

Systematic Review

# The centralization phenomenon of spinal symptoms—a systematic review

Alessandro Aina<sup>a</sup>, Stephen May<sup>b,\*</sup>, Helen Clare<sup>c</sup>

<sup>a</sup> *Studio di Riabilitazione DTM, Milano, Italy*

<sup>b</sup> *Sheffield Hallam University, Collegiate Crescent Campus, Collegiate Crescent, Sheffield S10 2BP, UK*

<sup>c</sup> *School of Physiotherapy, Sydney University, Australia*

Received 20 July 2003; received in revised form 3 December 2003; accepted 4 March 2004

## Abstract

The centralization phenomenon was first described 20 years ago. It refers to the abolition of distal pain emanating from the spine in response to therapeutic exercises. Since then a number of papers on the subject have been published. A review of current knowledge is appropriate.

Selection criteria were established prior to a computer-aided search for published papers. Two reviewers independently extracted data and checked quality; a third reviewer resolved any disagreements. A narrative review was conducted based on the findings. The review primarily considered prevalence, reliability of assessment, and prognostic significance. These have been most commonly reported, and are important to establish the clinical worth of this symptom response.

Fourteen studies were identified. Quality of studies varied; prognostic studies were given a mean score of 3.3 out of 6 by using established quality criteria. The prevalence rate of pure or partial centralization was 70% in 731 sub-acute back patients, and 52% in 325 chronic back patients. It is a symptom response that can be reliably assessed during examination ( $\kappa$  values 0.51–1.0). Centralization was consistently associated with a range of good outcomes, and failure to centralize with a poor outcome.

Centralization appears to identify a substantial sub-group of spinal patients; it is a clinical phenomenon that can be reliably detected, and is associated with a good prognosis. Centralization should be monitored in the examination of spinal patients.

© 2004 Elsevier Ltd. All rights reserved.

## 1. Introduction

Back and neck pain are extremely common symptoms, but they are symptoms only—the majority of spinal pain is said to be non-specific (Spitzer et al., 1987; AHCPR, 1994; CSAG, 1994). Non-specific spinal pain has no obvious diagnostic, prognostic or treatment indicators and treatment is something of a lottery dependent on which clinician is seen (Deyo, 1993; Cherkin et al., 1994; Van Tulder et al., 1997). Clinicians collect a range of clinical data during examination, including observational and palpatory findings supposedly to guide management. However there is evidence that this information is unable to indicate an appropriate management strategy (Dreyfuss et al., 1996; Levangie, 1999a; Hestboek and Leboeuf-Yde, 2000; Leboeuf-Yde et al., 2002). Identifying clinical findings that could guide treatment strategies is clearly desirable.

Finding predictors of chronicity and disability after an acute attack of back pain is an important aim of primary care research (Borkan et al., 1998). Identification of such patients at an early stage would allow for the targeting of more intensive and costly interventions to those more likely to benefit from them most (Haldorsen et al., 2002). There has been considerable research into factors that may predict poor outcomes (Chavannes et al., 1986; Lanier and Stockton, 1988; Goertz, 1990; Hasenbring et al., 1994; Burton et al., 1995; Gatchel et al., 1995; Hazard et al., 1996; Smedley et al., 1998; Thomas et al., 1999; Potter et al., 2000), which has explored a wide range of individual, clinical and psychosocial factors. The relative importance of different factors remains unclear, but psychosocial factors are thought to have a dominant role in predicting future disability and clinical factors a limited role (Pincus et al., 2002). Psychological factors that are thought to be important include attitudes, cognitions, fear-avoidance towards pain, depression, anxiety and distress (Linton, 2000).

\*Corresponding author. Tel./fax: +44-114-221-7303.

E-mail address: stephen@maymunro.f9.co.uk (S. May).

The phenomenon of centralization was first recognized by McKenzie in the 1950s and after much experimentation and verification was described in the literature (McKenzie, 1981). It is the process by which pain radiating from the spine is sequentially abolished, distally to proximally, in response to therapeutic positions or movements; and includes reduction and abolition of spinal pain. Centralization can occur in the lumbar, cervical, and thoracic spine (McKenzie, 1981, 1990). The opposite symptom response—a distal spread of pain into the limb is termed peripheralization.

The symptom modification of centralization, which may occur during physical examination, is potentially important, as this response may guide clinicians in their management of non-specific spinal problems. Centralization may indicate the appropriateness of specific, rather than generalized exercises, and allow for the confident selection of mobilization or manipulation techniques (Lisi, 2001). In the absence of other clinical indicators for management of non-specific spine pain centralization may provide practitioners with a simple, yet valuable tool. Furthermore the phenomenon of centralization may act as a clinical predictor of good or bad outcomes. This would allow the identification of those patients who may need more involved interventions.

Centralization was not evaluated in the scientific literature until 1990 (Donelson et al., 1990; Kilby et al., 1990), since which time there have been a number of studies. These have been of variable quality and therefore a systematic review of the available literature was deemed necessary in order to clarify the current documented knowledge regarding the value of this clinical phenomenon.

## 2. Methods

A literature search was made on Medline, CINAHL, Embase and PEDro, start dates varied on different databases up to 2002 (key words: centralization/centralization, and spine, and lumbar, and cervical). This was supplemented by a manual search of the references. Studies had to meet the following criteria:

1. The study involved centralization.
2. The study subjects suffered from back or neck pain.
3. The study was a controlled trial or a cohort design.
4. Published in peer-reviewed journal, and not an abstract.
5. The article was written in English.

