsequently receive two different types of manipulative treatment; standard side-lying manual

manipulation (Diversified) and mechanical manipulation with an Activator Instrument. The main outcome measures will be pain and function as measured by the Numeric Pain Rating Scale (NPRS) and Oswestry Low Back Pain Disability Index (ODI) at initial intake examination (baseline) and subsequent reexamination at four weeks.

1.5.5 Hypothesis 2

The null hypothesis is that there will be no significant difference between the two cohorts with respect to their NPRS and ODI difference scores compared from baseline to four weeks.

1.5.6 Specific Aim 3

To gather qualitative information from patients about their expectations and beliefs regarding the treatment they are about to receive. More specifically, patients will be asked about their feelings regarding manipulation in general, why they chose the particular clinician they are seeing, and how they feel about "cracking" manipulations.

1.5.7 Hypothesis 3.1

The null hypothesis is that the Activator and Diversified patient cohorts will not differ significantly in their underlying beliefs/expectations about the manipulation techniques they have chosen to receive, and those they have chosen not to receive.

1.5.8 Hypothesis 3.2

The null hypothesis is that the level of expectation about treatment effect will not be correlated with changes in pain and function scores from baseline to four weeks.

1.5.9 Specific Aim 4

To gather psychosocial data regarding fear avoidance beliefs in the two cohorts by use of the Fear Avoidance Beliefs Questionnaire (FABQ) at baseline and four-weeks.

1.5.10 Hypothesis 4

The null hypothesis is that there will be no significant difference between the two cohorts with respect to their FABQ scores at baseline and four weeks.

2.0 REVIEW OF THE LITERATURE

2.1 HISTORICAL PERSPECTIVES ON MANIPULATION

Manipulation of the lumbar spine has been practiced through-out recorded human history, with records as far back as the time of Hippocrates depicting manual manipulative interventions for low back pain. There is written and pictorial evidence that spinal manipulation has been performed by people within the cultures of Europe, Asia, and the Middle East for thousands of years (21). In many European countries, the art of manipulation was passed down through the generations as the practice of "bone-setting". During the first half of the twentieth century in the United States, Europe, and Australia, manipulation surfaced again as a more "scientific" method of manual treatment, chiefly performed by chiropractors, osteopaths, and physical therapists.

Each of these professions has authors who produced textbooks that described and illustrated the various manipulation methods favored by those professions, along with many theories and hypotheses about purported mechanisms of action for manipulation methods based upon mostly empirical clinical observation (22-27). The early manipulation literature (prior to 1970) consisted mostly of these textbooks along with published case reports, descriptive studies, anecdotal clinical vignettes, and some low quality observational studies. The first fully indexed and peer-reviewed chiropractic journal (Journal of Manipulative and Physiological Therapeutics) appeared

in 1978 and began to set more rigorous standard for chiropractic research publications. Gradually, during the latter part of the century, randomized clinical trials and high quality observational studies began to emerge in the chiropractic and health care literature regarding the use of manipulation for back and neck pain.

2.2 SYSTEMATIC REVIEWS OF THE MANIPULATION LITERATURE

As the number of higher quality research studies on manipulation and back pain grew in the medical literature, it became feasible for panels of researchers to perform meta-analyses and systematic reviews of these studies. At least a dozen systematic reviews of the manipulation literature for low back pain have now been published in the period from 1992 to 2004 (5-10, 12, 28-33), with two systematic reviews published in 2004 alone (7, 28) and the most recent systematic review published in 2008 (34).

The first meta-analysis of the spinal manipulative therapy (SMT) literature was published in 1992 by Anderson et al which systematically searched for all studies of SMT published up to June 1989 (5). The Index Medicus from 1980 was expanded by citation tracking. The Chiropractic Research Archives Collection was utilized as a regularly updated bibliographic source for the location of research publications. A hand search of professional chiropractic journals was also undertaken, since all of those journals were not indexed in the standard biomedical literature. Studies in English with concurrent controls treated by methods other than SMT, including sham, produced 23 randomized controlled clinical trials of the effectiveness of spinal manipulation. Most studies compared SMT to an alternative treatment. This probably

obscured the effectiveness of SMTs since the comparison treatments were presumably also effective. These authors concluded that SMT proved to be consistently more effective in the treatment of low back pain than were any of the other comparison treatments.

Also in 1992, Shekelle et al published a comprehensive and systematic review of the SMT literature for low back pain (12). The authors identified articles through a MEDLINE search, review of articles' bibliographies, and advice from expert orthopedists and chiropractors. All studies reporting use and complications of spinal manipulation and all controlled trials of the efficacy of spinal manipulation were analyzed. 58 articles, including 25 controlled trials, were retrieved. Data on the use and complications of spinal manipulation were summarized. Controlled trials of efficacy were critically appraised for study quality. Data from nine studies were combined using the confidence profile method of meta-analysis to estimate the effect of spinal manipulation on patients' pain and functional outcomes. They concluded that spinal manipulation was of short-term benefit particularly in those cases with uncomplicated, acute low-back pain. Data were insufficient concerning the efficacy of spinal manipulation for chronic low-back pain.

Lawrence et al published a meta-analysis of the SMT literature for LBP in 1993 (9). Articles were identified through Index Medicus and a MEDLINE search using the MeSH terms chiropractic, manipulation, and backache. Additional articles were identified by reviewing the bibliographies of retrieved articles and by consulting expert orthopedists and chiropractors. All RCTs evaluating the use of spinal manipulation in out patients were considered, along with important case series, textbooks, and reviews. All articles on complications of manipulation were

reviewed. 58 studies, including 25 randomized controlled trials, were reviewed. The authors concluded that the literature showed spinal manipulation increases the probability of recovery at 3 weeks in outpatients with acute and subacute low-back pain. Data were insufficient to evaluate the efficacy for chronic low-back pain and back pain with sciatica.

A 1995 article was published by Assendelft et al who performed a meta-analysis of all of the systematic reviews of the manipulation literature (35). A total of 51 systematic reviews were assessed, 17 of which were neutral and 34 positive, with respect to manipulation showing a benefit for back pain. The methodological quality of the systematic reviews was low, with a median score of 23%. Nine of the 10 methodologically best reviews were positive. Other factors associated with a positive reviewers' conclusion were review of spinal manipulation only, inclusion of a spinal manipulator in the review team, and a comprehensive literature search. Although, in particular, the vast majority of the reviews with a relatively high methodological quality had a positive conclusion, these authors cautioned that strong conclusions were precluded by the overall low quality of the reviews.

Koes, Assendelft, et al published a follow-up article in 1996, that instead of a meta-analysis of other systematic reviews (like their 1995 article), was their own direct systematic review of the manipulation literature(8). Thirty-six randomized clinical trials comparing spinal manipulation with other treatments were identified. Nineteen studies (53%) showed favorable results for manipulation. In addition, five studies (14%) reported positive results in one or more subgroups only. Twelve trials included patients with acute low back pain only. Of these, five reported positive results, four reported negative results, and three reported positive results in a subgroup

of the study population only. There were eight trials comparing manipulation with other conservative treatment modalities, focusing on patients with sub-acute or chronic low back pain. Of these, five reported positive results, two reported negative results, and in one study no conclusion was presented. There were 16 studies that included an effect measurement of at least 3 months. In six of these the authors report positive effects of manipulation. The authors concluded that there were indications that manipulation might be effective in some subgroups of patients with low back pain, but again emphasized the fact that the methodological quality of manipulation trials was suboptimal and remained a critical aspect that should be dealt with in future studies.

A systematic review of randomized controlled trials was published in 1997 by van Tulder, Koes, and Bouter (33). Their objective was to assess the effectiveness of the most common conservative types of treatment for patients with acute and chronic nonspecific low back pain, including spinal manipulation. A rating system was used to assess the strength of the evidence, based on the methodological quality of the randomized controlled trials, the relevance of the outcome measures, and the consistency of the results. The scores ranged from 20 to 79 points for acute low back pain and from 19 to 79 points for chronic low back pain on a 100-point scale, indicating the overall poor quality of the trials.

Overall, only 28 (35%) randomized controlled trials on acute low back pain and 20 (25%) on chronic low back pain had a methodological score of 50 or more points, and were considered to be of high quality. Strong evidence was found for the effectiveness of muscle relaxants and non-steroidal anti-inflammatory drugs and the ineffectiveness of exercise therapy for acute low back

pain; strong evidence also was found for the effectiveness of manipulation, back schools, and exercise therapy for chronic low back pain, especially for short-term effects. These authors suggested that the quality of the design, execution, and reporting of randomized controlled trials should be improved, to establish strong evidence for the effectiveness of the various therapeutic interventions for acute and chronic low back pain.

In 2003, Assendelft et al published an updated meta-analysis of the spinal manipulation literature, citing the fact that previous systematic reviews and practice guidelines had reached discordant results on the effectiveness of this therapy for low back pain (6). The purpose of their newer meta-analysis was an attempt to resolve the discrepancies related to use of spinal manipulative therapy and to update previous estimates of effectiveness by comparing spinal manipulative therapy with other therapies and then incorporating data from recent high-quality randomized, controlled trials (RCTs) into the analysis. MEDLINE, EMBASE, CINAHL, the Cochrane Controlled Trials Register, and previous systematic reviews were all systematically searched for RCTs of patients with low back pain that evaluated spinal manipulative therapy with at least 1 day of follow-up and at least one clinically relevant outcome measure. Thirty-nine such RCTs were identified and evaluated by meta-regression models for acute or chronic pain and short-term and long-term pain and function.

For patients with acute low back pain, spinal manipulative therapy was superior only to sham therapy or therapies judged to be ineffective or even harmful. Spinal manipulative therapy had no statistically or clinically significant advantage over general practitioner care, analgesics, physical therapy, exercises, or back school. These authors concluded that there was no evidence

that spinal manipulative therapy is superior to other standard treatments for patients with acute or chronic low back pain.

In contract to the conclusions of Assendelft et al, the most recent systematic review of the manipulation literature published by Bronfort et al in 2004 concludes that spinal manipulation is a viable option for the treatment of low back and neck pain (7). In this study, Bronfort et al searched for RCTs including 10 or more subjects per group receiving spinal manipulation or mobilization techniques and using patient-oriented primary outcome measures (e.g., patient-rated pain, disability, global improvement and recovery time). Articles in the English, Danish, Swedish, Norwegian and Dutch languages reporting on randomized trials were identified by a comprehensive search of computerized and bibliographic literature databases up to the end of 2002. Two reviewers independently abstracted data and assessed study quality according to eight explicit criteria. A best evidence synthesis incorporating explicit, detailed information about outcome measures and interventions was used to evaluate treatment efficacy. The strength of evidence was assessed by a classification system that incorporated study validity and statistical significance of study results. Sixty-nine RCTs met the study selection criteria and were reviewed and assigned validity scores varying from 6 to 81 on a scale of 0 to 100. Forty-three RCTs met the admissibility criteria for evidence, and were divided into low back and neck pain studies. These studies were further subdivided for meta-analysis based upon whether the patient population was acute, chronic, or mixed.

For acute LBP there was moderate evidence that manipulation provided more short-term pain relief than mobilization and detuned diathermy, and limited evidence of faster recovery than a

commonly used physical therapy treatment strategy. For chronic LBP there was moderate evidence that manipulation had an effect similar to an efficacious prescription non-steroidal anti-inflammatory drug, manipulation/mobilization is effective in the short term when compared with placebo and general practitioner care, and in the long term compared to physical therapy. There is limited to moderate evidence that manipulation is better than physical therapy and home back exercise in both the short and long term. There is limited evidence that manipulation is superior to sham manipulation in the short term and superior to chemonucleolysis for disc herniation in the short term. However, there is also limited evidence that mobilization is inferior to back exercise after disc herniation surgery. For a mix of acute and chronic LBP manipulation/mobilization seems to provide either similar or better pain outcomes in the short and long term when compared with placebo and with other treatments, such as McKenzie therapy, medical care, management by physical therapists, soft tissue treatment and back school.

Bronfort et al recently (Jan 2008) published an updated systematic review of their previous work, which included references that were searched through the year 2006 (34). The previous systematic review published in 2004 only covered the literature through the year 2002. In this most recent review, Bronfort et al performed an online search for RCTs evaluating the therapeutic efficacy of spinal manipulative therapy (SMT) or mobilization (MOB) for chronic LBP that was performed using the same strategy as their original systematic review (7). Additionally, citation tracking of references in relevant publications was used, including the non-indexed chiropractic, osteopathic, physical therapy, and medical journals. Abstracts from proceedings and unpublished trials were not included. To be included in this review, each study was required to have greater than or equal to 10 subjects receiving SMT or MOB and main

outcome measures had to be patient oriented (e.g., pain, global improvement, low-back disability, recovery time, work loss, medication use, and functional health status).

The search strategy identified 42 studies assessing SMT/MOB for CLBP, eight more than the previous systematic review. The studies were too dissimilar in terms of patient characteristics, outcome measures, time points, and type of treatment comparisons to allow for statistical pooling. These authors concluded that the literature provided moderate evidence for several conclusions regarding SMT and MOB for chronic LBP. In terms of patient-rated pain, SMT with strengthening exercise is similar in effect to prescription non-steroidal anti-inflammatory drugs with exercise in both the short term and long term. There is also moderate evidence that SMT/MOB is superior to usual medical care and placebo for patient improvement. High-dose SMT is superior to low-dose SMT for pain in the very short term and similar in the short term. SMT is superior to chemonucleolysis, medication, and acupuncture; and MOB is inferior to exercise for disc herniation. The evidence is inconclusive as to whether SMT is superior to sham SMT for pain in the short term, and whether MOB is similar in effect to exercise for pain in both the short term and long term.

2.3 DEVELOPMENT OF GUIDELINES FOR MANIPULATION

The first attempt to create some practice guidelines for the use of spinal manipulation came from a national chiropractic consensus conference that convened at the Mercy Conference Center in Burlingame, California in early 1992 (36). This consensus conference was commissioned by the Congress of Chiropractic State Associations to systematically review all of

the procedures in common use by chiropractors, including spinal manipulation, and make recommendations and ratings as to their effectiveness and safety based upon the evidence. The driving force for creation of a set of internal chiropractic practice guidelines was the establishment of the Agency for Health Care Policy and Research (AHCPR) in late 1989 by the U.S. federal government. The AHCPR was created out of the perceived need by the U.S. government for national guidelines on each health care specialty, with the message that either the various health care professions develop their own guidelines or have third parties impose them.

The panel members at the Mercy Conference used a six-tier rating system to evaluate clinical procedures; approved, established, promising, equivocal, investigational, doubtful and inappropriate (see Table 1 below).

Table 1: Mercy Conference Chiropractic Procedure Ratings (36)

Established: Accepted as appropriate by the practicing chiropractic community for the given indication in the specified patient population.	**Investigational:** Evidence is insufficient to determine appropriateness. Further study is warranted. Use for a given indication in a specified patient population should be confined to research protocols. As more experience and evidence accumulates, this rating will change.
Promising: Given current knowledge, this appears to be appropriate for the given indication in the specified patient population. As more experience and long-term follow-up accumulated, this interim rating will change. This connotes provisional acceptance, but permits a greater role for the current level of clinical use.	**Doubtful:** Given current knowledge, this appears to be inappropriate for the given indication in the specified patient population. As more experience and long-term follow-up are accumulated, this interim rating will change.
Equivocal: Current knowledge exists to support a given indication in a specified patient population, though value can neither be confirmed nor denied. As more evidence and experience accumulates this rating will change. Expert opinion recognizes a need for caution in general application.	**Inappropriate:** Regarded by the practicing chiropractic community as unacceptable for the given indication in the specified patient population.

Three Classes (I, II, III) of quality of evidence and five Types (A-E) of strength of recommendation ratings were also established (see Tables 2 and 3 below). Class I evidence consisted of RCTs, Class II consisted of well-designed observational studies, and Class III consisted of expert opinion, descriptive studies, and case reports. Type A recommendations

indicated the strongest positive recommendation based upon Class I or overwhelming Class II evidence, and Type B recommendations were given based upon Class II evidence only.

Table 2: Mercy Conference Quality of Evidence Ratings (36)

Class I: Evidence provided by one or more well-designed controlled clinical trials: or will-designed experimental studies that address reliability, validity, positive predictive value, discriminability, sensitivity, and specificity.
Class II: Evidence provided by one or more well-designed controlled 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, and specificity; and published in refereed journals.
Class III: Evidence provided by expert opinion, descriptive studies, or case reports.

Table 3: Mercy Conference Strength of Recommendation Ratings (36)

Type A: Strong positive recommendation. Based on Class I evidence or overwhelming Class II evidence when circumstances preclude randomized clinical trials.
Type B: Positive recommendation based on Class II evidence.
Type C: Strong positive recommendation based on strong consensus of Class III evidence.
Type D: Negative recommendation based on inconclusive or conflicting Class II evidence.
Type E: Negative recommendation based on evidence of ineffectiveness or lack of efficacy based on Class I or Class II evidence.

The Mercy Conference expert panel made an important distinction during its process of rating the effectiveness and safety of the various chiropractic clinical procedures. They clearly recognized that many "technique systems" existed in the profession that consisted of various combinations of analytic and treatment components. The panel members chose to rate each treatment procedure on its own merits, separating them from the "systems" under which they were practiced. Two categories of treatment procedures that were distinguished in this process included manual manipulative techniques that utilized high velocity low amplitude (HVLA) thrusts and mechanical force techniques that utilized spring loaded instruments giving a mechanical low amplitude thrust, such as the Activator Instrument (AI).

The proceedings of this consensus conference were published in 1993 as the Guidelines for Chiropractic Quality Assurance and Practice Parameters (36), with the panel of 35 consultants publishing ratings based upon the quality of evidence for each of the treatment procedures. HVLA spinal manipulation was given a rating of "Established" for the care of patients with mechanical low-back problems and other musculoskeletal conditions based upon Class I and strong Class II evidence, and mechanical force manipulation (such as Activator) was given a rating of "Promising to Established" based upon Class I, II, III evidence.

As noted above, concurrent with the chiropractic profession's internal systematic review and guideline development, the U.S. Congress had established the Agency for Health Care Policy and Research (AHCPR) in 1989 to enhance the quality, appropriateness, and effectiveness of various health care services for conditions that had important public health concerns. The AHCPR, was created as one of eight agencies of the U.S. Public Health Service within the Department of Health and Human Services, and by 1992 had selected a number of topics upon which to create systematic reviews and publish guidelines to assist clinicians and patients about the appropriate decisions for health care options regarding these specific clinical circumstances. One of these topics was acute low back pain, due to its importance as a public health concern.

In 1994, the Agency for Health Care Policy and Research (AHCPR) published clinical practice guidelines for the management of acute low back problems in adults based upon a careful review and meta-analysis of the literature on various treatments for low back pain (1). These guidelines suggested that the first step in the management of acute low back pain (LBP) is to rule out any serious underlying medical conditions by looking for clinical "red flags"; i.e. signs and

symptoms suggestive of spinal or non-spinal pathology. In the ACHPR model, patients with LBP are categorized into 3 categories: potentially serious conditions suggested by presence of red flags, sciatica suggested by positive nerve root tension signs, or non-specific LBP suggested by a lack of red flags or nerve root tension.

