Directional preference constructs for patients' low back pain in the absence of centralization

Richard Yarnzbowicz^a and Minjing Tao^b

^aCenter for Orthopedic and Sports Physical Therapy, Tallahassee, FL, USA; ^bDepartment of Statistics, Florida State University, Tallahassee, FL, USA

ABSTRACT

Objectives: A detailed description of how Directional Preference (DP) constructs are measured could accelerate research to practice translation and improve research findings for Mechanical Diagnosis and Therapy (MDT) stakeholders. A secondary analysis of a prospective, observational cohort study was conducted to understand (1) the type and prevalence of DP constructs at first examination and (2) the relationships between DP constructs and clinical outcomes at follow-up.

Methods: Data were collected and analyzed from 1485 consecutive patients who presented to outpatient, private practice clinics with primary complaints of non-specific low back pain (LBP); 400 patients met the inclusion criteria and completed first examination and follow-up data. Statistical analysis determined prevalence and the relationships between DP constructs at first examination and clinical outcomes at follow-up.

Results: The primary findings in this investigation were that (1) the most prevalent DP constructs at first examination were related to range of motion (ROM) and pain intensity (Patient Reported Improvement in ROM (74.8%), Increase in Spine ROM (29.5%), and Pain Intensity Change (17.3%)), (2) all groups improved and made clinically meaningful improvements in disability and pain intensity at follow-up, (3) no clinically significant differences in disability or pain intensity were found between the groups at follow-up, and (4) 26.5% and 6.5% of patients exhibited a relative increase in lumbar spine extension and flexion ROM, respectively, post repeated movement testing on the first examination.

Discussion: The findings in this study assist providers in making assessment and treatment decisions with their patients by offering insight regarding the most prevalent DP constructs typically found at the first examination and their subsequent association with outcome when Centralization (CEN) does not occur. Recommendations for researchers have been made to further explore the DP framework used in this study.

KEYWORDS

Lumbar spine; low back pain; directional preference; McKenzie; centralization; practice-based evidence; orthopedic; musculoskeletal

Introduction

DP assessment for patients with LBP syndromes has gained considerable interest [1–3]. In a recent clinical practice guideline, a recommendation was made, based on strong evidence, that clinicians should consider using 'repeated exercises in a specific direction determined by treatment response to improve mobility and reduce symptoms in patients with acute, subacute, or chronic low back pain with mobility deficits [2].' However the efficacy of DP assessment and management has been challenged [3] perhaps in part due to the misunderstanding between DP and CEN; DP is a broader finding with a characteristic change of clinical signs and symptoms during a dynamic physical examination process; CEN is one of the many criteria used to determine the presence of DP. CEN is the systematic abolition of symptoms in a distal to proximal direction toward the spine. Many previous studies had not delineated between DP and CEN for patients with LBP syndromes [4]. Many different clinical profiles exist relative to DP assessment and DP should not be viewed as a

homogeneous entity [5–7], nor can it be presumed that all patients with LBP will respond in the same way [8]. DP in the absence of CEN is common [9–11]; George et al. reported that 50% of patients exhibited a DP without CEN [9], and CEN has found to be as low as 19.6% at the first examination [10,11]. However, the manner in which DP has been recorded in many studies has been inconsistent which may be leading to inconclusive research findings.

The Treatment Based Classification system is a popular LBP classification approach which helps classify patients via information gathered from the history and physical examination into one of four categories: manipulation, specific exercise (i.e. flexion, extension, and lateral-shift), stabilization, and traction [12]. Although CEN is considered the hallmark examination criterion for identifying the specific-exercise category, DP is also considered an important finding for determining inclusion. DP for the specific exercise category has been defined as a situation in which movement in one direction improves pain and limitation of ROM, and movement in the opposite

direction causes signs and symptoms to worsen. However, no detailed description is provided on how change in pain and ROM is determined, what quantity of change designates inclusion, or whether one, both, or additional factors are required to indicate the presence of DP.

A randomized controlled trial by Long et al. [8] displays this same challenge. The study shows that subjects who received matched DP exercises significantly decreased pain, medication use, and other notable outcomes. Patients were included in the trial if DP was demonstrated during the mechanical assessment, and treatment was based on DP principles. However, no description was given as to what individual clinical findings constituted a DP. A more standardized definition of CEN has been developed [13,14] but not for DP. In a paper by Werneke et al. [11], patients were classified as having DP at intake, based on information found in the McKenzie lumbar spine textbook [5], if (1) the patient's most distal pain intensity decreased, abolished, or centralized and/or the patient's lumbar ROM improved in response to repeated end-range movement tests or positional loading strategies, and/or (2) the patient reported a specific preference for activities and movements, such as standing and walking (extension preference) or forward bending and sitting (flexion preference), requiring objective confirmation. If the initial physical examination findings were inconclusive, the patient might have been judged to have DP based on self-reported activity and postural preferences. This definition of DP is limited and lacks critical information such as how pain and ROM were measured and what test criteria designated inclusion. Furthermore, basing DP on subjective history rather than objective testing has been shown to be unreliable [15].

Lumbar spine CEN is one clinical finding that indicated the presence of DP. If trials do not delineate between DP and CEN, research findings can be erroneously reported and interfere with research to practice translation. A recent systematic review [3] showed mixed results among six studies investigating the efficacy of DP management. However, the definition of DP in these trials was either vaguely defined or considered in combination with CEN. Browder et al. [16] included patients who had CEN only and did not consider those who exhibited DP without CEN. Similarly, the studies by Paatelma et al. [17], Brennan et al. [18], and Long et al. [8] did not delineate between DP and CEN or analyze patients with DP without CEN. Miller et al. [19] and Schenk et al. [20] classified patients via McKenzie method syndromes (i.e. Derangement, Dysfunction, Posture, and Other), not by DP and/or CEN. The Derangement category, which is the only category that could include patients who exhibit DP [5], was considered;

however, no detailed description was provided for the criteria used in the determination of DP or if DP occurred in the absence of CEN.

