

# RELIABILITY OF MCKENZIE CLASSIFICATION OF PATIENTS WITH CERVICAL OR LUMBAR PAIN

Helen A. Clare, PT, MAppSc,<sup>a</sup> Roger Adams, PhD,<sup>b</sup> and Christopher G. Maher, PT, PhD<sup>c</sup>

## ABSTRACT

**Background:** In the McKenzie system, patients are classified first into syndromes, then into subsyndromes. At present, the reliability of classification with this system is unclear. No study has included patients with cervical pain, and the studies to date have reported conflicting results.

**Objective:** The aim of the study is to investigate the interexaminer reliability of the McKenzie classification system for patients with cervical or lumbar pain.

**Subjects:** Fifty patients with spinal pain (25 with lumbar pain and 25 with cervical pain) were included in the study.

**Method:** The patients were assessed simultaneously by 2 physical therapists (14 in total) trained in the McKenzie method. Agreement was expressed using the multirater  $\kappa$  coefficient and percent agreement for classification into (i) syndromes and (ii) subsyndromes.

**Results:** The reliability for syndrome classification was  $\kappa = 0.84$  with 96% agreement for the total patient pool,  $\kappa = 1.0$  with 100% agreement for lumbar patients, and  $\kappa = 0.63$  with 92% agreement for cervical patients. The reliability for subsyndrome classification was  $\kappa = 0.87$  with 90% agreement for the total patient pool,  $\kappa = 0.89$  with 92% agreement for lumbar patients, and  $\kappa = 0.84$  with 88% agreement for the cervical patients.

**Conclusion:** The McKenzie assessment performed by persons trained in the McKenzie method may allow for reliable classification of patients with lumbar and cervical pain. (*J Manipulative Physiol Ther* 2005;28:122-127)

**Key Indexing Terms:** *Spine; Pain; Physical Examination; Reproducibility of Results*

The McKenzie method for evaluation and treatment of patients is frequently used by clinicians.<sup>1,2</sup> A number of reviews have concluded that the method is effective for the treatment of low-back pain (LBP)<sup>3-5</sup>; however, at present, there are no reviews that have evaluated efficacy for cervical pain. There is some evidence that treatment given according to subclassification is more effective than treatment given to an unselected population.<sup>6,7</sup> A key aspect of the McKenzie approach is that the patients receive individualized treat-

ment based upon their clinical presentation. The McKenzie method uses an assessment process, which aims to identify subgroups of patients within the nonspecific spinal pain population whose symptoms behave in a similar way when subjected to mechanical forces within the physical examination. The classification into subgroups then directs treatment.<sup>8</sup>

Patients are classified into 1 of the 3 McKenzie syndromes (derangement, dysfunction, postural), and those whose presentation does not fit 1 of the 3 syndromes are classified in an "other" grouping. The derangement syndrome is the most prevalent treatment classification.<sup>8</sup> This syndrome is characterized by the centralization and peripheralization of symptoms, in response to repeated movements or sustained postures of the lumbar spine within the physical examination. For example, a patient whose leg and back pain is increased with repeated or sustained flexion movements and reduced with repeated or sustained lumbar extension movements would be placed into the category of derangement and the subcategory of posterior derangement. In this example, the patient's directional preference is for extension movements; therefore, he would be treated with exercises into this direction and encouraged to avoid flexion movements and postures. With the McKenzie approach, the emphasis is on self-management with

<sup>a</sup> Physiotherapist in private practice and PhD candidate at the School of Physiotherapy, The University of Sydney, 2141 Australia.

<sup>b</sup> Senior lecturer, School of Physiotherapy, University of Sydney, 2141 Australia.

<sup>c</sup> Associate professor, School of Physiotherapy, University of Sydney, 2141 Australia.

Sources of support: none declared.

Submit requests for reprints to: Helen A. Clare, PT, MAppSc, 16 Ayres Road, St Ives, NSW 2075, Australia (e-mail: [clare.ha@bigpond.com](mailto:clare.ha@bigpond.com)).

Paper submitted February 3, 2003; in revised form March 21, 2003.

0161-4754/\$30.00

Copyright © 2005 by National University of Health Sciences.

doi:10.1016/j.jmpt.2005.01.003

manipulative techniques reserved for those patients who do not respond to self-management measures.<sup>8-10</sup>

The dysfunction syndrome is characterized by intermittent spinal pain that is reproduced at the end range of a restricted movement. Treatment emphasizes mobilizing exercises in the direction of movement that reproduces pain. The treatment rationale is that the exercise will remodel the tissues limiting the movement. The postural syndrome is characterized by intermittent spinal pain, which is produced with static positioning of the spine and abolished by moving the patient out of the static position. Treatment consists of patient education and avoidance of the provocative postures. The “other” grouping is used for patients where, after several days of mechanical evaluation, a mechanical syndrome cannot be identified.<sup>8</sup> Referral for further medical review is then indicated.

