The Chiropractic Hospital-Based Interventions Research Outcomes Study: Consistency of Outcomes Between Doctors of Chiropractic Treating Patients With Acute Lower Back Pain

Jeffrey A. Quon, DC, MHSc, PhD, FCCS(C), a,b,c Paul B. Bishop, DC, MD, PhD, d,e and Brian Arthur, DC, MSc b,c

Abstract

Objective: The aim of this study was to determine if effectiveness differs between community-based doctors of chiropractic administering standardized evidence-based care that includes high-velocity low-amplitude spinal manipulative therapy (SMT) for acute low back pain (LBP).

Methods: A secondary analysis of randomized controlled trial and observational pilot study data was performed with nonrandom allocation to 4 DCs. Patients included those with Quebec Task Force categories less than or equal to 2 and acute LBP of 2 to 4 weeks’ duration. The intervention provided was clinical practice guidelines–based care including high-velocity low-amplitude SMT. Primary outcomes included changes from baseline in modified Roland Disability Questionnaire (RDQ) at 24 weeks. Comparisons of simple main effects at 24 weeks and of marginal main effects in repeated-measures analyses were performed.

Results: Between groups, adjusted point-specific differences in RDQ change were minimally clinically important but not statistically significant at 24 weeks (largest pairwise difference, −3.1; 95% confidence interval, −6.3 to 0.1; overall P = .10). However, in optimal analyses that considered the repeated nature of the measurements for each outcome, significant differences in marginal mean RDQ changes were found between groups (largest pairwise difference, −3.8; 95% confidence interval, −4.9 to 2.6; overall P = .03).

Conclusions: Overall, DCs differed modestly in their effectiveness in improving LBP-specific disability. The point estimates mirrored typically reported effect sizes from recent systematic reviews of SMT; however, confidence limits did not exclude clinically negligible effects.

Key Indexing Terms: Low Back Pain; Acute Pain; Evidence-Based Practice; Chiropractic; Manipulation; Spinal

Within mainstream health care, the customary management of low back pain (LBP) by primary care medical physicians is often not evidence based. Interestingly, clinical practice guidelines (CPG) for the treatment of acute mechanical LBP, for example, have been developed independently by multidisciplinary expert panels in 12 countries.1-12 The recommendations from those guidelines have been further accompanied by rigorous
systematic reviews of the evidence\textsuperscript{13-15} rather than expert consensus alone,\textsuperscript{1} and, to date, they have generally endorsed the use of the following conservative modalities: (1) reassurance about the favorable natural history of acute LBP, (2) early activation, (3) time-limited nonsteroidal anti-inflammatory medication (barring contraindications), and (4) spinal manipulative therapy (SMT).

Despite widespread dissemination of CPG for LBP, compliance with this knowledge in general and with the SMT component in particular has been limited among mainstream health care providers. This is particularly true among family medical physicians,\textsuperscript{16-18} whose personal beliefs about effective LBP care are often discordant with what is known from external research evidence.\textsuperscript{19,20} Yet, ironically, family medical physicians account for most office visits for LBP in many North American jurisdictions.\textsuperscript{21}

In the province of British Columbia, Canada, family medical physicians represent the most common portal of entry into the health care system for patients with LBP. In an earlier observational study of injured workers, only 6% of attending family physicians recommended guideline-concordant spinal manipulation for acute LBP, whereas 54% recommended guideline discordant passive physiotherapy even after 4 weeks postinjury.\textsuperscript{16} In a subsequent randomized controlled trial (RCT), only 17% of family physicians ended up recommending guideline concordant spinal manipulation, even after receiving a copy of CPG for the management of acute LBP as well as letters at 3 stages of the patient’s clinical course, specifically urging compliance with the distributed information.\textsuperscript{17}

As family medical physicians represent the initial contact point for many patients with LBP, they remain a key user group for evidence-based practice guidelines that promote the use of spinal manipulation. However, locally, referring physicians as well as staff physicians and surgeons within our own hospital-based spine clinic have routinely suggested that greater endorsement of doctors of chiropractic (DC) in general and spinal manipulation specifically is hindered by a lack of confidence in the consistency of quality and appropriateness of care between different providers in the community.

