Manual Therapy 18 (2013) 395–402

Contents lists available at SciVerse ScienceDirect

Manual Therapy

journal homepage: www.elsevier.com/math

Original article

Within- and between-day reliability of spinal stiffness measurements obtained using a computer controlled mechanical indenter in individuals with and without low back pain

Arnold Y.L. Wong [a], Greg Kawchuk [a,*], Eric Parent [a,b], Narasimha Prasad [c]

[a] Department of Physical Therapy, University of Alberta, Edmonton, Alberta, Canada
[b] Glenrose Rehabilitation Hospital, Alberta Health Services, Edmonton, Alberta, Canada
[c] Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Alberta, Canada

ARTICLE INFO

Article history:
Received 11 September 2012
Received in revised form
23 January 2013
Accepted 4 February 2013

Keywords:
LBP
Mechanical indentation
Reproducibility
Spinal stiffness

ABSTRACT

Instrumented spinal stiffness measurements have shown high test–retest reliability. However, factors that may affect reliability have yet to be investigated. The objective of this study was to compare the: 1) within- and between-day reliability of a mechanical indentation device (MID) in measuring spinal stiffness, 2) measurement precision of averaging multiple measurements, and 3) reliability of stiffness measurements between individuals with and without low back pain (LBP).

The spinal stiffness of 26 volunteers with and without LBP was measured 3 times by MID in each of two visits 1–4 days apart. Two stiffness measures were calculated from the resulting force-displacement data: global stiffness and terminal stiffness. Intraclass correlation coefficients (ICCs) were used to estimate reliability. Measurement precision was measured by minimal detectable changes, bias and 95% limits of agreement.

Using the mean of three spinal stiffness measurements, the measurement precision was improved by 33.7% over a single measurement. Averaging three measurements, the within- and between-day reliability point estimates of both global and terminal stiffness were 0.99 and 0.98, respectively. The reliability estimates of spinal stiffness measurement using MID were not significantly altered by the participants' LBP status across all circumstances (95% confidence intervals overlapped).

With our experimental protocol, averaging three spinal stiffness measurements using MID produces reliable stiffness measurements regardless of individuals' LBP status.

1. Introduction

Assessing spinal stiffness is suggested to reflect mechanical properties (Maitland, Hengeveld, Banks, English, 2005), which are related to the underlying pathologies and symptoms (Latimer, Lee, Adams, Moran, 1996b; Kawchuk et al., 2001). Clinically, a clinician manually applies a posteroanterior force to the skin overlying the spinous processes of a prone patient and perceives the corresponding soft tissues response and spinal stiffness (Maitland et al., 2005). The perceived spinal stiffness together with the reproduction of symptoms during the manual spinal stiffness assessment (MSSA) are used to guide treatment decisions (Fritz et al., 2005) and to appraise treatment effectiveness (Tuttle, 2009).

Notwithstanding the popularity of MSSA among clinicians (Abbott et al., 2009), the reported reliability of MSSA is low (Seffinger et al., 2004). Specifically, the reported between-clinician reliability estimates ($ICC_{1,1}$) of MSSA range from 0.03 to 0.37 (Maher and Adams, 1994) and such estimates are not improved with clinical experience (Binkley et al., 1995). Although intra-rater reliability of MSSA is slightly higher, it is still less reliable than other physical examinations (Seffinger et al., 2004).

The low reliability of MSSA can be attributed to the limits of palpation sensitivity and subjective perception (Nicholson et al., 2003), which can be affected by the manual loading velocity and force (Wong, 2012). Research has also showed that humans are not sensitive to detecting changes in elastic and viscous stiffness (Nicholson et al., 1997, 2003). Given the viscoelastic nature of human spine, it is conceivable that clinicians perceive different stiffness values in the same spine or perceive no change in stiffness although a change in stiffness may have occurred.

* Corresponding author. Physical Therapy, University of Alberta, 3-48 Corbett Hall, Edmonton, Alberta, Canada T6G 2G4. Tel.: +1 780 492 6891; fax: +1 780 492 4429.
E-mail address: greg.kawchuk@ualberta.ca (G. Kawchuk).