Broad assessment descriptions found in research trials inhibit assimilation of research findings into practice. The purpose of this article is to help MDT providers understand (1) the type and prevalence of DP constructs typically found at the first examination and (2) the relationships between DP constructs and clinical outcomes at follow-up. Our study aims to develop a practical framework to more accurately and completely define DP for patients with LBP in order to improve the accuracy of research findings and accelerate research to practice translation.

Methodology

Design

This is the second analysis of these data. The first study was a prospective, observational cohort study examining the association between Pain Pattern Classification (PPC), DP, and clinical outcomes (i.e. pain intensity and self-reported disability) [10]. The current study is a subgroup of patients who exhibited DP at the first examination in the absence of CEN. Data were collected and analyzed from 1485 consecutive patients who presented to outpatient, private practice clinics with primary complaints of non-specific LBP (Figure 1.). The Florida State University Institutional Review Board for Protection of Human Subjects and the Florida Agricultural and Mechanical University Institutional Review Board for Protection of Human Subjects approved the project. This study did not include any change in routine clinical practice, and patient-informed consent was not required for the analysis and reporting of data.

The managing providers were responsible for determining and reporting inclusion, exclusion, and discharge status. Patients were included in the study if they presented with complaints of non-specific LBP with or without radiation into the lower extremities and their LBP was classified as DP Non-CEN or DP NC (Non-Classifiable) at first examination. NC consists of patients who do not have symptoms at rest during the beginning of the physical examination. Patients were excluded from the study if they had lumbar spine surgery within the preceding 12 months. Ten providers participated in the study (mean age: 39 years, range: 27–64; 30% females; mean years of clinical experience: 13.5). All providers participated throughout five private practices, outpatient orthopedic settings in Florida, Alabama, Oklahoma, South Carolina, and New York. Six providers were certified in MDT and four held diplomas in MDT. All facilities were Certified McKenzie Clinics.

Figure 1. Flow of Patients.

Notes: DP = Directional Preference; Non-CEN = Non-Centralization; NC = Non-Classifiable

DP classification and physical examination variables

Patients were classified via PPC and DP. PPC has been described elsewhere [10]. DP was considered present if any of the following DP constructs (Appendix) were found at the first examination: (1) Pain Intensity Change, (2) Increase in Spine ROM, (3) Change in Aberrant Trunk Movement, (4) Patient Reported Improvement in ROM, (5) Improved Neurological Signs, and (6) Patient Reported Improvement in Functional Task.

An inclinometer was used to measure lumbar spine ROM in flexion and extension before and after repeated movement testing. Intratester reliability of lumbar spine motion using a single inclinometer has been found to be adequate with an ICC of 0.85 [21] and correlation coefficients ranging from 0.79 to 0.98 [22,23]. In a study by Saur et al. [24], inclinometer measurement was strongly correlated with radiographic measurement ($r = 0.93$ overall) and exhibited adequate reliability between providers (interrater reliability was $r = 0.88$ for lumbar spine flexion and $r = 0.42$ for extension). Observations of aberrant trunk movements were determined by the criteria set forth by Hicks et al. [25] and demonstrated moderate to good reliability ($k = 0.60$) in three pairs of raters evaluating 63 consecutive patients with low back pain. In a more recent study by Biely et al. [26], substantial agreement was found for observation of any aberrant movement

pattern ($k = 0.65$) and aberrant movement patterns were more frequently observed in patients with current complaints of LBP than those without. If relevant to the patient's clinical presentation, a standard neurological examination was conducted consisting of an assessment of motor function, sensation, reflexes, and lower limb neurodynamic tension testing. Myotome and dermatome assessments were conducted via manual muscle testing and light touch testing, respectively. Myotomes and dermatome deficits were assessed according to the criteria set by the American Spinal Injury Association (ASIA) [27]. The reliability of the motor and light touch assessment has been found to be excellent (myotome assessment ICC = 0.98–0.99; light touch assessment ICC = 0.76–0.80) [28]. In general, the intratester reliability of manual muscle testing has been found to be good with correlation coefficients ranging from 0.63 to 0.98 [29–31]. Lower extremity reflexes (Quadriceps (L3) and Gastrocnemius (S1)) were assessed using a standard reflex hammer. Although lower extremity reflex testing is advocated as part of a standard neurological examination [32], few studies have examined reliability. A study by Litvan et al. [33] showed the intraobserver reliability to be near perfect to perfect ($k = 0.77$ –0.91) among neurologists with varying levels of training. Lower limb neurodynamic tension was assessed via the Slump Test [32]. The Slump Test has shown high reliability ($k = 0.83$ –0.89) for 6 pairs of physical therapists of varying experience [34].

Patient characteristics and outcome measures

Patient characteristics were collected at the first examination by the managing providers through patient self-report methods. The following continuous variables were measured: Age, Medication for Condition, Symptom Acuity, Surgical History, Fear-Avoidance Beliefs at first examination, pain intensity at first examination, disability at first examination, and Number of Comorbid Conditions. The following categorical variables were measured: Gender, Previous History of Condition, Imaging, Payer, Exercise History, and Referral Type.