Until now, outcome inconsistency has not been regarded as a significant barrier to interdisciplinary referrals. However, guarded attitudes toward chiropractors for other reasons regarding quality of care have been confirmed in formal studies. In a survey of 487 Canadian and American orthopedic surgeons (including surgeons from our own hospital-based spine center), Busse et al\textsuperscript{22} found that approximately 71% held either a neutral (26%) or negative (45%) view of DCs. Most orthopedic surgeons felt that DCs provided unnecessary treatment (73%), were too aggressive in their marketing (63%), and made patients dependent on short-term relief (52%). In at least 1 other study, a sizeable proportion of Canadian spine surgeons said that they were reluctant to make a formal referral to a DC for fear of incurring liability in the (albeit remote) event of an adverse outcome.\textsuperscript{23}

The Chiropractic Hospital-based Interventions Research Outcomes study is a series of research investigations carried out at our center to evaluate the feasibility and effectiveness of chiropractic patient management when integrated into a continuum of care team model involving interrelated medical and surgical disciplines, including neurosurgical and orthopedic surgical spine, medical/nonoperative spine, neurology, and anesthesiology services.\textsuperscript{24} In an earlier randomized clinical trial,\textsuperscript{24} we demonstrated that hospital-based guideline-concordant care that included SMT was associated with significantly better functional improvements in comparison with family medicine-directed usual care. Similarly, 1 other research group has documented the tremendous feasibility and patient satisfaction associated with using DCs in a standardized hospital-based spine care pathway.\textsuperscript{25} These previous studies have highlighted the potential value of integrating evidence-based DCs into the rapidly evolving area of mainstream spine patient care.\textsuperscript{26} Yet, despite showing the effectiveness of SMT-based treatment at our own center, primary care physicians within our referral network still remain reluctant to work with DCs outside our facility due to concerns about the quality and therefore consistency of outcomes between providers in the greater community.

There is a scarcity of evidence in the literature about the consistency of outcomes between different DCs specifically. However, a previous study of the effects of individual physical therapists on outcomes for neck and LBP showed that 3% to 7% of the total variance in pain-related disability scores could be attributed to differences between practitioners.\textsuperscript{27} On the other hand, these practitioner effects were less (0%-3%) for patients receiving manual therapy and practically nonexistent (0.3%-0.5%) when the treatment (consisting of combined physical therapy and manual therapy) was applied in a standardized manner.

One of the broader aims of our ongoing research has been to identify and address the modifiable barriers to interdisciplinary cooperation and thereby facilitate greater utilization and integration of evidence-based chiropractic into the mainstream health care system. As one of the barriers to greater medical acceptance and utilization of DCs by partners within our own center (and referral network) is a lack of confidence in the consistency in the quality of care between different practitioners, we sought to determine whether 1 particular aspect of quality, that is, desirable clinical or patient outcomes, was consistent between different DCs who had at least administered a standardized version of SMT. To our knowledge, there are no analytic studies formally evaluating the consistency of or differences in outcomes between individual DCs complying to a standardized approach.
The primary objective of the current study was to determine whether clinical outcomes differ significantly between patients who are treated by different DCs when all are administering a standardized protocol of high-velocity low-amplitude (HVLA) SMT for acute LBP. Our primary a priori hypothesis was that changes from baseline in LBP-specific disability at 24 weeks (our longest follow-up point), as assessed on the modified Roland Disability Questionnaire (RDQ), would not differ significantly between groups. Our secondary hypotheses were as follows: (1) changes from baseline in RDQ scores would not differ significantly between DC-specific patient groups at other follow-up time points and (2) changes from baseline in general bodily pain (BP) and general physical functioning (PF), as assessed on the 36-Item Short-Form Health Survey (SF-36) questionnaire, also would not differ significantly between groups at any time points.

METHODS

Study Design

This was a secondary analysis of data from both a prospective observational pilot study and a corresponding RCT that compared CPG based treatment—which included HVLA SMT—to usual primary medical care for acute LBP.24 Participants were allocated consecutively to 1 of 4 treatment groups defined by attending DC.

Patients

Participants were recruited from patients referred to a hospital-based, quaternary care surgical spine program ambulatory clinic serving a large Canadian metropolitan center. Participants were aged 19 to 75 years with a chief complaint of acute lower back pain. All participants satisfied Quebec Task Force Classification of Spinal Disorders criteria categories 1 (LBP without leg pain) or 2 (LBP with radiation above the knee)28 and had acute symptoms of 2 to 4 weeks’ duration. Patients were excluded in the presence of spinal “red flag” conditions (eg, cauda equina syndrome, fracture, malignancy, systemic infection, and active inflammatory conditions), spinal nerve root irritation/deficit, or pregnancy. Patients were also excluded if they had concomitant pain in another region of their spine (eg, chronic neck pain), previous spinal surgery, or an ongoing personal injury insurance claim (with workers’ compensation or a motor vehicle injury insurer).