The ACHPR guidelines (1) indicate a yearly prevalence of back pain in working age adults at 50 percent, with 15-20 percent seeking medical attention, and stating that LBP is the most common cause of disability in persons under the age of 45 years. The guidelines strongly suggest that 90 percent of patients with acute LBP spontaneously recover within four weeks, and therefore the focus of care should be away from pain control exclusively, and more toward helping patients improve activity tolerance. To this end, the guidelines recommend that the initial management of acute LBP should provide assurance of a rapid recovery, to provide comfort by means of symptom control methods, and to recommend activity modifications.

The two recommended methods (Table 4) for symptom control according to ACHPR guidelines are non-steroidal analgesic medications and/or spinal manipulation. Optional methods (Table 5) include muscle relaxants, opioids, physical agents and modalities, shoe insoles, and rest from activity. The ACHPR guidelines state that manipulation is safe and effective for patients in the first month of acute LBP without radiculopathy, and that manipulation is probably safe for symptoms lasting greater than one month, but its efficacy is unproven.

Table 4: Agency for Health Care Policy Research (AHCPR) *recommended* treatments for non-specific low back pain (1).

Non-prescription analgesics:	**Prescribed pharmaceutical methods:**	**Prescribed physical methods:**
Acetaminophen (safest) NSAIDs (aspirin, ibuprofen)	Other NSAIDs	Spinal manipulation

Table 5: AHCPR *optional* treatments for non-specific low back pain (1)

Prescribed pharmaceutical methods:	**Prescribed physical methods:**
Muscle relaxants	Physical agents and modalities
Opioids	Shoe insoles
	A few days' rest

The AHCPR guidelines marked the first time that a non-official government health agency recommended manipulation as a primary treatment for patients with acute low back pain. However, unlike the Mercy Conference which rated individual manipulative procedures, the AHCPR guidelines did not differentiate between the various types of spinal manipulation nor did they differentiate between subsets of "non-specific LBP". However, the AHCPR guidelines did make the primary care physician community aware of the fact that spinal manipulation was an appropriate and effective treatment option for control acute low back symptoms.

In 1996 the Canadian Chiropractic Association published the "Clinical Guidelines for Chiropractic Practice in Canada" (37). The Canadian guidelines are commonly referred to as the Glenerin Guidelines, since the consensus conference was held 1993 at the Glenerin Inn in Mississauga, Ontario. 35 commission members were brought together in round-table format and guideline recommendations were debated and voted upon in plenary sessions, which in many ways paralleled the Mercy Conference Guidelines produced in the U.S. a few years previously. The Glenerin committee members used exactly the same procedure and strength of recommendation ratings and the same quality of evidence levels, as the Mercy committee members.

The Glenerin consensus panel reached similar conclusions regarding the rating of side posture HVLA and mechanical manipulation devices (Activator, etc) as did the Mercy Conference panel members; HVLA manipulation was rated as " Established" based upon Class 1, II, and III

evidence, and mechanical manipulation devices were rated as "Promising" based upon Class II and III evidence. They also concluded that while there was a sound body of scientific evidence on the efficacy of HVLA type spinal manipulation for various conditions, there was a paucity of literature on the comparative effectiveness of other specific types of manipulative techniques such as mechanical manipulation devices.

The AHCPR guidelines made the practice of manipulation appear more acceptable in the mainstream health care delivery system, but failed to address two fundamentally important issues. The first issue regards the differences (if any) in effectiveness and safety of the various types of manipulative procedures. Within the chiropractic profession alone there are thought to be between 100 to 200 technique systems that use at least several dozen different manipulative methods (13, 36). It is unlikely that all of these manipulative procedures are equally effective.

The second issue is the lack of differentiation between potential subsets of low back disorders. The terms "mechanical low back pain" or "non-specific low back pain" are often used in a general manner to lump all patients with certain symptoms into one all-inclusive category. The AHCPR clinical practice guidelines suggest that up to 85% of acute low back pain patients seen by primary care physicians fit into this vague diagnostic category of "non-specific low back pain" (1). There are a number of causes of "mechanical low back pain" including myofascial pain syndromes, spondylolisthesis, lumbar facet joint dysfunction, intervertebral disc lesions, sacroiliac joint dysfunction, and others. It is likely that different manipulative procedures would have different effects on each of these subsets of the general category "non-specific LBP".

In some respects the Mercy and Glenerin consensus conferences addressed these two issues, but did not resolve them. Both panels separated chiropractic manipulation procedures into several categories, including HVLA manipulation and mechanical manipulation devices, and rated these methods separately. However, there was no attempt to match up subsets of low back diagnoses with each of the manipulation methods and make clinical recommendations based upon the literature. In an attempt to help bridge this knowledge gap about the effectiveness of distinct manipulative procedures, a consensus panel of chiropractic experts was created in 1995 to review the spinal manipulation literature for low back pain, with the specific intent to search for articles with an actual description of the manipulative procedure used and a descriptive diagnosis other than "non-specific low back pain".

This expert panel combined the results of a qualitative review of the manipulation literature with their clinical experience/opinion, and developed ratings of each manipulative procedure for each different diagnosis, using consensus methodology. The panel members subdivided the general category of "spinal manipulation" into eight sub-categories; high velocity low amplitude (HVLA) side posture manipulation, HVLA manipulation with drop table mechanisms, pelvic blocking procedures, mechanical instrument manipulation, mobilization techniques, distraction table techniques, prone HVLA manipulation (without drop table) procedures, upper cervical manipulation for LBP, non-thrust low force/reflex techniques, lower extremity manipulation for LBP. "Low back pain" as a diagnosis was also divided into eight subcategories, four patho-anatomic and four symptom-based subcategories; herniated disk, spondylolisthesis, facet joint dysfunction (subluxation), sacroiliac joint dysfunction, low back pain, low back pain plus buttock-thigh pain, low back and leg pain plus neurologic deficits, and leg pain only.

Thus an 8 x 8 matrix was formed, with 8 columns for the subcategories of LBP diagnoses matched with 8 rows for the different types of spinal manipulation. The panel members rated each type of manipulation for its clinical efficacy relative to each of the subcategories of LBP. The results of this consensus panel's ratings were published as two articles in 2001 (38, 39) with side posture HVLA manipulation receiving the highest rating for most of the diagnostic subcategories based upon the fact that most RCTs and observational studies in the manipulation literature used some variation of the HVLA manipulation technique. Mechanical force manipulation (e.g. Activator) devices received a very low rating for all of the diagnostic subcategories, due to the absence of any well designed RCTs or any observational studies with large effect size in the literature at that time.

The panel commented that although mechanical manipulation devices did not have an evidence base comparable to HVLA manipulation, lack of evidence should not be construed as evidence of lack of effectiveness. The panel concluded that many chiropractic manipulative procedures were lacking in their respective evidence bases, and that a research priority for the future should be the development of clinical trials to sort out the issue of differences in effectiveness (if any) between the various manipulative methods.

The debate about mechanical manipulation devices versus standard manual HVLA manipulation reached a pinnacle when the Canadian province of Saskatchewan prohibited by law the use of mechanical adjusting (manipulation) devices (MAD). In 2001, the Chiropractors' Association of Saskatchewan (CAS) established a scientific committee to review the literature on MAD and

make recommendations on its efficacy, safety, and uses, based upon consensus methods. A comprehensive literature search was performed utilizing keywords and searching MEDLINE, MANTIS, CINAHL, and the INDEX TO CHIROPRACTIC LITERATURE. 55 articles were found pertaining to the Activator instrument, of which 13 were eliminated from the final study. The panel published their results in a two-part series of articles in 2004 (19, 20).

Of the 42 articles reviewed by the Panel, only 21 were related to clinical efficacy, of which five studies were identified as Class 1 evidence and three studies as Class 2 evidence. Only one RCT related to low back pain was found in the literature, comparing Activator with side posture HVLA manipulation with a total sample size of 30 patients and a conclusion of no statistical difference between the two treatments in their ability to reduce acute low back pain(18). One case series study of 10 patients with low back pain found a statistically significant difference in VAS scores and Oswestry Index scores after receiving Activator treatment, which remained stable at a one-year follow-up. Two cases reports, both dealing with lumbar disc herniation treating with Activator methods, provided empirical observation and opinion that Activator treatment may provide an alternative method of treating low back pain when manual side posture manipulation may be contraindicated.

The committee members reached consensus that, while all of the Activator studies were 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 manipulation in producing clinical benefit and biological change. Their concluding remarks were that more

research, particularly a larger scale randomized controlled trial, would be helpful in determining efficacy to a further degree.

2.4 CLASSIFICATION SYSTEMS FOR MANIPULATION

The earliest effort in the medical literature to create a classification system for low back pain was published in 1987 as the results of large Consensus Conference in Canada called the Quebec Task Force (QTF) (40). The goal of the QTF was to create a diagnostic classification that could help to stratify patients with low back pain into different subsets for purposes of providing surgical and nonsurgical treatment options, establish a prognosis, and conduct clinical research. The QTF classification system provides 11 categories based upon a combination of symptoms and pathoanatomy, as listed in Table 6 below:

Table 6: Quebec Task Force Diagnostic Classification System (40)

1. Low back pain without radiation to lower extremity.
2. Low back pain with radiation to proximal lower extremity (above knee).
3. Low back pain with radiation to distal lower extremity (below knee).
4. Low back pain with radiation to lower extremity and neurological findings.
5. Spinal fracture.
6. Spinal nerve root compression
7. Spinal stenosis.
8, 9. Post surgical status.

10. Chronic pain syndrome.
11. Other diagnoses

The QTF classification system appears to have reasonable predictive ability to discern between low back pain patients who are likely to require surgery, have higher baseline pain and functional impairments, and will have difficulty with return to work (41, 42). However, for clinicians such as chiropractors and physical therapists who perform only nonsurgical treatment, the QTF categories are not as useful in discerning between subsets of LBP that would respond to manipulation versus rehabilitative exercises, for example. The QTF categories 1, 2, and 3 would all likely be classified as "non-specific LBP" according to the AHCPR guidelines.

Within the physical therapy literature, this similar issue has been addressed with respect to "non-specific LBP" being too broad of a diagnostic term for scientific research of the clinical efficacy of various physical therapy interventions. One of the earliest attempts to create subsets of LBP was published in a textbook by McKenzie in 1981, in which he proposed three basic categories of back pain: postural syndrome, dysfunction, and derangement as listed in Table 7 below (43).

Table 7: McKenzie based Classification System for non-specific LBP

1. **Postural Syndrome**: Patient has pain as a result of prolonged poor posture causing strain on soft tissues.	**Treatment**: Advice and education about maintaining good posture during sitting, standing, and other ADLs.
2. **Dysfunction Syndrome:** Patient has pain due to chronically shortened periarticular tissues or adhesions.	**Treatment:** Specific exercises and mobilizations to stretch these tissues and restore normal mobility.
3. **Derangement Syndrome:** Patient exhibits symptoms suggestive of disc derangement.	**Treatment:** Specific exercises, mobilizations, manipulations, or surgery; based upon symptomatic responses of patient.

For each category of LBP, the proposed clinical intervention was different; i.e. for postural syndrome the patient was treated with home exercises and postural advice whereas the patient with lumbar derangement was treated with specific mobilization techniques. McKenzie further divided the derangement category into seven subsets, based upon the patient's symptoms and physical examination findings, especially the response to flexion versus extension movements.

Delitto et al took a group of LBP patients referred for physical therapy whose examination findings indicated they were likely to respond best to an extension-mobilization approach, and randomized them into two treatment groups (44). The experimental and comparison group subjects were treated with either mobilization and extension (a treatment matched to the category) or a flexion exercise regimen (an unmatched treatment). The subjects' rate of improvement, as indicated by the Oswestry questionnaire scores, was dependent on the treatment group to which they were assigned. Subjects treated with extension and mobilization positively responded at a faster rate than did those treated with a flexion-oriented program. This study illustrated that classification of selected patients with non-specific LBP into a treatment-based category of extension and mobilization and subsequently treating the patients accordingly with specified interventions can be an effective approach.

This simple study was one of the first published trials to distinguish a selected subset of non-specific LBP of patients (extension responders) and match the treatment approach to the functional diagnostic category. In 1995, Delitto and Erhard proposed a more comprehensive classification system for non-specific LBP that created four distinct categories which were treatment-based and would specifically guide conservative management (45). The four

treatment-based categories are immobilization, manipulation/ mobilization, specific exercise, and traction as listed in Table 8 below:

Table 8: Treatment-based Classification System for non-specific LBP (45)

1. **Immobilization**: Patient shows signs of clinical instability, such as spondylolisthesis, ligamentous laxity, or hypermobility.
2. **Manipulation/mobilization:** Patient has acute LBP without nerve root tension signs.
3. **Specific exercise:** Patient shows a directional preference for flexion/extension based exercises which centralize their symptoms.
4. **Traction:** Patient has acute LBP with antalgic posture and/or significant distal leg symptoms.

Fritz, Delitto, and Erhard published the results of a RCT that explored the effectiveness of using this four-tiered classification system with acute low back pain patients compared to a strategy of minimal intervention (46). Clinical practice guidelines such as those published by the AHCPR recommend minimal intervention during the first few weeks of acute LBP, but fail to identify subsets of patients who are likely to respond to different interventions. The Fritz et al study involved 78 subjects with work-related low back pain of less than three weeks duration who were randomized to receive therapy based on the classification system in Table 7 that attempts to match patients to specific interventions, or therapy based on the AHCPR guideline of minimal intervention. The results showed that subjects receiving classification-based therapy showed greater change on the Oswestry scale and the SF-36 physical component scale after four weeks. Patient satisfaction was greater and return to full-duty work status was more likely after four weeks in the classification-based group compared to the minimal intervention group.

Flynn et al published a prospective cohort study in which 71 patients with non-radicular low back pain underwent a standardized examination and then a standardized spinal manipulation

treatment program, without any a priori classification into subsets (47). Success with treatment was determined using percent change in disability scores over three sessions and served as the reference standard for determining the accuracy of examination variables. Of the 71 patients who participated, 32 had success with the manipulation intervention. Examination variables were first analyzed for univariate accuracy in predicting success and then combined into a multivariate clinical prediction rule.

A clinical prediction rule with five variables was identified; symptom duration, fear-avoidance beliefs, lumbar hypomobility, hip internal rotation range of motion, and no symptoms distal to the knee. The presence of four of five of these variables (positive likelihood ratio = 24.38) increased the probability of success with manipulation from 45% to 95%. The authors suggested that it appeared possible that patients with low back pain likely to respond to manipulation could be accurately identified before treatment, using this clinical prediction rule. Indeed, a subsequent clinical trial by Childs et al confirmed that patients who were treated with manipulation and were expected to be "manipulation responders" by using this clinical prediction rule responded better than patients whose treatment was not matched to the rule (48).

Yet another attempt to create a diagnostic classification system for non-specific low back pain was published in 2003 by Petersen et al who propose that division of the large heterogeneous group of patients who are labeled with "non-specific low back pain" (85% of all LBP patients seen in primary care medicine) into diagnostic homogeneous subcategories is thought to improve treatment outcomes (49). They propose a classification system that has a patho-anatomic orientation, rather than the treatment-based approach proposed by Delitto et al. Their proposed

classification system is made up of 12 categories that refer to the known possible pain-producing structures in the low back, such as discs, zygapophysial joints, sacroiliac joints, nerves, and muscles, as listed in Table 9. They also include a category called "abnormal pain syndromes" that includes patients that are characterized by abnormal illness behavior.

Table 9: Pathoanatomic and clinical classification system for non-specific LBP (49)

1. Disc syndrome
2. Adherent nerve root
3. Nerve root entrapment
4. Nerve root compression
5. Spinal stenosis
6. Zygapophysial joint syndrome
7. Postural syndrome
8. Sacroiliac joint
9. Dysfunction syndrome
10. Myofascial pain
11. Adverse neural tension
12. Abnormal pain syndromes
13. Inconclusive

The classification system proposed by Petersen et al is preliminary in nature and has not been subjected to inter-examiner reliability studies or used in any clinical trials to determine its clinical predictive value. However, its clinical usefulness may yet be shown in future clinical trials. Clearly the intent of this classification system is similar to all the previously discussed

systems; i.e. an attempt to subdivide the broad category of non-specific LBP into clinically manageable subsets in which the treatment approach is matched to the pathoanatomic or functional diagnosis.

The most recent attempt at development of a classification system for spinal pain was published by Murphy and Hurwitz in 2008 (50) which proposes a diagnosis-based clinical decision rule (DBCDR) for the diagnosis and non-surgical management of back pain. The DBCDR is based on what the authors refer to as the three essential questions of diagnosis:

- Are the patient's symptoms reflective of a visceral disorder or a serious/potentially life-threatening illness?
- From where is the patient's pain arising?
- What has gone wrong with this person as a whole that would cause the pain experience to develop and persist?

The first question of the DBCDR is the process of ruling out the "red flags" of serious pathology in the patient presenting with low back pain. The second question seeks to determine which the source of the patient's pain, whether it be from the intervertebral disc, lumbar facet or sacroiliac joints, the muscles or tendons, or other somatic or visceral tissues.

Under the second question above, the authors suggest four signs are of great importance when determining the source of back pain:

- Centralization signs: the observation that distal symptoms in the lower extremity or buttock "centralize" toward the spine.
- Neurodynamic signs: reproduction of symptoms with stress applied to neural structures.

- Segmental pain provocation signs: manual pressure applied to selective lumbar facet and sacroiliac joints to test for pain provocation.

- Muscle palpation signs: reproduction of pain upon direct palpation of myofascial tissues.

Under question number three above, the authors note that it is important to identify and manage those factors that place the acute or sub-acute spinal pain patient at risk of developing ongoing problems or, in the case of the chronic or recurrent spinal pain patient, that contribute to the perpetuation of pain and dysfunction. Those factors would include:

- Dynamic instability: impaired motor control
- Central pain hypersensitivity
- Oculomotor dysfunction
- Fear and catastrophizing
- Depression
- Passive coping

Like many of the other classification systems published in the literature, this recently proposed DBCDR has not been validated through randomized clinical trials. However, all of these proposed systems tend to point in the same direction, that low back pain is not one grandiose syndrome with one "magic bullet" solution. Rather, it appears that the causes of back pain and the processes which perpetuate it are multi-factorial and require a broader approach to diagnosis and treatment.

2.5 SUMMARY OF MANIPULATION/LOW BACK PAIN LITERATURE

Numerous systematic reviews of the literature show a modest amount of support for the intervention of spinal manipulation for acute non-specific low back pain (LBP). The large number of favorable manipulation clinical trials in literature was taken into account by the AHCPR in publication of their Clinical Guidelines, which recommended spinal manipulation as one of the chief interventions for acute LBP patients who did not exhibit signs of pathology. The AHCPR guidelines did not differentiate between subsets of LBP, and lumped 85% of acute LBP patients into one large group of "non-specific LBP".

Several attempts have been made to subdivide the broad category of "non-specific LBP" into various subsets, and several of these proposed classification systems were reviewed in this chapter previously. The common denominator of these proposed systems is the idea that patients are more likely to respond to treatment methods that are matched to their specific type of LBP. In the most extreme example, patients should be separated into those who needed surgical interventions from those who would respond to conservative non-surgical interventions. In the chiropractic and physical therapy clinics, a more salient question might be how to separate "non-specific LBP" patients into those who will respond better to manipulation and mobilization techniques from those who would respond better to traction methods or rehabilitative exercise protocols.