Patient-reported and clinical examination responses were collected by the providers at first examination, regularly throughout the care episode, and at discharge. Two primary patient-reported outcomes were assessed: pain intensity and patient self-reported disability via the Numeric Pain Rating Scale (NPRS) and the Roland Morris Disability Questionnaire (RMDQ), respectively. Pain intensity reported by the patient within 'the last few days' was assessed using an 11-point NPRS ranging from 0 ('No Pain') to 10 ('Worst Pain Imaginable'). Childs et al. [35] reported that a two-point pain intensity unit change represents clinically meaningful change that exceeds the bounds of measurement error. The RMDQ is a patient self-report questionnaire for LBP and is scored by summing the number of items checked by the patient (e.g. 'Because of my back and/or leg pain, I use a handrail to get upstairs'). The score can vary ranging from 0 to 24. Ostelo et al. reported a minimally important change in the RMDQ score of 5 points (out of 24) or 30% from the baseline score [36]. Both the 11-point NPRS and the RMDQ have been shown to be acceptable clinical measures for patients with non-specific LBP [2].

Data analysis

Our primary aim was to report the prevalence of DP constructs at the first examination and determine the relationships between DP constructs and clinical outcomes (i.e. pain intensity and RMDQ score) for patients with LBP. Descriptive statistics were generated including proportions, standard deviations, and ranges for patient characteristics collected at the first examination. Chi-square tests of independence (dichotomous and categorical data) or two-sample t-tests (continuous data) were performed to compare patient characteristics and determine equivalence between patients with first examination data only and patients with first examination and follow-up data. Constructs measured approximated a normal distribution and the significance level for all tests was set at $p < .05$. To assist in the clinical interpretation of the results, the following continuous variables were converted to categories: Age – < 45y, 45–65y,

and > 65y; Symptom Acuity – Acute (0–2 weeks), Sub-Acute (3–11 weeks), and Chronic (≥ 12 weeks); Surgical History – None and 1 or More; Medication for Condition – None and 1 or More; Number of Comorbid Conditions – None, 1, 2 or 3, and 4 or More; Imaging – None, 1, and 2 or More; Fear-Avoidance Beliefs – Low Fear (< 14) and High Fear (≥ 14). To report the prevalence of DP constructs at first examination, patients were examined who exhibited a PPC of Non-Cen or NC at first examination and the total number of patients who exhibited the particular DP construct of interest was divided by the total number of patients included in the sample; these numbers were then converted to percentages. To determine the relationships between DP constructs and clinical outcomes, a Wilcoxon Test was conducted and p-values calculated. The Wilcoxon Test is a non-parametric test to compare the means of two samples. The Wilcoxon Test does not rely on the assumption that the corresponding distribution is normal. For the DP construct Increase in Spine ROM, the minimal detectable change was considered for use of an inclinometer reported by Fritz et al. [37] for lumbar spine flexion (15.11) and extension (7.90) ROM. The minimal detectable change represents the amount of change in degrees that must occur to exceed the bounds of measurement error. All ROM values measured via an inclinometer were whole numbers, thus a threshold of 16 and 8 degrees was set for lumbar spine flexion and extension, respectively, to increase the chance that 'real' change in ROM actually occurred. ROM change was dichotomized as '+' (i.e. ROM change values in lumbar spine flexion and/or extension that meet or exceed the minimal detectable change) and '0' (i.e. ROM change values in lumbar spine flexion and extension that do not meet or exceed the minimal detectable change). The prevalence of both categories was analyzed for lumbar spine flexion and extension ROM at first examination by determining the total number of patients within each category and then dividing by the total number of patients in the sample; these numbers were then converted to percentages. The mean and median were calculated for each ROM category.

Results

A total of 1485 consecutive patients entered the various clinics during the study period. The characteristics of the study participants are displayed in Table 1 and reasons for not completing data collection can be found in Figure 1. 282 patients did not complete follow-up data collection resulting in a participation rate of 81.0%. Compared to patients with first examination and follow-up data ($n = 400$; 58.7% completion rate), patients with first examination data only ($n = 282$) had more imaging ($\chi^2 = 6.13$, $df = 2$,

Table 1. Patient Characteristics.

Characteristics	First Examination Only (n = 282)	First Examination and Follow-up Data (n = 400)
Age (y)	43.5 ± 17.4 (5 – 85)	45.1 ± 17.4 (13 – 79)
Missing	1.4%	0.0%
Gender		
Male	42.2%	43.5%
Female	55.7%	56.5%
Missing	2.1%	0.0%
Symptom Acuity		
Acute (0–2 weeks)	18.1%	23.5%
Sub-Acute (3–11 weeks)	25.9%	27.0%
Chronic (≥ 12 weeks)	55.7%	49.3%
Missing	0.4%	0.3%
Surgical History		
None	90.8%	94.5%
1 or More	6.4%	4.8%
Missing	2.8%	0.8%
Medication for Condition		
None	48.2%	48.0%
1 or More	48.2%	50.0%
Missing	3.5%	2.0%
Imaging		
None	47.2%	55.0%
1	36.5%	34.3%
2 or More	16.3%	10.8%
Missing	0.0%	0.0%
Exercise History		
1 or 2 Times per Week	23.0%	22.8%
At least 3 Times per Week	39.7%	43.8%
Seldom or Not at All	34.4%	31.3%
Missing	2.8%	2.3%
Referral Type		
Primary Care	53.5%	58.0%
Specialist	19.1%	15.3%
Self	6.4%	9.3%
Other	20.6%	17.3%
Missing	0.4%	0.3%
Number of Comorbid Conditions		
None	23.0%	23.0%
1	26.2%	23.0%
2 or 3	27.0%	35.0%
4 or more	19.5%	17.5%
Missing	4.3%	1.5%
Payer		
Auto	2.1%	2.5%
Health Maintenance Organization (HMO)	75.5%	62.0%
Medicaid	0.7%	0.8%
Medicare Part B	7.1%	7.3%
Patient Private Pay	0.0%	0.3%
Preferred Provider Organization (PPO)	12.1%	23.3%
Workers' Compensation	0.0%	1.5%
Other	2.5%	2.5%
Missing	0.0%	0.0%
Fear-Avoidance Beliefs		
High Fear	53.5%	44.5%
Low Fear	41.1%	49.8%
Missing	5.3%	5.8%
RMDQ First Examination	8.3 ± 5.8 (0 – 24)	7.9 ± 5.4 (0 – 22)
Missing	3.9%	5.8%