One author (AA) conducted the initial search; the pre-determined selection criteria were applied by a second author (SM). Data abstraction and quality assessment was conducted independently by two authors (AA, SM);

disagreements were settled by negotiation with the third author (HC).

The main focus of the review was upon prevalence rate, prognostic value and reliability of evaluation of centralization. The initial review of articles available revealed these to be issues that had been commonly investigated, but also issues that are important in establishing the clinical utility of this phenomenon. Despite the heterogeneity of study design most of the studies recruited a sample of back pain patients and estimated the prevalence rate of centralization in their population. Clinically therefore the studies were similar because of the patient sample, but dissimilar in study design. The only data extracted on which a meta-analysis could therefore be performed were the prevalence rates (mean and range) from these similar patient populations. Consideration was also given to the operational definition of centralization, the type of exercise, posture or technique used to induce centralization (the loading strategy), and the external diagnostic validity.

Quality issues to be considered were sample size and standardization of assessment. One element of this was the use of an overlay template, which is described elsewhere (Donelson et al., 1991; Long, 1995). Briefly, a body chart filled in by the patient marking all areas of pain is covered by a transparent body chart with a scoring grid; with most distal pain scoring highest. This allows blinded assessment of patients' pain patterns.

To assess methodological quality in prognostic studies criteria were adapted from Laupacis et al. (1994) (Table 1). There are no widely agreed quality criteria for assessing prognostic studies (Altman, 2001). However discussion of the issue focuses on the same aspects of study design: patient sample, follow-up of patients, outcome, prognostic variables and analysis; and much the same quality criteria have been used (Laupacis et al., 1994; Hudak et al., 1996; Altman, 2001). Laupacis et al. (1994) failed to provide explicit detail for 2 of the criteria (C and D), but more detail was provided by Hudak et al. (1996). These were made explicit in the following ways: sufficient length of follow-up was defined as 1 year; sufficient numbers of follow-up was defined as more than 85% of initial cohort. These cut off points were selected to ensure an adequate follow-up period that addressed the known natural history of back pain (Abbott and Mercer, 2002), and an adequate follow-up sample, and were derived from another set of criteria (Hudak et al., 1996). If studies came near to these set criteria half a point was given. Hudak et al. (1996) also provided levels of evidence: *strong* evidence partially or fully meeting all criteria; *moderate* evidence partially fulfilling most criteria; *weaker* evidence when studies failed to fulfil multiple criteria.

Assessing methodological quality in the other studies was not possible due to the range of study designs that

Table 1  
Method score for studies about centralization and prognosis

Reference	A	B	C	D	E	F	Total
Donelson et al. (1990)	1	0	0	0	1	0	2
Long (1995)	0.5	1	1	0.5	1	1	5
Karas et al. (1997)	0	0	0	0.5	1	1	2.5
Sufka et al. (1998)	0.5	0	0	0.5	0	0.5	1.5
Werneke et al. (1999)	1	1	0	0.5	1	0	3.5
Werneke and Hart (2001)	1	1	1	0.5	1	1	5.5
Total	4	3	2	2.5	5	3.5	
Mean (SD)							3.3 (1.63)

A. Was the sample representative of the underlying population? B. Were they at a well-defined point in the natural history? C. Was the follow-up of sufficient length—1 year? D. Was there follow-up of > 85% of the sample? E. Was there blinded assessment of outcome? F. Were other prognostic factors equal or accounted for in analysis?

were retrieved, which were all relevant to the issue being reviewed. Studies were cross-sectional, reliability studies; prospective, randomized trials; and prospective cohort studies; some studies were experimental and some were descriptive. Heterogeneity of study design discounted the use of uniform quality criteria.

After the initial review the following general principles were adopted relative to certain pertinent outcomes. For reliability studies Kappa gives a numerical value to clinicians' ability to reproduce a test result: 0–0.2 poor, 0.2–0.4 fair, 0.4–0.6 moderate, 0.6–0.8 good, 0.8–1.0 very good reliability (Altman, 1991). Negative values represent reliability that is worse than chance agreement. Kappa values take chance agreement into account and thus are more appropriate for reliability studies than percentage agreement only. For clinically important changes in outcome measures guidelines for the interpretation of effect size suggest that up to 0.5 is small, and greater than 0.8 is large (Cohen, 1987). It has been suggested that a clinically important change in an 11-point pain rating scale for chronic pain should be about 18% (Farrar et al., 2001). Odds ratio (OR) is a means of comparing the odds or likelihood of an event in one group compared to another, for instance centralizers and non-centralizers. An interpretation of OR = 1 is the same outcome in both groups, OR > 1 is higher probability in that group, and OR < 1 is lower probability in that group (Earl-Slater, 2002).

### 3. Results

A total of 14 papers (13 separate studies) were retrieved that investigated or referred to the phenomenon of centralization, none before 1990 (Table 2). Seven studies investigated patient samples of less than one hundred and six more than one hundred. An overlay body template was used to standardize the assessment procedure in only 4 studies (Table 2). Five studies (six papers) investigated the prognostic value of centraliza-

tion and could be given a method score (Table 1). Quality scores ranged from 1.5 to 5.5, mean 3.3 (SD 1.63). Two studies provided strong evidence (Long, 1995; Werneke and Hart, 2001), one moderate (Werneke et al., 1999), and 3 studies weak evidence (Donelson et al., 1990; Karas et al., 1997; Sufka et al., 1998) according to Hudak et al. (1996). Most common study limitations were short-term follow-up and patient attrition greater than 15%; also samples were not always at a well-defined point in natural history and sometimes failed to include other prognostic variables.