1356-689X/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.math.2013.02.003

To improve the reliability of spinal stiffness assessments, researchers have developed spinal stiffness-testing devices (Table 1). Latimer et al. used their device to assess the within-day test–retest reliability of spinal stiffness measurement ($ICC_{2,1} = 0.96$) on low back pain (LBP) patients using two sets of 4 averaged indentation cycles (Latimer et al., 1996a). Edmondston et al. developed a stiffness device and reported high within-day reliability ($ICC_{2,1} = 0.98$) of spinal stiffness measurement using six separate single indentation cycles on asymptomatic participants (Edmondston et al., 1998). Recently, mechanically-assisted indentation devices with test–retest ICC of 0.79 (Owens, DeVocht, Wilder, Gudavalli, Meeker, 2007b) and 0.93 (Stanton and Kawchuk, 2009) have been developed to accommodate the need of portability for clinical use. Although existing devices demonstrate high reliability, factors that may affect the reliability of a given device (e.g. same- and between-day comparisons, repeating spinal stiffness measurements, and participants' LBP status) have yet to be investigated.

The objectives of this study were to: 1) determine the test–retest reliability of spinal stiffness measurement using a mechanical indentation device (MID) in within- and between-day comparisons, 2) compare the improvements in measurement precision based on using a single indentation cycle, or an average of two or three indentation cycles, and 3) compare the test–retest reliability of MID spinal stiffness measurements in participants with and without LBP in an exploratory analysis. We hypothesized that MID measurements were reliable ($ICC > 0.90$) for quantifying spinal stiffness in all testing situations regardless of the individuals' LBP status. We also hypothesized that measurement precision would be improved by increasing the number of averaged measurements.

2. Methods

2.1. Sample size calculation

Although little research has estimated the within- and between-day reliability of instrumented spinal stiffness measurements (intraclass correlation coefficients (ICCs)) in individuals with LBP, the reported between-day reliability of spinal stiffness measurement in individuals with LBP was lower than that of the asymptomatic counterparts (Table 1). As such, we hypothesized the between-day ICC value for spinal stiffness measurements in LBP and asymptomatic participants to be 0.85 and 0.95, respectively. Using the formula derived from Shoukri et al. (2004) and Rosner (2000) to detect this difference as statistically significant (power = 0.70, alpha = 0.05) the calculated sample size for each participant group was 13.

2.2. Participants

Thirty volunteers aged 18–60 years, with and without nonspecific LBP were recruited from the University of Alberta and surrounding area. Nonspecific LBP was defined as pain between the lowest costal margin and above the gluteal sulcus, with or without leg pain. From the LBP symptoms in the first visit, participants were classified into symptomatic (–LBP) and asymptomatic (–LBP). Participants were excluded if they were pregnant, could not lay prone for 20 min, had a diagnosis of scoliosis, congenital spinal disorders, prior lumbar surgery, spondylolisthesis, a history of severe trauma, or medical 'red flag' conditions such as cancer, spinal infection, fracture, or systemic disease. This study was approved by the University of Alberta Health Research Ethics Board.

2.3. Examiner

A physiotherapist with 7-year clinical experience collected spinal stiffness measurements. The examiner had used a mechanically-assisted indentation device for spinal stiffness measurement in previous research (Hu, Wong, Kawchuk, 2009).

2.4. Mechanical indentation device

Spinal stiffness was measured by using the MID as described in previous research (Fig. 1) (Fritz et al., 2011). The MID consists of a motorized indentation probe supported by an external frame. An electromechanical stepping motor (Dual Motion Motor, Waterbury, USA) is used to extend the linear probe. The resulting indentation force is measured by a compressive-tension load cell transducer (Entran, Fairfield, USA) attached in-series with the probe. The displacement of the indenter is measured by a rotary encoder (Dual Motion Motor, Waterbury, USA). The indentation velocity (2.5 mm/s) is controlled by customized LabVIEW 8.6 software (National Instruments, Austin, USA) while signals from the load cell and rotary encoder are collected at a rate of 200 Hz. To ensure the safety of the participants, the indentation process could be terminated by the examiner or the participant using hardware panic switches.

