

## HEALTH SERVICES RESEARCH

## McKenzie Lumbar Classification

*Inter-rater Agreement by Physical Therapists With Different Levels of Formal McKenzie Postgraduate Training*

Mark W. Werneke, PT, MS, Dip. MDT,\* Daniel Deutscher, PT, PhD,† Dennis L. Hart, PT, PhD,‡§ Paul Stratford, PT, MSc,§ Joel Ladin, PT,¶ Jon Weinberg, PT, Dip. MDT,|| Scott Herbowy, PT, Dip. MDT,\*\* and Linda Resnik, PT, PhD‡‡

**Study Design.** Inter-rater chance-corrected agreement study.

**Objective.** The aim was to examine the association between therapists' level of formal precredential McKenzie postgraduate training and agreement on the following McKenzie classification variables for patients with low back pain: main McKenzie syndromes, presence of lateral shift, derangement reducibility, directional preference, and centralization.

**Summary of Background Data.** Minimal level of McKenzie postgraduate training needed to achieve acceptable agreement of McKenzie classification system is unknown.

**Methods.** Ratets (N = 47) completed multiple sets of 2 independent successive examinations at 3 different stages of McKenzie postgraduate training (levels parts A and B, part C, and part D). Agreement was assessed with  $\kappa$  coefficients and associated 95% confidence intervals. A minimum  $\kappa$  threshold of 0.60 was used as a predetermined criterion for level of agreement acceptable for clinical use.

**Results.** Ratets examined 1662 patients (mean age = 51  $\pm$  15; range, 18–91; females, 57%). Data distributions were not even and

were highly skewed for all classification variables. No training level studied had acceptable agreement for any McKenzie classification variable. Agreements for all levels of McKenzie postgraduate training were higher than expected by chance for most of the classification variables except parts A and B training level for judging lateral shift and centralization and part D training level for judging reducibility. Agreement between training levels parts A and B, part C, and part D were similar with overlapping 95% confidence intervals.

**Conclusion.** Results indicate that level of inter-rater chance-corrected agreement of McKenzie classification system was not acceptable for therapists at any level of formal McKenzie postgraduate training. This finding raises concerns about the clinical utility of the McKenzie classification system at these training levels. Additional studies are needed to assess agreement levels for therapists who receive additional training or experience at the McKenzie credentialed or diploma levels.

**Key words:** inter-rater, chance-corrected, agreement, reliability, McKenzie system, syndrome classification, lumbar spine, centralization, directional preference, lateral shift, reducibility, derangement.

**Level of Evidence:** 2

**Spine 2014;39:E182–E190**

From the \*CentraState Medical Center, Freehold, NJ; †Physical Therapy Service, Maccabi Healthcare Services, Tel Aviv, Israel; ‡Focus On Therapeutic Outcomes, Inc., White Stone, VA; §School of Rehabilitation Science, Institute of Applied Sciences and Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada; ¶Physical Therapy Service, Maccabi Healthcare Services, Tel Aviv, Israel; ||Team Care Physical Therapy, Oxford, NC; \*\*Physical Therapy Department, St David's Hospital Austin, TX; and ††Providence VA Medical Center, Providence, RI, and Department of Health Services, Policy and Practice, Brown University, Providence, RI.

Acknowledgment date: June 24, 2013. First revision date: September 10, 2013. Second revision date: October 12, 2013. Acceptance date: October 21, 2013.

†Died April 11, 2012.

The manuscript submitted does not contain information about medical device(s)/drug(s).

No funds were received in support of this work.

There were no relevant financial activities outside the submitted work: board membership, consultancy, employment, payment for lecture.

Address correspondence and reprint requests to Mark W. Werneke, PT, MS, Dip. MDT, CentraState Medical Center, 901 West Main St, Freehold, NJ 07728; E-mail: mwernke@centrastate.com

DOI: 10.1097/BRS.0000000000000117

Classifying patients with nonspecific low back pain (LBP) to direct treatment and improve patient outcomes is recognized as an important clinical and research priority.<sup>1–7</sup> Before a classification system is recommended for clinical decision making, acceptable inter-rater chance-corrected agreement (agreement) should be demonstrated.<sup>8</sup>

One popular classification method is the McKenzie classification system (McKenzie).<sup>6</sup> McKenzie classifies patients with LBP into 1 of the 3 main syndromes (derangement, dysfunction, and posture).<sup>6</sup> Classifications are based on patient's subjective history and objective examinations of a patient's posture, trunk alignment, and specific clinical signs and symptoms (*i.e.*, centralization, peripheralization, and directional preference) from repeated end-range lumbar movements and manual and or static positioning techniques during physical examination. These pain responses have been well described in the literature<sup>6,9</sup> and centralization and directional

preference have been consistently reported to be associated with favorable patient pain and functional outcomes.<sup>10</sup>