#### 3.1. Definitions of centralization

There was consensus around the core definition of centralization—the abolition of distal pain in response to the deliberate application of movements or postures. If pain is only in the back this is centralized and then abolished. Some authors added additional components to McKenzie's original definition—Fritz et al. (2000) included a change in neurological signs and symptoms, and several studies included a *reduction* in intensity of the most distal symptoms in the definition (Delitto et al., 1993; Erhard et al., 1994; Karas et al., 1997). Werneke et al. (1999) applied a stricter definition: centralization occurred only in the clinic, and progressed sequentially toward the spine on each occasion until all symptoms were abolished. Werneke et al. (1999) also described a partial centralization group, in which changes in distal symptoms occurred, but less completely or not on each visit.

#### 3.2. Prevalence

In a meta-analysis of 1056 patients in ten studies (Table 3) centralization occurred in 681 patients (64.5%). In 731 patients with predominantly acute or sub-acute back pain (less than 7 weeks) 511 centralized (70%); in 325 patients with chronic back pain 170 centralized (52%). In individual studies prevalence of

Table 2  
Studies related to centralization

Reference	Study design	Sample	Sample frame	Purpose related to centralization	Use of template
Kilby et al. (1990)	Cross-section	41	Single PT dept	Reliability	No
Donelson et al. (1990)	Descriptive, retrospective	87	Private orthopaedic practice	Prognosis. Prevalence	No
Williams et al. (1991)	Prospective, randomised	207	Multi-centre PT clinics	Loading strategy	Yes
Donelson et al. (1991)	Prospective, randomised	145	International multi-centre PT clinics	Loading strategy	Yes
Delitto et al. (1993)	Prospective, randomised	24	PT clinic	Selection criterion for trial	No
Erhard et al. (1994)	Prospective, randomised	24	PT clinic	Selection criterion for trial	No
Long (1995)	Descriptive, prospective	223	Work hardening programme	Prevalence. Prognosis	Yes
Karas et al. (1997)	Descriptive, prospective	126	Rehabilitation programme	Prevalence. Prognosis	No
Donelson et al. (1997)	Comparison study	63	Tertiary centre for invasive testing	Criterion validity	No
Sufka et al. (1998)	Descriptive, prospective	36	PT clinic	Prevalence. Prognosis. Reliability	No
Werneke et al. (1999)	Descriptive, prospective	289	2 PT clinics McKenzie assessment methods	Prevalence. Prognosis. Reliability	Yes
Fritz et al. (2000)	Cross-section	NP&LBP Video	Parts of 12 patient assessments	Reliability	No
Werneke and Hart (2001)	Descriptive, prospective	187 LBP (84%)	Long-term follow-up (12 months)	Prognosis	Yes
Kilpikoski et al. (2002)	Cross section	39	Hospital patients from wider study	Reliability	No

NP=neck pain, LBP=low back pain. PT=physical therapy.

centralization ranged from 31% (Werneke et al., 1999) to 87% (Donelson et al., 1990; Kilpikoski et al., 2002). In the study with the lowest rate 77% of all patients were classified in either the complete or partial centralization categories (Werneke et al., 1999). So the true range of complete or partial centralization across the studies is between 47% and 87%.

Only one study included cervical problems in their cohort. Werneke et al. (1999) classified 25% of 66 neck pain patients in the centralization category and 46% in the partial centralization category.

### 3.3. Centralization and prognosis

Six papers investigated the prognostic value of centralization by comparing the outcomes of centralizers with non-centralizers, which demonstrated several significant differences (Table 3). Centralization was correlated with good/excellent overall outcomes, greater reduction in pain intensity, higher return to work rates, greater functional improvement, and less continued healthcare usage (Donelson et al., 1990; Long, 1995; Sufka et al., 1998; Karas et al., 1997; Werneke et al., 1999; Werneke and Hart, 2001).

Two studies found a lack of significant difference in Oswestry scores between centralizers and non-centralizers (Long, 1995; Sufka et al., 1998). Oswestry is a validated, self-reported, back specific functional disability questionnaire (Fairbank et al., 1980). One study

found a high Waddell score to be a more significant predictor of return to work than centralization (Karas et al., 1997). This score is derived from a collection of eight signs that are said to be indicative of non-organic pathology (Waddell et al., 1980).

Werneke et al. (1999) found a significant difference in the number of treatment sessions between the centralization group (3.9) and the partial centralization (7.7) and non-centralization groups (8). However both centralization and partial centralization groups had significantly greater improvements in pain and function than the non-centralization group. Multiple independent variables were gathered at baseline, including demographic, historical, work and psychosocial factors; and patients were reviewed at 1 year (Werneke and Hart, 2001). In a multivariate analysis that included all independent significant variables only failure to achieve centralization and leg pain at intake were significant for predicting chronic back pain, disability, return to work, and healthcare usage at 1 year (Table 3).