2.5. Study procedures

Participants were assessed in two visits 1–4 days apart. Both visits were scheduled at the same time of day. After giving written consent, participants completed self-reported measures, including a demographic and medical history information sheet, an 11-point numeric pain rating (Farrar et al., 2003), a body pain diagram (Werneke et al., 1999), and a modified Oswestry low back pain disability questionnaire (mODI) (Fritz and Irrgang, 2001). The examiner then performed a standardized physical examination to confirm the eligibility of participants. Standardized instructions including precautions were given to the eligible participants.

Spinal stiffness assessments were conducted with the participant prone. In the first visit, the examiner manually identified the participant's L3 spinous process and marked it with ink to guide the indenter placement in both visits. The spinal level was verified by ultrasound whenever there was uncertainty. The MID was then positioned over the identified L3 spinous process and the participant was instructed to relax the back muscles and minimize body movement. Participants were instructed to hold their breath at the end of normal exhalation for approximately 10 s throughout the indentation (Shirley, Hodges, Eriksson, Gandevia, 2003). All participants underwent an indentation familiarization procedure to precondition their soft tissues and minimize their anxiety. Specifically, the indentation involved advancement of the indenter from a preload of 5 N to a final load that stayed on the spinous process for 1 s before load removal. The final load was gradually increased by a magnitude of 10 N from 20 N to 60 N. Following the familiarization procedure, four experimental indentations were performed with the target load of 60 N. A maximum indentation force of 60 N was chosen because it lies within the recommended target loads of spinal stiffness testing (45 – 135 N) (Kumar and Stoll, 2011). Further, research showed that 20% of patients with LBP could not tolerate 80 N indentation force (Owens et al., 2007b) while 60 N indentation load was well tolerated among LBP participants without any reported side effects (Fritz et al., 2011). The first trial of the experimental indentations was discarded given its differential stiffness behavior from the subsequent indentations (Latimer et al., 1996c). The stiffness values of the remaining three indentations in each visit were used for data analysis. A 2-min rest was given between indentations. Upon completion of the data collection,

Table 1
The reliability and accuracy of measurements derived from human participants using indentation devices previously described in the literature.

	Device	Level/participants	Reliability type	Averaged measurements	Reliability (CI)	Accuracy (SD)	Stiffness coefficient	Relative MDC95[a]
Mechanical device								
Lee and Svensson (1990)	Spinal physiotherapy simulator	L3/11 −LBP	Between-day	5	$ICC_{2,1} = 0.88$	Not reported	Not reported	Not applicable
Lee and Evans (1992)	Spinal mobilizer	L3 to L5/10 −LBP	Between-day	1	ICC = 0.95 to 0.99	Maximum error range ± 0.8 mm	Not applicable because it measured relative displacement between spinal levels	
Latimer et al. (1996a)	Portable stiffness device	L2 to L5/22 +LBP	Within-day	4	$ICC_{2,1} = 0.96$ (0.91–0.98)	90% CI = ± 1.8N/mm	15.33N/mm	3.96%
Edmondston et al. (1998)	Spinal posteroanterior mobilizer	L5/8 −LBP	Within-day	1	$ICC_{2,1} = 0.979$	SEM = 0.515N/mm	10N/mm	14.28%
Shirley et al. (2002)	Portable stiffness device	L4/18 −LBP	Within-day	4	$ICC_{2,1} = 0.96$ (0.90–0.99)	SEM = 0.57N/mm SEM = 1.20N/mm	13.16N/mm	12.01%
			Between-day	4	$ICC_{2,1} = 0.85$ (0.64–0.94) and $ICC_{2,1} = 0.83$ (0.53–0.93)		13.18N/mm	25.24%
Mechanically assisted								
Owens et al. (2007b)	Posterior-to-anterior spinal stiffness device	L1 to L5/36 +LBP	Combined within- and between-day	4 (90% data) 2 to 3 (8% data) 1 (2% data)	$ICC_{3,1} = 0.79$ (0.739–0.832)	SEM = 1.62N/mm	11.2N/mm	40.01%
Tuttle et al. (2008)	Passive movement assessment device	Cervical segments/ 10 −LBP	Within-day	1	CMD = 0.99	Not reported	Not applicable	Not applicable
			Between-day	1	CMD = 0.97			
			Inter-rater	1	CMD = 0.96			
Stanton and Kawchuk (2009)	Assisted indentation device	L4/23 −LBP	Within-day	1	$ICC_{3,1} = 0.91$ for GS $ICC_{3,1} = 0.93$ for MMS	Inter-trial inconsistency: 6.23% (4.52%) for the GS and 7.71% (5.33%) for MMS	GS: 6.23N/mm MMS: 7.71N/mm	Not applicable