McKenzie is commonly used by physical therapists to evaluate, clinically diagnose, and manage patients with lumbar impairments.<sup>11–14</sup> Despite the popularity and prevalence of using McKenzie, scant data have been reported on its inter-rater reliability. Two prior studies by Riddle and Rothstein<sup>15</sup> and Kilby *et al*<sup>16</sup> who reported  $\kappa$  under 0.60 involved therapists with limited training in the McKenzie classification system, that is, no formal training or partially trained therapists who attended only parts A and B. In contrast, 3 additional studies by Razmjou *et al*,<sup>17</sup> Kilpikoski *et al*,<sup>18</sup> and Clare *et al*<sup>19</sup> reported acceptable to almost perfect agreement with  $\kappa = 0.70, 0.60$ , and  $1.0$ , respectively, for judging McKenzie main classification syndromes for patients with lumbar impairments. Physical therapists participating in these studies had completed all postgraduate McKenzie training levels, that is parts A to D, and achieved credentialed status in the McKenzie classification system. However, the design of these studies did not follow recommended blinded assessment and number of raters and patient guidelines for reporting reliability,<sup>20,21</sup> thereby threatening internal validity of their results. No studies have examined agreement for therapists who had advanced training, for parts C and D, and elected not to take the credentialed examination.

The minimal level of formal McKenzie postgraduate training needed to achieve acceptable agreement levels remains unknown. Considering the wide variations in training levels for physical therapists using the McKenzie method during routine practice,<sup>13–16</sup> it would be of clinical and financial interest to determine the minimal level of McKenzie educational training required to classify patients according to McKenzie classification system, in a reliable and clinically useful manner.

Thus, the purpose of this study was to examine the association between therapist level of McKenzie postgraduate training and agreement of McKenzie syndrome classification variables for patients with LBP. Specific aims were to assess agreement between therapists as they progressed through McKenzie postgraduate precredential training for identifying: (1) main McKenzie syndromes, that is, posture, dysfunction, derangement, or “other” category, (2) presence of lateral shift and reducibility of derangements, and (3) directional preference and centralization. We hypothesized that acceptable agreement would be observed only for trainees completing all levels of the McKenzie postgraduate education, that is, parts A and B, C, and D.

## MATERIALS AND METHODS

### Design

We conducted a prospective study investigating the ability of 47 physical therapists to agree on main McKenzie syndromes, that is, derangement, dysfunction, posture, and other. For those patients judged to have a derangement by both raters, agreement analyses were calculated for presence of lateral shift and derangement reducibility, that is, reducible *versus* irreducible derangement. For patients judged to have

a reducible derangement by both raters, agreement analyses were calculated for directional preference and centralization (Figure 1). The institutional review board for the Protection of Human Subjects of Maccabi Healthcare Services (Maccabi), Israel, approved this study.

### Procedures

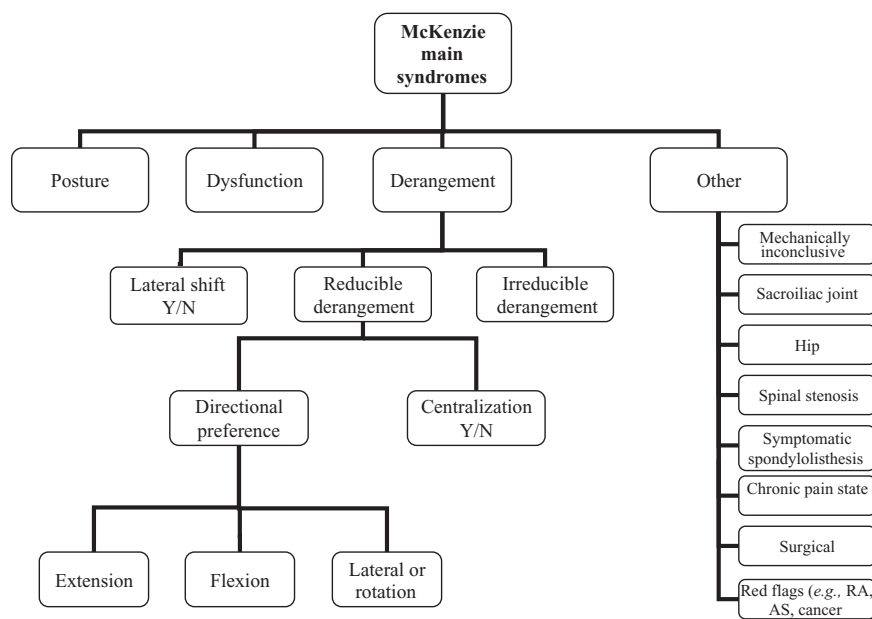
A formal McKenzie postgraduate educational program as recommended by the International McKenzie Institute was implemented, including four 28-hour courses (parts A, B, C, and D). Part A and B educational courses consist of (1) lecture format augmented by actual examination and treatment by the instructor on several real-time patients who were experiencing lumbar (part A) or cervical (part B) pain and (2) open discussions throughout the course to enhance participants' understanding of the practical application of the McKenzie approach. Parts C and D are considered advanced training in McKenzie methods for those who have completed parts A and B. Part C emphasizes problem-solving case studies and clinical reasoning for judging all patient classification levels using the McKenzie methods. Part D specifically focuses on demonstrating and practicing manual spinal mobilization techniques and progression of mobilization forces. For the study purposes, 3 training levels defining 3 study stages were level 1: parts A and B, level 2: part C, and level 3: part D, with each stage lasting up to 1 year during a 3-year period from August 2008 to June 2011.