Study Maneuvers and Treatments

Patients were screened for eligibility by a nonoperative spine medical physician (PB) in our hospital spine clinic. After completing informed consent, participants were assigned consecutively to the attending DC on rotation at the time of enrolment. Four DCs in total treated participants during the study. Initially, 2 DCs (providers A and D) were selected based on their long-standing working relationship with one of the authors (PB) and known proficiency in administering HVLA lumbar SMT. Two additional DCs were subsequently selected based on recommendations from provider A. All participating DCs had to be actively working in a community-based private practice as well as be proficient in the use of HVLA lumbar SMT. Otherwise, criteria such as years of experience, philosophical persuasion, and attitude toward full-spine treatment were intentionally overlooked so as to permit later generalizability of our results to a broader cross-section of DCs. Each participating DC ended up being a member of either the chiropractic association board or the chiropractic regulatory college in British Columbia; however, this was not intentional.

Three DCs each worked 2 separate 8-week-long rotations in our hospital-based clinic. Because of a personal scheduling conflict, 1 DC (provider C) completed only 1 rotation. The order of each DC’s rotation was not randomly determined but instead based on convenience and personal availability during the study period.

Each DC administered CPG-based care that consisted of the following: (1) reassurance of the natural history of acute lower back pain; (2) avoidance of passive treatment (eg, bed rest, heat, or the use of back supports/corsets/braces); (3) continued activation through a progressive walking program (2 walks a day, each initially for 5-15 minutes depending on tolerance, increasing by 2 minutes each week); 4) acetaminophen, 650 mg every 6 to 8 hours as required for pain for 2 to 4 weeks (unless contraindicated by allergy, compromised liver function, or acute porphyria); and (5) a maximum of 4 weeks of side posture. HVLA lumbar SMT.29 Patients received HVLA lumbar SMT exclusively in the painful region of the lumbosacral spine. Based on a systematic review of the relevant literature, 4 weeks was thought to constitute a proper minimum challenge of SMT.30 Spinal manipulative therapy was therefore administered 2 to 3 times a week for 4 to 6 weeks at the discretion of the treating provider. Each DC was granted hospital privileges specifically for the study. All patients were regularly reminded to avoid guideline-discordant treatments, including the following: (1) muscle relaxants and opioid-class medications, (2) passive physiotherapeutic modalities, (3) bed rest, and (4) “special” back exercise programs (eg, core stabilization).

Outcomes

The primary outcome was the change from baseline in back pain-specific functioning as measured on the modified RDQ31 specifically at 24 weeks. Both Roland and Fairbank31 and Bombardier et al32 recommended the use of 2 to 3 points as the minimal clinically important difference on the RDQ in clinical trials of LBP. The secondary outcomes were the changes in RDQ at 8 and 16 weeks and normalized BP and PF domain scores of the SF-36 questionnaire at 8, 16, and
24 weeks. In a comparison of the reliability and responsiveness of 5 different scales for measuring low back disability, Davidson and Keating showed that a minimum clinically important difference on the SF-36 BP and PF scales was 14 and 15, respectively. Baseline historical, demographic, RDQ, and SF-36 questionnaires were administered by an independent study coordinator at enrolment. Follow-up RDQ and SF-36 questionnaires were readministered by mail. Patients were asked about cointervention use during each survey.

**Statistical Analysis**

The distributions of baseline demographic and clinical variables were summarized using means and SDs for continuous variables and frequency counts and percentages for categorical variables. Summary statistics were tabulated for patients overall and by attending provider. To test for between-group differences in outcomes, mixed-effects models were used in which subjects were included as a random effect, and repeated outcome measurements over time were handled as correlated observations. In adjusted analyses, “chiropractor,” “follow-up time,” and the interaction between these variables (chiropractor by time) were included in the models. Age, sex, and any baseline variables that were identified through purposeful selection were included as covariates. For the few patients with incomplete follow-up, missing outcome data were imputed using the last observation carried forward. Analyses were
implemented in PROC MIXED, in SAS, version 9.1, software (SAS, Cary, NC). All hypothesis tests were conducted at $\alpha = .05$ (2 sided) for main effects and at $\alpha = .10$ for treatment-by-time interactions. $P$ values were not corrected for multiple comparisons. For our primary outcome, we examined the simple main effect of treatment group (“chiropractor”) on RDQ change at 24 weeks exclusively; however, in the absence of a significant treatment-by-time interaction, the “marginal” main effect of treatment across all time points was also examined. For our secondary outcomes, we tested for differences in both SF-36 PF and SF-36 BP change scores also at 24 weeks and differences in changes between groups on each of these outcomes at other time points (6 weeks, then 12 weeks). Again, given that no significant treatment-by-time interactions were detected, we also examined the marginal main effects of treatment group on both PF and BP across all time points.

Ethics approval was obtained from the University of British Columbia Clinical Research Ethics Board (certificate no. H04-70588). The clinical trial registration number is NCT00135239.