### 3.4. Reliability of detection of centralization

Five studies considered whether clinicians were reliable in detecting centralization (Kilby et al., 1990; Sufka et al., 1998; Werneke et al., 1999; Fritz et al., 2000; Kilpikoski et al., 2002). Three studies did so in a small group of patients, with only two or three therapists (Kilby et al., 1990; Werneke et al., 1999;

Table 3  
Prevalence and prognostic association of centralization

Reference	N	Patient description	% C	Prognostic association of centralization
Donelson et al. (1990)	87	Acute 61%, sub-acute 17%, chronic 22%	87%	Correlation between centralization and good/excellent outcome ( $P < 0.001$ ), non-centralization and poor/fair outcome ( $P < 0.001$ )
Donelson et al. (1991)	145	Acute 23%, sub-acute 38%, chronic 39%	47%	
Delitto et al. (1993)	24	Acute 100%	61%	
Erhard et al. (1994)	24	Sub-acute 100%	55%	
Long (1995)	223	Chronic 100%, Not working 100%	47%	Greater reduction in pain intensity ( $< 0.05$ ), higher return to work rate ( $P = 0.034$ )
Karas et al. (1997)	126	Acute and chronic, Not working 100%	73%	More frequent return to work ( $P = 0.038$ )
Donelson et al. (1997)	63	Chronic 100%, Not working 70%	49%	
Sufka et al. (1998)	36	Acute 16%, sub-acute 42%, chronic 42%	69%	Greater functional improvement ( $P = 0.015$ )
Werneke et al. (1999)	289	Back pain 77%, neck pain 23%. Acute 100%, Not working 37%	A: 31% B: 46%	1: Fewer visits ( $P < 0.001$ ). 1 + 2: Greater improvements in pain ( $P < 0.001$ ), and function ( $P < 0.001$ )
Werneke and Hart (2001)	187	Reviewed at 1 year—back pain only	77% (A + B)	Non-centralization predicted work status, function, health care use ( $P < 0.001$ ), and pain intensity ( $P = 0.004$ )
Kilpikoski et al. (2002)	39	Chronic 100%	87%	
Total-N (%)	1056		681 (64.5%)	

% C = proportion in which centralization occurred. Werneke et al. (1999), Werneke and Hart (2001): A = centralization, B = partial centralization (see text).

Kilpikoski et al., 2002); one study did not provide details (Sufka et al., 1998). Kilby et al. (1990) asked a broader question, ‘Do repeated movements decrease, abolish or centralize the pain?’ Fritz et al. (2000) used a large number of therapists who watched the same video containing edited parts from 12 examinations. Percentage agreement was high in all studies (88% to 100%). Kappa values were 0.51 (Kilby et al., 1990), 0.92 and 1.0 (Werneke et al., 1999), 0.82 and 0.76 for graduate and student therapists (Fritz et al., 2000), and 0.7 (Kilpikoski et al., 2002).

### 3.5. Loading strategy and centralization

Centralization occurs during the physical examination in response to exercises, postures or mobilization. The type of loading strategy that was used to induce centralization was the focus of three studies (Donelson et al., 1990, 1991; Williams et al., 1991). When using exercises or postures only in the sagittal plane extension loading is significantly more effective at achieving centralization than flexion, and flexion is more likely to cause peripheralization (Table 4) (Donelson et al., 1991; Williams et al., 1991). Donelson et al. (1990) utilized exercises in sagittal and frontal planes and achieved centralization in 87%.

Table 4  
Loading strategy associated with centralization

Reference	N	Patient population	Extension/ lordosis	Flexion/ kyphosis
Williams et al. (1991)	210	Acute 24%, sub-acute 41%, chronic 35%	C 56% <sup>a</sup> P 6% <sup>a</sup>	C 10% <sup>a</sup> P 24% <sup>a</sup>
Donelson et al. (1991)	145	Acute 23%, Sub-acute 38%, Chronic 39%	C 40%	C 7%

C = centralization, P = peripheralization.

<sup>a</sup>Only pain below knee.

### 3.6. Diagnostic implications of centralization

Only one study examined the criterion validity of centralization with a blinded comparison between a mechanical evaluation and discography (Donelson et al., 1997). Patients displaying centralization or peripheralization were significantly more likely to have a positive discography than patients with non-centralization. A competent annulus was recorded in 91% ( $P < 0.001$ ) of those with centralization and a positive discogram, compared to 54% of those with peripheralization and a positive discogram.

#### 4. Discussion

In this review the authors evaluated the evidence concerning the centralization phenomenon, which was first reported by McKenzie (1981). Publications began in 1990, since when 14 studies have documented or referred to centralization. McKenzie (1981, 1990) reported that the phenomenon could be used to guide therapeutic management, and suggested a good prognosis. Potentially this clinical phenomenon may have both therapeutic and prognostic implications. If this potentially important symptom modification is to be used in assessment certain attributes must be documented. For it to be of clinical utility it must occur in a substantial proportion of those with spinal pain, its definition must be consistent across groups, the phenomenon must be identified reliably, and its identification must be worthwhile for management, prognostic or diagnostic purposes.

From this review it appears that centralization is a common clinical occurrence in patients with acute or chronic lumbar symptoms, which has been documented in 47% to 87% of cohorts in 10 separate studies. The studies were consistent in the core definition—the abolition of distal limb pain in response to therapeutic loading strategies. Four out five studies found that evaluation of centralization can be performed with good or very good reliability. These findings suggest that centralization can be used to identify a sub-group of the back pain population consistently and reliably. These are important attributes for any clinical phenomenon that is to be used in practice.

This review found the kappa values for centralization to range from 0.51 to 1.0. Historically palpation findings have been used to guide management. A selection of studies investigating the reliability of various palpation techniques for the lumbar spine and sacro-iliac joint demonstrated kappa values between  $-0.025$  and  $0.30$  (Van Deursen et al., 1990; Binkley et al., 1995; Lindsay et al., 1995; McKenzie and Taylor, 1997; Meijne et al., 1999; Vincent-Smith and Gibbons, 1999; O'Haire and Gibbons, 2000). Such a comparison is clearly far from systematic or comprehensive, but it does appear to highlight dramatic differences in the reliability of 2 approaches to assessment.