There is now some literature supporting the notion of a clinical predication rule for LBP patients who are likely to be manipulation responders. However, most of the clinical trials reported in the literature describe only one type of manipulation; a specific variation of high velocity low

amplitude (HVLA) thrust technique rendered with the patient in a side-lying posture. Very few clinical trials have been published that compare different types of manipulation methods on the same cohort of LBP patients. An important question within the chiropractic profession is whether or not mechanical manipulation devices are as clinically effective as manual HVLA manipulation methods. This dissertation project will explore this yet unanswered question with a prospective cohort design that tracks the clinical outcomes of LBP patients with the same inclusion criteria treated by these two different types of manipulation methods.

3.0 METHODS

3.1 EXPERIMENTAL DESIGN

The experimental design of this study consisted of two phases; (1) inter-examiner reliability studies for the prone leg length analysis and prone springing palpation over the lumbar facets and sacroiliac joints and (2) an observational prospective cohort study, with two types of manipulation defining the cohorts; mechanical assisted manipulation (Activator) and manual manipulation (Diversified). For both of these studies, research volunteers with acute and sub-acute low back pain (LBP) were recruited from the private practices of three chiropractic offices in the Greater Pittsburgh metropolitan region. One office had two Activator Methods proficiency certified chiropractors who exclusively used the Activator Instrument as the sole manipulative treatment for LBP. The other two offices utilized standard side posture manipulation (Diversified) for treatment of LBP, and consisted of a solo practitioner rendering all treatments at each respective office.

The primary research question was to explore the differences in treatment effect on LBP, if any, between mechanical assisted manipulation with the Activator Instrument and standard side posture Diversified lumbar manipulation. The Activator Method protocol requires the use of the prone leg length analysis to determine where to position the Activator Instrument and apply the

mechanical impulse treatment. The Diversified Method relies on springing or motion palpation to determine at which level of the lumbar spine or sacroiliac joints to apply the manual thrust technique. Since these physical examination procedures were utilized on every visit in these respective chiropractic offices, it was feasible to perform inter-examiner reliability studies on these clinical testing procedures as a secondary research question. The reliability studies were performed prior to initiation of the cohort study, and separate IRB approval was obtained.

3.2 RESEARCH PARTICIPANT CHARACTERISTICS

3.2.1 Inclusion Criteria

(1) Onset of acute low back pain symptoms within 0 – 12 weeks from initial baseline visit. Acute and sub-acute back pain patients are more responsive to manipulation intervention than chronic patients.

(2) Age 18 years or older. This is the age group most often reported in spinal manipulation clinical trials and the peak incidence of low back pain in adults.

(3) Oswestry Low Back Pain Disability Index score of at least 20 points, which represents mild to moderate self-perceived disability in activities of daily living. This minimum was chosen to prevent a floor effect, which might occur with subjects having minimal disability.

(4) Numeric Pain Rating Score of at least four points out of maximum scale of ten points, with the intent of avoiding a similar floor effect with patients reporting minimal pain.

3.2.2 Exclusion Criteria

(1) Age less than 18 years old. This represents a pediatric population that has not been carefully studied in the spinal manipulation literature.

(2) Low back pain onset greater than 12 weeks prior to baseline evaluation. Chronic back pain patients are a different cohort than acute and sub-acute patients, and may respond less favorably to manipulation.

(3) Prior history of lumbar spine surgery, unstable spondylolisthesis, spinal stenosis, or scoliosis greater than 20 degrees.

(4) Signs or symptoms suggestive of severe neurological deficit such as lower extremity motor weakness, paresthesia distal to the knee, strong nerve root tension signs, bowel or bladder dysfunction.

(5) Any female patient who is pregnant and has trouble lying prone.

(6) History of metastatic cancer, osteoporosis, long-term corticosteroid use, or any other medical condition that would contraindicate thrust-type spinal manipulative treatment.

(7) Patients who have received physical therapy, chiropractic, or any other manual therapy for low back pain within the past 3 months.

(8) Oswestry score greater than 70 and/or Numerical Pain Score greater than 8/10. This cohort of patients typically has severe symptoms that may indicate underlying serious pathology, which may not be suitable for manipulative therapy.

(9) Oswestry score less than 20 and/or Numerical Pain Score less than 4/10. This cohort of patients typically has a low level of pain and disability that might resolve through natural history, without any clinical intervention.

3.2.3 Modifications of Inclusion/Exclusion Criteria

Two modifications were made to the above inclusion and exclusion criteria after the cohort study commenced at three private chiropractic clinics in the Metropolitan Pittsburgh region. Both of these modifications were based upon practical concerns, and were approved by the University of Pittsburgh's Institutional Review Board (IRB).

Originally the research design had excluded patients who were older than 65 years of age. As the study progressed, it became apparent that many senior citizens were being excluded from participation in the study that were in good physical health and did not have any other specific contraindications to spinal manipulative therapy. A modification was made to remove the upper age restriction from the exclusion criteria, based upon the fact that older patients would still be screened for the presence of specific contraindications to spinal manipulation such as spinal stenosis, history of osteoporosis, prolonged use of steroid medication, etc. Senior citizens who had any of these conditions that would be considered contraindications would be excluded from participation based upon medical reasons, and not merely because of their age.

Another modification was made regarding the amount of time from onset of low back symptoms. Originally the research design called for inclusion of only those patients whose back pain commenced within two weeks of baseline, which corresponded with one of the key predictor variables in a recently published clinical prediction rule for manipulation success (51). However, this inclusion criterion became a stumbling block to recruitment of research participants in a private practice setting. Many potential research participants were found to have an acute onset

of back pain within the previous month or two, but had to be turned away from our study based upon the strict inclusion criterion of two weeks onset.

3.3 RECRUITMENT PROCEDURES

Research subjects were recruited from three local chiropractic offices in the Pittsburgh area by two general strategies; staff inquiry procedures and office brochures/posters. Patients who had been experiencing acute or sub-acute low back pain that made an appointment by phone or in person were given an informational pamphlet by the office staff or attending clinician with information about this research study and asked if they would consider participating in it. In addition, each chiropractic office was given posters to display in a prominent location in their waiting rooms and treatment areas, along with tri-fold pamphlets describing the research study.

Patients who expressed interest volunteering for the study were given a more detailed explanation of the study by either the Principal Investigator (PI) or support staff persons who were trained by the PI. This detailed explanation included a brief screening to make an assessment about whether the patient appeared eligible to participate in the study by meeting the inclusion/exclusion criteria. Checklists were provided that outlined all of the inclusion/exclusion criteria for screening purposes, which were used by the chiropractors and their staff members to screen potential research subjects. The PI held several training sessions with the participating chiropractors and their respective office staff personnel at all three chiropractic offices, to make sure consistency in presentation of the research information was preserved. These meetings also

included training in the informed consent procedure, privacy protection, and research integrity. The entire informed consent document was reviewed with all staff persons.

Research volunteers who passed the screening process for inclusion/exclusion criteria were taken to a private room where the informed consent document was thoroughly explained, and written consent to participate in this study was obtained before any clinical examination or treatment was provided. The informed consent document and the research protocol were approved by the Institutional Review Board at the University of Pittsburgh. We emphasized the volunteer nature of this study and told all participants that they were free to cancel their participation at any time if that was their choice. The informed consent process typically occurred on the same day of the initial baseline examination. Low back pain is the most common reason for patients to seek examination and treatment at a chiropractic office, and therefore it was anticipated that the pool of potentially eligible patients in these chiropractic offices would be rather large.

3.4 PRELIMINARY RELIABILITY STUDIES

3.4.1 Prone leg length analysis

The two chiropractors who exclusively used Activator Methods in their office examined a series of 45 patients with a history of acute or sub-acute LBP for leg length differences, according to the Activator Methods protocol (16). Each patient was lowered from the upright standing position to the prone position on an Activator table, and the following protocol was followed by each chiropractor in sequence with each patient:

1) Patient was lowered from standing position to prone position passively on an Activator table. Clinician inspects patient position on table and visually observes leg length by inspection of the feet.

2) Clinician grabs feet of patient and gently pushes into plantar-flexion and external rotation, until mild resistance is felt.

3) The leg that appears shortened is recorded as the "short leg".

4) Clinician again grabs feet of patient and gently pushes into plantar-flexion and lifts the legs by raising the feet and flexing the knees to approximately 90°.

5) The short leg side is observed for one of three possible outcomes;

(a) Short leg appears to have lengthened.

(b) Short leg appears to have shortened further, relative to the previous observation.

(c) No change in leg length occurs.

Patients were instructed not to communicate with the chiropractors during this study, and each clinician was blinded to the outcome of the other clinician. The principal investigator was the impartial observer, and recorded all of the examination findings and monitored the process of blinded examinations by the two clinicians. The PI also recorded the side of low back pain reported by the patient verbally after both clinicians had completed their leg length tests. This

data was used to see if there was any correlation between the side of reported short leg and the painful side as reported by the patient. All of the data regarding each patient was recorded by the PI on a leg length analysis form (Appendix A).

Two one-hour training sessions and orientations were scheduled prior to the initiation of the study, in order that both clinicians were briefed on the examination form/questions, the leg length analysis protocol, and instructions on how to follow the same procedures with all patients. Both chiropractors practiced the protocol on each other with the PI observing them during these training sessions, which were designed to minimize measurement error due to inconsistencies in the examination process.

3.4.2 Facet joint and sacroiliac joint palpation

In another private chiropractic practice where Diversified Technique was chiefly utilized to treat low back pain, two clinicians were asked to manually palpate a series of 50 patients for lumbar facet joint and sacroiliac joint pain/segmental mobility dysfunction, and record their findings on an examination form. The patients were placed in the prone position on an examination table by the PI, who then marked the skin overlying each spinous process and right/left facet joints from L1 through L5. The sacrum was also marked over S2, and each sacroiliac joint right and left was marked by the PI.

These skin markings were performed in order to provide the examining clinicians with common reference points for their palpation. Each of the two clinicians was asked to enter the room separately, and apply posterior to anterior manual pressure over each of the sacroiliac joints, the

spinous process of each lumbar vertebra, and the right/left lumbar facets joints from L1 through L5. Three sets of springing palpation were performed on each patient, by each clinician who was blinded to other's clinical findings:

(1) The first palpation was for clinician-perceived segmental joint restriction, while the patients were instructed not to provide any verbal or non-verbal communication about pain. Facet joint mobility at each lumbar level bilaterally, and bilateral sacroiliac joint mobility were recorded as 0 or 1, to indicate normal motion and hypomobility respectively.

(2) The second palpation was applied over the same bony landmarks, this time while asking the patients to tell the examiner whenever the palpatory pressure was painful over a particular joint. Pain responses were recorded as "painful" = 1 or "non-painful" = 0.

(3) The third palpation was performed only on patients who reported pain provocation over one or more joints upon palpation. These patients were repositioned on the exam table such that their feet touched the floor, and then were asked to lift their legs a few centimeters off the floor to engage the erector spinae muscles. The clinicians then re-applied springing palpation over any previously painful joints and reported whether there was a change in the pain provocation pattern. This procedure has been described elsewhere as a test for lumbar segmental hypermobility known generally as the prone instability test (52).

The following protocol is a summary of the specific steps that were followed by each chiropractor in sequence with each patient:

(1a) Clinician placed the heel of his hand over the right sacroiliac joint and applied PA pressure and then repeated this procedure over the left sacroiliac joint. Their perceptions of joint mobility were recorded as 0="no restriction" or 1="segmental restriction" over each joint, while the patients were told not to verbalize any pain they were experiencing.

(1b) Clinicians contacted the spinous process of L5 and applied springing pressure in a posterior to anterior (PA) direction, and then repeated this manual springing pressure over each subsequent spinous process at L4, L3, L2, and L1. Their perceptions of joint mobility were recorded as 0 or 1 over each joint, while the patients were told not to verbalize any pain they were experiencing.

(2a) Same examination process as (1a) above, but this second manual assessment was used to determine the presence of any pain provocation responses at each SI joint as patients were now instructed to tell the examiner which joints were painful upon palpation. Responses were recorded by the PI as 0 = no pain produced upon palpation or 1 = pain produced upon palpation. The clinicians were not asked to report their perceptions of joint mobility on this second set of palpation.

(2b) Same procedure as (2a) above, but assessing patient response to pain at each segmental level. Record as 0 = no pain, or +1 = pain.

(3) If pain is provoked at any segments in 2a or 2b above, then repeat palpation at these levels with the patient repositioned with slight contraction of the back extensor muscles (per prone instability test). Record as 0 = no pain, or +1 = pain.

The PI recorded the segmental mobility and pain provocation responses from each clinician on each patient on a springing palpation analysis form (Appendix B). Two one-hour training sessions and orientations were scheduled prior to the initiation of the study, in order that both clinicians were briefed on the overall examination process, springing palpation protocols, and instructions on how to follow the same procedures with all patients. The two clinicians practiced their palpation skills on the PI, who gave feedback to each of them regarding the perceived amount of applied force and other aspects of the palpation process.

The training sessions were designed to minimize inconsistencies between the examiners with respect to the amount of force applied duration the palpation and variations in the palpation methodology.

3.5 PRIMARY RESEARCH: OBSERVATIONAL COHORT STUDY

3.5.1 Drug/Device Information

The only medical device used in this study was the Activator IV Instrument, FDA approval # K003185. Manufacturer: Activator Methods International Ltd., 2950 N 7th Street, Phoenix, AZ 85014. This instrument is a hand-held device that contains a spring-loaded

mechanism that delivers a mechanical impulse over bony prominences and is used by Activator Methods clinicians as an alternative to manual manipulation.

3.5.2 Overview of research design

This was an observational study that employed a prospective cohort design. The study looked at the clinical outcomes of two cohorts of low back pain (LBP) patients in three separate chiropractic clinics that utilize two different types of spinal manipulation methods. One clinic uses the Activator Instrument exclusively for treatment of LBP and the other two clinics use manual manipulation methods exclusively. The main outcome measures were pain and function, as measured by the Numeric Pain Rating Scale (NPRS) (53) and the Oswestry Low Back Pain Disability Index Questionnaire (ODI) (54) respectively. Samples of the NPRS and ODI forms are attached as Appendix C and D. Both of these outcome measures have been widely used in previously published low back pain clinical trials.

The NPRS utilized in this study had patients report three levels of pain on an 11 point scale ranging from 0 (no pain) to 10 (worst pain); 1) Pain level right now, (2) Worst pain level in past 24 hours, and (3) Best level of pain in past 24 hours. The primary outcome measure utilized in this study took the average of the "pain right now" and "worst pain in past 24 hours" scores.

The ODI form consists of a series of ten questions that each have six possible responses that are graded from 0 to 5 points, based upon the severity of self-perceived disability regarding each question. Therefore the total possible number of points is 10 x 5 = 50 points, which would

indicate crippling disability. It is customary to report ODI scores as a percentage, which is derived by dividing the number of total points by 50.

Approximately 95 patients with low back pain of 0-12 weeks duration and who met the other inclusion/exclusion criteria outlined previously were recruited to participate in this prospective cohort study over the course of a year and a half. The ideal goal was to recruit a grand total of 100 patients (50 in each cohort) with the actual final number meeting 95 % of this initial recruitment goal.

After giving informed consent, the two cohorts of patients were each followed as they received these two types of chiropractic manipulation. No experimental treatments or placebo groups were involved with this study, since this was an observational study of treatment-as-usual in these three private chiropractic clinics. In addition to the NPRS and ODI questionnaires which were the primary and secondary outcome measures, several additional patient self-reported questionnaires were administered to the patients in order to assess their level of self-perceived disability, fear avoidance beliefs, and self-reported change in clinical status.

These self-reported questionnaires included the Roland-Morris Questionnaire (Appendix E) (55), the Fear Avoidance Beliefs Questionnaire (Appendix F) (56), and Patient Global Index of Change form (Appendix G) (57). The Roland Morris Disability Questionnaire is similar to the ODI, and is a validated instrument for assessing self-reported disability from low back pain. The Fear Avoidance Beliefs Questionnaire is a validated instrument that consists of two sub-scales, one for assessing the level of self-perceived fear about normal activities of daily (Physical

activities subscale) and the other for assessing fear about work related activities (Work subscale). For the purposes of this study, only the physical activities subscale was utilized. The Patient Global Index of Change is a validated tool for assessing a patient's level of self-perceived change in their condition during treatment. It consists of a single sentence asking what level of change (worsening or improvement) has occurred during treatment, using a seven point Likert scale ranging from "much worse" to "much better", with "no change" as the middle point.

All of these forms were administered prior to the initiation of treatment at the baseline visit, and then subsequently at four weeks following the baseline visit. The NPRS, ODI, and PGIC were also administered at one and twelve weeks. Two additional forms were given to the patients only at the baseline visit, a treatment credibility/expectation form (Appendix H) and a socio-demographic questionnaire (Appendix I). Table 10 below gives a visual outline of all these questionnaires and the timetable of their administration during the study.

Table 10: Timetable of questionnaire administration

	Baseline	Week 1	Week 4	12 weeks
Numeric pain rating scale (NPRS)	X	X	X	X
Oswestry (ODI)	X	X	X	X
Fear avoidance beliefs questionnaire	X		X	
Roland-Morris	X		X	
Patient global index of change (PGIC)		X	X	X
Treatment credibility/expectation	X			
Sociodemographic data	X			

At the initial baseline visit, the treating chiropractor ascertained that consenting patients were candidates for spinal manipulative therapy, did not have any red flags of serious pathology, and met the general inclusion criteria for the study. The above self-report questionnaires were administered prior to the physical examination, at the same office visit. The initial examination

also included lumbar ranges of motion, prone leg length analysis (Activator office) and spinal palpation for facet and sacroiliac joint mobility and pain provocation (Diversified office).

After the initial examination, the patients received the normal mode of chiropractic care as determined by their attending chiropractor at each of the three offices participating in this study. It was not feasible for the PI to monitor all of the treatments at each private clinic and therefore this study relied on the veracity of the written medical records. The participating chiropractors were instructed to treat the research subjects the same as they would with any other regular LBP patient at their office. However, in order to avoid potential confounding, they were asked to refrain from providing axial lumbar traction or specific rehabilitative exercises during the four week period of this study. Lumbar traction and specific exercises may have an independent treatment effect that would cloud statistical evaluation of the main variable of interest; i.e. mechanical vs. manual manipulation methods.

The chiropractors were permitted to apply adjunctive physical agents, such as electrical stimulation, ultrasound, hot/cold packs, as these modalities have previously been shown not to add any significant additional treatment effect to manipulation for the treatment of acute LBP (58, 59). In addition, the clinicians would occasionally give postural advice and general exercise instructions to their patients. It was not considered ethical to prohibit the treating chiropractors from using these adjunctive therapies, postural advice, and general exercise advice. All of these adjunctive procedures were recorded in the medical records and the utilization rates of each modality were compared between the two cohorts at the end of the study.

Research subjects were treated by their attending chiropractors for a maximum of four weeks or eight treatment sessions, whichever occurred first. At either end point, the research study was terminated and patients were told that if additional treatment was required, they would be treating outside the parameters of the research study. All patients were informed that they would receive twelve week follow-up paperwork by mail and would not be required to return for another physical examination at that time. Since this was a treat-as-usual study, all the subjects were private patients at the participating chiropractic offices and paid for their treatments with private insurance or self-pay. No auto accident or worker's compensation patients were recruited in either cohort.

3.6 STATISTICAL DESIGN

3.6.1 Reliability Studies

There were two separate and distinct inter-examiner reliability studies that were performed, both of which relied chiefly upon calculation of raw percentages of agreement and Kappa statistics for determination of the level of reliability. However, each study had unique attributes that required slightly different statistical analysis.