### Results

#### Descriptive Analysis

A patient flow chart is depicted in Figure 1. The dropout rate was low for our study overall (7%); however, as chiropractor C completed only 1 clinical rotation, that particular treatment group was smaller ($n = 14$) and most severely affected by an otherwise small number of dropouts (29%).

Table 1 shows the baseline characteristics of participants separated according to treating chiropractor. Most of the differences between groups were small and not clinically meaningful. However, statistically significant overall differences between treatment groups were present in terms of SF-36 mental component summary (MCS) ($P < .0001$), RDQ ($P = .03$), and SF-36 BP scores ($P < .0001$); and subsequent pairwise comparisons revealed a few clinically significant differences. Compared with participants treated by chiropractor A (arbitrary reference group), those treated particularly by chiropractor C had significantly worse BP (difference in BP, >14 points), back pain-specific functional disability (difference in RDQ, >2 points), and mental health (difference in MCS, >2.5 points). Furthermore, patients treated by chiropractor B had clinically significantly higher (better) baseline MCS scores.

The plots in Figures 2, 3, and 4, respectively, depict mean point estimates with 95% confidence limits for RDQ, SF-36 PF, and SF-36 BP scale scores over time for each treatment group. Between groups, these figures generally showed less overlap in the confidence bars and correspondingly greater variability in the distributions of the RDQ (Fig 2) and PF (Fig 3) change scores. In contrast, there was comparatively greater overlap in the confidence bars and therefore less variability in the distributions of the BP (Fig 4) change scores. Furthermore, within each of the plots, the overall slopes of the lines were roughly similar with all groups showing a general pattern of monotonic improvement over time.

#### Repeated-Measures Analysis

In adjusted repeated-measures analyses, “time” exhibited a significant main effect on RDQ ($P = .0002$), PF ($P < .001$), and BP ($P = .0001$) change scores, merely indicating that patients improved significantly within treatment groups on each of these measured outcomes over time. More importantly, however, “treatment” exhibited a significant main effect, meaning that significant overall differences “between” groups in terms of RDQ ($P = .03$) and PF ($P = .04$) but not BP change scores ($P = .4$) were present. There were no significant treatment-by-time interactions in our models (RDQ, $P = .6$; PF, $P = .16$; and BP, $P = .8$), meaning that these differences between groups did not vary significantly across time points.

### Table 1. Baseline Characteristics of Enrolled Patients, by Attending DC

<table>
<thead>
<tr>
<th>Characteristic a</th>
<th>Total Sample</th>
<th>Treating Chiropractor</th>
<th>$P$ b</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>97</td>
<td>23</td>
</tr>
<tr>
<td>Age (y)</td>
<td></td>
<td>34.4 (12.7)</td>
<td>34.0 (11.9)</td>
</tr>
<tr>
<td>Female, n (%)</td>
<td></td>
<td>53 (55)</td>
<td>15 (65)</td>
</tr>
<tr>
<td>Working, n (%)</td>
<td></td>
<td>62 (64)</td>
<td>15 (65)</td>
</tr>
<tr>
<td>MCS</td>
<td></td>
<td>42.1 (3.6)</td>
<td>42.0 (3.2)</td>
</tr>
<tr>
<td>RDQ</td>
<td></td>
<td>12.7 (5.1)</td>
<td>11.7 (4.9)</td>
</tr>
<tr>
<td>PF</td>
<td></td>
<td>39.6 (12.6)</td>
<td>40.2 (14.2)</td>
</tr>
<tr>
<td>BP</td>
<td></td>
<td>38.0 (12.2)</td>
<td>45.3 (13.1)</td>
</tr>
</tbody>
</table>

BP, general bodily pain; MCS, mental component summary; PF, general physical functioning; RDQ, Roland Disability Questionnaire.

a Means with SD in parentheses unless otherwise specified.

b Fisher exact test for categorical variables and 1-way analysis of variance test for continuous variables.

c Significantly different from group A (arbitrary control group) assuming a minimum clinically significant difference of 2 points on the RDQ scale.

d Significantly different from group A (arbitrary control group) at $\alpha$ level of .05, 2 sided, unadjusted, assuming a minimum clinically significant difference of 14 points on the SF-36 BP scale.
Given the absence of significant treatment-by-time interactions, we could have legitimately focused exclusively on the differences in marginal effects between groups. However, as our primary a priori hypothesis was aimed at testing for differences in RDQ change scores at 24 weeks specifically, we first examined the simple effects (i.e., point-specific means) of treatment on RDQ change at 24 weeks (Table 2). Furthermore, for secondary purposes, we examined the point-specific effects of treatment on RDQ change at 8 and 16 weeks, followed by an assessment of the point-specific effects on PF and BP change scores at each time point.