It would appear then that centralization allows reliable recognition of a substantial sub-group of the back pain population. However does its detection during assessment provide an important marker for outcome, management or diagnosis? This review demonstrated that centralization is consistently associated with a range of good outcomes in 6 studies relating to pain, function, return to work, and decreased healthcare usage (Donelson et al., 1990; Long, 1995; Karas et al., 1997; Sufka et al., 1998; Werneke et al., 1999; Werneke and Hart, 2001). We might question if *statistically*

significant differences between centralization and non-centralization groups are *clinically* significant. One study (Donelson et al., 1990) did not use established outcome measures, and 2 (Werneke et al., 1999; Karas et al., 1997) did not report absolute changes in pain, disability, or return to work thus making it difficult to evaluate the clinical importance of any change. The authors of this review found that the effect sizes for the partial and complete centralization groups (Werneke et al., 1999) ranged from 2.06 to 2.61 for pain and disability, compared to 0.51 to 1.02 for the non-centralization group (Werneke, 2003). Long (1995) reported pain changes and return to work rates as 16% and 68% in the centralization group compared to 6% and 52% in the non-centralization group. Sufka et al. (1998) reported mean improvements in the Spinal Function Sort of 51 in the centralization group and 15 in the non-centralization group, with 250 representing maximum functional disability. These studies suggest that there may be clinically important differences in outcome between centralizers and non-centralizers. However some studies do not allow a judgement about clinical importance to be made, some studies are ambiguous about the size of difference, and this issue requires further research.

Only one study (Werneke and Hart, 2001) provided 1-year follow-up, and provided odds ratios (OR) for different outcomes. Non-centralization gave an OR of 3.0 for high pain intensity, 9.4 for failure to return to work, 5.2 for interference with daily activities, and 4.4 for further health care. These substantial OR demonstrated a definite poorer long-term outcome using multiple measures; all of clinical importance, in patients who fail to demonstrate centralization (Werneke and Hart, 2001). Other studies did not report OR, nor provide data in a format that allowed their calculation, except for return to work (RTW) rates in Long (1995). The authors of this review calculated from Long (1995) the OR for RTW as 1.9 for the centralization group in a chronic back pain sample initially all off work. Further research needs to be conducted in this area using OR and long-term follow-up to confirm the relative value of centralization, but initial studies identify a clinical phenomenon that appears to have prognostic importance.

Failure to achieve centralization has been shown to have greater prognostic significance than certain psychosocial factors, including job satisfaction, fear avoidance behaviour and depression, in one study (Werneke and Hart, 2001). This study is unusual in demonstrating a clinical factor to be more significant in predicting long-term outcomes than psychosocial factors. However it is important to bear in mind that multiple large-scale studies have confirmed the importance of psychosocial variables as prognostic factors (Chavannes et al., 1986; Lanier and Stockton, 1988; Goertz, 1990; Hasenbring et al., 1994; Burton et al., 1995; Gatchel et al., 1995;

Hazard et al., 1996; Smedley et al., 1998; Thomas et al., 1999; Potter et al., 2000).

Although these 6 studies (Donelson et al., 1990; Long, 1995; Karas et al., 1997; Sufka et al., 1998; Werneke et al., 1999; Werneke and Hart, 2001) are consistent in identifying the prognostic value of centralization it must be noted that 3 of them provided only weak evidence, and one of moderate evidence, as defined by the criteria used to judge methodological quality (Hudak et al., 1996). Weaker studies suffered from short-term follow-up, lack of well-defined point in natural history and failure to include other prognostic variables; substantial patient attrition was a common failing in all studies. Two high-quality studies confirmed the prognostic importance of centralization (Long, 1995; Werneke and Hart, 2001).

In the updated *levels of evidence* suggested for evaluating the quality of intervention studies from the Cochrane Collaboration Back Review Group (Van Tulder et al., 2003) the following levels are suggested:

- *Strong*—consistent findings among multiple high-quality randomized clinical trials (RCTs).
- *Moderate*—consistent findings among multiple low-quality RCTs and/or controlled clinical trials (CCTs) and/or one high RCT.
- *Limited*—one low-quality RCT and/or CCT.
- *Conflicting* inconsistent findings among multiple trials (RCTs and/or CCTs).
- No evidence—no trials.

These types of study designs are not transferable to a review of centralization; but using the equivalent levels of evidence for the studies we have reviewed would suggest *moderate* to *strong* evidence for the prognostic value and prevalence of centralization that we have described.

Another important attribute for a clinical phenomenon is its ability to provide an indication of appropriate management. Contained within the definition of centralization is the assumption that it is induced during the physical examination, rather than occurring because of natural history. Ensuring centralization *was* clinically induced and that the initial symptom modification remained better was a reason for the stricter definition employed by Werneke et al. (1999). They identified 2 groups—centralization and partial reduction. In the first group changes were strictly clinically induced, but in the second changes occurred, but less completely or not on each visit. It is unknown if the partial reduction group improved due to natural history in an acute group of patients or due to the effect of exercises continued at home in between clinic visits. However as two studies (Long, 1995; Donelson et al., 1997) involving chronic patients recorded centralization prevalence of 48% it seems unlikely that natural history was entirely the reason. Further investigation is needed to confirm this.

In fact the distinction between the centralization and partial reduction groups although valid in terms of number of treatment sessions (3.9 versus 7.7), was not material to changes in pain and disability, which were similar—both significantly better than the non-centralization group. Failure to achieve symptom modification by the seventh visit identified patients who will do poorly in the long-term (Werneke et al., 1999). It would seem that it is the occurrence of centralization, rather than its speed or completeness that is important to outcome.