### Subjects

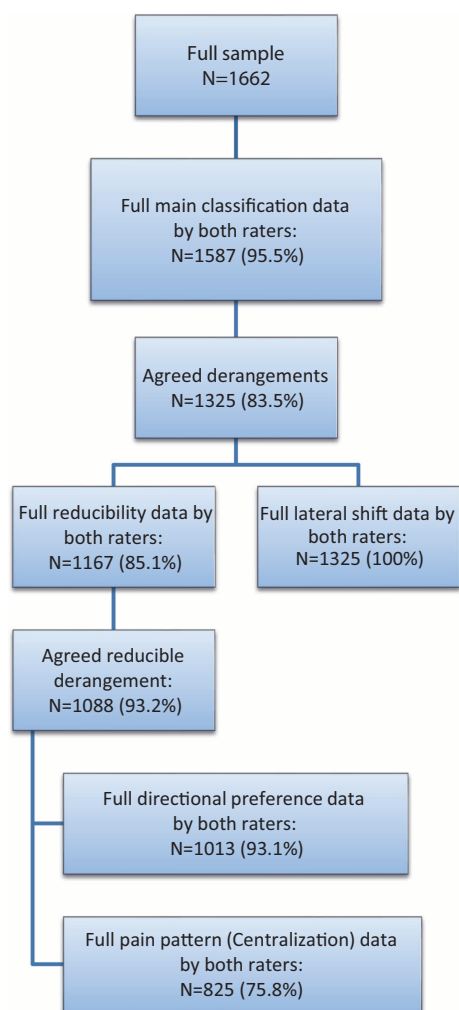
Adult patients (mean age, 51; SD [standard deviation] = 15; range, 18–91; 57% females; 57% patients with chronic LBP; 8% elected lumbar surgery; 37% were taking medications related to lumbar impairment; and 12% had 2 or more medical comorbidities) seeking rehabilitation for LBP with or without referred lower extremity symptoms in Maccabi's physical therapy outpatient clinics were asked to participate in the study and sign a consent form. This sample is representative of the regular population treated in Maccabi's physical therapy service due to lumbar impairments.<sup>22</sup> Patients were asked to participate if they were fluent in Hebrew, not pregnant, did not have previous spinal or hip surgery within the past year, and were not involved in work compensation or car insurance litigations. Subjects were recruited from 25 clinics throughout Israel including all 5 districts defined nationally by geographical regions. All participating clinics had 2 or more physical therapists participating in the study who normally evaluated and treated patients with LBP as part of their customary clinical practice. Four clinics had more than 2 participating physical therapists. When an odd number of physical therapists participated, 3 examiners (E1, E2, and E3) created 3 pairs (E1–E2, E2–E3, and E1–E3).

### Examiners

All examiners were physical therapists who signed a consent form agreeing to participate and follow study procedures. Participating therapists (N = 47) had an average of 14 years (SD = 6; range, 5–32) experience treating patients with LBP



**Figure 1.** Variables for agreement analyses: main McKenzie syndromes, derangement's reducibility, directional preference, and centralization. RA indicates rheumatoid arthritis; AS, ankylosing spondylitis.



**Figure 2.** Patient participation and classification.

syndromes. Mean age was 43 (SD = 8; range, 30–65), 76% were females, all had a bachelor's degree in physical therapy, 17% also obtained a master's degree and none had a doctoral degree. Average caseload during the study period defined as percent of all patients treated due to lumbar impairments was 24.5%.

## Patient Examinations

Patient examination followed the recommended guidelines and the Standards for Reporting Diagnostic Accuracy criteria for improving the design of agreement studies.<sup>20,21,23</sup> Each pair of raters was instructed to perform independent and consecutive evaluations for 25 to 30 patients. Patients were scheduled for independent evaluations by 2 examiners during a single clinic visit. Total examination time to complete 2 independent successive examinations was estimated to be 1.5 hours including a washout period of 10 minutes between examinations. To maximize clinician productivity, coordinate clinicians' working hours, and control costs, managerial preorganization of examination sessions between raters was necessary. Therefore, it was not clinically feasible to assign paired raters randomly. Paired examiners were alternately sequenced, that is, examiner number 1 or examiner number 2, so that each would be examiner number 1 in 50% of the patient cases.