At present, the reliability of the McKenzie classification for patients with spinal pain has not been clearly established. To date, no study has included subjects with cervical pain. This has clinical relevance because cervical pain is a common complaint in primary care. Although 4 studies have investigated lumbar patients, none provides convincing estimates of the reliability to be expected when trained therapists classify patients into the syndromes and subsyndromes described by McKenzie. Problems include (i) the use of minimally trained therapists,<sup>11,12</sup> (ii) the use of a small number of therapists thus limiting the ability to generalize,<sup>11,13,14</sup> (iii) not using a chance-corrected measure of agreement,<sup>11</sup> (iv) not evaluating classification into all subsyndromes,<sup>11,12</sup> and (v) use of categories not part of the McKenzie system.<sup>11</sup>

The aim of the current study was to address this gap in the literature and to clearly establish the intertester reliability of the McKenzie system. We studied formally trained therapists classifying patients with cervical or lumbar pain into both syndromes and subsyndromes.

## METHODS

This study was approved by the University of Sydney Human Ethics Committee.

### Subjects

Patients attending private physical therapy clinics in Australia for treatment of cervical or back pain, with or without radiation into the limb, were invited to participate in the study.

### Examiners

All examiners (n = 14) were physical therapists who had completed the McKenzie credentialing examination. In addition, 7 had also completed the McKenzie diploma qualification. Each rater examined between 2 and 8 patients.

**Table 1.** Patient characteristics at baseline

Characteristics of subjects	
Number of subjects	50
Age (y)	48.6 (SD 12)
Pain intensity (VAS cm)	5.6 (SD 1.7)
Quebec disability score	47.3 (SD 19)
Female sex	48%
Past LBP	82%
Frequency of pain (% constant)	46%
Duration of symptoms	
Acute (=7 d)	20%
Subacute (>7 d to 7 wk)	46%
Chronic (>7 wk)	34%
Radiation into leg/arm	56%
Radiation below the knee/elbow	28%
Working normal duties	48%

VAS, visual analogue scale.

### Procedure

Treating therapists invited their patients to participate in the study. Patients received an explanation of what was required of them and signed an informed consent form. Before clinical assessment, information was collected from the patients regarding their sex, age, weight, height, location of symptoms, duration of symptoms, working status, previous history of LBP, pain intensity, pain frequency, and functional status (Table 1).

The patients underwent an assessment as described by McKenzie<sup>9,10</sup> with a standard McKenzie assessment form (cervical or lumbar) used to record the findings. Pairs of therapists simultaneously assessed the patient. One of the examiners was the treating therapist and directed the assessment, and the other was the researcher who observed the assessment and did not speak with the patient beyond the initial greetings.

At the completion of the assessment, the 2 therapists moved to separate rooms and independently recorded their classification of the patient on a form provided by the researcher (Appendix A). The forms were then placed inside an envelope together with a copy of the patient assessment form and sealed. The 2 sealed envelopes were then placed inside a larger envelope together with the patient demographic forms, which were posted to the investigator's office. The classifications were then coded to allow for data analysis (Appendix B).

Interrater reliability was estimated by calculating percent agreement and multirater  $\kappa$  (an unweighted form of  $\kappa$ ) using the MKAPPASC.SPS macro in SPSS 10.0 (SPSS, Chicago, Ill). Also calculated were 95% confidence intervals for  $\kappa$ .

## RESULTS

The ratings for classification into syndromes and subsyndromes are shown in Tables 2 and 3. The prevalence of the derangement syndrome was 88% for the first rater

**Table 2.** Distribution of syndrome judgments made by the 2 raters\*

Classification made by rater 1	Classifications made by rater 2				Total
	Derangement	Dysfunction	Postural	Other	
Derangement	42	2			44
Dysfunction					
Postural					
Other				6	6
Total	42	2		6	50

**Table 4.**  $\kappa$  Values and percent agreement for syndromes and subsyndromes

Pooled patients (lumbar/cervical)	$\kappa$	% Agreement
Syndromes	.84	96%
Subsyndromes	.87	90%
Lumbar patients		
Syndromes	1.0	100%
Subsyndromes	.89	92%
Cervical patients		
Syndromes	.63	92%
Subsyndromes	.84	88%