(Continued)

Table 1. (Continued).

Characteristics	First Examination Only (n = 282)	First Examination and Follow-up Data (n = 400)
NPRS First Examination	5.6 ± 4.4 (0 – 10)	5.5 ± 2.4 (0 – 10)
Missing	0.4%	0.0%

Notes: RMDQ = Roland Morris Disability Questionnaire; NPRS = Numeric Pain Rating Scale

p -value = 0.0467), had higher levels of fear-avoidance beliefs ($\chi^2 = 5.09$, $df = 1$, p -value = 0.0241), and less HMO health insurance ($\chi^2 = 15.62$, $df = 5$, p -value = 0.0080).

The following prevalence of DP constructs (Table 2) were found at first examination for patients with first examination and follow-up data: Pain Intensity Change (17.3%), Increase in Spine ROM (29.5%), Change in Aberrant Trunk Movement (6.3%), Patient Reported Improvement in ROM (74.8%), Improved Neurological Signs (10.3%), and Patient Reported Improvement in Functional Task (6.3%).

The effects of DP constructs on disability and pain intensity were analyzed (Table 2). Difference scores were calculated, or the difference between the clinical outcome measure (i.e. disability or pain intensity) at first examination and follow-up. All groups made clinically meaningful improvements in disability and pain intensity at follow-up; however, no clinically significant differences were found between the groups.

Lumbar spine extension and flexion ROM changes were divided into two categories (Table 3): '+' (i.e. ROM change values in lumbar spine flexion and/or extension that meet or exceed the minimal detectable change) and '0' (i.e. ROM change values in lumbar spine flexion and extension that do not meet or exceed the minimal detectable change). 29.5% exhibited an increase in lumbar spine extension ROM post repeated movement testing on the first examination. 6.5% of patients exhibited an increase in lumbar spine flexion ROM post repeated movement testing on the first examination. The mean change for extension and flexion lumbar spine ROM on first examination was 5.65 and 3.53, respectively. The median change for extension and flexion lumbar spine ROM on first examination was 5 and 0, respectively. The average time to discharge was 36.1 days (SD = 24.17; range = 1–124).

Discussion

The findings in this investigation were that (1) the most prevalent DP constructs at first examination were related to ROM and pain intensity, (2) all groups improved and made clinically meaningful improvements in disability and pain intensity at follow-up, (3) no clinically significant differences in disability or pain intensity were found between the groups at

Table 2. First Examination DP Constructs and Follow-up Disability and Pain Intensity Comparison ($n = 400$).

	Status	Counts	Frequency	Mean Pain at First Examination	Mean Pain at Follow-up	Pain Difference	Test for Pain Difference (p-value)	Mean Disability at First Examination	Mean Disability at Follow-up	Disability Difference	Test for Disability Difference (p-value)
Pain Intensity Change	Present	69	17.3%	6.45	1.86	4.59	0.00*	9.31	3.80	5.50	0.06
	Absent	331	82.8%	5.36	2.04	3.31		7.58	3.17	4.41	
	Missing	0	0.0%								
Increase in Spine ROM	Present	118	29.5%	5.93	1.69	4.24	0.01*	7.95	2.69	5.26	0.14
	Absent	232	58.0%	5.48	2.02	3.46		8.43	3.80	4.63	
	Missing	50	12.5%								
Change in Aberrant Trunk Movement	Present	25	6.3%	5.72	0.80	4.92	0.02*	7.20	1.64	5.56	0.73
	Absent	375	93.8%	5.54	2.09	3.44		7.93	3.39	4.54	
	Missing	0	0.0%								
Patient Reported Improvement in ROM	Present	299	74.8%	5.65	1.99	3.66	0.11	8.06	3.33	4.73	0.28
	Absent	101	25.3%	5.25	2.08	3.17		7.33	3.12	4.21	
	Missing	0	0.0%								
Improved Neurological Signs	Present	41	10.3%	6.39	1.70	4.70	0.00*	9.67	3.48	6.19	0.14
	Absent	359	89.8%	5.45	2.05	3.40		7.67	3.25	4.42	
	Missing	0	0.0%								
Patient Reported Improvement in Functional Task	Present	25	6.3%	6.44	1.84	4.60	0.06	11.35	2.65	8.70	0.00*
	Absent	375	93.8%	5.49	2.02	3.46		7.65	3.31	4.34	
	Missing	0	0.0%								

Notes: Values are proportions (percent) or prevalence rates of the DP construct of interest. Level of disability was determined by the Roland Morris Disability Questionnaire (RMDQ). Level of pain intensity was determined by the Numeric Pain Rating Scale (NPRS). Difference scores were calculated by subtracting the mean value (i.e. NPRS or RMDQ score) at first examination from the mean value at follow-up. *Statistically significant

Table 3. Lumbar Spine ROM, RMDQ and NPRS Analysis ($n = 400$).