Data collection was completed for eligible patients during the initial evaluation. Each examiner used a standard and recommended McKenzie Lumbar Spine Assessment Form (Form) during the evaluation (see Supplemental Digital Content Appendix available at <http://links.lww.com/BRS/A843>).<sup>24</sup> Examiner 1 recorded the evaluation findings on the Form, documented classification data, and placed all completed data in a sealed envelope. Examiner 2 re-evaluated the patient using an identical, but separate Form and recorded the examination

**TABLE 1. Classifications of the First and Second Examiners for Main McKenzie Syndrome (N = 1587)**

First Examiner	Second Examiner			
	Posture Syndrome	Dysfunction	Derangement	Other
Part A + B				
Posture syndrome	8 (1.4)	0 (0)	8 (1.4)	1 (0.2)
Dysfunction	0 (0)	2 (0.3)	8 (1.4)	1 (0.2)
Derangement	7 (1.2)	3 (0.5)	479 (82.0)	27 (4.6)
Other	1 (0.2)	2 (0.3)	18 (3.1)	19 (3.3)
Part C				
Posture syndrome	2 (0.3)	0 (0)	5 (0.8)	0 (0)
Dysfunction	0 (0)	2 (0.3)	3 (0.5)	1 (0.2)
Derangement	1 (0.2)	4 (0.7)	512 (86.6)	16 (2.7)
Other	0 (0)	2 (0.3)	23 (3.9)	20 (3.4)
Part D				
Posture syndrome	0 (0)	0 (0)	2 (0.5)	0 (0)
Dysfunction	0 (0)	1 (0.2)	6 (1.5)	0 (0)
Derangement	1 (0.2)	5 (1.2)	334 (81.1)	17 (4.1)
Other	1 (0.2)	1 (0.2)	24 (5.8)	20 (4.9)

Values represent number (% of total N within the study stage).  
Total percentage within study stage may range between 99.9 and 100.1 due to rounding of percentages.

and classification data electronically in Maccabi's electronic health record. The first examiner was instructed to avoid discussing the specific examination and classification results with the patient. To avoid expectation bias between examiners, examiners were not allowed to observe the other examiner's evaluation nor discuss any classification findings. All eligible and consenting patients completed the double examination process.

### Data Analyses

Inter-rater chance-corrected agreement for pairs of physical therapist raters were calculated using generalized  $\kappa$  values

with 95% confidence intervals estimated using a 1000 bootstrapping with replacement procedure using STATA software (version 10.1) (StataCorp LP, College Station, TX). The generalized  $\kappa$  accommodates 2 nonunique raters. Separate  $\kappa$  estimates were performed for 3 levels of postgraduate education (parts A and B, C, and D). In addition, we calculated data distribution for all classification variables.

Kappa values of 0.60 to 0.79, 0.80 to 0.90, and above 0.90 are interpreted as moderate, strong, and almost perfect, respectively.<sup>25</sup> Although there is no clear consensus on a single threshold, most experts recommend  $\kappa$  values of 0.60, 0.70, or 0.85 as minimum standards for reliability coefficients for use in health care or in clinical research.<sup>21,23,25</sup> For this study, we selected  $\kappa = 0.60$  as a predetermined criterion for acceptable agreement.

### RESULTS

The numbers of patients analyzed for each study variable are shown in Figure 2. Agreement data could be calculated only for patients who had full (no missing) data from both raters for the relevant classification categories.

Not all physical therapists participating in the study completed data collection at all stages of the study, that is parts A and B, part C, and part D, for a variety of personal reasons. Sixteen physical therapists (8 paired raters) completed only the first 2 stages and 1 paired rater completed only the first

**TABLE 2. Agreement Data for Main McKenzie Syndrome (N = 1587)**

	Parts A and B	Part C	Part D
N	584	591	412
Observed agreement (%)	87	91	86
$\kappa$	0.40	0.44	0.37
95% CI	0.28–0.50	0.32–0.56	0.24–0.49

CI indicates confidence interval.

**TABLE 3. Classifications of the First and Second Examiners for Presence of Lateral Shift for Patients With Agreed Derangement (N = 1325)**

First Examiner	Second Examiner	
	No Lateral Shift	Lateral Shift
Parts A + B		
No lateral shift	433 (90.4)	16 (3.3)
Lateral shift	26 (5.4)	4 (0.8)
Part C		
No lateral shift	448 (87.5)	20 (3.9)
Lateral shift	27 (5.3)	17 (3.3)
Part D		
No lateral shift	287 (85.9)	8 (2.4)
Lateral shift	24 (7.2)	15 (4.5)

Values represent N (% of total N within the study stage).  
Total percentage within study stage may range between 99.9 and 100.1 due to rounding of percentages.

stage, and then dropped out of the study. The average number of patients examined per paired raters at each stage was 26.8 (SD = 7.8; range, 2–36).