**Table 3.** Distribution of subsyndrome classifications made by raters\*

Classification made by rater 1	Subsyndrome	Classification made by rater 2																			Total
		Postural		Dysfunction					Derangement					Other							
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	
Postural	1																				
Dysfunction	2																				
	3																				
	4																				
	5																				
	6																				
	7																				
Derangement	8								5												5
	9																				
	10				1					1			17	1	1						21
	11												2								2
	12														13	1					14
	13														2						2
	14																				
	15																				
Other	16																				
	17																				
	18																	5			5
	19																				
	20																				
Total					1				1	5			17	3	14	3			5		50

\* Rater 1 classified a patient as subsyndrome 10 (line 10 on the vertical axis), whereas rater 2 classified the patient as subsyndrome (column 3 on the horizontal axis). Codes for the subsyndrome are in Appendix B.

and 84% for the second rater; dysfunction syndrome was 0% and 4%, respectively; postural syndrome was 0% for both raters; and “other” was 12% and 12%, respectively (Table 2). Within the classification of derangement, 86% of the patients were classified as subsyndromes D1, D3, and D5 (Table 3).

The percent agreement for assignment into syndromes for the total patient pool was 96%, with the point estimate and 95% confidence interval for  $\kappa$  being 0.84 (95% confidence interval [CI] 0.35-1.0). For subsyndromes, the percent agreement was 90% with  $\kappa = 0.87$  (95% CI 0.71-1.0). For the 25 lumbar patients, the percent agreement was 100% with  $\kappa = 1.0$  (95% CI 0.35-1.0) for

syndrome classification, and 92% agreement with  $\kappa = 0.89$  (95% CI 0.66-1.0) for subsyndromes. For the 25 cervical patients, the percent agreement was 92% with  $\kappa = 0.63$  (95% CI -0.11 to 1.0) for syndromes, and 88% agreement with  $\kappa = 0.84$  (95% CI 0.60-1.0) for subsyndromes (Table 4).

## DISCUSSION

The principal finding of this study is that trained McKenzie therapists are able to classify patients with spinal pain into the categories described by McKenzie with good

**Table 5.** Summary of previous studies of reliability of McKenzie classification lumbar spine

Reference	Study design	No. of therapists	Minimum training	No. of patients	Syndrome classification		Reliability of McKenzie subsyndrome classification	
					% Agreement	$\kappa$	% Agreement	$\kappa$
Kilby et al <sup>11</sup>	Simultaneous assessment	2	#Part B	41 Lumbar patients	No	No	No	No
Riddle and Rothstein <sup>12</sup>	Consecutive assessment	49	#16 to Part A	363 Lumbar patients	60%*	0.33*	39%	0.26#
Razmjou et al <sup>14</sup>	Simultaneous assessment	2	Credentialed	45 Lumbar patients	93%	0.70	91%*	0.88*
Kilpikoski et al <sup>13</sup>	Simultaneous—history, consecutive—objective	2	Credentialed	39 Lumbar patients	95%	0.60	74%	0.70
Clare et al 2002	Simultaneous assessment	14	Credentialed	25 Cervical and 25 lumbar patients	96%	0.84	90%	0.87

\* Reliability figures calculated by current authors from original data in manuscripts. Riddle and Rothstein's patients were classified into the following categories: postural, dysfunction, the 7 derangement subsyndromes, none. Razmjou et al data from manuscript were analyzed to determine reliability for subsyndromes.

**Table 6.** Prevalence of syndrome (%)

Reference	No. of patients	% Derangements	% Dysfunction	% Postural	% Other
Kilby et al <sup>11</sup>	41	42.7	22.0	2.4	32.9
Riddle and Rothstein <sup>12</sup>	363	52.9	34.7	9.6	2.8
Razmjou et al <sup>14</sup>	45	86.7	4.4	2.2	6.7
Kilpikoski et al <sup>13</sup>	39	90	2	Nil	8
Clare et al 2005 (lumbar and cervical patients)	50	86	2	Nil	12

reliability. The reliability measures for lumbar patients ( $\kappa = 0.89$ ) and cervical patients ( $\kappa = 0.84$ ) were both acceptably high for subsyndrome classification. Reliability for syndrome classification was less clear, with a  $\kappa$  value and percent agreement clearly acceptable for lumbar patients (1.0 and 100%, respectively), whereas for cervical patients, the percent agreement was acceptably high, but the  $\kappa$  value was not. We attribute this pattern of results to the behavior of  $\kappa$  when prevalence is low (the low base-rate problem), where only 1 or 2 disagreements between raters can markedly reduce the  $\kappa$  value. Although there is some ambiguity about the reliability of cervical syndrome classification (because of the conflicting  $\kappa$  and agreement scores), there is no ambiguity about the reliability of cervical subsyndrome classification.