**Primary Outcome: RDQ Change at 24 Weeks**

As Table 2 shows, after adjusting for potential confounders, the overall difference in RDQ change scores between groups at 24 weeks was nearly but not clearly statistically significant \( P = .10 \). On the other hand, the difference in the mean RDQ change scores between the most improved (group B, \( -4.4 \)) and least improved (group C, \( -1.4 \)) patients was clinically important (mean difference, \( -3.1 \); 95% confidence interval [CI], \(-6.3 \) to \( 0.1 \)).

**Secondary Outcomes**

At other time points, a statistically significant difference in RDQ change scores was observed between groups at 8 weeks \( (P = .02) \). Subsequent pairwise comparisons showed that the largest difference at 8 weeks was between groups B and C, for which the adjusted mean difference in RDQ change was \(-4.0 \) (95% CI, \(-7.1 \) to \(-0.9 \)).

Table 2 also shows the marginal mean RDQ change scores, the largest of which was for group B (mean effect, \(-3.8 \); 95% CI, \(-4.9 \) to \( 2.6 \)) and the smallest of which was for group C (mean effect, \(-0.8 \); 95% CI, \(-3.1 \) to \( 1.5 \)). As shown in Table 3, the difference in the marginal means between this pair of groups was \(-3.0 \) (95% CI, \(-5.9 \) to \(-0.1 \)), which was clinically important. Furthermore, minimum clinically important differences were evident in pairwise comparisons of the marginal means of groups A vs B, groups A vs D, and groups C vs D (Table 3).

Within individual groups, significant time-related improvements in mean RDQ scores were observed for 3 of the groups \( (P < .05 \) for groups A, B, and D) but not for group C which was the only group that did not exhibit a clearly monotonic improvement in RDQ scores over time (mean RDQ change across all time points, \(-0.8 \); 95% CI, \(-3.1 \) to \( 1.5 \)).

Table 2 also shows the simple effects (point-specific changes) and marginal effects (overall changes) for the SF-36 PF and BP scores. Between treatment groups, a significant overall difference in PF change scores was detected but again, only at 8 weeks \( (P = .008) \). The largest pairwise difference in PF change scores at 8 weeks occurred between groups A and D (mean difference, \( 7.0 \); 95% CI, \(2.6 \) to \( 11.4 \)), which was statistically significant but not clinically important. Similarly, statistically significant differences were detected in comparisons of the marginal means of groups A vs B and groups A vs D; however, none of these differences was clinically important.

In contrast, BP change scores were neither statistically nor clinically significantly different between groups at any time points (all \( P \) values > .3). As the main effect of treatment on this outcome was not significant, no pairwise comparisons of the marginal effects were conducted.

Within individual treatment groups, statistically significant time-related improvements from baseline were detected.
for all 4 DCs in terms of both PF and BP scores (P < .0001 for all groups). No adverse events from HVLA SMT were observed throughout the entire study period.

**DISCUSSION**

Despite having previously shown the effectiveness of including community-based DCs in the routine management and medical comanagement of patients with acute LBP, primary care medical physicians and specialists in our referral network had expressed their reluctance to refer patients for SMT to nonhospital-affiliated DCs due to uncertainty over the consistency and therefore quality of outcomes between DCs in the broader community. Given the absence of previous literature on this subject, we undertook this study to formally test for differences in outcomes specifically between community-based DCs administering a protocol of otherwise standardized care. The present study therefore investigated the consistency of clinical outcomes as a potential barrier to the integration of DCs into any newly defined spine treatment pathway. Our a priori expectation was that patients’ outcomes would be similar between different DCs. In an analysis of simple effects, our results showed that most patients were likely to experience similar levels of improvement in both back pain-specific and general PF over 24 weeks regardless of treating practitioner and that only a minority of patients do not improve significantly (or worsen) in terms of back pain-specific functioning. However, we also found that a standardized approach to administering SMT, in and of itself, was unlikely to ensure consistency of outcomes between different practitioners. The greatest disparity between groups was evident early on in terms of the change in back pain-specific PF at 8 weeks. At 24 weeks, the initial difference was slightly diminished and no longer statistically significant yet still clinically important. By comparison, the change in “general” PF was also significantly different between 1 pair of groups at 8 weeks but neither statistically nor clinically significant at later time points.

In an optimal analysis that considered the repeated nature of our measurements for each outcome, we found that the simple effects of treatment on the RDQ change were not significantly modified by time. This meant that the differences in the marginal effects (overall changes) appropriately represented the differences in the simple effects (point-specific changes) between DCs across all time points. After detecting a significant main effect of treatment group in the repeated-measures analysis, we found both statistically and clinically significant pairwise differences in the marginal changes between different DCs. The size of the marginal effects was as large as −3.1 RDQ points (95% CI, −6.0 to −0.2) between the most disparate pair of groups. Although these confidence limits did not exclude a clinically negligible effect, they clearly included magnitudes of effects on RDQ scores that equaled or exceeded those reported in published systematic reviews of RCTs comparing SMT with other treatments for acute LBP.