Distribution of classifications of the first and second examiners, and consequent observed and chance-corrected agreement with  $\kappa$  coefficients with associated 95% confidence interval for each study stage are presented in Tables 1–10. Data distributions were not even and were highly skewed for all classification variables. Agreements for levels of McKenzie postgraduate training were higher than expected by chance for most of the classification variables except lateral shift and centralization (Tables 4, 10) for training level parts A and B and reducibility (Table 6) for training level part D. Level of agreement between therapists at training levels part A and B, part C, and part D was similar and below acceptable agreement levels for all McKenzie classification variables assessed.

**TABLE 4. Agreement Data on Presence of Lateral Shift for Patients With Agreed Derangement (N = 1325)**

	Parts A and B	Part C	Part D
N	479	512	334
Observed agreement (%)	91	91	90
$\kappa$	0.11	0.37	0.43
95% CI	−0.01 to 0.27	0.22–0.51	0.24–0.60

CI indicates confidence interval.

**TABLE 5. Classifications of the First and Second Examiners for Reducible Versus Irreducible for Patients With Agreed Derangement (N = 1167)**

First Examiner	Second Examiner	
	Reducible Derangement	Irreducible Derangement
Parts A + B		
Reducible derangement	352 (92.4)	10 (2.6)
Irreducible derangement	14 (3.7)	5 (1.3)
Part C		
Reducible derangement	433 (94.7)	13 (2.8)
Irreducible derangement	7 (1.5)	4 (0.9)
Part D		
Reducible derangement	303 (92.1)	15 (4.6)
Irreducible derangement	9 (2.7)	2 (0.6)

Values represent N (% of total N within the study stage).  
Total percentage within study stage may range between 99.9 and 100.1 due to rounding of percentages.

## DISCUSSION

Our hypothesis that clinicians completing all formal McKenzie training levels would demonstrate acceptable agreement on McKenzie classification was not supported. Although we found that agreement for most levels of trainees was greater than expected by chance, all education levels had below acceptable agreement for all classification variables assessed. Our findings are limited to therapists who completed all formal McKenzie educational training and cannot be generalized to therapists who have passed a 1 day credentialing examination. However, it is not clear why therapists in our study, who completed the same formal training prerequisites as credentialed therapists in prior studies,<sup>17–19</sup> had unacceptable reliability. There are no additional formal postgraduate training courses for therapists selecting credentialed accreditation.

**TABLE 6. Agreement Data on Reducible Versus Irreducible for Patients With Agreed Derangement N = 1167)**

	Parts A and B	Part C	Part D
N	381	457	329
Observed agreement (%)	94	96	93
$\kappa$	0.26	0.26	0.11
95% CI	0.04–0.45	0.04–0.48	−0.05 to 0.29

CI indicates confidence interval.



**TABLE 7. Classifications of the First and Second Examiners for Directional Preference for Patients With Agreed Reducible Derangement (N = 1013)**

First Examiner	Second Examiner		
	Extension	Flexion	Lateral
Parts A + B			
Extension	231 (74.5)	20 (6.5)	6 (1.9)
Flexion	17 (5.5)	16 (5.2)	3 (1.0)
Lateral	8 (2.6)	2 (0.6)	7 (2.3)
Part C			
Extension	291 (71.3)	8 (2.0)	35 (8.6)
Flexion	10 (2.5)	6 (1.5)	6 (1.5)
Lateral	31 (7.6)	3 (0.7)	18 (4.4)
Part D			
Extension	204 (69.2)	9 (3.1)	19 (6.4)
Flexion	11 (3.7)	7 (2.4)	3 (1.0)
Lateral	22 (7.5)	4 (1.4)	16 (5.4)

Values represent N (% of total N within the study stage).  
Total percentage within study stage may range between 99.9 and 100.1 due to rounding of percentages.