To date, 4 studies have assessed the reliability of the McKenzie classification for patients with lumbar pain. The 2 first studies reported low reliability,<sup>11,12</sup> whereas the 2 more recent studies have reported acceptably high reliability.<sup>13,14</sup> However, direct comparison between studies is difficult because each of the 4 studies used different design and analysis strategies. Studies differed in terms of (i) the

number of raters, (ii) the number of patients, (iii) whether raters assessed the patient simultaneously or consecutively, (iv) the level of training of the raters, (v) the classification categories used, and (vi) the coefficient of agreement presented. These differences between the studies and the current study are summarized in Table 5.

The 2 most recent studies<sup>13,14</sup> provide the highest estimates of the reliability of classification into the syndromes and subsyndromes. The design and analysis of these 2 studies were similar to the present study; however, we believe that the use of a larger pool of raters and the inclusion of both cervical and back pain patients in our study allow greater generalizability of our results. Interestingly, the reliability observed here for the total patient pool (which included patients with symptoms of lumbar and cervical origin) was of a similar magnitude to that reported by Razmjou et al<sup>14</sup> for syndromes and by Kilpikoski et al<sup>13</sup> for subsyndromes, for patients with pain of lumbar origin.

The 2 studies that have reported unacceptably low reliability both recruited raters with minimal training in the McKenzie assessment method. In the Riddle and Roth-

stein<sup>12</sup> study, two thirds of the raters had not completed any formal McKenzie education, a feature which may explain the low reliability observed in that study. The raters in the Kilby et al<sup>11</sup> study had completed only the basic training (parts A and B) in the McKenzie education program; however, it is not clear whether the low reliability found here was the result of insufficient training or the use of a different classification system to that advocated by McKenzie. It would seem that, from a consideration of the studies conducted to date, raters need to have at least reached the level of credentialing to use the classification system with acceptable reliability. As yet, no study has directly addressed this issue, and it would be of interest to conduct further research to establish the minimal level of education required to use the system with good reliability.

This study is the first that has assessed the reliability of the McKenzie classification for cervical pain patients. Obtained point estimates for  $\kappa$  values for the patients with lumbar pain were higher than those for cervical pain: syndromes  $\kappa = 1.0$  versus 0.63; subsyndromes  $\kappa = 0.89$  versus 0.84; however, in each case, the 95% confidence intervals overlap, so the difference in reliability could have arisen by chance. To resolve this, a more precise estimate of reliability is required, and further study using a larger pool of raters is recommended.

The prevalence of the syndrome classifications found in our study is similar to that reported by both Razmjou et al<sup>14</sup> and Kilpikoski et al<sup>13</sup> (Table 6). The majority of the patients were classified as derangement in all 3 studies, a finding that is consistent with McKenzie<sup>9</sup> who states "most patients develop pain and seek assistance as the result of derangement." Within the derangement classification, 86% of the patients fell into the subsyndromes of D1, D3, and D5. McKenzie<sup>9</sup> also states that approximately 89% of derangements will fall into the subsyndromes of D1, D3, and D5. In the patient population of Razmjou et al,<sup>14</sup> 90% were classified in these subsyndromes. This is in contrast to the findings of Kilpikoski et al<sup>13</sup> who found that the majority (56%) of the patients were classified into derangement D4. Differences in the health care setting for the provision of the patients for the study may provide an explanation for their result.

One limitation of this study was that the 95% confidence intervals for  $\kappa$  were quite broad. A lack of precision is a feature of the traditional reliability study that uses a small number of raters (typically 2) assessing a larger number of patients. The problem with this design is that with only 2 raters, increasing the number of patients provides a very inefficient method of increasing power. A more efficient method is to expand the number of raters beyond 2; however, this may not be practical for clinical assessments. For example, the measure may be reactive (the potential for the attribute being measured to change with repeated measurement),<sup>15</sup> or the time involved in repeated measure-

ments may make participation unattractive to a patient. We are currently considering other designs to address these problems in clinical reliability studies.

## CONCLUSION

The McKenzie assessment, when performed by therapists with training in the McKenzie method, allows for reliable classification of patients with lumbar and cervical pain.