We acknowledge that none of our effect sizes was large; however, half of our pairwise comparisons of the marginal effects on RDQ change met the threshold for a minimum clinically important difference. Moreover, our modest effect sizes should be given due consideration in light of the fact that typically reported effect sizes for SMT are already only generally moderate at best. On the other hand, we strongly emphasize that our confidence limits do not exclude clinically negligible effects and that therefore...
our study is not definitive and requires further replication using both larger comparison groups and randomized allocation to treating practitioners. For now, however, we feel our findings are strong enough to at least mildly unsettle our prior assumption—and prior declaration to our referring primary medical physicians—that outcomes are likely consistent between different evidence-based providers of SMT.

Also in the meantime, our finding that pain-specific disability varied significantly between DCs somewhat conflicts with the findings of Lewis et al. Upon reanalyzing data drawn from 2 earlier randomized trials, Lewis et al found that disability outcomes did not vary significantly between different physical therapists administering either “exercise plus advice and manual therapy” or “physiotherapy including manual therapy techniques.” However, in the study of Lewis et al, the effect estimates were not adjusted for differences in patients’ baseline characteristics, whereas in our study, we carefully controlled for clinical characteristics through the use of restrictive inclusion/exclusion criteria and statistical adjustments for age, sex, and baseline outcome measurements. Neither the study of Lewis et al nor our study controlled for differences in the quality of the practitioner-patient relationship or for other unmeasured nonspecific effects that could have easily confounded the true treatment effects. We speculate that underlying differences in the effects of treatment expectation and the practitioner-patient relationship may account for at least some of the discrepancy between our findings and those of Lewis et al.

Although practitioner effects have not been previously documented among DCs, they have been studied among other clinicians. According to Simon et al, most studies of practitioner effects are from psychotherapy and clinical psychology, where therapists account for up to 18% of the unique variability in patients’ outcomes. Documented practitioner effects are sometimes attributed to factors that are extrinsic (eg, age, training institution, and experience) or intrinsic (eg, personal preferences or values about treatment) to the clinician. In an earlier systematic review of the medical, psychologic, and sociologic literature, Blasi et al found evidence drawn from RCTs that practitioners who combined cognitive care (ie, raising patients’ expectations about treatment through a clear diagnosis and/or positive prognosis) with emotional care (ie, being “warm and friendly” or “firm and reassuring”) were more effective than providers of neutral, formal, or uncertain consultations in decreasing pain and increasing the speed of recovery.

In the spine literature, traditional exemplars of clinical expertise (eg, years in practice and specialty certification in manual therapy) have been found not to account for significant variability in LBP outcomes among physical therapists. In contrast, treatment type may be important as practitioner effects appear to be stronger with predominantly psychosocial-based interventions—such as “brief psychologically oriented pain management”—in comparison with predominantly physical-based interventions—such as manual therapy.

Our finding has potential methodological implications for future as well as previously conducted, observational, and randomized controlled studies of SMT. One particularly concerning implication is the possibility that practitioner effects may be an important confounder for the
overall treatment effect of SMT. Admittedly, in many pragmatic studies, practitioner variability "within" arms does not constitute a threat to external validity, especially when the intentional goal is to estimate the customary effect of SMT as administered under real-world conditions (by the broader chiropractic community). However, in trials of SMT in which intrinsic practitioner or contextual effects (eg, personal preference and enthusiasm for the treatment being administered) are conceivably greater in the experimental arm, the true effect of SMT could be exaggerated.

Lewis et al emphasize that a practitioner effect can give rise to a systematic bias in the estimated effect of treatment whenever the distribution of influential

Table 2. Adjusted Mean Changes in Outcomes (With 95% CIs for Mean Change), by Attending DC Over Time