The different agreement results between previous studies and our data maybe partially explained by weak study designs used in prior reliability studies. For example, independent history and physical examination have been reported to be a preferred study design for examining agreement between raters to account for rater bias and interactions that occur between patients and raters.<sup>15,26-30</sup> Prior studies used either simultaneous assessment<sup>17,19</sup> or simultaneous subjective history assessment followed by independent objective assessment.<sup>18</sup> In addition, prior studies investigating inter-rater agreement of McKenzie classifications involved only 1 paired raters<sup>17,18</sup> or only a few patients examined by paired raters.<sup>19</sup> These limitations in study design threaten internal validity and

**TABLE 9. Classifications of the First and Second Examiners for Centralization Versus Noncentralization for Patients With Agreed Reducible Derangement (N = 825)**

First Examiner	Second Examiner	
	Centralization	Noncentralization
Parts A + B		
Centralization	175 (78.1)	23 (10.3)
Noncentralization	20 (8.9)	6 (2.7)
Part C		
Centralization	232 (69.7)	39 (11.7)
Noncentralization	30 (9.0)	32 (9.6)
Part D		
Centralization	139 (51.9)	34 (12.7)
Noncentralization	40 (14.9)	55 (20.5)

Values represent N (% of total N within the study stage).  
Total% within study stage may range between 99.9 and 100.1 due to rounding of percentages.

agreement results cannot be clinically interpreted with confidence. Our study used a multicentered approach, recruiting large numbers of pairs of raters and large number of patients per paired raters for each training stage. In addition, we used independent subjective and objective physical examination assessments to control for test, diagnostic and clinical review biases.<sup>20</sup> We think our study has strong internal and external validity enhancing the generalizability, comprehensiveness, and accuracy of our agreement data.

Although our study addressed limitations of prior studies by including therapists with formal McKenzie training, our data are consistent with earlier findings of inadequate agreement data for judging main syndrome McKenzie classification categories.<sup>15,16</sup> Because our findings were unexpected, we also examined agreement levels for each paired raters (data

**TABLE 8. Agreement Data on Directional Preference for Patients With Agreed Reducible Derangement (N = 1013)**

	Parts A and B	Part C	Part D
N	310	408	295
Observed agreement (%)	82	77	77
$\kappa$	0.39	0.27	0.33
95% CI	0.27–0.52	0.17–0.37	0.22–0.45

CI indicates confidence interval.

**TABLE 10. Agreement Data on Centralization Versus Noncentralization for Patients With Agreed Reducible Derangement (N = 835)**

	Parts A and B	Part C	Part D
N	224	333	268
Observed agreement (%)	81	79	72
$\kappa$	0.11	0.35	0.39
95% CI	–0.04 to 0.27	0.22–0.46	0.28–0.51

CI indicates confidence interval.

available upon request). Briefly, for main syndrome, lateral shift, reducibility, and centralization, 2, 4, 1, 2 paired raters, respectively, achieved agreement levels with  $\kappa > 0.6$  at training level part D. Therefore it seems possible that some paired raters achieve acceptable levels of agreement. Although this is encouraging, ways to improve inter-rater agreement in classifying patients with LBP syndromes using the McKenzie system for groups of therapists undertaking formal McKenzie education are yet to be identified.

We examined the association between level of McKenzie training and inter-rater reliability but did not ascertain therapists' competency with the educational materials, which was beyond the scope of our study. McKenzie International Institute does not evaluate a therapist's competence after completing each part of postgraduate training and only offers a basic competency credential examination for therapists who completed all the 4 formal postgraduate courses. Also, we did not examine if the below acceptable agreement levels found resulted from disagreement on information derived from patient's subjective history or physical examination findings, both needed to establish a working classification.

Therefore, we suggest several avenues for future research to explore ways to improve agreement. First, it would be interesting to study if inter-rater classification exercises during formal McKenzie training courses could contribute to improved agreement. Second, we recommend future research examining the therapists' level of competency after taking each of the MDT continuing educational courses, that is, parts A to D. There may be differences in classification agreement based upon competency achieved. Third, because ongoing education has been reported to be superior to traditional short-term intensive courses only, we recommend research into the effect of postcourse implementation and ongoing education.<sup>31,32</sup> Fourth, to improve our understanding of possible sources of disagreement, we recommend evaluating level of agreement on interpretation of either patient's subjective history or physical examination findings, so future educational emphasis could be adapted accordingly. Fifth, we suggest additional rigorously designed research to study inter-rater agreement of McKenzie when used by therapists at the credential or diploma accreditation levels. Finally, we recommend research into whether or not there should be modification to the McKenzie classification categories. The highly skewed data distributions of classification variables found in our study, elevates the chance of reaching observed inter-rater agreement by chance only, lowering the ability to reach acceptable levels of chance-corrected agreement. The lowest chance-corrected agreement levels in our study were for derangements' reducibility, which also had the most skewed distributions and the highest observed agreement due to raters agreeing on 92% to 95% of patients having a reducible derangement (Table 5). Our highly skewed distributions are not an isolated observation; many other authors have also reported high prevalence rates for judging derangement syndromes, directional preference and centralization when using McKenzie methods.<sup>5,17,18,24,33–35</sup>

## Limitations

Not all physical therapists participating in the study were able to examine the recommended number of patients (*i.e.*, 25–30 patients per pair of raters), although those who did provided 94% of our data. The recommendation of rating 25 patients per pair was clinically intuitive but arbitrary. Thus, to determine whether the inclusion of therapists with fewer than 25 patients biased the results, we conducted a secondary analysis including data only from paired raters who had examined at least 25 patients. Results were practically identical to those reported here (data available upon request), strengthening generalizability of our results.