The prone leg length study involved collection of data regarding the clinician-reported side of short leg and the patient-reported side of pain. These observations were analyzed using a simple 2 x 2 table and Chi square cross-tabulation. Kappa statistics were calculated for the remaining data, which included the level of agreement between the two clinicians for side of short leg with

knees extended, amount of leg length difference, changes in short leg with head rotation right/left, and changes in short leg with knee flexion. We also calculated the prevalence and bias indices for each variable, and used these values to determine the respective Prevalence-Adjusted Bias-Adjusted Kappa (PABAK) coefficients. The methodology and rationale for using PABAK values is well described in an article on the subject by Sim and Wright (60).

Sample size was calculated using various possible values of coefficients of determination (rho) or Kappa scores of .40 or higher, with the reasoning that any value less than .40 would indicate poor reliability and would be clinically irrelevant. Choosing an alpha level of .05 and beta level of .20, sample size calculation indicates a minimum sample size of n = 38 subjects in order to capture a level of reproducibility in the fair to moderate range ($K \geq .40$). Therefore, the final sample size of n = 45 had 80% power to detect a significant Kappa value of .40 or higher.

For the springing palpation reliability study, the data was collected in the form of dichotomous variables for each of the three parts of the palpatory examination and analyzed with the Kappa statistic, the most appropriate reliability coefficient for this type of data. We also calculated the prevalence and bias indices for each variable, and used these values to determine the respective Prevalence-Adjusted Bias-Adjusted Kappa (PABAK) coefficients. Percentages of agreement between the two examiners were also reported for mobility, pain provocation, and the prone instability test, as well as the raw numbers of positive and negative examination findings.

Although we recorded raw data at each lumbar segmental level, the raw data were collapsed for purposes of statistical analysis into two broad categories of "upper lumbar" and "lower lumbar"

as follows; the upper three lumbar facet joints (L1-2, L2-3, L3-4) on each side were coded into the variables "upper left" and "upper right", and the lower two facet joints and sacroiliac joints into the variables "lower right" and "lower left". The spinous processes from L1, L2, and L3 were coded into the variable "upper spinous" and those from L4 and L5 into "lower spinous".

For the prone instability test, we considered any spinous process that was painful on the initial prone position of palpation to reflect a response of "pain provoked" regardless of which level elicited the pain response, and when that pain was relieved with elevation of the feet we recorded this as a "positive test". Therefore, we arbitrarily divided the prone instability test into two portions and calculated Kappa values on each separately. The first portion of the test was the pain provocation upon spinous process palpation with the patient lying prone on the exam table with their feet touching the floor. The second portion of the test occurred when the feet were elevated off the floor, while palpation was again performed over the spinous processes.

Our rationale for collapsing the segmental data into upper and lower lumbar variables fits with the reality of normal clinical practice, in which clinicians are often making judgments about upper or lower lumbar dysfunctions. Although we collapsed the sacroiliac data into the "lower lumbar" category, we decided to also analyze the sacroiliac data separately and report those values by themselves. The rationale for this decision was based upon typical clinical practice in which many clinicians purport that the sacroiliac joint has special importance in the manual palpation examination and spinal manipulation procedures.

3.6.2 Cohort Study

The cohort study data was first analyzed with descriptive statistics of the various baseline socio-demographic variables including age, gender, employment, smoking status, level of education, marital status, income level, previous chiropractic experience, etc. The categorical variables were tested with Chi square analysis of the respective row by column tables, and continuous variables were analyzed with a simple t-test of the means. This baseline analysis was performed in order to explore any differences between the two cohorts, in consideration that any significant differences might require consideration as potential covariates in the primary analysis.

The primary and secondary outcome measures were analyses of the differences between baseline and four-week scores for the Numeric Pain Rating Scale (NPRS) and the Oswestry Low Back Pain Disability Index Questionnaire (ODI) respectively. There were two possible ways to perform a statistical analysis of these scores; (1) Taking the four-week final scores as the dependent variable and "group" as the independent variable while controlling for baseline score or (2) Calculating difference scores to be used as the dependent variable with "group" as the independent variable. There is no consensus on which method is preferable, with various authors advocating one method or the other (61, 62).

For this study, it was decided that the primary form of statistical analysis would be an ANCOVA design, using the four-week pain scores as the dependent variable with group, expectation, and age as the independent variables while controlling for baseline pain score as the covariate. However, purely for sake of an exercise in statistical design, a secondary analysis was performed using the difference scores as the dependent variable. Difference scores were calculated by

subtracting the four-week NPRS and ODI scores from their respective baseline scores for each of the two treatment cohorts. An analysis of variance (ANOVA) was then performed using the NPRS difference scores as the dependent variable with cohort, expectation and age as the independent variables, without any adjustment for baseline pain. A similar ANOVA was performed with the ODI difference scores as the dependent variable, and the same independent variables.

The measurement of treatment expectancy came from a questionnaire that asked the research subjects to report on a Likert scale (0-7) how much worse/better they expected to be after one month of treatment. The expectation score was added to the ANOVA models noted previously as another independent variable, along with the cohort by expectation interaction term. The purpose of this analysis was to explore the main effect of treatment expectation on the outcome measures, and more specifically to see if this effect differed across the two cohorts.

Lastly, scores from the Fear Avoidance Belief Questionnaire (FABQ) were obtained at baseline and four-weeks. To test for any significant differences between the two cohorts on FABQ scores simple independent sample t-tests at baseline and at four-weeks were performed, looking for any significant differences between the mean values of the two cohorts at these two time points.

4.0 RESULTS

4.1 SPECIFIC AIM 1: RELIABILTY STUDIES

4.1.1 Prone leg length analysis

The frequency distributions of the side of reported pain are none=9, right=9, left=10, and central or bi lateral=17. Note that there are a wide variety of observations, with most of the patients reporting central pain (n=17) and almost equal distributions of right (n=9) and left (n=10) sided pain. A number of patients (n=9) also reported feeling no pain on the day of examination.

Chi square (χ^2) analysis of these frequencies using cross-tabulation of "side of short leg" with "side of reported pain" did not show any statistically significant correlation above chance observation. This cross-tabulation was performed for the results of each of the two examiners separately. For the first clinician χ^2=.55 (p=.91) and for the second clinician χ^2=1.55 (p=.67). Note that in the charts of data from both clinicians there is almost equal distribution of right and left short legs with both right and left side of pain, no pain, and central pain, without any apparent pattern or correlation between the variables.

When comparing the two examiners' observations about the side of short leg, they showed 82.2% raw agreement with a Kappa value of .65 that would be interpreted as "good" to "substantial" inter-examiner reliability, with the clinical inference that two clinicians can agree on the side of short leg reasonably well.

The two examiners were asked to determine the approximate amount of leg length difference, using four categories of <¼", ¼-½", ½-¾", or >¾". No ruler or tape measure was used in this study; the examiners were asked to "eyeball" the perceived amount of leg length difference. There were no cases of either clinician reporting a difference of >¾" and only two reported cases of ½-¾" differences. Therefore it was decided to collapse the three higher categories into one global category of "¼" or greater", and analyze the data using a cross-tabulation between two new categories of "less than ¼" versus "¼"or greater". These were coded with the dichotomous variables "0" and "1" respectively. There was 66.7% agreement on the amount of leg length difference which gave a Kappa value of .28 that can be interpreted as "fair" reliability.

Kappa values for head rotation to the right and left were calculated separately (n=22), with no significant findings found for this portion of the leg length analysis. For head rotation to the left inter-examiner agreement was 50% (Kappa=.04), and for head rotation to the left agreement was 45.5% (Kappa= -.19). These Kappa values indicate virtually no inter-examiner reliability above chance observation. The negative Kappa value is not a mistake; a negative value denotes that the recorded raw agreement was less than 50% or chance observation. These data indicated that with head rotation to the left, the raw agreement was less than chance at 45%.

The last part of the statistical analysis involved the inter-examiner reliability for Position Two of the leg length analysis, in which the clinicians flexed the patients' knees to 90° and observed for any changes on the short leg side. It was interesting to find that the two examiners had 93.3% agreement that the short leg side got "longer" during this portion of the analysis. Despite this level of raw agreement, a Kappa statistic could not be calculated on this portion of the test, due to the extremely high prevalence of cases in the concordant cell "short leg gets longer" (42 of 45 cases) and no reported cases of "short leg gets shorter" by either examiner. The results of all these data analyses are presented in Table 11 below:

Table 11: Analysis of data from leg length reliability study.

Clinical Procedure	Kappa	95% CI	% agreement	Prevalence Index	Bias Index
Determination of which is the short leg side	**.65**	**.43, .87**	**82%**	**.02**	**.04**
Estimation of the amount of difference (¼", ½", etc)	**.22**	**.01, .55**	**62%**	**.31**	**.16**
Change in short leg with head rotation to left	**.04**	**-.25, .33**	**50%**	**.41**	**.36**
Change in short leg with head rotation to right	**-.20**	**-.30, .38**	**45%**	**.45**	**.41**
Change in short leg with knees flexed	**0.0***	**-1.0, 1.0**	**93%***	**.93**	**.00**
Observation of a short leg	**0.0***	**-1.0, 1.0**	**100%***	**1.0**	**.00**

4.1.2 Facet joint and sacroiliac joint palpation

Data were collected from 39 subjects in the form of dichotomous variables for each of the three parts of the palpatory examination and analyzed with the Kappa statistic, the most appropriate reliability coefficient for this type of data (63). Kappa values reflect the percentage of agreement between examiners that is above chance agreement. Simple percentages of agreement between the two examiners were reported for mobility, pain provocation, and the prone instability test, as

well as the raw numbers of positive and negative examination findings. The prevalence and bias indices for each variable were calculated, and these values were used to determine the respective Prevalence-Adjusted Bias-Adjusted Kappa (PABAK) coefficients. The methodology and rationale for using PABAK values is well described in an article on the subject by Sim and Wright (60).

Although raw data was collected at each segmental level, the raw data were collapsed for purposes of statistical analysis into two broad categories of "upper lumbar" and "lower lumbar" as follows; the upper three lumbar facet joints (L1-2, L2-3, L3-4) on each side were coded into the variables "upper left" and "upper right", and the lower two facet joints and sacroiliac joints into the variables "lower right" and "lower left". The spinous processes from L1, L2, and L3 were coded into the variable "upper spinous" and those from L4 and L5 into "lower spinous".

For the prone instability test, any spinous process that was painful on the initial prone position of palpation was considered to reflect a response of "pain provoked" regardless of which level elicited the pain response, and when that pain was relieved with elevation of the feet it was recorded as a "positive test". Therefore, the prone instability test was arbitrarily divided into two portions, with Kappa values calculated on each portion separately. The first portion of the test was the pain provocation upon spinous process palpation with the patient lying prone on the exam table with their fee touching the floor. The second portion of the test occurred when the feet were elevated off the floor, while springing palpation was again performed over the spinous processes.

The rationale for collapsing the segmental data into upper and lower lumbar variables fits with the reality of normal clinical practice, in which clinicians are often making judgments about upper or lower lumbar dysfunctions. Although we collapsed the sacroiliac data into the "lower lumbar" category, we decided to also analyze the sacroiliac data separately and report those values by themselves. The rationale for this decision was based upon typical clinical practice in which many clinicians purport that the sacroiliac joint has special importance in the manual palpation examination and spinal manipulation procedures.

Sample size was calculated using various possible values of coefficients of determination (rho) or kappa values of .40 or higher, with the reasoning that any value less than .40 would indicate poor reliability and would be clinically irrelevant. Choosing an alpha level of .05 and beta level of .20, sample size calculation indicated a minimum sample size of n = 38 subjects in order to capture a level of reproducibility in the fair to moderate range (K≥ .40). Therefore, the final sample size of n = 39 would appear to have 80% power to detect a significant kappa value of .40 or higher.

The data analysis included calculations of standard kappa values with their respective 95% confidence intervals (CI), prevalence-adjusted bias-adjusted kappa (PABAK) values, raw percentages of agreement, prevalence index, bias index, and raw numbers of positive (+ve) / negative (-ve) findings for each examiner. A positive finding was defined as either pain provoked during palpation (pain tests) or examiner determination of "joint restriction" (mobility tests) during palpation. All of these data are summarized in Table 12 below:

Table 12: Data analysis from palpation reliability study. "High" is the group variable for L1-2-3-4 facets and "low" is the grouping variable for L4-5-S1 facets. Abbreviations: PABAK= Prevalence-adjusted Bias-adjusted Kappa; Positive = "+ve", Negative = "-ve".

Region of Palpation	**Kappa** un-adjusted κ	**95%CI** un-adjusted κ	**PABAK** adjusted κ	**%** agreement	**prevalence index**	**bias index**	**Examiner 1** -ve/+ve	**Examiner 2** -ve/+ve
Low mobility left	-.17	-.41, .06	.08	54%	.49	.15	26/13	32/7
Low spinous mobility	-.05	-.36, .27	.11	56%	.41	.03	28/11	27/12
Low mobility right	-.12	-.41, .18	-.09	46%	.36	.03	26/13	27/12
High mobility left	.17	-.14, .48	.44	72%	.46	.13	26/13	31/8
High spinous mobility	.02	-.27, .32	.07	54%	.28	.15	28/11	22/17
High mobility right	-.01	-.33, .30	.44	72%	.67	.03	32/7	33/6
Sacroiliac mobility right	-.10	-.18,-.02	.64	82%	.82	.03	36/3	35/4
Sacroiliac mobility left	-.11	-.21,-.01	.54	77%	.77	.08	33/6	36/3
Low pain left	.73	.51, .95	.74	87%	.26	.08	23/16	26/13
Low spinous pain	.57	.32, .83	.58	79%	.23	.10	13/26	17/22
Low pain right	.52	.25, .79	.54	77%	.21	.03	23/16	24/15
High pain left	.46	.17, .75	.48	74%	.23	.05	23/16	25/14
High spinous pain	.21	-.10, .53	.34	67%	.15	.03	23/16	22/17
High pain right	.38	.06, .69	.54	77%	.31	.03	25/14	26/13

Sacroiliac pain right	.14	-.19, .47	.38	69%	.54	.05	29/10	31/8
Sacroiliac pain left	.33	0.0, .66	.54	77%	.56	.08	29/10	32/7
Prone instability test (1)	.54	.27, .81	.58	79%	.33	.05	12/27	14/25
Prone instability test (2)	.46	.15, .77	.58	79%	.49	.05	28/11	30/9

As a general rule, the results showed that the pain provocation tests were more reliable (κ range, .21 to .73) than the segmental mobility tests (κ range, -.17 to .17) with respect to the unadjusted kappa values. There was little change in the pain provocation values when comparing the unadjusted Kappa values (κ range, .21 to .73) with the adjusted values (PABAK range, .34 to .74). The prone instability test showed moderate reliability with both unadjusted (κ range, .46 to .54) and adjusted Kappa values (PABAK range, .58 to .58). Landis and Koch (63) have suggested the following standards for strength of agreement for Kappa coefficients:

- ≤ 0 = poor reliability
- .01-.20 = slight reliability
- .21-.40 = fair reliability
- .41-.60 = moderate reliability
- .61-.80 = substantial reliability
- .81-1.0 = almost perfect reliability

One interesting observation is the large increase in the reliability for the sacroiliac joint (SIJ) data with respect to both pain and mobility testing when using the adjusted Kappa value. The unadjusted Kappa range for pain provocation over the SIJ is .14 to .33, which is considered poor to fair reliability. However, when adjusted for prevalence and bias, the PABAK value for SIJ pain provocation rises substantially to a range of .54 to .56, which is considered good reliability. The SIJ mobility testing data is even more affected by Kappa adjustment; unadjusted Kappa values range from -.10 to -.11 and rise to a PABAK range of .77 to .82. Essentially, if one looks at the unadjusted Kappa values for SIJ mobility testing the conclusion is that the procedure is completely unreliable and actually less than chance (negative values), but an extremely reliable procedure when looking at the adjusted Kappa values. This large differential between adjusted

and unadjusted Kappa values causes some concern about which values reflect the true reliability of the procedure.

4.2 SPECIFIC AIM 2: OBSERVATIONAL COHORT STUDY

4.2.1 Socio Demographic Baseline Characteristics

All research subjects were asked to complete a socio-demographic questionnaire on the baseline visit that captured a number of variables that might potentially affect their clinical outcomes. A copy of this questionnaire is located in the Appendix section (Appendix I). The compliance with participants completing this questionnaire was good except for the question regarding income, which a large number of subjects did not complete.

The socio-demographic data were analyzed with the Chi Square test that was performed on all of the categorical variables to compare the frequency distributions between the two cohorts. A t-test was used to compare the mean ages (continuous variable) between the two cohorts. The results of this basic socio-demographic analysis revealed that the two cohorts did not differ significantly on any of these variables other than age, which was significantly higher in the manipulation cohort. Age was therefore used as a covariate in a subsequent regression analysis, but was not found to have a main effect on changes in pain or Oswestry scores (see RESULTS section). Age was not significantly correlated with either NPRS or ODI scores at baseline or four-weeks. Table 13 below provides a summary of the socio-demographic data analysis:

Table 13: Baseline socio-demographic characteristics of research subjects

Socio-Demographic Variable	Activator	Manipulation	p-value (Chi-square)
	n = 55	n = 38	
Gender (Males)	56%	54%	.81
Age (mean/standard deviation)	38.4 yrs (15.1)	49.7 yrs (14.6)	.001** (t-test)
Race (Caucasian)	100%	100%	n/a
Marital Status			.21
Single	35%	18%	
Married	56%	68%	
Other	9%	14%	
Education			.25
≤ High School	49%	36%	
College	51%	64%	
Income			.25
≤ $35,000/ yr	56%	47%	
$35,001 to $70,000	30%	23%	
> $70,000/ yr	15%	30%	
Employment Status			.99
Working Full-time	65%	63%	
Working Part-time	13%	13%	
Other (not working)	22%	24%	
Smoker	30%	26%	.73
Medications			.14
None	59%	78%	
NSAIDs	32%	19%	
Prescription pain meds	9%	3%	

4.2.2 Analysis of Numeric Pain and Oswestry Scores

The primary outcome measure in this study was the change in numeric pain rating scores (NPRS) from baseline to four weeks, with the secondary outcome measure being the change in Oswestry low back pain disability (ODI) scores during the same time period. The NPRS scores at baseline and four weeks were recorded as the average of the "pain right now" and "worst pain

in past 24 hours" questions on the triple numeric pain scale. The Oswestry scores were reported as percentages, by taking the raw total score and dividing by 50.