	ROM	Counts	Frequency	Mean Disability at First Examination	Mean Disability at Follow-up	Disability Difference	Test for Disability Difference (p-value)	Mean Pain at First Examination	Mean Pain at Follow-up	Pain Difference	Test for Pain Difference (p-value)
Extension	(+)	106	26.5%	7.45	2.44	5.01	0.34	5.90	1.68	4.22	0.01*
	(0)	241	60.2%	8.63	3.86	4.77		5.53	2.00	3.53	
	Missing	53	13.3%								
Flexion	(+)	26	6.5%	10.15	3.53	6.62	0.03*	6.00	1.61	4.39	0.10
	(0)	321	80.2%	8.13	3.43	4.70		5.60	1.93	3.67	
	Missing	53	13.3%								

Notes: ROM = range of motion; RMDQ = Roland Morris Disability Questionnaire; NPRS = Numeric Pain Rating Scale. Values are proportions (percent) or prevalence rates for the ROM related Directional Preference (DP) constructs. Level of disability was determined by the RMDQ. Level of pain intensity was determined by the NPRS. Difference scores were calculated by subtracting the mean value (i.e. NPRS or RMDQ score) at first examination from the mean value at follow-up. (+) corresponds with ROM change that meets or exceeds the minimal detectable change. (0) corresponds with ROM change that does not meet or exceed the minimal detectable change.

*Statistically significant

follow-up, and (4) more patients exhibited a relative increase in lumbar spine extension ROM than flexion ROM on the first examination.

The most prevalent DP constructs were related to pain intensity and ROM change. This finding aligns with reports by McKenzie [5] in that pain and ROM change are typically associated with DP. McKenzie describes DP as 'the propensity of mechanical back and referred pain to lessen if movements or positions in one direction are performed and to worsen if movements or postures in the opposite direction are performed.' McKenzie also describes DP to be associated with the Derangement Syndrome and that the mechanical presentation (i.e. movement and ROM-related findings) changes depending on the type of DP [5]. The findings of this study corroborate these reports by McKenzie [5] and show that pain intensity and ROM change are key determinants of DP when

CEN does not occur at the first examination. MDT providers should consider tailoring their physical examinations to assess these constructs and feel confident that their patients will attain clinically significant improvements at follow-up when an MDT examination is conducted in the manner presented in this study.

All groups improved and made clinically meaningful improvements in disability and pain intensity at follow-up. However, no clinically significant differences in disability or pain intensity were found between the groups at follow-up (Table 2). Several considerations may explain the outcomes observed in this investigation. Firstly, the clinically significant improvements found at follow-up may be due to the individualized MDT assessment, classification, and treatment that each patient received for their LBP complaints. Patient classification and individually tailored exercise programs have been shown to result

in superior outcomes [8]. However, the design of this study precludes analysis and understanding of whether the outcomes achieved were due to the treatments rendered or other factors related to the clinical encounter, such as patient and/or provider related characteristics. Additionally, the timeframe in which the outcomes were assessed may have attributed to the lack of difference in clinically significant outcomes. The average time to follow-up was 36.1 days (SD = 24.17; range = 1–124). As patients who exhibit DP typically display rapid improvement during the physical examination [5], meaningful differences in outcomes may be better assessed during the first week of treatment rather than at discharge. Further studies should consider assessing outcomes at multiple time points with a focus on short-term outcomes for patients who exhibit a DP in the absence of CEN. Finally, patients were classified at the first examination only and it was not possible to determine if patients' LBP classifications changed on follow-up assessment. Non-CEN and NC groups could have subsequently converted to CEN throughout the clinical encounter and further studies should consider the stability of classification in this cohort.

Lumbar spine flexion and extension ROM was measured via an inclinometer before and after repeated movement testing. An increase in lumbar spine extension ROM was more common (26.5%) than an increase in lumbar spine flexion ROM (6.5%). Furthermore, no change or a loss in lumbar spine flexion ROM was common (80.2%). This was an expected finding being that the most common DP is extension [38] and, although detailed interventions were not recorded, many patients were likely performing repeated lumbar spine extension and/or lateral movements and instructed to abstain from lumbar spine flexion in an attempt to elicit DP during the examination. It could be that most patients were performing lumbar spine extension and/or lateral procedures for the majority of the session and only had the opportunity to move into lumbar spine flexion for the initial measurement and then the retake measurement. Thus, it is not surprising that no change or a loss of lumbar spine flexion ROM occurred. In the authors' opinion, it is not uncommon that patients report that lumbar spine flexion ROM feels 'stiffer' after performing repeated lumbar spine extension movements. Further research should consider the detailed intervention performed to understand whether repeated and/or sustained movements in one direction have an effect on the ROM in the same direction, the opposite direction, or both.

There are several limitations to the findings in this study. The DP constructs used in this study were adopted from McKenzie [5] and a previous investigation [11] and expanded upon in focus group meetings with diploma and certified providers in MDT who worked in Certified McKenzie Clinics. Efforts were made to develop a DP framework that reflected actual MDT clinical practice; however, the number of

constructs used may be excessive or limited and studies need to investigate these considerations. Further studies should examine the comprehensiveness, mutual exclusivity, or discriminant validity of the DP assessment framework used in this investigation.