CI, confidence interval; CMD, adjusted coefficient of multiple determination; CS, cervical spine; GS, global stiffness; ICC, intra-class correlation coefficient; MDC95, minimal detectable change at 95% confidence interval; MMS, mean maximal stiffness; SEM, standard error of measurement; −LBP, asymptomatic participants; +LBP, participants with low back pain.

[a] Relative MDC95 is a ratio between MDC95 of spinal stiffness in a given reliability study and the corresponding average spinal stiffness in that study. Since previous studies did not report the MDC95 of spinal stiffness measurements, the relative MDC95 of various studies in this table were calculated from their reported SEMs, ICCs and mean stiffness values.

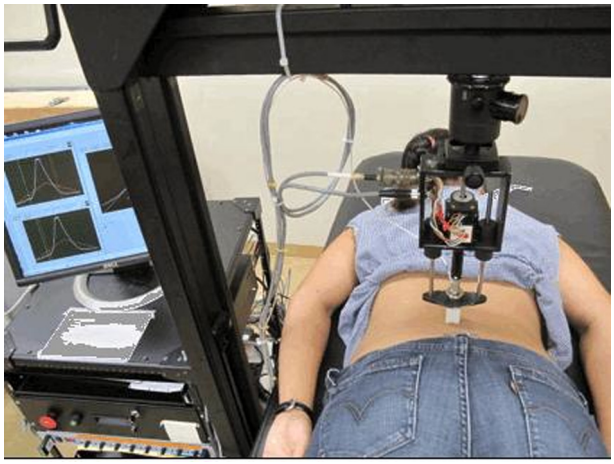

Fig. 1. Spinal stiffness testing by the mechanical indentation device.

participants rated their indentation-related pain intensity. Participants were advised to remain active between visits. The same procedures were repeated in the second visit.

2.6. Data analysis of spinal stiffness

The raw force-displacement data of indentations of participants was saved in an Excel file and identified by a random number to prevent bias during stiffness calculation. Two types of spinal stiffness values were calculated from the force–displacement data of each indentation: global stiffness and terminal stiffness. Global stiffness was calculated from the slope of force–displacement curve between 5 N and 60 N, representing the stiffness of underlying tissues throughout the indentation (Fritz et al., 2011). Terminal stiffness was a ratio between the maximal applied force (60 N) to the maximal resultant displacement, representing stiffness at the end of indentation (Fig. 2) (Fritz et al., 2011).

2.7. Statistical analyses

Statistical Package for the Social Sciences version 17.0 software (SPSS Inc., Chicago, USA) was used for all analyses.