The two cohorts had similar baseline mean pain scores, with the manual manipulation group ending up with a slightly lower four-week mean pain score. However, after controlling for baseline pain and expectation, the difference in four-week pain scores between the two cohorts was not statistically significant (p=.07, 95% CI; -1.68, .068). The change scores were 3.5 points for the Activator cohort and 4.3 points for the manipulation cohort which was also not significant (p=.12, 95% CI; -.21, 1.81). A graphic display of the mean baseline and four-week pain scores in each cohort, as well as their respective change scores is depicted in Figure 3 below:

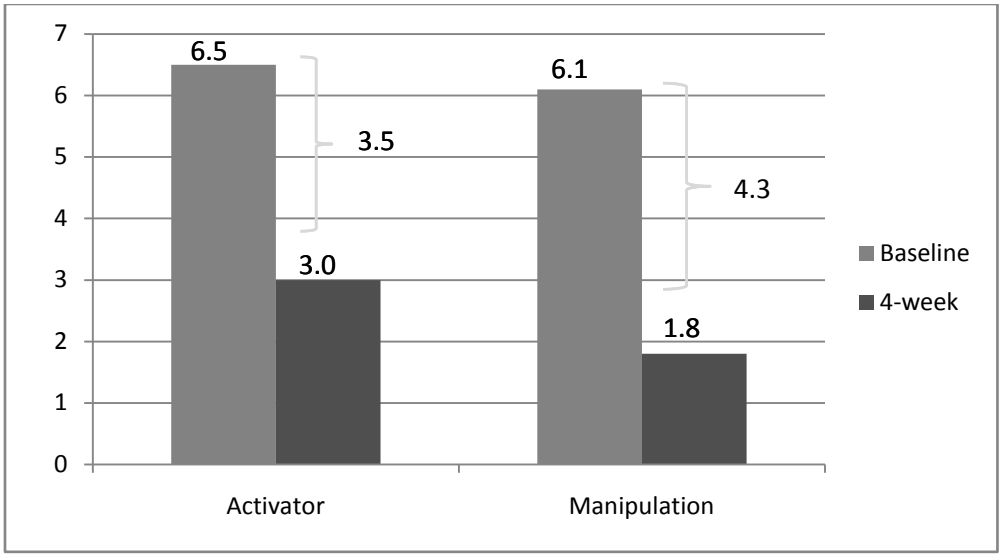

Figure 3: Baseline and four-week mean pain scores by cohort

The data were analyzed using a regression model with pain score at four-weeks as the dependent variable, cohort as the key independent variable, while controlling for centered baseline pain score as a covariate. The baselines scores were centered by subtracting each individual scores from the mean baseline scores (within-group mean) in each cohort. Since age was the only socio-

demographic variable that was statistically different between the two cohorts at baseline, age and its interaction term (age*cohort) were added to the model. Lastly, another hypothesis of this study was that treatment expectation might be a strong predictor of clinical success and therefore expectation and its interaction term (expectation*cohort) were added to the model as well.

The full model started with pain score at four-weeks as the dependent variable, with these independent variables: cohort, centered baseline pain, age, age*cohort interaction, expectation, and expectation*cohort interaction. This model showed that there was a significant main effect for expectation, but that this effect was not significantly different between the two cohorts (no significant interaction effect) and therefore expectation was retained in the model, but the interaction term was dropped. There was no main effect for age and no significant interaction effect, and therefore these two variables were dropped from the model. The final model retained three variables; cohort, centered baseline pain, and expectation.

Difference scores were also created by subtracting the four-week pain scores from their respective baseline values, and a subsequent regression analysis was performed using the difference scores as the dependent variable with cohort and expectation as the main independent variables. The results of this alternative regression model were not substantially different from the original model using the four-week pain scores as the dependent variable and controlling for baseline pain as a covariate. The ANOVA tables, regression models, and STATA syntax for all of these regression models are listed in Appendix J.

The secondary outcome analysis was performed using a similar regression model, but this time substituting the four-week Oswestry (ODI) scores as the dependent variable. The same step-wise process of different regression models was used with these covariates: centered baseline Oswestry score, age, expectation, and their respective interaction terms. As with the previous regression models using pain scores, age and its interaction term did not show any significant effects and were dropped from the final model. Expectation did not show a main effect and its interaction term was also non-significant, therefore both variables were dropped. The final model compared four-week Oswestry scores between the two cohorts while controlling for baseline Oswestry score.

Figure 4 below is a graphic display of the mean baseline and four-week Oswestry scores in each cohort. The two cohorts had similar baseline mean ODI scores, with the manual manipulation group ending up with a non-significant lower four-week mean ODI score with a regression model that controlled for baseline Oswestry score (p=0.6, 95% CI; -6.9, 4.0).

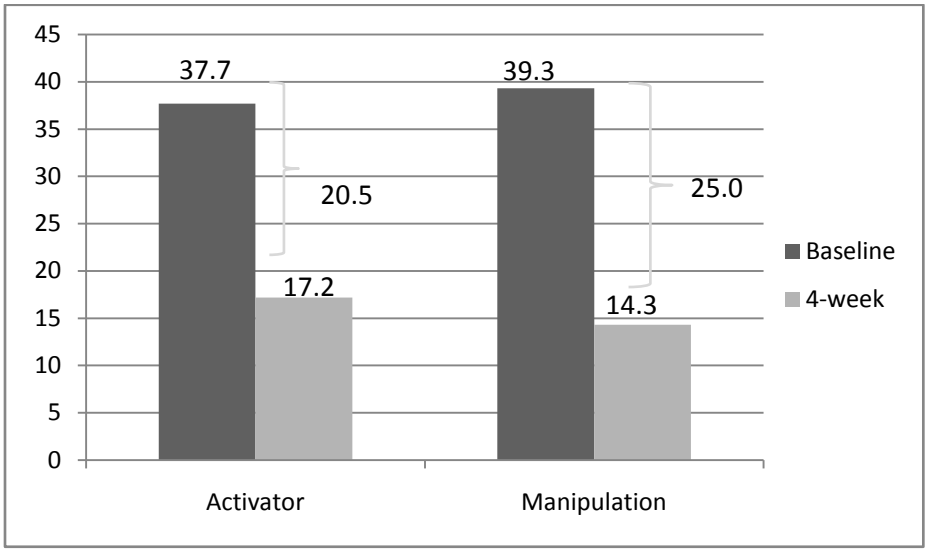

Figure 4: Baseline and four-week mean Oswestry scores by cohort. The four-week scores were not statistically significant after controlling for baseline pain and expectation (p=0.6, 95% CI; -6.9, 4.0). The difference in pain scores between baseline and four-weeks (20.5 vs. 25.0) were also not statistically significant (p=.20, 95% CI; -2.4, 11.12).

Difference scores were also calculated by subtracting the four-week scores from the baseline scores, and an alternative regression model was run using the difference scores as the dependent variable. The results did not reveal any statistically significant main effect for ODI difference scores between the two cohorts (p=.20, 95% CI; -2.4, 11.12). The mean change in ODI score was 20.5 points for the Activator cohort and 25 points for the manual manipulation cohort. The ANOVA tables, regression models, and STATA syntax for these Oswestry regression analyses can be found in Appendix J.

4.3 SPECIFIC AIM 3: EXPECTATIONS/PATTERNS OF CARE

4.3.1 Treatment Expectations Questionnaire

In addition to recording basic socio-demographic information, research subjects were also asked to complete a treatment expectations questionnaire. This questionnaire was developed by the PI for the purpose of this study, and was not subjected to any formal reliability or validity testing. However, it was thought that gathering this information would be important for exploring the possible differences in treatment expectation from patients who were self-selecting two very different types of chiropractic manipulative treatment. A copy of this expectations questionnaire is included in the Appendix section (Appendix H).

The Activator Methods protocol is marketed to the general public as being a safer and gentler alternative to manual manipulation, and many patients seek treatment with the Activator Method because they are fearful or concerned about thrust manipulation. Conversely, patients who have

experienced effective relief with manual manipulation are dubious about a mechanical device which they perceive as delivering a minimal level of force and not comparable with manual manipulation. It was a specific aim of this study to explore some of these differences in attitudes and beliefs about chiropractic treatment methods, and therefore this treatment expectations questionnaire was designed with the intent of capturing some of this information.

A few challenges arose with the implementation of this questionnaire including a high rate of non-response to several questions, lack of experience by the manipulation cohort with Activator and vice versa with the Activator cohort. Despite these challenges, an overview of the questionnaire responses and their respective Chi Square values are listed in Table 14 below:

Table 14: Analysis of expectations at baseline. Some groups of questions do not add up to 100% due to non-response.

Expectations Variable	**Activator**	**Manipulation**	**p-value (Chi-square)**
Had previous chiro care? (yes)	**69%**	**84%**	**.09**
Type of chiro received in past:			**.002****
Activator	**13%**	**0%**	
Manual manip	**26%**	**62%**	
Both	**28%**	**21%**	
Past experience: Activator			**.25**
Not helpful to somewhat helpful	**11%**	**3%**	
Helpful to extremely helpful	**35%**	**28%**	
Past experience: Manual Manip			**<.001****
Not helpful to somewhat helpful	**24%**	**5%**	
Helpful to extremely helpful	**35%**	**80%**	
One month expectation			**.66**
Much worse to a little better	**8%**	**6%**	

Moderately better to completely better	**91%**	**94%**	
Do you expect to hear a "pop"?			**.008****
Yes	**8%**	**19%**	
No	**62%**	**37%**	
Not sure	**23%**	**44%**	
How do you feel about hearing a "pop"?			**.009****
Good	**26%**	**54%**	
Not Good	**11%**	**3%**	
Not sure	**61%**	**36%**	
Expectations about Activator			**.18**
Not helpful at all to probably will help	**11%**	**3%**	
Will help a little to will help greatly	**35%**	**28%**	
Expectations Re: Manual Manip			**<.001****
Not helpful at all to probably will help	**24%**	**5%**	
Will help a little to will help greatly	**35%**	**80%**	

The majority of patients in each cohort had previous experience with chiropractic care, with the manual manipulation cohort reporting significantly more previous experience with manual manipulation (62%) than the Activator cohort reported with previous Activator experience (13%). However, these data are confounded by the large number of patients who did not complete this question.

There was no significant difference between groups regarding past Activator experience, however there was a significant difference regarding past manipulation experience with 80% of manipulation patients reporting good previous results and only 5% reporting poor previous results with manipulation, compared to 35% of Activator patients reporting good previous results and 24% reporting poor previous results with manipulation.

When asked about their general expectations regarding how they would be feeling in one month, the majority of patients in both cohorts reported that they expected good to excellent results with treatment. There was no significant difference between the two groups on overall treatment expectancy. However, there was a significant difference with respect to expectations about manual manipulation with 80% of manipulation patients reporting they expect good results with manipulation compared to only 35% of Activator patients reporting good expectations about manual manipulation. However, 41% of Activator patients did not complete this question. 35% of Activator patients and 28% of manipulation patients reported a positive expectation about Activator treatment, with 54% of Activator and 69% of manipulation patients not completing this question. These differences on Activator expectations were not statistically significant.

There were also significant differences between the groups about the expectation of hearing a "pop" during the treatment. 19% of the manipulation patients and 8% of Activator patients expected to hear a "pop". 64% of the Activator patients did not expect to hear a pop, compared with only 37% of the manipulation patients. However, a large number of the manipulation patients did not answer this question which may confound these results. 54% of manipulation patients reported that they "felt good" about hearing a pop and 39% were "not sure" or "did not

feel good", whereas only 26% of Activator patients felt good about hearing a pop and 72% reported "not sure" or "did not feel good". The Chi Square tests showed all of the above differences to be significant.

4.3.2 Analysis of usage of physical agents/modalities

Due to the treatment-as-usual design of this study, the participating chiropractors were told to treat the research subjects in the same manner as they normally would treat any other low back pain patient in their private clinics. Chiropractic management of low back pain often includes the use of adjunctive physical agents, posture education, and home exercises. Since we did not perform a randomized trial with rigorous control on the use of these adjunctive procedures, the usage of these procedures was recorded and analyzed in the chart below. Percentages of use for each procedure within each cohort are listed in Table 15 below, along with their respective p-values determined by Chi Square analysis.

Table 15: Group comparisons of physical agents/modalities usage

Type of Modality	Activator	Manipulation	p-value (Chi-square)
Electrical stimulation	96%	74%	.002**
Ultrasound	2%	5%	.38
Laser	9%	0%	.05**
Lumbar belt	0%	0%	n/a
Roller table (intersegmental mobs)	74%	0%	<.001**
Myofascial release	0%	3%	.24
Posture education	54%	31%	.03**
Home exercises	60%	56%	.70
Heat/Ice	0%	8%	.038**

The Activator cohort had a significantly higher percentage of usage of most physical agents, posture education, and home exercises, with the exception of heat/ice application which showed a significantly higher percentage of usage by the manual manipulation cohort.

4.3.3 Analysis of Clinical Patterns of Care

The Activator Methods clinic and manual manipulation clinics appeared to have some significant differences in their respective clinical patterns of care, and these data are analyzed and presented in Table 16 below:

Table 16: Group comparisons of clinical patterns of chiropractic care

Clinical Variable	Activator	Manipulation	p-value (Chi-square)
Study terminated at:			**<.001**
8 visits	**70%**	**15%**	
4 weeks	**17%**	**23%**	
< 4 weeks	**13%**	**62%**	
Continued with care after study? (yes)	**78%**	**18%**	**<.001**
Diagnostic studies ordered during study			**<.001**
None	**41%**	**100%**	
Lumbar x-rays	**54%**	**0%**	
Lumbar MRI	**5%**	**0%**	
Number of visits at 4 weeks			
(Mean / standard deviation)	**9.2** (2.7)	**4.5** (2.3)	**<.001** (t-test)
Location of Symptoms			**.07**
LBP only	**50%**	**74%**	

LBP and buttock	**28%**	**16%**	
LBP, buttock, and thigh	**22%**	**10%**	
Onset of LBP			**.016**
≤ 14 days	**52%**	**81%**	
15 days to 12 weeks	**48%**	**19%**	
Antalgic Lean Present?			**.01**
None	**94%**	**74%**	
Flexion	**0%**	**18%**	
Lateral list	**6%**	**8%**	
Limitation of Lumbar Flexion			**<.001**
None	**2%**	**31%**	
Mild (> 41º)	**48%**	**45%**	
Moderate (20º - 40º)	**41%**	**21%**	
Severe (< 20º)	**9%**	**3%**	

There was a significant difference between the two cohorts with respect to the number of patients in each group that required the maximum number of treatment sessions (eight visits), 70% in the Activator group compared to only 15% in the manual manipulation group. 78% of the Activator patients continued with additional chiropractic care after study termination whereas only 18% in the manual manipulation groups received additional chiropractic treatment. A t-test of the mean number of visits at four weeks showed a significant difference between the two cohorts, with the Activator group having a mean of 9.2 visits which was about twice as high as the manual manipulation group mean of 4.5 visits.

The manual manipulation cohort had a significantly higher number of acute patients (onset < 14 days) with more patients displaying a flexion antalgic lean. The Activator patients were more likely to receive lumbar x-rays (54%) or MRI (5%) since none (0%) of the manual manipulation patients received any of these diagnostic tests during the course of this study. There was no significant difference in the location of symptoms between these two cohorts.

4.4 SPECIFIC AIM 4: EXPLORE FEAR AVOIDANCE BELIEFS

4.4.1 Analysis of Fear Avoidance Beliefs Questionnaire

The chief hypothesis of this specific aim was whether a difference in Fear Avoidance Beliefs Questionnaire (FABQ) scores would be found between the Activator and manipulation cohorts at baseline and four-weeks. The FABQ is divided into two subscales; the physical activities (PA) and work subscales. For the purpose of this cohort study, only the PA subscale of the FABQ was used, because none of the research subjects had any work related injury. The PA subscale consists of four questions that are each graded from 0-6, with higher scores indicating more self-perceived limitation of physical activities due to back pain. The maximum score on the PA subscale is 4 x 6 = 24 points.

The analysis of this hypothesis involved two simple t-tests for differences in the means at baseline and at four-weeks. The output of the t-test for mean differences at baseline is listed in Table 17 below. There was no significant difference between the Activator and manipulation group FABQ means at baseline, with or without the assumption of equal variances.

Table 17: T-test of baseline FABQ scores, with and without the assumption of equal variances

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
fabqpa	Equal variances assumed	.336	.564	-.508	90	.613	-.58285	1.14689	-2.86134	1.69565
	Equal variances not assumed			-.496	72.316	.622	-.58285	1.17602	-2.92703	1.76134

A second t-test was performed comparing the four-week FABQ scores between the two groups, and again no statistically significant difference between the mean values was observed. The output for this t-test is listed in Table 18 below:

Table 18: T-test of four-week FABQ scores with and without the assumption of equal variances

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
fabqpa	Equal variances assumed	.039	.844	.393	90	.696	.53119	1.35301	-2.15679	3.21917
	Equal variances not assumed			.389	76.759	.699	.53119	1.36722	-2.19142	3.25380

It is interesting to note that the four-week mean FABQ scores in both the Activator and manipulation groups had dropped by approximately 25% from the baseline scores. However, the difference between the two group means was not statistically significant based upon the previously noted t-tests. The conclusion was that FABQ scores were not significantly different between the Activator and manipulation cohorts at either baseline or four-weeks. Figure 5 below shows the mean FABQ scores for both cohorts at baseline and four weeks.

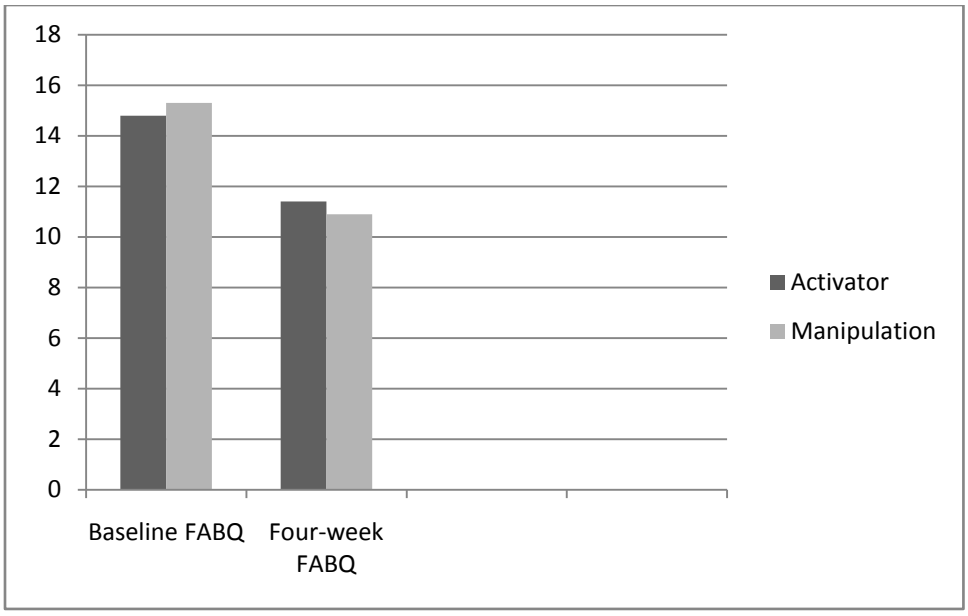

Figure 5: Mean values of baseline and four-week physical activities (PA) Fear Avoidance Beliefs Questionnaire (FABQ) scores by cohort. Note that both cohorts had a drop in their mean scores from about 15 to 11 points over four-weeks. The maximum score for the FABQ - PA subscale is 24 points.

5.0 DISCUSSION

5.1 LEG LENGTH RELIABILITY STUDY

The clinical observation of leg length inequality has been a routine part of the physical examination procedure for assessing symmetry of the pelvis and spine. However, the use of the prone leg length analysis is a central component of the Activator Methods protocol, as it is the main determinant of where the chiropractor practicing this method will apply a mechanical force to the patient's spine.