Does centralization have diagnostic implications? Only one study has attempted to match centralization with a tissue-specific diagnosis in a group with chronic back pain (Donelson et al., 1997). This study demonstrates an association between centralization and discogenic pain, but obviously requires further verification. The link between symptom response generated by specific loading strategies and discogenic pain is reviewed in more detail elsewhere (Wetzel and Donelson, 2003).

Centralization thus describes the abolition of distal symptoms in response to specific loading strategies, involving exercise, posture or mobilization. This symptom response is commonly found in acute and chronic back pain populations, and is consistently and reliably interpreted. It appears to be most helpful in guiding management and understanding prognosis. Limitations exist in the current literature on this topic; primarily the moderate quality of most studies, and several suggestions can be made to improve the quality of research. Past studies have involved small numbers, lack of consistent reporting mechanisms, lack of clear definition of sample, short-term follow-up, and lack of multiple prognostic variables. In acute populations it may be necessary to ensure that this symptom response is clinically induced, and not a product of natural history. Furthermore it is important to ensure that initial improvements are sustained and developed over the period of treatment. A strict operational definition of centralization that guaranteed these features was only described in one study (Werneke et al., 1999; Werneke and Hart, 2001). However, as discussed earlier, the distinction between partial and full centralizers impacted on number of treatment sessions, but not clinical outcomes.

Follow-up has mostly been short-term, but longer follow-up has shown the prognostic importance of centralization after 1 year (Long, 1995; Werneke and Hart, 2001). None of the studies used a control group, other than the non-centralization group. Variations in treatment, initial symptom location, and duration of problems occurred across the different studies, and may be confounding factors when considering outcomes. Further research is needed into the comparative value of other prognostic indicators. Further descriptive studies

could also help to delineate more clearly the therapeutic forces needed to achieve centralization. Only one study (Werneke et al., 1999) has investigated this symptom response in the cervical spine. This demonstrated that it may be as common an occurrence in the cervical as in the lumbar spine, but further trials are needed to confirm this. As most of the work so far has been conducted in lumbar spine patients it would be inappropriate to automatically extrapolate this information to cervical spine patients.

An issue in exploring this phenomenon is the operational definition that is used, and the fact that although the core definition has been consistent, there have been some discrepancies in previous research. Differences in definition include the addition of changes in symptom severity, rather than site, the addition of changes in neurological symptoms, and a distinction between *centralising* and full *centralization*. Werneke et al. (1999) gave clear-cut operational definitions of complete and partial centralization categories, which focus on a clinically induced change in symptom location, without consideration of symptom intensity, until all symptoms were abolished. The stricter definitions make research into the phenomenon more rigorous and should be used in future studies, as well as the overlay template, which allows for blinded judgement about symptom changes.

Despite weaknesses in the quality of the literature all studies investigating this phenomenon are consistent in the direction of outcome, and the most rigorous study to date (Werneke et al., 1999; Werneke and Hart, 2001) confirms this trend with stricter operational definitions, longer follow-up, the inclusion of multiple potential prognostic variables, and appropriate data analysis. What is apparent from the current literature is that centralization is a common phenomenon that can be reliably assessed, and because of its association with good outcomes can be used to guide treatment strategies in certain patients. This is in marked contrast to other commonly used assessment procedures involving palpation or observation, which have frequently been found to be of poor reliability (Van Deursen et al., 1990; Binkley et al., 1995; Lindsay et al., 1995; McKenzie and Taylor, 1997; Meijne et al., 1999; Vincent-Smith and Gibbons, 1999; O'Haire and Gibbons, 2000). Tests of palpation use data that are subjectively determined by clinicians, which has not been shown to differentiate the back pain population from normal individuals (Dreyfuss et al., 1996; Egan et al., 1996; Levangie, 1999b; Lebouef-Yde et al., 2002), and has been unable to establish a lesion that is amenable to manipulation (Hestboek and Lebouef-Yde, 2000). In contrast to centralization commonly used tests based on palpation and observation of the spine appear to be poor determinants of management of spinal problems. A systematic review of these contrasting methods of

physical examination is required to provide a more definitive evaluation of the comparative reliability of these different approaches.

Finally it should be noted that the level of training and experience of the assessing clinician might well be important to maximize outcome. Reliability studies have shown that whereas therapists who were naïve to the McKenzie system are unreliable in determining symptom response (Riddle and Rothstein, 1993), those who are trained and experienced in the system can reliably assess symptom response (Razmjou et al., 2000, Kilpikoski et al., 2002).

## 5. Conclusion

This review has found that the methodological quality of the literature in this area is mostly weak or moderate, however the findings are consistent and confirmed by 2 high-quality studies. Centralizing symptoms and full centralization are common clinical occurrences. It is a phenomenon that can be evaluated with good reliability, and is consistently associated with better outcomes than its absence. Centralization occurs frequently with extension exercises or postures, but also with other loading strategies. Non-centralization has been shown to be an important predictor of poor outcomes at one year in one study. This symptom response thus has important therapeutic and prognostic implications. In the light of the reliability with which centralization can be assessed, and its common occurrence and clinical importance it is recommended that it should be monitored routinely during spinal assessments and be used to guide treatment strategies.