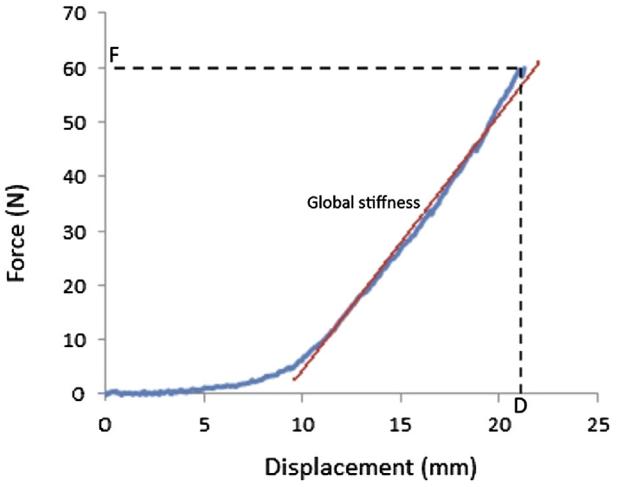

Fig. 2. Calculating global and terminal stiffness from a typical force–displacement curve.*D, terminal displacement; F, terminal load; *Global stiffness = slope of the linear regression line between 5 N and 60 N; Terminal stiffness = F/D.

The within- and between-day test–retest reliability estimates of global and terminal stiffness of the participants were assessed using ICCs. Specifically, ICCs (model 3,k) were calculated for global and terminal stiffness measurements where k indicated 1, 2 and 3 indentations. $ICC_{3,k}$ was chosen because one examiner was involved in this study. An ICC value 0.90 is the acceptable reliability for clinical use or individual comparisons (Nunnally, 1978). Standard errors of the measurement (SEM = pooled standard deviation $\times \sqrt{(1 - ICC)}$), which estimate the "true" variation of all possible measurement errors, were calculated for all testing conditions to quantify the measurement precision (Portney and Watkins, 2008). Minimal detectable changes ($MDC95 = 1.96 \times \sqrt{2} \times SEM$) were also calculated to indicate the minimal change in spinal stiffness that one should note before being 95% confident that the observed differences exceed measurement error (Roebroeck et al., 1993). The level of significance was 0.05 for all statistical tests.

The within-day variability of spinal stiffness from replicated measurements was assessed by repeatability coefficients (Bland and Altman, 1999). The repeatability coefficient of spinal stiffness in a given session was calculated using the formula: $1.96 \times \sqrt{2} \times$ the standard deviation of within-participant spinal stiffness in that session. The resulting calculation represents the expected difference between 2 within-session measurements in 95% of the participants. The repeatability coefficients of the two visits were compared.

The between-day agreement in spinal stiffness (global and terminal stiffness) was quantified by 95% Limits of agreement (LOA), bias and 95% CI for the bias (Bland and Altman, 1999). Bias is equal to the mean intra-individual differences in spinal stiffness between the two visits. The 95% CI for the bias was calculated from the standard error of the mean difference in spinal stiffness between the two visits. If the 95% CI for the bias contains zero, it implies no significant difference in between-day spinal stiffness.

The effect of averaging multiple indentation cycles on measurement precision improvement was analyzed by comparing the SEM reductions resulting from using a single indentation cycle, or an average of the first two or three indentation cycles.

To ensure that the test–retest reliability of spinal stiffness measurements was estimated from participants with stable physical conditions (Portney and Watkins, 2008), participant with a between-session pain or LBP-related disability exceeding the clinically important difference thresholds (NRS > 2.4/10 ((Maughan and Lewis, 2010) or mODI scores >12% (Fritz and Irrgang, 2001)) would be excluded from the data analysis. The systematic effect resulting from repeated spinal stiffness measurements was analyzed using the repeated measures analysis of variance (ANOVA).

3. Results

Thirty participants were recruited. Three participants had incomplete stiffness data due to technical errors. One participant was excluded because of indentation-related discomfort. No participant was excluded due to a between-session change of NRS > 2.4/10 or mODI >12%. This resulted in 14 asymptomatic and 12 symptomatic participants (Table 2).

3.1. Effect of averaging multiple measures on improvement in reliability estimates

While the primary objective of this study was to compare the within- and between-day reliability estimates of spinal stiffness measurements using MID, we reported the effect of averaging multiple indentation cycles on measurement precision (SEMs reduction) to guide the choice of the reliability estimates reporting in this paper. With reference to the SEM of a single indentation

Table 2
Description of participants.