To maximize generalizability, providers in the study were instructed to maintain usual care and integrate their clinical expertise relative to MDT to determine which DP constructs were applicable for a given patient presentation. For instance, neurological assessment is typically not indicated for axial LBP and providers may have excluded the corresponding DP tests from their examination. Thus, it was not required that all DP constructs be assessed for each patient and some constructs which may have been present could have been excluded if the provider felt assessment of that DP construct was irrelevant. Further studies need to examine the differences and similarities in practice patterns related to DP assessment in the absence of CEN among MDT providers in varying geographical locations and practice settings. As providers in the current study had advanced levels of training in MDT, the results may only be generalizable to providers with this type of education. The role of MDT training level and its subsequent effect on the DP assessment process warrants investigation.

Conclusion

The primary findings in this investigation were that (1) the most prevalent DP constructs at first examination were related to ROM and pain intensity, (2) all groups improved and made clinically meaningful improvements in disability and pain intensity at follow-up, (3) no clinically significant differences in disability or pain intensity were found between the groups at follow-up, and (4) more patients exhibited an increase in lumbar spine extension ROM than flexion ROM post repeated movement testing on the first examination. The findings in this study assist providers in making assessment and treatment decisions with their patients by offering insight regarding the most prevalent DP constructs typically found at the first examination and their subsequent association with outcome when CEN does not occur. Recommendations for researchers have been made to further explore the DP framework used in this study.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Richard Yarznbowicz obtained his doctor of physical therapy from the University of the Sciences in Philadelphia in 2010 before earning his diploma in Mechanical Diagnosis and Therapy in 2013. He is also a board certified Orthopedic

Clinical Specialist. He has authored and co-authored several publications regarding the assessment and treatment of patients with musculoskeletal disorders.

Minjing Tao is an Assistant Professor of Statistics at The Florida State University. She earned her Ph.D. in Statistics from UW-Madison in 2013. Her research interests have included high dimensional data analysis, financial time series, and statistical inference for high-frequency financial data.

References

- [1] Cook C, Ej H, Ramey K. Physical therapy exercise intervention based on classification using the patient response method: a systematic review of the literature. *J Man Manip Ther.* 2005;13:152–162.
- [2] Delitto A, George SZ, van Dillen LR, et al. Low back pain. *J Orthop Sports Phys Ther.* 2012;42:A1–57.
- [3] Surkitt LD, Ford JJ, Hahne AJ, et al. Efficacy of directional preference management for low back pain: a systematic review. *Phys Ther.* 2012;92:652–665.
- [4] Werneke MW. “Centralization” and “directional preference” are not synonymous. *J Orthop Sport Phys Ther.* 2009;39:827.
- [5] McKenzie RA, May S. The lumbar spine: mechanical diagnosis and therapy. Waikanae, New Zealand: Spinal Publication, Ltd; 2003.
- [6] Hall H, McIntosh G, Boyle C. Effectiveness of a low back pain classification system. *Spine J.* 2009;9:648–657.
- [7] Fairbank J, Se G, Jc F, et al. The role of classification of chronic low back pain. *Spine (Phila Pa 1976).* 2011; S19–42.
- [8] Long A, Donelson R, Fung T. Does it matter which exercise? *Spine (Phila Pa 1976).* 2004;29:2593–2602.
- [9] George SZ, Bialosky JE, Donald DA. The centralization phenomenon and fear-avoidance beliefs as prognostic factors for acute low back pain: a preliminary investigation involving patients classified for specific exercise. *J Orthop Sport Phys Ther.* 2005;35:580–588.
- [10] Yarnsbowicz R, Tao M, Owens A, et al. Pain pattern classification and directional preference are associated with clinical outcomes for patients with low back pain. *J Man Manip Ther.* 2018;26:18–24.
- [11] Werneke MW, Hart DL, Cutrone G, et al. Association between directional preference and centralization in patients with low back pain. *J Orthop Sport Phys Ther.* 2011;41:22–31.
- [12] Fritz JM, Cleland J, Jd C. Subgrouping patients with low back pain: evolution of a classification approach to physical therapy. *J Orthop Sports Phys Ther.* 2007;37:290–302.
- [13] Werneke M, Hart DL. Discriminant validity and relative precision for classifying patients with nonspecific neck and back pain by anatomic pain patterns. *Spine (Phila Pa 1976).* 2003;28:161–166.
- [14] Werneke MW, Hart DL, Resnik L, et al. Centralization: prevalence and effect on treatment outcomes using a standardized operational definition and measurement method. *J Orthop Sport Phys Ther.* 2008;38:116–125.
- [15] Wand BM, Hunter R, O’Connell NE, et al. The self-reported aggravating activities of people with chronic non-specific low back pain do not involve consistent directions of spinal movement: an observational study. *Aust J Physiother.* 2009;55:47–51.
- [16] Browder DA, Childs JD, Cleland JA, et al. Effectiveness of an extension-oriented treatment approach in a subgroup of subjects with low back pain: a randomized clinical trial. *Phys Ther.* 2007;87:1608–1618.
- [17] Paatelma M, Kilpikoski S, Simonen R, et al. Orthopaedic manual therapy, Mckenzie method or advice only for low back pain in working adults: a randomized controlled trial with one year follow-up. *J Rehabil Med.* 2008;40:858–863.
- [18] Brennan GP, Fritz JM, Hunter SJ, et al. Identifying subgroups of patients with acute/subacute “nonspecific” low back pain results of a randomized clinical trial. *Spine (Phila Pa 1976).* 2006;31:623–631.
- [19] Miller ER, Schenk RJ, Karnes JL, et al. A comparison of the Mckenzie approach to a specific spine stabilization program for chronic low back pain. *J Man Manip Ther.* 2005;13:103–112.
- [20] Schenk RJ, Jozefczyk C, Kopf A. A randomized trial comparing interventions in patients with lumbar posterior derangement. *J Man Manip Ther.* 2003;11:95–102.
- [21] Rondinelli R, Murphy J, Esler A, et al. Estimation of normal lumbar flexion with surface inclinometry: A comparison of three methods. *Am J Phys Med Rehabil.* 1992;71:219–224.
- [22] Mellin G. Measurement of thoracolumbar posture and mobility with a mvrin inclinometer. *Spine (Phila Pa 1976).* 1986;11:759–762.
- [23] Mellin G, Kiiski R, Weckstrom A. Effects of subject position on measurements of flexion, extension, and lateral flexion of the spine. *Spine (Phila Pa 1976).* 1991;16:1108–1110.
- [24] Saur PMM, Ensink FBM, Frese K, et al. Lumbar range of motion: reliability and validity of the inclinometer technique in the clinical measurement of trunk flexibility. *Spine (Phila Pa 1976).* 1996;21:1332–1338.
- [25] Hicks GE, Fritz JM, Delitto A, et al. Interrater reliability of clinical examination measures for identification of lumbar segmental instability. *Arch Phys Med Rehabil.* 2003;84:1858–1864.
- [26] Biely SA, Silfies SP, Smith SS, et al. Clinical observation of standing trunk movements: what do the aberrant movement patterns tell us? *J Orthop Sport Phys Ther.* 2014;44:262–272.
- [27] Sc K, Waring W, Biering-Sorensen F, et al. Reference for the 2011 revision of the international standards for neurological classification of spinal cord injury. *J Spinal Cord Med.* 2011;34:547–554.
- [28] Clifton GL, Donovan WH, Dimitrijevic MM, et al. Omental transposition in chronic spinal cord injury. *Spinal Cord.* 1996;34:193–203.
- [29] Wadsworth CT, Krishnan R, Sear M, et al. Intrarater reliability of manual muscle testing and hand-held dynamometric muscle testing. *Phys Ther.* 1987;67:1342–1347.
- [30] Jm F, Pandya S, Wm K, et al. Intrarater reliability of manual muscle test (Medical research council scale) grades in Duchenne’s muscular dystrophy. *Phys Ther.* 1992;72:115–122.
- [31] Barr A, Diamond B, Wade C, et al. Reliability of testing measures in Duchenne or Becker muscular dystrophy. *Arch Phys Med Rehabil.* 1991;72:315–319.
- [32] Magee D. Orthopedic physical assessment. 5th ed. St. Louis, Missouri: Saunders Elsevier; 2008.