Seventeen (36%) physical therapists did not complete all study stages for a variety of reasons we could not control. To determine if loss of therapists at more advanced training levels influenced our results, we conducted an additional agreement analysis including only data from therapists that completed all study stages and had examined at least 25 patients per pair. Results were very similar, with identical clinical interpretation (data available upon request), which again strengthens generalizability of our results.

A common criticism of independent and consecutive examinations used in our study is that back-to-back examinations may aggravate or change the severity of a patient's symptoms affecting agreement judging syndrome classification. In our study, raters disagreed on whether the derangement syndrome was reducible or irreducible for only 3.5% (41 of 1167) of patients judged for reducibility (Table 5). For instance, 30 patients classified as irreducible by the first examiner and as reducible by the second examiner might have improved between examinations, and 11 patients classified as reducible by the first examiner and as irreducible by the second examiner (might have aggravated between examinations). Therefore, even if disagreement was due to symptom lability between consecutive examinations, we do not think that this substantially biased our findings.

## CONCLUSION

Inter-rater chance-corrected agreements for judging main McKenzie syndromes, presence of lateral shift, derangement reducibility, directional preference, and centralization in patients experiencing LBP did not reach an acceptable agreement level at educational levels parts A and B, C, and D. These findings suggest limited clinical utility of the classification system for therapists with these levels of training. Future research is warranted to study if a modified education process, ongoing implementation processes, additional education at the credential or diploma accreditation levels, or modifications in the classification categories, lead to improved inter-rater chance-corrected agreement in the McKenzie classification system.

### ➤ Key Points

- ❑ Results indicate below acceptable agreement for judging the McKenzie main syndrome

classification, presence of lateral shift, derangement, reducibility, directional preference, and centralization between physical therapist raters completing McKenzie postgraduate training levels including parts A and B, C, and D.

- ❑ The unacceptable agreement levels found to be associated with postgraduate McKenzie educational levels parts A to D raise questions about the clinical utility of the McKenzie classification system at these training levels.
- ❑ The study did not examine agreement levels for physical therapist raters who were credentialed in the McKenzie classification system. There have been no rigorously conducted studies of agreement of raters with McKenzie credentialed accreditation. Future research is warranted to examine the implementation and educational processes required to obtain adequate agreement in the McKenzie classification system.

## Acknowledgments

The authors thank Dr. Yosi Azuri, head of clinical research unit in Maccabi Healthcare Services, for supporting and enabling this study.

The authors also thank the physical therapists and their clinic managers from Maccabi Healthcare Services who participated in this study. Their efforts during the 3-year process enabled this project. They also acknowledge Mr. Dvir Chen, BPT, MManipTher, for helping set up the study and manage the data collection process during the first year of the project. Finally, a special gratitude is dedicated to the research and development team of the Maccabi physical therapy service, for their continuous dedication and excellence, which made this long-term study possible.

A special acknowledgment is dedicated to our coauthor Dr. Dennis L. Hart, who recently passed away, for his guidance and expertise in the study's design and preliminary analyses of the data. Dennis—you are greatly missed, but your contributions to our research group and research initiatives as well as to the physical therapy profession will be long lasting.

Supplemental digital content is available for this article. Direct URL citation appearing in the printed text is provided in the HTML and PDF version of this article on the journal's web site ([www.spinejournal.com](http://www.spinejournal.com)).