## References

- Abbott JH, Mercer SR. The natural history of acute low back pain. *New Zealand Journal of Physiotherapy* 2002;30:8–16.
- AHCPR. (Agency for Health Care Policy and Research). In: Bigos S, Bowyer O, Braen, et al., editors, *Acute low back problems in adults*. USA: Department of Health and Human Services; 1994.
- Altman DG. *Practical statistics for medical research*. London: Chapman and Hall; 1991. p. 404 [chapter 14].
- Altman DG. Systematic reviews of evaluations of prognostic variables. *British Medical Journal* 2001;323:224–8.
- Binkley J, Stratford PW, Gill C. Interrater reliability of lumbar accessory motion mobility testing. *Physical Therapy* 1995;75:786–95.
- Borkan JM, Koes B, Reis S, et al. A report from the Second International Forum for Primary Care research on Low Back Pain. Re-examining priorities. *Spine* 1998;23:1992–6.
- Burton AK, Tillotson KM, Main CJ, Hollis S. Psychosocial predictors of outcome in acute and subchronic low back trouble. *Spine* 1995;20:722–8.
- Chavannes AW, Gubbels J, Post D, Rutten G, Thomas S. Acute low back pain: patients' perceptions of pain four weeks after initial



- diagnosis and treatment in general practice. *Journal of Royal College of General Practitioners* 1986;36:271–3.
- Cherkin DC, Deyo RA, Wheeler K, Ciol MA. Physician variation in diagnostic testing for low back pain. What you see is what you get. *Arthritis & Rheumatism* 1994;37:15–22.
- Cohen J. *Statistical power analysis for the behavioral sciences*, Revised ed. London: Lawrence Erlbaum Associates; 1987.
- CSAG. (Clinical Standards Advisory Group). *Back pain*. London: HMSO; 1994.
- Delitto A, Cibulka MT, Erhard RE, Bowling RW, Tenhula JA. Evidence for use of an extension-mobilization category in acute low back syndrome: a prescriptive validation pilot study. *Physical Therapy* 1993;73:216–28.
- Deyo RA. Practice variations, treatment fads, rising disability. *Spine* 1993;18:2153–62.
- Donelson R, Silva G, Murphy K. Centralization phenomenon. Its usefulness in evaluating and treating referred pain. *Spine* 1990;15:211–3.
- Donelson R, Grant W, Kamps C, Medcalf R. Pain response to sagittal end-range spinal motion. A prospective, randomized, multicentered trial. *Spine* 1991;16:S206–12.
- Donelson R, Aprill C, Medcalf R, Grant W. A prospective study of centralization of lumbar and referred pain. A predictor of symptomatic discs and annular competence. *Spine* 1997;22:1115–22.
- Dreyfuss P, Dreyer S, Griffin J, Hoffman J, Walsh N. Positive sacroiliac screening tests in asymptomatic adults. *Spine* 1994;19:1138–43.
- Dreyfuss P, Michaelsen M, Pauza K, McLarty J, Bogduk N. The value of medical history and physical examination in diagnosing sacroiliac joint pain. *Spine* 1996;21:2594–602.
- Earl-Slater A. *The handbook of clinical trials and other research*. Oxford: Radcliffe Medical Press; 2002.
- Egan D, Cole J, Twomey L. The standing forward flexion test: an inaccurate determinant of sacroiliac joint dysfunction. *Physiotherapy* 1996;82:236–42.
- Erhard RE, Delitto A, Cibulka MT. Relative effectiveness of an extension program and a combined program of manipulation and flexion and extension exercises in patients with acute low back syndrome. *Physical Therapy* 1994;74:1093–100.
- Fairbank JCT, Couper J, Davies JB, O'Brien JP. The Oswestry low back pain disability questionnaire. *Physiotherapy* 1980;66:271–3.
- Farrar JT, Young JP, LaMoreaux L, Werth JL, Poole RM. Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. *Pain* 2001;94:149–58.
- Fritz JM, Delitto A, Vignovic M, Busse RG. Interrater reliability of judgements of the centralization phenomenon and status change during movement testing in patients with low back pain. *Archives of Physical Medicine and Rehabilitation* 2000;81:57–61.
- Gatchel RJ, Polatin PB, Mayer TG. The dominant role of psychosocial risk factors in the development of chronic low back pain disability. *Spine* 1995;20:2702–9.
- Goertz MN. Prognostic indicators for acute low-back pain. *Spine* 1990;15:1307–10.
- Haldorsen EMH, Grasdal AL, Skouen JS, Risa AE, Kronholm K, Ursin H. Is there a right treatment for a particular patient group? Comparison of ordinary treatment, light multidisciplinary treatment, and extensive multidisciplinary treatment for long-term sick-listed employees with musculoskeletal pain. *Pain* 2002;95:49–63.
- Hasenbring M, Marienfeld G, Kuhlendahl D, Soyka D. Risk factors of chronicity in lumbar disc patients. A prospective investigation of biologic, psychologic, and social predictors of therapy outcome. *Spine* 1994;19:2759–65.
- Hazard RG, Haugh LD, Reid S, Preble JB, MacDonald L. Early prediction of chronic disability after occupational low back injury. *Spine* 1996;21:945–51.
- Hestboek L, Leboeuf-Yde C. Are chiropractic tests for lumbo-pelvic spine reliable and valid? A systematic critical literature review. *Journal of Manipulative and Physiological Therapeutics* 2000;23:258–75.
- Hudak PL, Cole DC, Haines AT. Understanding prognosis to improve rehabilitation: the example of lateral elbow pain. *Archives of Physical Medicine Rehabilitation* 1996;77:586–93.
- Karas R, McIntosh G, Hall H, Wilson L, Melles T. The relationship between nonorganic signs and centralization of symptoms in the prediction of return to work for patients with low back pain. *Physical Therapy* 1997;77:354–60.
- Kilby J, Stigant M, Roberts A. The reliability of back pain assessment by physiotherapists, using a 'McKenzie algorithm'. *Physiotherapy* 1990;76:579–83.
- Kilpikoski S, Airaksinen O, Kankaanpaa M, Leminen P, Videman T, Alen M. Interexaminer reliability of low back pain assessment using the McKenzie method. *Spine* 2002;27:E207–14.
- Lanier DC, Stockton P. Clinical predictors of outcomes of acute episodes of low back pain. *Journal of Family Practice* 1988;27:483–9.
- Laupacis A, Wells G, Richardson WS, Tugwell P. Users' guide to the medical literature v. How to use an article about prognosis. *Journal of American Medical Association* 1994;272:234–7.
- Leboeuf-Yde C, van Dijk J, Franz C, et al. Motion palpation findings and self-reported low back pain in a population-based study sample. *Journal of Manipulative and Physiological Therapeutics* 2002;25:80–7.
- Levangie PK. The association between static pelvic asymmetry and low back pain. *Spine* 1991a;24:1234–42.
- Levangie PK. Four clinical tests of sacroiliac joint dysfunction: the association of test results with innominate torsion among patients with and without low back pain. *Physical Therapy* 1999b;79:1043–57.
- Lindsay DM, Meeuwisse WH, Mooney ME, Summersides J. Interrater reliability of manual therapy assessment techniques. *Physiotherapy Canada* 1995;47:173–80.
- Linton S. A review of psychological risk factors in back and neck pain. *Spine* 2000;25:1148–56.
- Lisi AJ. The centralization phenomenon in chiropractic spinal manipulation of discogenic low back pain and sciatica. *Journal of Manipulative and Physiological Therapeutics* 2001;24:596–602.
- Long AL. The centralization phenomenon. Its usefulness as a predictor of outcome in conservative treatment of chronic low back pain (a pilot study). *Spine* 1995;20:2513–21.
- McKenzie RA. *The lumbar spine. Mechanical diagnosis and therapy*. Waikanae: Spinal Publications; 1981.
- McKenzie RA. *The cervical and thoracic spine. Mechanical diagnosis and therapy*. Waikanae: Spinal Publications; 1990.
- McKenzie AM, Taylor NF. Can physiotherapists locate lumbar spinal levels by palpation? *Physiotherapy* 1997;83:235–9.
- Meijne W, van Neerbos K, Aufdemkampe G, van der Wurff P. Intraexaminer and interexaminer reliability of the Gillet test. *Journal of Manipulative and Physiological Therapeutics* 1999;22:4–9.
- O'Haire C, Gibbons P. Inter-examiner and intra-examiner agreement for assessing sacroiliac anatomical landmarks using palpation and observation: pilot study. *Manual Therapy* 2000;5:13–20.
- Pincus T, Burton AK, Vogel S, Field AP. A systematic review of psychological factors as predictors of chronicity/disability in prospective cohorts of low back pain. *Spine* 2002;27:E109–20.
- Potter RG, Jones JM, Boardman AP. A prospective study of primary care patients with musculoskeletal pain: the identification of predictive factors for chronicity. *British Journal of General Practice* 2000;50:225–7.
- Razmjou H, Kramer JF, Yamada R. Interrater reliability of the McKenzie evaluation in assessing patients with mechanical low-back pain. *JOSPT* 2000;30:368–89.