- [33] Litvan I, Mangone CA, Werden W, et al. Reliability of the NINDS myotatic reflex scale. *Neurology*. 1996;47:969–972.
- [34] Philip K, Lew P, Matyas T. The Inter-Therapis Reliability of the slump test. *Aust J Physiother*. 1989;35:89–94.
- [35] Childs J, Piva S, Fritz J. Responsiveness of the numeric pain rating scale in patients with low back pain. *Spine (Phila Pa 1976)*. 2005;30:1331–1334.
- [36] Ostelo RWJG, Deyo RA, Stratford P, et al. Interpreting change scores for pain and functional status in low back pain: towards international consensus regarding minimal important change. *Spine (Phila Pa 1976)*. 2008;33:90–94.
- [37] Fritz JM, Piva SR. Physical impairment index: reliability, validity, and responsiveness in patients with acute low back pain. *Spine (Phila Pa 1976)*. 2003;28:1189–1194.
- [38] May S, Aina A. Centralization and directional preference: a systematic review. *Man Ther*. 2012;17:497–506.
- [39] Werneke MW, Hart DL, Cook D, et al. Study of the centralization phenomenon. *Spine (Phila Pa 1976)*. 1999;24:676–683.
- [40] Cyriax J. *Diagnosis of soft tissue lesions*. 6th ed. Baltimore: Williams & Wilkins; 1976.
- [41] Delitto A, Erhard RE, Richard W, et al. Approach to low back syndrome : identrfying and staging patients for conservative treatment. *Phys Ther*. 1995;75:470–485.
- [42] Nachemson A. Lumbar spine instability: a critical update and symposium summary. *Spine (Phila Pa 1976)*. 1985;10:290–291.
- [43] Ogon M, Bender BR, Hooper DM, et al. A dynamic approach to spinal instability: part II: hesitation and giving-way during interspinal motion. *Spine (Phila Pa 1976)*. 1997;22:2859–2866.
- [44] Macnab I, McCulloch J. *Backache*. 2nd ed. Baltimore: Williams & Wilkins; 1990.