## References

- 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* 2006;31:623–31.
- Delitto A, Erhard RE, Bowling RW. A treatment-based classification approach to low back syndrome: identifying and staging patients for conservative treatment. *Phys Ther* 1995;75:470–85; discussion 485–9.
- Fritz JM, Cleland JA, Childs JD. Subgrouping patients with low back pain: evolution of a classification approach to physical therapy. *JOSPT* 2007;37:290–302.
- Henry SM, Fritz JM, Trombley AR, et al. Reliability of a treatment-based classification system for subgrouping people with low back pain. *J Orthop Sports Phys Ther* 2012;42:797–805.
- Long A, Donelson R, Fung T. Does it matter which exercise? A randomized control trial of exercise for low back pain. *Spine* 2004;29:2593–602.
- McKenzie R, May S. *The Lumbar Spine: Mechanical Diagnosis and Therapy*. 2nd ed. Waikanae, New Zealand: Spinal Publication Ltd; 2003.
- Riddle DL. Classification and low back pain: a review of the literature and critical analysis of selected systems. *Phys Ther* 1998;78:708–37.
- Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. *Fam Med* 2005;37:360–63.
- Donelson R, Aprill C, Medcalf R, et al. A prospective study of centralization of lumbar and referred pain. A predictor of symptomatic discs and annular competence. *Spine (Phila Pa 1976)* 1997;22:1115–22.
- May S, Aina A. Centralization and directional preference: a systematic review. *Man Ther* 2012;1–10.
- Battie MC, Cherkin DC, Dunn R, et al. Managing low back pain: attitudes and treatment preferences of physical therapists. *Phys Ther* 1994;74:219–26.
- Byrne K, Doody C, Hurley DA. Exercise therapy for low back pain: a small-scale exploratory survey of current physiotherapy practice in the Republic of Ireland acute hospital setting. *Man Ther* 2006;11:272–8.
- Foster NE, Thompson KA, Baxter GD, et al. Management of nonspecific low back pain by physiotherapists in Britain and Ireland. A descriptive questionnaire of current clinical practice. *Spine* 1999;24:1332–42.
- Gracey JH, McDonough SM, Baxter GD. Physiotherapy management of low back pain: a survey of current practice in northern Ireland. *Spine* 2002;27:406–11.
- Riddle DL, Rothstein JM. Interrater reliability of McKenzie's classifications of the syndrome types present in patients with low back pain. *Spine* 1993;18:1333–44.
- Kilby J, Stigant M, Roberts A. The reliability of back pain assessment by physiotherapists, using a McKenzie Algorithm. *Physiotherapy* 1990;76:579–83.
- Razmjou H, Kramer JF, Yamada R. Interrater reliability of the McKenzie evaluation in assessing patients with mechanical low back pain. *J Orthop Sports Phys Ther* 2000;30:368–83; discussion 384–9.
- Kilpikoski S, Airaksinen O, Kankaanpää M, et al. Interexaminer reliability of low back pain assessment using the McKenzie method. *Spine* 2002;27:E207–14.
- Clare HA, Adams R, Maher CG. Reliability of McKenzie classification of patients with cervical or lumbar pain. *J Manipulative Physiol Ther* 2005;28:122–7.
- Bossuyt PM, Reitsma JB, Bruns DE, et al. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Ann Intern Med* 2003;138:W1–2.
- Kottner J, Audige L, Brorson S, et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *J Clin Epidemiol* 2011;64:96–106.
- Deutscher D, Horn SD, Dickstein R, et al. Associations between treatment processes, patient characteristics, and outcomes in outpatient physical therapy practice. *Arch Phys Med Rehabil* 2009;90:1349–63.
- May S, Littlewood C, Bishop A. Reliability of procedures used in the physical examination of non-specific low back pain: a systematic review. *Aust J Physiother* 2006;52:91–102.
- Clare H, Adams R, Maher CG. Reliability of the McKenzie spinal pain classification using patient assessment forms. *Physiotherapy* 2004;90:114–9.
- McHugh ML. Inter-rater reliability: the kappa statistic. *Biochem Med (Zagreb)* 2012;22:276–82.
- Cleland JA, Childs JD, Fritz JM, et al. Inter-rater reliability of the history and physical examination in patients with mechanical neck pain. *Arch Phys Med Rehabil* 2006;87:1388–95.
- Dankaerts W, O'Sullivan PB, Straker LM, et al. The inter-examiner reliability of a classification method for non-specific chronic low back pain patients with motor control impairment. *Man Ther* 2006;11:28–39.



28. Fersum KV, Dankaerts W, O'Sullivan PB, et al. Integration of subclassification strategies in randomised controlled clinical trials evaluating manual therapy treatment and exercise therapy for non-specific chronic low back pain: a systematic review. *Br J Sports Med* 2009;44:1054–62.
29. Martin RL, Sekiya JK. The inter-rater reliability of 4 clinical tests used to assess individuals with musculoskeletal hip pain. *J Orthop Sports Phys Ther* 2008;38:71–7.
30. Waddell G, Somerville D, Henderson I, et al. Objective clinical evaluation of physical impairment in chronic low back pain. *Spine (Phila Pa 1976)* 1992;17:617–28.
31. Brennan GP, Fritz JM, Hunter SJ. Impact of continuing education interventions on clinical outcomes of patients with neck pain who received physical therapy. *Phys Ther* 2006;86:1251–62.
32. Cleland JA, Fritz JM, Brennan GP, et al. Does continuing education improve physical therapists' effectiveness in treating neck pain? A randomized clinical trial. *Phys Ther* 2009;89:38–47.
33. Aina A, May S, Clare H. The centralization phenomenon of spinal symptoms—a systematic review. *Man Ther* 2004;9:134–43.
34. Donelson R, Silva G, Murphy K. Centralization phenomenon. Its usefulness in evaluating and treating referred pain. *Spine (Phila Pa 1976)* 1990;15:211–3.
35. Hefford C. McKenzie classification of mechanical spinal pain: profile of syndromes and directions of preference. *Man Ther* 2006.