The data analysis did not show any correlation between the side of the observed short leg and the side of pain as reported by the patient. The clinical relevance of this finding is not completely surprising, since clinicians practicing the Activator Methods protocol have empirically noted that the symptomatic side of back pain does not seem to directly correlate with the side of short leg. They argue that the observation of leg length inequality is merely a screening test that indicates potential spinal dysfunction. However, the data from this study showed 100% prevalence of leg length inequality in all 45 participants regardless of whether they were symptomatic or not which raises the question of clinical relevancy for the observation of leg length inequality.

Two clinicians appear to be able to reliably determine the side of short leg, as noted by the Kappa value of .65 that indicated a reasonable level of reliability. Although the side of short leg does not seem to correlate directly with the side of low back pain, there may be other forms of clinical application to this knowledge of leg length inequality.

Determining the magnitude of leg length inequality by "eyeballing" the amount of difference does not appear to be very reliable according to our data analysis, with a small Kappa value of .28 for this reliability test. The goal standard procedure for determining the precise amount of leg length inequality is a standing plain film x-ray with the central ray directed parallel with the top of the femur heads. It is possible that other examination procedures such as measuring leg length inequality with a tape measure might be more reliable than the simplistic "eye-balling" method used in our study.

The chiropractic variation of the prone leg length analysis using head rotation (Derifield test) appears to be completely unreliable based upon our data analysis. The Kappa values for head rotation right and left were .04 and -.19 respectively, which is essentially chance observation. This head rotation procedure is still quite commonly utilized within the chiropractic profession as a screening test for cervical versus lumbo-pelvic spinal dysfunction as related to observed leg length inequality.

A surprising result of our data analysis indicated that flexing the knees almost always results in the observation of a short leg getting longer, and that in no case did our two clinicians agree on the observation of a short-leg getting shorter or staying the same length. Due to this extreme

prevalence bias of one exclusive finding, it was not possible to calculate a Kappa value for this reliability test. The Activator clinicians have empirically noted that this is the most common observation found when performing knee flexion, and stated that it was extremely rare to see the short-leg get shorter during this procedure. The clinical relevance of this finding is unknown.

The results of this leg length reliability study were summarized and published as an article in the Journal of Manipulative and Physiological Therapeutics (JMPT) in the Fall of 2007 (64). The full text article gives a more detailed discussion of the issues regarding the clinical relevancy of these findings.

5.2 SPRINGING PALPATION RELIABILITY STUDY

This study was performed to explore the reliability of springing palpation used over the lumbar facet and sacroiliac joints as a screening physical examination procedure for low back pain and joint dysfunction. Many variations of springing palpation are used within the chiropractic and physical therapy professions to determine areas of spinal joint dysfunction and pain. The clinicians in the manual manipulation cohort utilize springing palpation routinely in their practices, making the study feasible.

As a general rule, our results showed that the pain provocation tests were more reliable (κ range, .21 to .73) than the segmental mobility tests (κ range, −.17 to .17) with regard to the unadjusted κ values. There was little change in the pain provocation values when comparing the unadjusted κ values (κ range, .21 to .73) with the adjusted values (PABAK range, .34 to .74). The prone

instability test showed moderate reliability with both unadjusted (κ range, .46 to .54) and adjusted κ values (PABAK range, .58 to .58). The sacroiliac joint (SIJ) data analysis was overall more reliable than the lumbar facet joint reliability data analysis.

One interesting observation is the large increase in the reliability for the SIJ data with respect to both pain and mobility testing when using the adjusted κ value. The unadjusted κ range for pain provocation over the SIJ is .14 to .33, which is considered poor to fair reliability. However, when adjusted for prevalence and bias, the PABAK value for SIJ pain provocation rises substantially to a range of .54 to .56, which is considered good reliability. The SIJ mobility testing data are even more affected by κ adjustment; unadjusted κ values range from −.10 to −.11 and rise to a PABAK range of .77 to .82.

This springing palpation data analysis serves as a good example of the well known issue that Kappa values are very sensitive to influence by the prevalence of the clinical attribute being tested. This is due to the fact that if a particular observation only occurs rarely, it can violate the basic "chance agreement" assumption of the κ statistic. In our data, it is apparent that there is a much higher proportion of agreed-upon negative examination findings (normal mobility or no pain provoked), compared with the proportion of agreed-upon positive findings (restriction or pain provocation). The disproportionately high number of negative examination findings raises the "chance" of a negative finding, which confounds the κ calculations. When the κ and corresponding PABAK coefficients are widely different in magnitude, it indicates the possibility of an underlying confounding situation with disproportionate prevalence findings.

Another issue confounding the Kappa calculations was the extreme number of negative exam findings at most lumbar facet levels, and the sheer number of levels to analyze. With 39 patients, no meaningful statistical analysis could be performed that would involve 15 possible data points: 10 facet joint levels (5 on each side for each patient) and five spinous process levels. If we analyzed all 15 of these possible combinations of reliability, it would have produced a 15 × 15 contingency table that would contain so many empty cells as to render any statistical analysis completely meaningless. Therefore, although the data were collected at each lumbar facet joint bilaterally and over each spinous process, this data was collapsed into two dichotomous categories: "upper" and "lower" lumbar.

It is also important to note that spinal palpation for mobility is often confounded by some sort of nonverbal or verbal communication by patients. It is very difficult to separate the "subjective" pain provocation component of springing palpation from the "objective" palpatory sensation reported by the clinicians. Our study design very carefully separated the pain provocation and segmental mobility tests by giving explicit instructions to the research subjects not to verbally communicate during the segmental mobility portion of the test.

The results of this springing palpation reliability study were summarized and published as an article in the Journal of Manipulative and Physiological Therapeutics (JMPT) in the Summer of 2008 (65). The full text article gives a more detailed discussion of the issues regarding the clinical relevancy of these findings.

5.3 COHORT STUDY

The main focus of this cohort study was to explore the differences in treatment effect between mechanical (Activator) and manual manipulation methods in patients with acute and sub-acute back pain. The differences between numeric pain and Oswestry scores from baseline to four-weeks served as the primary and secondary outcome measures. In addition to exploring the treatment effect of these methods, this study was also designed to capture data regarding baseline socio-demographic characteristics, fear avoidance beliefs, expectations, and clinical patterns of care in the respective patient populations of the two cohorts.

The primary outcome measure of pain scores and the analysis between the differences in baseline and four-week scores showed a larger drop in pain scores in the manual manipulation cohort, but this difference did not reach statistical significance when controlling for baseline pain. It is interesting to note that the 95% confidence interval for this drop in pain score ranged from .068 to -1.68, which shows a trend toward more reduction of pain in the manipulation cohort. The lower bound of this confidence interval is a pain reduction of 1.68 points, which compares well to the minimal detectable change of 2 points and standard deviation of 1.02 points reported in the pain score literature (53). The small sample size and lack of control over treatment protocol may have been important factors in the width of the confidence interval calculated from these data.

The secondary outcome measure of Oswestry (ODI) scores and the differences between four-week and baseline also showed that the manipulation group had lower Oswestry scores at four weeks compared to the Activator group, but that this difference was also not statistically significant. In the case of Oswestry scores, the 95% confidence interval is very wide, ranging

from -6.9 points to 4.0 points. This wide confidence interval makes interpretation of the data very difficult and imprecise, considering that the minimal clinically important difference for Oswestry scores is 6 points as reported in the literature (66).

These Oswestry results suggested the possibility of a Type II error due to the small sample size of this pilot study. Therefore a post hoc power analysis was therefore performed on the Oswestry difference scores, using the means, standard deviations, and sample sizes of the two cohorts. The estimated power based upon the ODI data analysis was only 24%, which confirms the strong likelihood of a Type II error in this observational study.

The a priori power analysis performed at the onset of this study was based upon the effect sizes and minimal clinically important differences for NPRS (2 points, standard deviation = 1) and ODI (8 points, standard deviation = 14) values reported in previous low back pain manipulation trials (53, 66). There was no realistic way to ascertain what magnitude of treatment effect size would be found in this observational study, since this is a new area of clinical research. In fact, one of the ancillary goals of this study was to obtain preliminary pilot data on treatment effect size between manual and mechanical types of manipulation.

Post hoc power analysis using the NPRS effect size and standard deviations obtained from the data in this cohort study (MCD of 1 point; standard deviation of 1.8 points) with alpha of .05 and beta of .80 gives a projected sample size of 51 per group. A similar power analysis using the ODI data from this study (MCD of 6 points; standard deviation of 14) gives a projected sample size of 86 per group.

One might argue that if the study was underpowered from the onset, why bother moving forward with the study at all? This is the challenge with most exploratory observational studies in which the true treatment effect size is literally unknown. It was not possible to know a priori what level of effect size would be found with respect to Activator vs. manual manipulation treatments because this type of research and data had not been previously reported in the literature. Therefore the a priori power analysis was based upon the MCIDs as reported in the previous low back pain literature as the best reasonable estimate of effect size.

5.4 FEAR AVOIDANCE BELIEFS QUESTIONNAIRE

One hypothesis was that there might be a difference in the Fear Avoidance Beliefs Questionnaire (FABQ) scores at baseline between the Activator and manual manipulation patient populations. A t-test of the baseline FABQ scores did not reveal any significant differences by the two cohorts. However, the average FABQ scores in both the Activator and manual manipulation groups had dropped by about 25% at four-weeks, from about 15 points at baseline to approximately 11 points at four-weeks (maximum score=24 points). A subsequent t-test of the four-week FABQ scores did not reveal any significant difference between the two groups.

There is a wealth of information in the low back pain literature regarding the issue of fear avoidance beliefs. Waddell was responsible for the development of the FABQ and he (56) demonstrated that, after controlling for pain intensity and pain location, fear-avoidance beliefs about work explained a significant amount of variance in disability (23%) and work loss (26%). More recently, George et al (67) showed that the resultant disability from low back pain at 4

weeks and 6 months after treatment was dependent on an interaction between the type of physical therapy treatment received and the initial level of fear avoidance beliefs. The nature of the interaction suggests that patients with elevated fear-avoidance beliefs benefit from fear-avoidance–based physical therapy, whereas patients with lower fear-avoidance beliefs do not benefit from fear-avoidance–based physical therapy. At 4 weeks and 6 months, a significant decrease in fear-avoidance beliefs about physical activity was observed in patients receiving the fear-avoidance–based physical therapy.

It is presently unknown if a similar interaction exists between the type of manipulative treatment received and the level of fear-avoidance-beliefs. This cohort study showed that these two patient populations did not differ significantly at baseline or four-weeks with respect to their overall FABQ scores, but we did not test the hypothesis that a subset of patients with elevated fear-avoidance-beliefs at baseline might have a differential treatment effect with manual manipulation vs. Activator type treatment. Also, FABQ scores dropped substantially in both cohorts, the meaning of which remains unclear. These issues could be explored in a future randomized trial by examining the FABQ scores with treatment response after randomization.

5.5 CLINICAL PATTERNS OF CARE

The data analysis of clinical patterns of care showed many significant differences between the Activator and manual manipulation groups with respect to the use of physical agents. The Activator clinic had significantly higher utilization of electrical muscle stimulation, laser,

postural advice, and mechanical inter-segmental mobilization table ("roller-massage" table) compared to the manual manipulation clinics.

Although electrical stimulation has not been shown to have an independent treatment effect, the literature is starting to show some preliminary evidence for the effectiveness of laser therapy. The treatment effect manipulation plus the addition of postural advice and ten minutes of intersegmental mobilization on a mechanized roller table is unknown. These differences in the utilization rates of these physical agents and postural advice may be potential confounders to the main effect of type of manipulation.

There were also significant differences in the clinical parameters of care, with respect to frequency and duration of treatment, x-ray utilization, and clinical presentation of the patient populations. Seventy percent of the patients receiving Activator treatment terminated the research study because they had received the maximum of eight treatment sessions, with an average of 9.2 treatment sessions at four-weeks. This is contrasted with only 13% of the manual manipulation patients reaching the maximum of eight treatment sessions during the study, with an average of 4.5 treatment sessions at four-weeks.

X-ray utilization rates were significantly higher in the Activator clinic, with 54% of patients receiving lumbar x-rays during the four-weeks of the research study. Neither of the manual manipulation clinics reported any patient (0%) receiving lumbar x-rays during the study period. This contrast in x-ray utilization rates may be somewhat attributable to differences in practice management or philosophy, as some chiropractic techniques are more likely to require x-rays for

analysis of vertebral misalignment or dysfunction. Another practical difference was the fact that the Activator clinic had an x-ray machine on premises, and it was easier for the Activator clinicians to simply take an x-ray if they thought such films were clinically warranted. Neither manual manipulation clinic had an x-ray machine on site, therefore those clinicians had to refer their patients to an out-patient x-ray facility to get lumbar radiographs if warranted. This additional step of referring to an outside facility may have lead to a decrease in x-ray utilization by those clinicians. However it is still interesting to note that no manual manipulation patient received a lumbar x-ray during the course of this study.

The manual manipulation patient population also differed from the Activator population with respect to clinical presentation at baseline. Eighty-one percent of the manipulation patients had an onset of pain ≤ 14 days with 18% exhibiting a flexion antalgic posture, as contrasted with 52% of Activator patients presenting with the same level of acute onset and 0% with flexion antalgia reported on the initial examination. However, there was more reported "moderate" and "severe" limitation of flexion range of motion in the Activator patients than the manual manipulation patients. These data were derived from the medical records only, and therefore may be subject to recall and/or reporting bias. However, a potential confounding situation arises when the population of patients in one cohort differs significantly with respect to duration of symptoms and the presence of antalgic lean, which is considered to be potentially indicative of more severe underlying facet or disc dysfunction.

To summarize the analysis of clinical patterns of care, the Activator patients received twice as many treatment sessions as the manual manipulation patients with significantly greater

utilization of physical agents and postural advice. The increased number of office visits and physical agents may have contributed to the overall treatment effect within the Activator cohort, but this cannot be ascertained from the present data. Also, there is the possibility of attention bias due to the increased amount of inter-personal interaction time within the Activator clinic relative to the manipulation clinics.

5.6 FINAL CONCLUSIONS

5.6.1 Reliability studies

The reliability studies resulted in some interesting findings. Visual observation of leg length inequality in the prone position seems to be a reliable procedure, but not the quantification of the magnitude of the difference in length. Rotation of the head while observing for changes in leg length is an unreliable procedure, while flexing the knees and observing such changes is indeterminate. The springing palpation findings were basically the same as reported in previous studies; i.e. that the reliability of palpation for segmental mobility is poor, while better for pain provocation.

Another important side-benefit of these reliability studies was the experience gained by designing and implementing them. The process of writing informed consent documents and clearing the research methodology through the Institutional Review Board (IRB) at the University of Pittsburgh was an educational experience. The practical aspects of implementing the studies was also educational, such as the need for providing training sessions with the

clinicians to make sure they were consistent with their examination methods and writing a procedural manual with definitions and detailed descriptions of all the examination techniques.

Lastly, experiencing first-hand the complications of computing Kappa statistics when there is a low prevalence of negative findings was also an educational and instructive process. These complications lead to further reading and subsequent understanding of the nuances of prevalence bias in reliability studies, and the important differences in using dichotomous versus continuous variables in reliability studies.

5.6.2 Cohort study

There are several important observations and conclusions to be made regarding the results of this observational cohort study, which utilized a treatment-as-usual design. Many significant differences were found between the patient populations in the two cohorts, especially with respect to their respective treatment frequency, treatment expectations, clinical presentation at baseline, and utilization rate of physical agents. Any or all of these differences may have contributed to the main treatment effects, and it is not possible to determine if significant confounding occurred as a result of these differences.

In addition to these obvious differences between the two groups, the other major limitations of this study were the relatively small sample size and the observational nature of the study design. Sample size and study design were mostly constrained by the lack of external funding to recruit a larger number of research participants and to employ a research coordinator that would be integral part of coordinating a randomized clinical trial.

The multiple and significant differences in clinical patterns of care are an intrinsic challenge with any observational research design, which permits treatment-as-usual by different clinicians. Since the participants in this cohort study were observed over the course of their care in private chiropractic clinics, it would not have been considered ethical to constrain the clinicians by imposing experimental conditions on them, and their patients.

Despite these limitations, this cohort study provided valuable information regarding some important differences between the characteristics of clinics providing Activator and manual manipulation methods of treatment. These results are important for consideration in designing a future randomized trial which can eliminate many of these potential confounders by imposing a more rigid treatment protocol that would include:

- A defined number of treatment sessions.
- The same utilization rate for physical agents, exercises, and posture education.
- Minimization of attention bias between groups.
- Randomization of patients into groups vs. self-selection.

This study highlighted the important challenges that are inherent within an observational cohort design, with the intrinsic issue of lack of control over treatment parameters between cohorts and other significant differences between the patient populations in each cohort. These issues are mostly resolved with the use of a randomized controlled trial design, and were an instructional experience in understanding the powerful nature of the experimental design vs. the observational treatment-as-usual design.

5.6.3 Expectations and Clinical Patterns of Care

This study confirmed previous reports that patient expectations are a strong predictor of clinical outcome. It was interesting to note that there was no general difference in treatment expectation between the two cohorts in this study. However, the data did show rather large differences in expectations about the effectiveness of manual manipulation and Activator and beliefs about hearing a "pop", in the respective cohorts. These data will be important to consider in a future RCT that will randomize patients to Activator vs. manual manipulation, taking beliefs and expectations at baseline into consideration as predictor variables of the outcome after randomization. More specifically, there will be a natural 2 x 2 design to this issue with two cohorts of Activator vs. manual manipulation and two "expectation groups" with one group having high expectation for Activator (low expectation for manipulation) and the other group having high expectation for manipulation (low expectation for Activator).

Randomization will therefore create four groups:

1. High expectation for manipulation – randomized to manipulation
2. Low expectation for manipulation – randomized to manipulation
3. High expectation for Activator – randomized to Activator
4. Low expectation for Activator – randomized to Activator

This design will allow for statistical evaluation of any main effect of treatment expectation on the outcomes of pain and Oswestry scores. The observational nature of this cohort study did not

allow for this type of analysis, because all the research subjects had self-selected their respective type of treatment.

It was interesting to observe a dramatic difference in the clinical patterns of care between the Activator clinic and the two manipulation clinics involved with this cohort study. The Activator patients were seen for twice as many visits as the manual manipulation patients, and were more likely to continue with their chiropractic care after the four-week termination of the study. Both groups had reductions of pain and Oswestry scores that were not statistically different, but there was no control group with which to compare for a measure of natural history without any treatment. However, from a patient-centered and cost-effectiveness perspective it would seem that to achieve the same clinical outcome, Activator treatment is more costly and time consuming.

The results of this small and underpowered cohort study should be interpreted with caution, however, and may not be generalizable to the larger population of all chiropractic patients treated with these two methods. The observational nature of this study did not allow for any experimental control over the treatment frequency and duration, and the differences in clinical patterns of care reported in this study may be due to differences in practice management styles rather than the intrinsic efficacy of the treatment procedures themselves. This particular Activator office had a pattern of seeing patients three times a week for the first four weeks of treatment, whereas the two manipulation offices participating in this study did not have this policy. Some of the difference in clinical patterns of care may relate to the clinicians' belief system in the treatments they provide, and this may need to be explored in future studies.

5.6.4 Fear Avoidance Beliefs

The data analysis showed no significant difference in Fear Avoidance Beliefs Questionnaire (FABQ) scores between the two cohorts at baseline or four-weeks. Yet both cohorts showed a reduction their respective FABQ scores (with-in group) from baseline to four weeks. This observation is very interesting, because it suggests that FABQ scores are not a static phenomenon and are subject to fluctuation over time. This suggests that there may be a correlation between successful clinical outcome and reduction of FABQ score, regardless of the specific type of treatment provided. This is an area worthy of future investigation in future RCTs involving different types of manipulation for the treatment of LBP.