Characteristic[a]	All participants ($n = 26$)	Asymptomatic participants ($n = 14$)	Low back pain participants ($n = 12$)
Age mean (year)	28.7 (9.8)	25.9 (5.4)	31.9 (12.8)
Sex (percent male)	30.8%	35.7%	28.6%
Body mass index (kg/m^2)	22.6 (3.6)	22.8 (3.1)	22.3 (4.3)
Modified Oswestry disability index (/100)	5.1 (8.3)	0.0 (0.0)	11.5 (9.3)
Numeric pain rating scale (/10)	1.2 (1.9)	0.1 (0.3)	2.6 (2.0)
Prior history of low back pain[b]	11 out of 26	1 out of 14	10 out of 12

[a] Values are reported as mean (SD) unless indicated.
[b] Prior history of low back pain is defined as history of low back pain that might or might not need to see physicians.

cycle, averaging of two repeated indentation measurements reduced the SEM by a mean of 15.2% over all measurement conditions, while averaging three indentation measurements reduced the SEM by a mean of 33.7% (Table 3). Since the 18.5% additional mean precision improvement using three indentation cycles was relatively large, reliability results calculated from the average of three indentation cycles are reported in this paper.

3.2. Within- and between-day reliability

The within- and between-day reliability point estimates ($ICC_{3,3}$) of both global and terminal stiffness were 0.99 and 0.98, respectively (Table 4). The lower bounds of the 95% CIs were equal or greater than 0.90 for all reliability estimates.

Repeated measures ANOVA results showed no significant difference in within-participant spinal stiffness measurements in within- and between-session comparisons. The corresponding repeatability coefficients of global and terminal stiffness were similar in the first and second visit (Table 5). The bias estimates of the between-session spinal stiffness measurements were not significantly different from 0 (Table 5). The LOAs, bias and the corresponding 95% CIs are shown in Table 5.

3.3. Reliability estimates in −LBP and +LBP

The exploratory analysis revealed no significant difference in the reliability estimates between +LBP and −LBP groups across all testing conditions (95% CIs overlapped). In both groups, the within- and between-day point estimates ($ICC_{3,3}$) of spinal stiffness (global and terminal stiffness) were 0.99 and 0.98, respectively (Table 4).

4. Discussion

This is the first study to compare the reliability of instrumented spinal stiffness measurements: 1) in within- and between-day conditions, 2) using averages of different multiples of indentation measurements, and 3) from individuals with and without LBP. Our results demonstrated excellent within- and between-day reliability estimates (Roebroeck et al., 1993). Averaging three indentation measurements obtained the highest measurement precision of spinal stiffness. −LBP and +LBP participants showed no significant difference in the reliability estimates of spinal stiffness measurements across all circumstances.

4.1. Within- and between-day reliability

Our within-day reliability estimates of spinal stiffness measurements were concordant with those of previous devices (Table 1). The reported within-day ICCs of spinal stiffness measurements using portable stiffness device (Latimer et al., 1996a) and spinal posteroanterior mobilizer (Edmondston et al., 1998) were 0.96 and 0.98, respectively.

On the contrary, our within-day reliability was higher than those of mechanically-assisted indentation devices (Table 1). The reported test–retest reliability point estimate ($ICC_{3,1}$) of Owens et al.'s manually assisted indentation device was 0.79 (Owens et al., 2007b) and that of the assisted indentation device ranged from 0.91 to 0.93 (Stanton and Kawchuk, 2009). The higher reliability of measurements using our MID may be attributed to not only the averaging of multiple measurements but also the more precise control over loading velocity and force as compared with mechanically-assisted indentation devices.

The repeatability coefficients of both global and terminal stiffness in the two visits were comparable, implying similarity in the within-session variability of repeated spinal stiffness measurements. Additionally, the repeatability coefficients of global and terminal stiffness were comparable to the corresponding between-day 95% LOAs. These results imply that the lack of agreement between visits were due to the lack of repeatability between measurements rather than any systematic effects between visits (Bland and Altman, 1999).