Appendix

DP Constructs		
Construct	Measurement Method	DP Inclusion Criteria
Pain Intensity Change	Before and after repeated movement testing, the patient was asked by the managing provider to report his or her distal most relevant pain intensity. The patient was instructed to shade in the body region on a body diagram where he or she was currently experiencing spinal pains and referred symptoms in a standing position [39]. The patient rated the distal most relevant symptoms using the 11-point NPRS ranging from 0 ("No Pain") to 10 ("Worst Pain Imaginable") [35].	A decrease in the patient's most distal relevant symptom by at least two pain intensity units
Increase in Spine ROM	Before and after repeated movement testing, the managing provider measured the patient's lumbar spine ROM in flexion and extension. The managing provider placed and zeroed an inclinometer at the thoracolumbar junction while the patient was standing in neutral [24]. The patient was asked to bend forward maximally and the motion was recorded at the thoracolumbar junction. The patient was then asked to bend backward maximally and the motion was recorded at the thoracolumbar junction.	An increase in lumbar spine ROM by at least 16 degrees in flexion and/or 8 degrees in extension
Change in Aberrant Trunk Movement	Before and after repeated movement testing, the managing provider recorded the presence of aberrant trunk movement according to the following criteria: Painful Arc in Flexion: symptoms felt during the movement at a particular point in the motion (or through a particular portion of the range) that are not present before or after this point [40] Painful Arc on Return: symptoms occur only during the return from the flexed to the erect position [40] Gower Sign ("thigh climbing"): pushing on the thighs or another surface with the hands for assistance during the return from the flexed to the erect position [41] Instability Catch: any sudden acceleration or deceleration of trunk movement or movement occurring outside the primary plane of motion (e.g., lateral bending or rotation during trunk flexion) [42,43] Reversal of Lumbopelvic Rhythm: on attempting to return from the flexed position, the patient bends the knees and shifts the pelvis anteriorly before returning to the erect position [44]	An abolition of any of the aberrant trunk movements
Patient Reported Improvement in ROM	During the initial lumbar spine movement testing, the provider asked the patient if he or she was experiencing any pain during or at the end of the movement. The provider assessed lumbar spine flexion, extension, and left and right side glide. After repeated movement testing, each lumbar spine movement was reassessed and the provider asked the patient "Does this movement feel better, worse, or the same?"	The patient states that the movement feels better relative to the initial assessment
Improved Neurological Signs	The managing provider determined if the patient's clinical presentation warranted a neurological examination. If relevant to the patient's clinical presentation, a standard neurological examination was conducted before and after repeated movement testing. The examination included assessment of lower extremity myotomes, dermatomes, reflexes, and the Slump Test (please see descriptions below)	Improvement in the patient's neurological signs including any of the following: improvement in lower extremity myotomes by $\geq$ one manual muscle test grade, lower extremity dermatomes by $\geq$ one grade, abnormal hypotonic or hypertonic lower extremity reflexes by $\geq$ one grade, and the Slump Test by a patient-reported reduction in concordant symptoms during the test
Patient Reported Improvement in Functional Task	Before repeated movement testing, the managing provider asked the patient to perform a functional task in an attempt to elicit the concordant symptom. The patient was asked to either lift a box from the floor to a surface or perform a squat. The patient was then asked to report the level of pain he or she was experiencing while performing the functional movement. The 11-point NPRS [35] ranging from 0 ("No Pain") to 10 ("Worst Pain Imaginable") was used to quantify the intensity of pain during the functional task. After repeated movement testing, the patient's pain intensity was reassessed during the functional movement using the NPRS.	A decrease in the patient's pain intensity with the functional task by at least two pain intensity units
Neurological Tests		
Test	Description	Grading
Myotome Testing	Before repeated movement testing, the managing provider assessed the following lower extremity myotomes via manual muscle testing: L2 – Hip flexors (Iliopsoas), L3 – Knee extensors (Quadriceps), L4 – Ankle dorsiflexors (Tibialis Anterior), L5 – Long toe extensors (Extensor Hallucis Longus), and S1 – Ankle plantar flexors (Gastrocnemius, Soleus) [27]	Myotome testing was graded via 0 = total paralysis, 1 = palpable or visible contraction, 2 = active movement, full range of motion (ROM) with gravity eliminated, 3 = active movement, full ROM against gravity, 4 = active movement, full ROM against gravity and moderate resistance in a muscle-specific position, 5 = (normal) active movement, full ROM against gravity and full resistance in a functional muscle position expected from an otherwise unimpaired person [27]

(Continued)

(Continued).

Neurological Tests		
Test	Description	Grading
Dermatome Testing	Before repeated movement testing, the managing provider assessed the following lower extremity dermatomes via light touch testing: L1 – Midway distance between the key sensory points for T12 and L2, L2 – On the anterior-medial thigh at the midpoint drawn connecting the midpoint of inguinal ligament (T12) and the medial femoral condyle, L3 – Medial femoral condyle above the knee, L4 – Medial malleolus, L5 – Dorsum of the foot at the third metatarsal phalangeal joint, S1 – Lateral heel (calcaneus), and S2 – Mid-point of the popliteal fossa [27]	Dermatome testing was graded via 0 = absent, 1 = altered (either decreased/impaired sensation or hypersensitivity, 2 = normal [27]
Reflex Testing	Before repeated movement testing, the managing provider assessed the following lower extremity reflexes via a standard reflex hammer: the Quadriceps (L3) and Gastrocnemius (S1) muscle stretch reflexes [32]	Reflex testing was graded via 0 = no response (always abnormal), 1+ = a slight but definitely present response (may or may not be normal), 2+ = a brisk response (normal), 3+ = a very brisk response (may or may not be normal), and 4+ = a tap elicits a repeating reflex or clonus (always abnormal) [32]
Slump Test	Before repeated movement testing, the managing provider asked the patient to sit in a slumped position with the knees flexed over the treatment table. The managing provider sequentially introduced cervical flexion, knee extension, and ankle dorsiflexion and assessed for the onset of the patient's concordant symptoms. Relief of symptoms was assessed via cervical spine extension, ankle plantarflexion, and/or knee flexion [34].	The test was considered positive if the patient's report of concordant symptoms were increased in the slumped position and decreased with cervical spine extension, ankle plantarflexion, and/or knee flexion