APPENDIX A

LEG LENGTH ANALYSIS FORM

Date _____ Recorded by _____

Patient Number _____ Doctor _____

Side of Pain: Rt Lt Both Central

1. Prone Leg Length – Knees Extended (Position #1)

Short Leg Present?	+1	0		
Short Leg Side	Lt	Rt	None	
Amount of difference	≤ ¼”	½”	¾”	≥ 1”

(For differences ≤ ¼”, perform again with head rotation to right and left)

	Head Lt.			Head Rt.		
Does short leg change?	+1	0	-1	+1	0	-1

2. Prone Leg Length – Knees Flexed (Position #2)

Does short leg change? +1 0 -1

Key: +1 = “yes” or “short leg gets longer”
0 = “no” or “no change is observed”
-1 = “short leg gets shorter”

APPENDIX B

SPINAL PALPATION DATA FORM

Date _____ Recorded by _____

Patient Number _____ Doctor _____

Side of Pain: Rt Lt Both Central Key: +1 = HYPER mobility -1 = HYPO mobility
Y = Pain produced 0 = No mobility dysfunction N = No pain produced

Palpation Step 1 (Patient prone on table)

	Mobility Tests						***Pain Provocation Tests***					
	Left			Spinous	Right		Left		Spinous	Right		
SIJ	+1	0	-1	----	+1	0 -1	Y	N	---	Y	N	
L5-S1	+1	0	-1	+1 0 -1	+1	0 -1	Y	N	Y N	Y	N	
L4-5	+1	0	-1	+1 0 -1	+1	0 -1	Y	N	Y N	Y	N	
L3-4	+1	0	-1	+1 0 -1	+1	0 -1	Y	N	Y N	Y	N	
L2-3	+1	0	-1	+1 0 -1	+1	0 -1	Y	N	Y N	Y	N	
L1-2	+1	0	-1	+1 0 -1	+1	0 -1	Y	N	Y N	Y	N	

Palpation Step 2 (Patient at edge of table, feet on floor, measure table height)

	Pain Provocation Tests		
	Left	Spinous	Right
L5-S1	Y N	Y N	Y N
L4-5	Y N	Y N	Y N

L3-4	Y N	Y N	Y N
L2-3	Y N	Y N	Y N
L1-2	Y N	Y N	Y N

Palpation Step 3 (Patient at edge of table, feet elevated from floor, give a rest between joint and spinous tests)

Pain Provocation Tests

	Left	Spinous	Right
L5-S1	Y N	Y N	Y N
L4-5	Y N	Y N	Y N
L3-4	Y N	Y N	Y N
L2-3	Y N	Y N	Y N
L1-2	Y N	Y N	Y N

APPENDIX C

OSWESTRY LOW BACK PAIN DISABILITY INDEX QUESTIONNAIRE

Revised Oswestry Questionnaire

PLEASE READ: This questionnaire is designed to enable your health care provider to understand how much your low back pain has affected your ability to manage everyday activities. Please answer each section by circling the ONE choice that most applies to you. We realize you may feel that more than one statement may relate to you, but PLEASE JUST CIRCLE THE ONE CHOICE WHICH MOST CLOSELY DESCRIBES YOUR PROBLEM RIGHT NOW.

SECTION 1 – PAIN INTENSITY A The pain comes and goes and is very mild. B The pain is mild and does not vary much. C The pain comes and goes and is moderate. D The pain is moderate and does not vary much. E The pain comes and goes and is severe. F The pain is severe and does not vary much.	Section 6 – Standing A I can stand as long as I like without pain. B I have some pain while standing but it does not increase with time. C I cannot stand for longer than one hour without increasing pain. D I cannot stand for longer than ½ hour without increasing pain. E I cannot stand for longer than 10 minutes without increasing pain. F I avoid standing because it increases the pain straight away.
SECTION 2 – PERSONAL CARE A I would not have to change my way of washing or dressing in order to avoid pain. B I do not normally change my way of washing and dressing even though it causes some pain. C Washing and dressing increase the pain but I manage not to change my way of doing it. D Washing and dressing increase the pain and I find it necessary to change my way of doing it. E Because of the pain, I am unable to do *some* washing and dressing without help F Because of the pain I am unable to do *any* washing and dressing without help.	Section 7 – Sleeping A I get no pain in bed. B I get pain in bed but it does not prevent me from sleeping well. C Because of pain my normal night's sleep is reduced by less than D ¼. E Because of pain my normal night's sleep is reduced by less than F ½. Because of pain my normal night's sleep is reduced by less than ¾. Pain prevents me from sleeping at all.
SECTION 3 – LIFTING A I can lift heavy weights without extra pain. B I can lift heavy weights but it causes extra pain. C Pain prevents me from lifting heavy weights off the floor. D Pain prevents me from lifting heavy weights off the floor, but I can manage if they are conveniently positioned (e.g.	Section 8 – Social Life A My social life is normal and gives me no pain. B My social life is normal but increases the degree of pain. C Pain has no significant effect on my social life apart from limiting my more energetic interests, e.g. dancing, etc. D Pain has restricted my social life and I do not go out very often.

E on a table). F Pain prevents me from lifting heavy weights but I can manage light to medium weights if they are conveniently positioned. I can only lift very light weights, at the most.	E Pain has restricted my social life to my home. F I have hardly any social life because of the pain.
SECTION 4 – WALKING A Pain does not prevent me from walking any distance. B Pain prevents me from walking more than one mile. C Pain prevents me from walking more than ½ mile. D Pain prevents me from walking more than ¼ mile E I can only walk while using a cane or on crutches. F I am in bed most of the time and have to crawl to the toilet.	Section 9 – Traveling A I get no pain while traveling. B I have some pain while traveling but none of my usual forms of travel make it any worse. C I have extra pain while traveling but it does not compel me to seek alternate forms of travel. D I get extra pain while traveling that compels me to seek alternative forms of travel. E Pain restricts all forms of travel. F Pain prevents all forms of travel except that done lying down.
SECTION 5 – SITTING A I can sit in any chair as long as I like without pain B I can only sit in my favorite chair as long as I like. C Pain prevents me from sitting more than 1 hour. D Pain prevents me from sitting more than ½ hour. E Pain prevents me from sitting more than ten minutes. F Pain prevents me from sitting at all.	Section 10 – Changing Degree of Pain A My pain is rapidly getting better. B My pain fluctuates but overall is definitely getting better. C My pain seems to be getting better, but improvement is slow at present. D My pain is neither getting better nor worse. E My pain is gradually worsening. F My pain is rapidly worsening.

Print name:	Signature:	Date:
Comments:		Oswestry #

APPENDIX D

NUMERIC PAIN RATING SCALE

Please use the 3 scales below to rate your pain.
Circle one number on each line.

Rate your Pain	**0 = No Pain**										**10 = Severe, intense pain**
Right Now	0	1	2	3	4	5	6	7	8	9	10
Worst in past 24 hours	0	1	2	3	4	5	6	7	8	9	10
Best in past 24 hours	0	1	2	3	4	5	6	7	8	9	10

_____	_____	_____
Patient Name (printed)	Patient Signature	Date

APPENDIX E

ROLAND-MORRIS LOW BACK PAIN QUESTIONNAIRE

When your back hurts, you may find it difficult to do some of the things you normally do.

This list contains some sentences that people have used to describe themselves when they have back pain. When you read them, you may find that some stand out because they describe you *today*. As you read the list, think of yourself *today*. When you read a sentence that describes you *today*, mark the box next to it. If the sentence does not describe you, then leave the space blank and go on to the next one. **Remember, only mark the sentence if you are sure that it describes you *today***.

1. I stay at home most of the time because of the pain in my back.
2. I change position frequently to try and make my back comfortable.
3. I walk more slowly than usual because of the pain in my back.
4. Because of the pain in my back, I am not doing any of the jobs that I usually do around the house.
5. Because of the pain in my back, I use a handrail to get upstairs.
6. Because of the pain in my back, I lie down to rest more often.
7. Because of the pain in my back, I have to hold on to something to get out of a reclining chair.
8. Because of the pain in my back, I ask other people to do things for me.
9. I get dressed more slowly than usual because of the pain in my back.
10. I only stand up for short periods of time because of the pain in my back.
11. Because of the pain in my back, I try not to bend or kneel down.
12. I find it difficult to get out of a chair because of the pain in my back.
13. My back hurts most of the time.
14. I find it difficult to turn over in bed because of the pain in my back.
15. My appetite is not very good because of the pain in my back.
16. I have trouble putting on my socks (or stockings) because of the pain in my back.

17. I only walk short distances because of the pain in my back.
18. I sleep less because of the pain in my back.
19. Because of the pain in my back, I get dressed with help from someone else.
20. I sit down for most of the day because of the pain in my back.
21. I avoid heavy jobs around the house because of the pain in my back.
22. Because of the pain in my back, I am more irritable and bad tempered with people.
23. Because of the pain in my back, I go upstairs more slowly than usual.
24. I stay in bed most of the time because of the pain in my back.

APPENDIX F

FEAR AVOIDANCE BELIEFS QUESTIONNAIRE

FABQ – Physical Activity

Here are some of the things which other patients have told us about their pain. For each statement please circle any number from 0 to 6 to say how much physical activity such as bending, lifting, walking or driving affect or would affect your back pain.

		Completely Disagree			Unsure			Completely agree
1.	My pain was caused by physical activity	0	1	2	3	4	5	6
2.	Physical activity makes my pain worse	0	1	2	3	4	5	6
3.	Physical activity might harm my back	0	1	2	3	4	5	6
4.	I should not do physical activities which (might) make my pain worse	0	1	2	3	4	5	6
5.	I cannot do physical activities which (might) make my pain worse	0	1	2	3	4	5	6

FABQ – WORK

The following statements are about how your normal work affects or would affect your back.

		Completely Disagree			Unsure			Completely Agree
6.	My pain was caused by my work or by an accident at work	0	1	2	3	4	5	6
7.	My work aggravated my pain	0	1	2	3	4	5	6
8.	I have a claim for compensation for my pain	0	1	2	3	4	5	6
9.	My work is too heavy for me	0	1	2	3	4	5	6

10.	My work makes or would make my pain worse	0	1	2	3	4	5	6
11.	My work might harm my back	0	1	2	3	4	5	6
12.	I should not do my regular work with my present pain	0	1	2	3	4	5	6
13.	I cannot do my normal work with my present pain	0	1	2	3	4	5	6
14.	I cannot do my normal work until my pain is treated	0	1	2	3	4	5	6
15.	I do not think that I will be back to my normal work within 3 months	0	1	2	3	4	5	6
16.	I do not think that I will ever be able to go back to work	0	1	2	3	4	5	6

Print Name _____ Signature_____

Date_____ FABQ-PA_____ FABQ-WORK_____

APPENDIX G

PATIENT GLOBAL INDEX OF CHANGE FORM

Since the start of my care at this office, my overall status is:
(check one box only)

1. □ Very Much Improved
2. □ Much Improved
3. □ Minimally Improved
4. □ No Change
5. □ Minimally Worse
6. □ Much Worse
7. □ Very Much Worse

Please complete the sentence below:

Overall, since I first started treatment at this office I feel about _ _____ % improved.

Name _____ Signature _____

Date _____

APPENDIX H

TREATMENT EXPECTATIONS QUESTIONNAIRE

1. Have you ever seen a doctor of chiropractic for back or neck pain before today's visit?

□ YES □ NO

2. If yes, what type of chiropractic adjustment or manipulation did you receive in the past?

□ Activator Methods □ Hands-on adjustment □ Both

□ Other _____

3. What was your previous experience with this type of chiropractic treatment?

Activator adjustment:

1	2	3	4	5	6	7
Not helpful at all	Not sure if Helpful	Slightly Helpful	Somewhat Helpful	Helpful	Very Helpful	Extremely Helpful

Hands-on adjustment:

1	2	3	4	5	6	7
Not helpful at all	Not sure if Helpful	Slightly Helpful	Somewhat Helpful	Helpful	Very Helpful	Extremely Helpful

4. One month from now, how do you expect your back pain to be?

1	2	3	4	5	6	7
Much worse	A little worse	About the same	A little better	Moderately better	Much better	Completely gone

5. Do you expect to hear or feel a “pop” or “click” in your back during your chiropractic treatment?
□ YES □ NO □ NOT SURE

6. How do you feel about hearing your back pop or click during chiropractic treatment?
□ GOOD □ NOT GOOD □ NOT SURE

7. How helpful do believe that the following chiropractic methods are for relieving back pain?

Activator adjustment:

1	2	3	4	5	6	7
Don’t expect it to help at all	Low expectation that it will help	Not sure if it will help	Probably will be helpful	Expect it to help at least a little	Will definitely be helpful	Expect it will help greatly

Hands-on adjustment:

1	2	3	4	5	6	7
Don’t expect it to help at all	Low expectation that it will help	Not sure if it will help	Probably will be helpful	Expect it to help at least a little	Will definitely be helpful	Expect it will help greatly

APPENDIX I

SOCIO-DEMOGRAPHIC QUESTIONNAIRE

What is your gender?
□ Male □ Female

What is your current age? I am _____ years old. My date of birth is: _____

What is your ethnic or racial background?
□ White Caucasian
□ African American
□ Latin American
□ Asian American
□ Other (specify) _____

What is your current marital status?
□ Single (never married)
□ Married
□ Living with significant other
□ Divorced/separated
□ Widowed

What level of education have you completed?
□ Less than high school
□ Graduated from high school
□ Some college
□ Graduated from college
□ Some post graduate work
□ Post graduate degree

What is your approximate yearly household income?

□ Less than $20,000
□ $20,000 to 35,000
□ $35,001 to $50,000
□ $50,001 to 70,000
□ Greater than 70,000

What is your current employment status? (mark the **one** category that best describes your current status)

□ Working regular duty full time
□ Working regular duty part time
□ Working light duty or modified position full time
□ Working light duty or modified position part time
□ Temporarily unable to work due to health status
□ Permanently unable to work or retired due to health status
□ Unemployed
□ Home maker (not working outside the home)
□ Student (not currently working)

Do you currently smoke or use tobacco products?

□ Yes □ No

How did your back pain develop?

□ Activities of daily living
□ Motor vehicle accident
□ Work related
□ Sports related
□ Following surgery
□ Slip or fall
□ Other (specify) _____

Are you currently taking any medications, whether over-the-counter or prescribed?

□ Yes □ No

If yes, please list the medications you are taking: _____
_____
_____

As a result of your back pain are you presently receiving worker's compensation?

□ Yes □ No

Is your current condition the subject of any legal action?

□ Yes □ No

Have you ever had back surgery?

□ Yes □ No

Have you ever been told by a doctor that you have a significant spinal problem such as scoliosis (curvature of the spine), stenosis (severe arthritis of the spine), or spondylolisthesis (slippage of vertebrae)?

□ Yes □ No

APPENDIX J

STATISTICS OUTPUT AND COMMAND SYNTAX (STATA)

```
regress pain4 cohort paincenter expect0 (FINAL MODEL)
```

```
      Source |       SS       df       MS              Number of obs =      92
-------------+------------------------------           F(  3,    88) =    7.02
       Model |  85.1761505     3  28.3920502           Prob > F      =  0.0003
    Residual |  356.149936    88  4.04715837           R-squared     =  0.1930
-------------+------------------------------           Adj R-squared =  0.1655
       Total |  441.326087    91  4.84973722           Root MSE      =  2.0118
------------------------------------------------------------------------------
       pain4 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      cohort |  -.8075978   .4406673    -1.83   0.070    -1.683331    .0681359
  paincenter |   .2717994   .1289979     2.11   0.038     .0154431    .5281556
     expect0 |  -.6150679   .2043601    -3.01   0.003    -1.021191   -.2089451
       _cons |    6.62339   1.223084     5.42   0.000     4.192768    9.054013
------------------------------------------------------------------------------
```

```
regress paindiff cohort expect0
```

```
      Source |       SS       df       MS              Number of obs =      92
-------------+------------------------------           F(  2,    89) =    4.65
       Model |  50.8000494     2  25.4000247           Prob > F      =  0.0120
    Residual |  485.816798    89  5.45861571           R-squared     =  0.0947
-------------+------------------------------           Adj R-squared =  0.0743
       Total |  536.616848    91  5.89688844           Root MSE      =  2.3364

------------------------------------------------------------------------------
    paindiff |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      cohort |   .4335398   .5117719     0.85   0.399    -.5833399     1.45042
     expect0 |   .6158241   .2373353     2.59   0.011     .1442439    1.087404
       _cons |  -.1375407   1.420438    -0.10   0.923    -2.959921     2.68484
------------------------------------------------------------------------------
```

```
. regress paindiff cohort age0

      Source |       SS       df       MS              Number of obs =      92
-------------+------------------------------           F(  2,    89) =    1.42
       Model |  16.6209224     2  8.31046119           Prob > F      =  0.2463
    Residual |  519.531252    89  5.83742979           R-squared     =  0.0310
-------------+------------------------------           Adj R-squared =  0.0092
       Total |  536.152174    91  5.89178213           Root MSE      =  2.4161

------------------------------------------------------------------------------
    paindiff |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      cohort |   .6706146   .5510676     1.22   0.227    -.4243449    1.765574
        age0 |   .0106939   .0172832     0.62   0.538    -.0236475    .0450353
       _cons |   3.028957    .747516     4.05   0.000     1.543659    4.514255
------------------------------------------------------------------------------


. regress paindiff cohort age0 expect0

      Source |       SS       df       MS              Number of obs =      92
-------------+------------------------------           F(  3,    88) =    3.09
       Model |  51.0983919     3  17.0327973           Prob > F      =  0.0312
    Residual |  485.053782    88   5.5119748           R-squared     =  0.0953
-------------+------------------------------           Adj R-squared =  0.0645
       Total |  536.152174    91  5.89178213           Root MSE      =  2.3478

------------------------------------------------------------------------------
    paindiff |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      cohort |   .4239625   .5444913     0.78   0.438    -.6580996    1.506024
        age0 |   .0018726   .0171609     0.11   0.913    -.0322311    .0359762
     expect0 |   .6094826   .2436953     2.50   0.014     .1251893    1.093776
       _cons |  -.1825724   1.475307    -0.12   0.902    -3.114436    2.749291
------------------------------------------------------------------------------


. regress paindiff cohort age0 expect0  Interaction3

      Source |       SS       df       MS              Number of obs =      92
-------------+------------------------------           F(  4,    87) =    2.46
       Model |  54.5430312     4  13.6357578           Prob > F      =  0.0510
    Residual |  481.609143    87  5.53573727           R-squared     =  0.1017
-------------+------------------------------           Adj R-squared =  0.0604
       Total |  536.152174    91  5.89178213           Root MSE      =  2.3528

------------------------------------------------------------------------------
    paindiff |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      cohort |  -2.001068   3.122258    -0.64   0.523    -8.206895    4.204758
        age0 |    .001734   .0171987     0.10   0.920    -.0324504    .0359183
     expect0 |   .4667342   .3039582     1.54   0.128    -.1374157    1.070884
Interaction3 |   .3914841   .4962837     0.79   0.432    -.5949334    1.377902
       _cons |   .6550473   1.820285     0.36   0.720    -2.962967    4.273061
```