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Treating Chiropractor</th>
<th>P (Over All Groups)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>RDQ change</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 wk</td>
<td>−0.8 (−2.3 to 0.6)</td>
<td>−3.0 (−4.4 to −1.7)</td>
</tr>
<tr>
<td>16 wk</td>
<td>−1.3 (−3.3 to −0.3)</td>
<td>−3.9 (−5.3 to −2.6)</td>
</tr>
<tr>
<td>24 wk</td>
<td>−2.3 (−3.9 to −0.8)</td>
<td>−4.4 (−5.8 to −3.0)</td>
</tr>
<tr>
<td>Marginal mean</td>
<td>−1.6 (−2.9 to −0.4)</td>
<td>−3.8 (−4.9 to −2.6)</td>
</tr>
<tr>
<td>P (over all time points)</td>
<td>.01</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>SF-36 PF change</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 wk</td>
<td>15.0 (11.8-18.3)</td>
<td>9.0 (6.0-12.0)</td>
</tr>
<tr>
<td>16 wk</td>
<td>17.0 (13.7-20.3)</td>
<td>14.0 (11.0-17.1)</td>
</tr>
<tr>
<td>24 wk</td>
<td>19.4 (16.0-22.8)</td>
<td>15.1 (12.0-18.2)</td>
</tr>
<tr>
<td>Marginal mean</td>
<td>17.2 (14.2-20.1)</td>
<td>12.7 (9.9-15.5)</td>
</tr>
<tr>
<td>P (over all time points)</td>
<td>&lt;.0001</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>SF-36 BP change</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 wk</td>
<td>8.8 (5.4-12.2)</td>
<td>5.6 (2.6-8.6)</td>
</tr>
<tr>
<td>16 wk</td>
<td>10.9 (7.5-14.3)</td>
<td>9.3 (6.2-12.3)</td>
</tr>
<tr>
<td>24 wk</td>
<td>12.6 (9.1-16.1)</td>
<td>9.9 (6.8-13.0)</td>
</tr>
<tr>
<td>Marginal mean</td>
<td>10.8 (7.7-13.9)</td>
<td>8.2 (5.5-11.0)</td>
</tr>
<tr>
<td>P (over all time points)</td>
<td>&lt;.0001</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

* From random-effects models controlling for age, sex, baseline outcome score, and correlated measurements over time. Baseline MCS score was weakly associated with RDQ (P = .19) and SF-36 BP (P = .07) scores and was therefore included as a covariate in those corresponding models. P values are not adjusted for multiple comparisons.

b Type 3 test of main effect of "chiropractor" in fully adjusted model.

BP, general bodily pain; PF, general physical functioning; RDQ, Roland Disability Questionnaire; SF-36, Short Form 36 Health Survey.

Table 3. Adjusted Pairwise Differences in Marginal Mean Changes (With 95% CIs) for Outcomes Associated With a Significant Main Effect of Treatment

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Treating Chiropractor</th>
<th>P (Over All Groups)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>RDQ change</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>2.2 (0.4-3.9)</td>
<td>P = .01</td>
</tr>
<tr>
<td>C</td>
<td>−0.9 (−3.6 to 1.7)</td>
<td>−3.1 (−6.0 to −0.2)</td>
</tr>
<tr>
<td>D</td>
<td>1.9 (0.2-3.6)</td>
<td>−0.3 (−1.8 to 1.3)</td>
</tr>
<tr>
<td>PF change</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>4.44 (0.34-8.52)</td>
<td>P = .03</td>
</tr>
<tr>
<td>C</td>
<td>5.11 (−0.95 to 11.16)</td>
<td>0.67 (−5.95 to 7.28)</td>
</tr>
<tr>
<td>D</td>
<td>5.11 (1.28-9.34)</td>
<td>0.67 (−2.85 to 4.59)</td>
</tr>
</tbody>
</table>

PF, general physical functioning; RDQ, Roland Disability Questionnaire.
practitioner attributes is imbalanced between treatment arms. In the case of pragmatic studies aimed at measuring the customary effectiveness of SMT under every day conditions, that bias can be controlled for either prospectively in the design, through “expertise-based randomization” or retrospectively in the analysis through statistical methods that adjust for any clustering of outcomes of patients treated by the same practitioner.

In other settings, such as in explanatory studies aimed at measuring the best possible efficacy of SMT under ideal circumstances, researchers might legitimately aim to harness the practitioner effect by, for example, intentionally loading the experimental arm with the most physically and contextually “expert” practitioners available.

In the meantime, it is worth noting that despite the existence of more than 100 published RCTs of SMT for LBP in the literature, the strength of conclusions from recent systematic reviews is generally weak, especially about the effectiveness of SMT for acute LBP. Rubinstein et al conclude that due to considerable heterogeneity in outcomes and methods between SMT trials for acute LBP, the value of conducting more RCTs—simply to improve the precision of our estimates of only mild-to-moderate effect sizes—is questionable. As an alternative, Rubinstein et al suggest a moratorium on simple SMT effectiveness trials and a greater focus on studies aimed at improving the identification of responsive subgroups. Clearly, simultaneous research to identify the modifiable determinants of skill, expertise, and other practitioner-based determinants of effectiveness and outcomes would be very complementary to the call of Rubinstein et al to better identify patient-based predictors of treatment responsiveness.