## REFERENCES

1. Battie MC, Cherkin DC, Dunn R, Ciol MA, Wheeler K. Managing lumbar pain: Attitudes and treatment preferences for physical therapists. *Phys Ther* 1994;74:219-96.
2. Foster NE, Thompson KA, Baxter GD, Allen JM. Management of non specific lumbar pain by physiotherapists in Britain and Ireland. *Spine* 1999;24:1332-42.
3. Rebbeck T. Position statement on the efficacy of physiotherapy interventions for the treatment of low back pain. Melbourne: Australian Physiotherapy Association; 2002.
4. Maher C, Latimer J, Refshauge K. Prescription of activity for low-back pain: What works? *Aust J Physiother* 1999;45:121-32.
5. Danish Institute for Health and Technology Assessment. Low back pain frequency, management and prevention from an HTA perspective. *Danish Health Technology Assessment* 1999;1:1-106.
6. Fritz JM, Delitto A, Erhard RE. Comparison of classification-based physical therapy with therapy based on clinical practice guidelines for patients with acute low back pain. A RCT. *Spine* 2003;28:1363-72.
7. Long A, Donelson R. Does it matter which exercise? A multi-centred RCT of low back pain subgroups. Proceedings of the McKenzie Institute 8th International Conference; 2003 Sep 12-14; Rome. September 12-14th, 2003. Waikanae (New Zealand): McKenzie Institute International; 2003.
8. McKenzie RA, May S. The lumbar spine: Mechanical diagnosis and therapy. 2nd ed. Waikanae (New Zealand): Spinal Publications; 2003.
9. McKenzie RA. The cervical and thoracic spine: Mechanical diagnosis and therapy. Waikanae (New Zealand): Spinal Publications; 1990.
10. McKenzie RA. The lumbar spine: Mechanical diagnosis and therapy. Waikanae (New Zealand): Spinal Publications; 1981.
11. Kilby J, Stigant M, Roberts A. The reliability of back pain assessment by physiotherapists, using a "McKenzie algorithm". *Physiotherapy* 1990;76:579-83.
12. Riddle D, Rothstein J. Intertester reliability of McKenzie's classifications of the syndrome types present in patients with lumbar pain. *Spine* 1993;18:1333-44.
13. Kilpikoski S, Airaksinen O, Kankaanpaa M, Leminen P, Videman T, Alen M. Interexaminer reliability of lumbar pain assessment using the McKenzie Method. *Spine* 2002;27:E207-14.
14. Razmjou H, Kramer J, Yamada R. Intertester reliability of the McKenzie evaluation in assessing patients with mechanical low-back pain. *J Orthop Sports Phys Ther* 2000;30:384-6.
15. Campbell D, Stanley J. Experimental and quasi-experimental designs for research. Chicago (Ill): Rand McNally & Company; 1963.

APPENDIX A. RELIABILITY EVALUATION OF MCKENZIE SPINAL PAIN CLASSIFICATION

---

Date: \_\_\_\_\_ Treating therapist: Yes/No Researcher: Yes/No

Lumbar spine: Yes/No Cervical spine: Yes/No

Classification: Please circle the classification and subgroup.

(1) Postural syndrome: Yes/No

(2) Dysfunction syndrome: Yes/No

If yes,

(a) Flexion  
 (b) Extension  
 (c) Lateral:  
     Lumbar spine Side glide  
     Cervical spine Rotation  
     Lateral flexion  
 (d) Adherent nerve root  
 (e) Multidirectional

(3) Derangement: Yes/No

If yes,

(a) 1  
 (b) 2  
 (c) 3  
 (d) 4  
 (e) 5  
 (f) 6  
 (g) 7  
 (h) Irreducible

(4) Other

Lumbar spine:      Entrapment                      Yes/No  
                           Spondylolisthesis                      Yes/No  
                           Canal stenosis                              Yes/No  
                           Nonmechanical                              Yes/No  
                           Inconclusive                              Yes/No

Cervical spine:    Entrapment                      Yes/No  
                           Cervical headache                      Yes/No  
                           Nonmechanical                              Yes/No  
                           Inconclusive                              Yes/No

---

APPENDIX B.

Coding of syndromes

- (1) Postural
- (2) Dysfunction
- (3) Derangement
- (4) Other

Coding of subsyndromes

- (1) Postural
- (2) Dysfunction—Flexion
- (3) Dysfunction—Extension
- (4) Dysfunction—Lateral (lumbar spine = lateral glide; cervical spine = lateral flexion).
- (5) Dysfunction—Rotation
- (6) Dysfunction—Adherent nerve root
- (7) Dysfunction—Multi directional
- (8) Derangement 1
- (9) Derangement 2
- (10) Derangement 3
- (11) Derangement 4
- (12) Derangement 5
- (13) Derangement 6
- (14) Derangement 7
- (15) Derangement—Irreducible
- (16) Other—Entrapment
- (17) Other—Nonmechanical
- (18) Other—Inconclusive
- (19) Other—Canal stenosis
- (20) Other—Lumbar spine = spondylolisthesis; cervical spine = headache