ANOVA tables of output from different regression analyses with pain difference scores from baseline to four-weeks as the dependent variable. Key: paindiff = four-week pain score subtracted from baseline pain score, cohort = Activator (coded "0") or manipulation (coded "1"), Age0 = age in years, expect0 = treatment expectation score at baseline (0-7 scale), and Interaction3 = expect0 x cohort.

regress odi4 cohort odicenter expect0

```
      Source |       SS       df       MS              Number of obs =      92
-------------+------------------------------           F(  3,    88) =    3.59
       Model |  1696.18102     3  565.393672           Prob > F      =  0.0168
    Residual |  13863.6451    88  157.541421           R-squared     =  0.1090
-------------+------------------------------           Adj R-squared =  0.0786
       Total |  15559.8261    91    170.9871           Root MSE      =  12.552

------------------------------------------------------------------------------
        odi4 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      cohort |  -1.468771   2.749729    -0.53   0.595     -6.93328    3.995737
   odicenter |   .2542185   .1030776     2.47   0.016     .0493735    .4590635
     expect0 |   -2.55532   1.277345    -2.00   0.049    -5.093775   -.0168642
       _cons |   32.10602   7.644142     4.20   0.000      16.9149    47.29715
------------------------------------------------------------------------------
```

regress odidiff cohort

```
      Source |       SS       df       MS              Number of obs =      92
-------------+------------------------------           F(  1,    90) =    1.66
       Model |  429.457568     1  429.457568           Prob > F      =  0.2007
    Residual |  23257.4446    90  258.416051           R-squared     =  0.0181
-------------+------------------------------           Adj R-squared =  0.0072
       Total |  23686.9022    91  260.295628           Root MSE      =  16.075

------------------------------------------------------------------------------
     odidiff |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      cohort |   4.372037   3.391434     1.29   0.201     -2.36564    11.10971
       _cons |   20.67925   2.208116     9.37   0.000     16.29244    25.06605
------------------------------------------------------------------------------
```

regress odidiff cohort expect0

```
      Source |       SS       df       MS              Number of obs =      92
-------------+------------------------------           F(  2,    89) =    2.75
       Model |  1378.67582     2  689.337911           Prob > F      =  0.0694
    Residual |  22308.2264    89  250.654229           R-squared     =  0.0582
-------------+------------------------------           Adj R-squared =  0.0370
       Total |  23686.9022    91  260.295628           Root MSE      =  15.832

------------------------------------------------------------------------------
     odidiff |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      cohort |   2.556593   3.467949     0.74   0.463    -4.334147    9.447333
     expect0 |    3.12971   1.608269     1.95   0.055    -.0658854    6.325306
       _cons |   2.432444   9.625397     0.25   0.801    -16.69301     21.5579
------------------------------------------------------------------------------
```

regress odidiff cohort age0

```
      Source |       SS       df       MS              Number of obs =      92
-------------+------------------------------           F(  2,    89) =    0.84
       Model |  440.709122     2  220.354561           Prob > F      =  0.4335
    Residual |  23246.1931    89   261.19318           R-squared     =  0.0186
-------------+------------------------------           Adj R-squared = -0.0034
       Total |  23686.9022    91  260.295628           Root MSE      =  16.161

------------------------------------------------------------------------------
     odidiff |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      cohort |   4.662785   3.686166     1.26   0.209    -2.661548    11.98712
        age0 |   -.023995   .1156098    -0.21   0.836    -.2537091    .2057192
       _cons |   21.60916   5.000237     4.32   0.000      11.6738    31.54453
------------------------------------------------------------------------------
```

regress odidiff cohort age0 expect0

```
      Source |       SS       df       MS              Number of obs =      92
-------------+------------------------------           F(  3,    88) =    1.95
       Model |  1476.64612     3  492.215374           Prob > F      =  0.1274
    Residual |  22210.2561    88  252.389273           R-squared     =  0.0623
-------------+------------------------------           Adj R-squared =  0.0304
       Total |  23686.9022    91  260.295628           Root MSE      =  15.887

------------------------------------------------------------------------------
     odidiff |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      cohort |   3.310763   3.684451     0.90   0.371    -4.011309    10.63283
        age0 |  -.0723491   .1161239    -0.62   0.535    -.3031209    .1584227
     expect0 |   3.340875   1.649032     2.03   0.046      .063771    6.61798
       _cons |   4.005182   9.983075     0.40   0.689    -15.83408    23.84444
------------------------------------------------------------------------------
```

regress odidiff cohort age0 expect0 Interaction3

```
      Source |       SS       df       MS              Number of obs =      92
-------------+------------------------------           F(  4,    87) =    1.48
       Model |  1508.49427     4  377.123568           Prob > F      =  0.2154
    Residual |  22178.4079    87  254.924229           R-squared     =  0.0637
-------------+------------------------------           Adj R-squared =  0.0206
       Total |  23686.9022    91  260.295628           Root MSE      =  15.966

------------------------------------------------------------------------------
     odidiff |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      cohort |  -4.062971   21.18784    -0.19   0.848    -46.17609    38.05015
        age0 |  -.0727705   .1167117    -0.62   0.535    -.3047476    .1592066
     expect0 |   2.906824   2.062679     1.41   0.162    -1.192975    7.006623
Interaction3 |   1.190377   3.367812     0.35   0.725    -5.503514    7.884267
       _cons |   6.552113   12.35257     0.53   0.597    -17.99995    31.10418
------------------------------------------------------------------------------
```

ANOVA tables of output from four regression analyses with mean Oswestry (ODI) difference scores from baseline to four-weeks. Key: odidiff = four-week ODI score subtracted from baseline ODI score, cohort = Activator (coded “0”) or manipulation (coded “1”) group, age0 = age in years, expect0 = treatment expectation at baseline (0-7 scale), Interaction3 = expect0 x cohort.

BIBLIOGRAPHY

1. Bigos S, Bowyer O, Braen G. Acute Low Back Problems in Adults. Clinical Practice Guideline No. 14. Rockville, MD: Agency for Health Care Policy and Research, Public Health Service, U.S. Dept. of Health and Human Services.; 1994.
2. Cunningham LS, Kelsey JL. Epidemiology of musculoskeletal impairments and associated disability. Am J Public Health 1984;74:574-9.
3. Walker BF, Muller R, Grant WD. Low back pain in Australian adults: prevalence and associated disability. Journal of Manipulative & Physiological Therapeutics 2004;27(4):238-44.
4. Hestbaek L, Leboeuf-Yde C, Engberg M, Lauritzen T, Bruun NH, Manniche C. The course of low back pain in a general population. Results from a 5-year prospective study. Journal of Manipulative & Physiological Therapeutics 2003;26(4):213-9.
5. Anderson R, Meeker WC, Wirick BE, Mootz RD, Kirk DH, Adams A. A meta-analysis of clinical trials of spinal manipulation. Journal of Manipulative & Physiological Therapeutics 1992;15(3):181-94.
6. Assendelft WJ, Morton SC, Yu EI, Suttorp MJ, Shekelle PG. Spinal manipulative therapy for low back pain. A meta-analysis of effectiveness relative to other therapies. Annals of Internal Medicine 2003;138(11):871-81.
7. Bronfort G, Haas M, Evans R L, Bouter LM. Efficacy of spinal manipulation and mobilization for low back pain and neck pain: a systematic review and best evidence synthesis. Spine Journal: Official Journal of the North American Spine Society 2004;4(3):335-56.
8. Koes BW, Assendelft WJ, van der Heijden GJ, Bouter LM. Spinal manipulation for low back pain. An updated systematic review of randomized clinical trials. Spine 1996;21(24):2860-71.
9. Lawrence V. Spinal Manipulation for Low-Back Pain: A Meta-analysis. ACP Journal Club Mar-April 1993;118:49.
10. Mohseni-Bandpei M A, Stephenson R, Richardson B. Spinal manipulation in the treatment of low back pain: a review of the literature with particular emphasis on randomized controlled clinical trials. Physical Therapy Reviews 1998;3(4):185-94.
11. Shekelle PG, Adams A. The Appropriateness of Spinal Manipulation for Low-Back Pain: Indications and Ratings by a Multidisciplinary Expert Panel. In; 1991; Santa Monica: RAND (R-4025/2-CCR/FCER); 1991.
12. Shekelle PG, Adams AH, Chassin MR, Hurwitz EL, Brook RH. Spinal manipulation for low-back pain. Annals of Internal Medicine 1992;117(7):590-8.
13. Bergmann TF. Various forms of chiropractic technique. J Chiro Tech 1993;5(2):53 - 55.
14. Christensen MG, Kerkhoff D, Kollash MW. Job Analysis of Chiropractic. Greeley, CO: National Board of Chiropractic Examiners; 2000.

15. Christensen M G, M organ D R. J ob A nalysis o f C hiropractic. Greeley, C O: N ational Board of Chiropractic Examiners; 1993.
16. Fuhr AW, Colloca CJ, Green JR, Keller TS. Activator Methods Chiropractic Technique. St. Louis: Mosby; 1997.
17. Shearar K A, Colloca C J, W hite H L. A R andomized C linical T rial of M anual V ersus Mechanical Force Manipulation in the Treatment of Sacroiliac Joint Syndrome. J Manip Physiol Ther 2005;28(7):493-501.
18. Gemmell HA, Jacobson BH. The immediate effect of activator vs. meric adjustment on acute low back pain: a randomized controlled trial. J Manipulative Physiol Ther 1995;18(7):453-6.
19. Taylor SH, Arnold ND, Biggs L, Colloca CJ, Mierau DR, Symons BP, et al. A review of the l iterature p ertaining to the ef ficacy, safety, educ ational r equirements, us es, a nd us age of mechanical a djusting de vices. P art 2 of 2. Journal of t he C anadian C hiropractic A ssociation 2004;48(2):152-179.
20. Taylor SH, Arnold ND, Biggs L, Colloca CJ, Mierau DR, Symons BP, et al. A review of the lite rature p ertaining t o t he e fficacy, s afety, e ducational r equirements, us es, a nd us age of mechanical a djusting de vices. P art 1 of 2. Journal of t he C anadian C hiropractic A ssociation 2004;48(1):74-88.
21. Schiotz E , C yriax J . M anipulation P ast a nd P resent. London: W illiam H einemann Medical Books, LTD.; 1975.
22. Maitland G. Vertebral Manipulation, 5th Edition. London: Butterworths; 1986.
23. Maigne R. Low back pain of lumbar origin. Arch Phys Med Rehabil 1980;61:389-395.
24. Lewit K. Manipulative therapy in rehabilitation of the motor system. Oxford: Butterworth Heinemann; 1991.
25. Greenman PE. Principles of M anual M edicine. Baltimore, MD: W illiams & W ilkins; 1989.
26. Cox JM. M echanism, DIagnosis, a nd T reatment of Low B ack P ain. Baltimore, M D: Williams and Wilkins; 1980.
27. Schafer RC, Faye LJ. Motion Palpation and Chiropractic Technique. Huntington Beach, CA: The Motion Palpation Institute; 1989.
28. Assendelft WJJ, Morton SC, Yu Emily I, Suttorp MJ, Shekelle PG. Spinal manipulative therapy for low-back pain. Cochrane Database of Systematic Reviews 2004;2:2.
29. Koes B W, B outer LM, va n de r H eijden G J. M ethodological qua lity of r andomized clinical trials on treatment efficacy in low back pain. Spine 1995;20(2):228-35.
30. Mathews J A, Mills S B, Jenkins V M, Grimes S M, Morkel M J, Mathews W , et al. Back pain and sciatica: controlled trials of manipulation, traction, sclerosant and epidural injections. British Journal of Rheumatology 1987;26(6):416-23.
31. Pengel HM, Maher CG, Refshauge KM. Systematic review of conservative interventions for subacute low back pain. Clinical Rehabilitation 2002;16(8):811-20.
32. Twomey L, Taylor J. Exercise and spinal manipulation in the treatment of low back pain. Spine 1995;20(5):615-9.
33. van T ulder M W, K oes BW, B outer LM. C onservative t reatment of acu te and chronic nonspecific l ow ba ck p ain. A s ystematic r eview of r andomized c ontrolled t rials of t he m ost common interventions. Spine 1997;22(18):2128-56.

34. Bronfort G, Haas M, Evans R, Kawchuk G, Dagenais S. Evidence-informed management of chronic low back pain with spinal manipulation and mobilization. The Spine Journal 8 (2008) 213–225 2008;8:213-225.
35. Assendelft W J, K oes B W, K nipschild P G, Bouter LM. T he r elationship be tween methodological qu ality a nd c onclusions i n r eviews o f s pinal manipulation. J AMA 1995;274(24):1942-8.
36. Haldeman S, Chapman-Smith D , Petersen D M, e ditors. G uidelines f or C hiropractic Quality Assurance and Practice Parameters. Gaithersburg, MD: Aspen Publishers, Inc.; 1993.
37. Henderson DJ, M ior S, Vernon H, editors. C linical G uidelines for C hiropractic P ractice in Canada. Toronto, Ontario: The Canadian Chiropractic Association; 1996.
38. Cooperstein R, Perle SM, Gatterman MI, Lantz C, Schneider MJ. Chiropractic technique procedures for specific low back conditions: Characterizing the literature. J Manip Physiol Ther 2001;24(6):407-24.
39. Gatterman M I, C ooperstein R , P erle S M, Lantz C , S chneider M J. R ating s pecific chiropractic t echnique p rocedures f or c ommon low ba ck c onditions. J Manip P hysiol T her 2001;24(7):449-56.
40. Spitzer W O, L eBlanc FE, D upuis M . S cientific a pproach t o t he a ssessment a nd management of activity-related s pinal di sorders. A m onograph f or c linicians. R eport of t he Quebec Task Force on Spinal Disorders. Spine 1987;12(7 Suppl):S1-59.
41. Atlas S J, Deyo RA, P atrick D L, C onvery K, K eller R B, S inger D E. T he Q uebec T ask Force cl assification for Spinal D isorders and the s everity, treatment, and outcomes of s ciatica and lumbar spinal stenosis. Spine 1996;21(24):2885-92.
42. Loisel P, Vachon B, Lemaire J, Durand MJ, Poitras S, Stock S, et al. Discriminative and predictive validity assessment of the quebec task force classification. Spine 2002;27(8):851-7.
43. McKenzie R A. The Lumbar S pine: M echanical D iagnosis and T herapy. N ew Zealand: Spinal Publications, Inc.; 1981.
44. Delitto A , C ibulka M T, Erhard R E, B owling R W, T enhula JA. E vidence f or us e of a n extension-mobilization c ategory i n a cute l ow b ack s yndrome: a pr escriptive va lidation pi lot study. Physical Therapy 1993;73(4):216-22.
45. Delitto A , E rhard R E, B owling R W. A tr eatment-based c lassification a pproach t o l ow back s yndrome: i dentifying a nd s taging pa tients f or c onservative t reatment. P hysical T herapy 1995;75(6):470-85.
46. Fritz JM, Delitto A, Erhard RE. Comparison of classification-based physical therapy with therapy based on clinical practice guidelines for patients with acute low back pain: a randomized clinical trial. Spine 2003;28(13):1363-72.
47. Flynn T, Fritz J, Whitman J, Wainner R, Magel J, Rendeiro D, et al. A clinical prediction rule for cl assifying pa tients w ith l ow back pa in w ho de monstrate s hort-term imp rovement w ith spinal manipulation. Spine 2002;27(24):2835-43.
48. Childs JD, Fritz JM, Piva SR, Erhard RE. Clinical decision making in the identification of pa tients like ly to benefit f rom s pinal m anipulation: a t raditional ve rsus a n e vidence-based approach. Journal of Orthopaedic and Sports Physical Therapy 2003;33(5):259-72.
49. Petersen T , Laslett M , Thorsen H , M anniche C, E kdahl C , Jacobsen S . D iagnostic classification of non -specific l ow back pa in. A ne w s ystem i ntegrating p atho-anatomic and clinical categories. Physiotherapy Theory and Practice 2003;19:213-237.

50. Murphy DR, Hurwitz EL. A theoretical model for the development of a diagnosis-based clinical decision rule for the management of patients with spinal pain. BMC Musculoskeletal Disorders 2007;8:75.
51. Childs J, Fritz J, Flynn T, Irrgang J, Johnson K, Majkowski G, et al. A Clinical Prediction Rule To Identify Patients with Low Back Pain
Most Likely To Benefit from Spinal Manipulation: A Validation Study. Ann Intern Med 2004;141:920-928.
52. McGill S M. Low Back Disorders: Evidence-Based Prevention and Rehabilitation. Champaign: Human Kinetics; 2002.
53. Childs J, Piva SR, Fritz J. Responsiveness of the Numeric Pain Rating Scale in Patients with Low Back Pain. Spine 2005;30(11):1331-1334.
54. Fairbank J C, Couper J, Davies J B, O'Brien J P, Fairbank J C, Couper J, et al. The Oswestry low back pain disability questionnaire. Physiotherapy 1980;66(8):271-3.
55. Stratford PW, Binkley JM. Measurement properties of the RM-18: A modified version of the Roland-Morris Disability Scale. Spine 1997;22:2416-2421.
56. Waddell G, Newton M, Henderson I, Somerville D, Main CJ. A Fear-Avoidance Beliefs Questionnaire (FABQ) and the role of fear-avoidance beliefs in chronic low back pain and disability. Pain 1993;52(2):157-68.
57. Farrar J T, Young J JP, LaMoreaux L, Werth J L, Poole R M. Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. Pain 2001;94(2):149-158.
58. Albright J, Allman R, Bonfiglio R P, Conill A, Dobkin B, Guccione A A. Philadelphia Panel evidence-based clinical practice guidelines on selected rehabilitation interventions for low back pain. Phys Ther 2001;81(10):1641-74.
59. Hurwitz E L, Morgenstern H, Harber P, Kominski G F, Belin T R, Yu F, et al. The Effectiveness of Physical Modalities Among Patients With Low Back Pain Randomized to
Chiropractic Care: Findings From the UCLA Low Back Pain Study. J Manip Phys Ther 2002;25(1):10-20.
60. Sim J, Wright C C. The Kappa statistic in reliability studies: Use, interpretation, and sample size requirements. Physical Therapy 2005;85(3):257-268.
61. Lee J. A note on the comparison of group means based on repeated measurements of the same subject. J Chron Dis 1980;33:673-675.
62. Norman GR. Issues in the use of change scores in randomized trials. J Clin Epidemiol 1989;42(11):1097-1105.
63. Landis J R, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33:159-174.
64. Schneider MJ, Homonai R, Moreland B, Delitto A. Interexaminer reliability of the prone leg length analysis procedure. J Manip Physiol Ther 2007;30(7):514-21.
65. Schneider MJ, Erhard R, Brach J, Tellin W, Imbarlina F, Delitto A. Spinal palpation for lumbar segmental mobility and pain provocation: An interexaminer reliability study. J Manip Physiol Ther 2008;31(6):465-73.
66. Fritz J, Irrgang J. A Comparison of a Modified Oswestry Low Back Pain Disability Questionnaire and the Quebec Back Pain Disability Scale. Physical Therapy 2001; 81(2):776-788.

67. George S Z, F ritz J M, B ialosky J E, D onald D A. T he e ffect of a f ear-avoidance-based physical t herapy i ntervention f or pa tients w ith acute l ow ba ck pa in: r esults of a r andomized clinical trial. Spine 2003;28(23):2551-60.