**Study Strengths**

An important strength of this study is that it involved a very homogenous sample of patients. By design, we used restriction methodology to ensure that participants had LBP of similar location, duration, and neurologic normalcy, meaning we did not have to control for these traditional confounding variables statistically. Similarly, our use of a standardized treatment protocol meant that other potentially influential variables, such as duration of each visit, frequency and overall duration of treatment, SMT technique, and prescription of cointerventions were controlled by the design of the study. Furthermore, we measured other potential confounders that we were able to control for in the analysis, so long as this was indicated by our purposeful selection process. Finally, DCs in this study were very representative of DCs in the community, as they did not receive specialized clinical training apart from being asked to adhere to a standardized HVLA SMT protocol.

**Limitations**

There are important limitations to our study that potentially explain the observed disparity in outcomes, particularly between group C patients (who improved the least in terms of condition-specific and general PF) and those of the other 3 groups. Although we standardized the technical aspects of SMT practice between DCs, we did not measure patient satisfaction or control for other related variables such as the effectiveness of doctor-to-patient communication skills and the ability of each DC to accommodate patients’ treatment expectations and preferences—which have been shown to be associated with clinical outcomes in other studies. With respect to such contextual variables, it is noteworthy that group C was unique in that it accounted for all unsolicited reports of apparent dissatisfaction with chiropractic treatment during the study and/or comments that their treating DC was markedly lacking in some or all of the above described aspects. We emphasize that we did not interpret these anecdotal remarks as evidence that chiropractor C was “necessarily” different in terms of clinical skill or conduct from our other 3 practitioners, only that patients in that group likely “did” differ from the other groups in terms of their overall “perceptions” and level of satisfaction with treatment.

In terms of external validity, our study population did not include patients who were having concomitant pain in another region of their spine, and, therefore, our results are not generalizable to those with multiple axial pain conditions. Similarly, our results are not applicable to patients with either chronic pain, previous spinal surgery, or an ongoing insurance claim (eg, with workers’ compensation or motor vehicle insurer). Furthermore, patients attending a hospital-based spine program ambulatory clinic may have better outcomes than patients attending a private community-based clinic simply due to systematic differences between patients who do and do not gain access to prominent specialists (centripetal bias) and tertiary care centers (referral filter bias).

As chiropractic treatment in this study consisted specifically of HVLA SMT, our findings are not generalizable to community-based care involving other chiropractic techniques or systems of treatment. It is worth noting that, although “chiropractic” SMT most commonly includes one form or another of treatment administered by hand, some chiropractic patients with acute lower back pain do not even receive treatment to the lumbar spine, let alone a technique as well validated as HVLA SMT.

Another methodological limitation of our study is that it was based on a secondary analysis of data from 2 previous studies, neither of which was specifically powered to test for a significant difference between smaller subgroups. As reflected in the wider confidence limits around the point estimates of some of our outcomes, particularly for group C, a lack of sufficient power likely hindered our ability to detect a statistically significant yet otherwise clinically
important, point-wise difference in back pain-specific PF between groups at 24 weeks exclusively.

**CONCLUSION**

The findings of this study show that regardless of the treating DC, most patients with acute LBP without radiculopathy appear to experience consistent levels of improvement in terms of BP and general PF after receiving guidelines-based treatment that includes a component of standardized HVLA SMT. However, although our analysis did not show a statistically significant difference between groups in our primary outcome at 24 weeks, there was a trend toward significance with at least 1 pairwise comparison showing a minimum clinically important effect. A more optimal analysis of our repeated measurements revealed differences in the marginal changes in back pain-specific PF that were both statistically significant and clinically important. However, the confidence limits around our point estimates did not exclude clinically negligible effects. Therefore, these initial findings should be interpreted cautiously, as they require replication in additional studies using larger comparison groups and randomized allocation to different DCs.

**FUNDING SOURCES AND POTENTIAL CONFLICTS OF INTEREST**

No funding sources or conflicts of interest were reported for this study.

**CONTRIBUTORSHIP INFORMATION**

Concept development (provided idea for the research): P.B.B.
Design (planned the methods to generate the results): J.A.Q., P.B.B.
Supervision (provided oversight, responsible for organization and implementation, and writing of the manuscript): J.A.Q., P.B.B.
Data collection/processing (responsible for experiments, patient management, organization, or reporting data): J.A.Q., P.B.B.
Analysis/interpretation (responsible for statistical analysis, evaluation, and presentation of the results): J.A.Q.
Literature search (performed the literature search): J.A.Q., P.B.B., B.A.
Writing (responsible for writing a substantive part of the manuscript): J.A.Q., P.B.B., B.A.
Critical review (revised manuscript for intellectual content, this does not relate to spelling and grammar checking): J.A.Q., P.B.B., B.A.

**REFERENCES**

8. RTreatment guideline—backache. Drug Committee of the German Medical Society. in GermanZ Arztl Fortbild Qualitatsch 1997;91:457-60.