- Riddle D L, Rothstein JM. Intertester Reliability of McKenzie's classification of the type of the syndrome types present in patients with low back pain. *Spine* 1993;18:1333–44.
- Smedley J, Inskip H, Cooper C, Coggon D. Natural history of low back pain. A longitudinal study in nurses. *Spine* 1998;23:2422–6.
- Spitzer WO, LeBlanc FE, Dupuis M, et al. Scientific approach to the activity assessment and management of activity-related spinal disorders. *Spine* 1987;12:S1–55.
- Sufka A, Hauger B, Trenary M, et al. Centralization of low back pain and perceived functional outcome. *JOSPT* 1998;27:205–12.
- Thomas E, Silman AJ, Croft PR, Papageorgiou AC, Jayson MIV, Macfarlane GJ. Predicting who develops chronic low back pain in primary care: a prospective study. *British Medical Journal* 1999;318:1662–7.
- Van Deursen LLJM, Patijn J, Ockhuysen AL, Vortman BJ. The value of some clinical tests of the sacroiliac joint. *Journal of Manual Medicine* 1990;5:96–9.
- Van Tulder MW, Koes BW, Bouter LM, Metsemakers JFM. Management of chronic nonspecific low back pain in primary care: a descriptive study. *Spine* 1997;22:76–82.
- Van Tulder M, Furlan A, Bombardier C, et al. Updated method guidelines for systematic reviews in the Cochrane Collaboration Back Review Group. *Cochrane Library*, Issue 4. Oxford: Update Software; 2003.
- Vincent-Smith B, Gibbons P. Inter-examiner and intra-examiner reliability of the standing flexion test. *Manual Therapy* 1999;4:87–93.
- Waddell G, McCulloch JA, Kummel E, Venner RM. Nonorganic physical signs in low-back pain. *Spine* 1980;5:117–25.
- Werneke M, Hart DL, Cook D. A descriptive study of the centralization phenomenon. *Spine* 1999;24:676–83.
- Werneke M, Hart DL. Centralization phenomenon as a prognostic factor for chronic low back pain and disability. *Spine* 2001;26:758–65.
- Werneke M. Personal communication, 2003.
- Wetzel FT, Donelson R. The role of repeated end-range/pain response assessment in the management of symptomatic lumbar discs. *The Spine Journal* 2003;3:146–54.
- Williams MM, Hawley JA, McKenzie RA, van Wijmen PM. A comparison of the effects of two sitting postures on back and referred pain. *Spine* 1991;16:1185–91.