Regarding the between-day reliability point estimates, our findings were higher than those reported in previous research

Table 3
Changes in standard error of measurement (SEM) using the average of 2 or 3 measurements with reference to a single measurement.[a]

	Global stiffness, GS (N/mm) ($n = 26$)					Terminal stiffness, TS (N/mm) ($n = 26$)				
	Single measure	Mean of 2 measures (% decrease from 1 measure)		Mean of 3 measures (% decrease from 1 measure)		Single measure	Mean of 2 measures (% decrease from 1 measure)		Mean of 3 measures (% decrease from 1 measure)	
Within-day SEM										
Day 1	0.2	0.2	(23.7%)	0.1	(42.4%)	0.2	0.1	(12.7%)	0.1	(41.8%)
Day 2	0.2	0.1	(25.6%)	0.1	(42.2%)	0.1	0.1	(17.7%)	0.1	(42.5%)
Average % decrease			**24.7**		**42.3**			**15.2**		**42.2**
Between-day SEM										
Day 1 vs. Day 2	0.2	0.2	(5.7%)	0.2	(28.8%)	0.2	0.2	(14.9%)	0.2	(21.2%)
Overall avg % decrease			**15.2**		**35.6**			**15.1**		**31.7**

Avg, Average.
[a] The Overall mean decrease in SEM resulting from using the average of two measures: [(15.2 + 15.1)/2]% = 15.2%. The mean decrease in SEM resulting from using the average of three measures: [(35.6.0 + 35.6.)/2]% = 33.7%. The difference in mean improvement in precision by using three measures rather than two measures = (33.7 − 15.2)% = 18.5.

Table 4
Within- and between-day test–retest reliability of global stiffness and terminal stiffness measurements.

	All participants ($n = 26$)			Asymptomatic participants ($n = 14$)			Participants with LBP ($n = 12$)		
	Mean (SD)	$ICC_{3,3}$ (95% CI)	MDC95	Mean (SD)	$ICC_{3,3}$ (95% CI)	MDC95	Mean (SD)	$ICC_{3,3}$ (95% CI)	MDC95
Within-day comparisons									
Global stiffness (N/mm)									
Day 1	6.0 (1.3)	0.99 (0.98–1.00)	0.3	6.0 (1.3)	0.99 (0.99–1.00)	0.2	6.0 (1.4)	0.99 (0.96–1.00)	0.4
Day 2	6.0 (1.3)	0.99 (0.99–1.00)	0.3	5.9 (1.4)	0.99 (0.99–0.99)	0.3	6.0 (1.4)	0.99 (0.98–1.00)	0.3
Terminal stiffness (N/mm)									
Day 1	5.6 (1.1)	0.99 (0.98–1.00)	0.3	5.6 (1.2)	0.99 (0.99–1.00)	0.2	5.6 (1.1)	0.99 (0.97–1.00)	0.3
Day 2	5.7 (1.2)	0.99 (0.99–1.00)	0.2	5.6 (1.2)	0.99 (0.99–1.00)	0.2	5.8 (1.2)	0.99 (0.99–1.00)	0.2
Between-day comparisons									
Global stiffness (N/mm)									
Day 1 vs Day 2	6.0 (1.3)	0.98 (0.95–0.99)	0.5	6.0 (1.3)	0.98 (0.93–0.99)	0.5	6.0 (1.4)	0.98 (0.94–0.99)	0.5
Terminal stiffness (N/mm)									
Day 1 vs Day 2	5.6 (1.2)	0.98 (0.96–0.99)	0.5	5.6 (1.2)	0.98 (0.95–1.00)	0.4	5.7 (1.3)	0.98 (0.90–0.99)	0.5

ICC, intra-class correlation coefficient; LBP, low back pain; MDC95, minimal detectable change at the 95% confidence interval; SD, standard deviation; vs, versus.

(Table 1). While the lower ICC of Owen et al.'s device might be attributed to potential inconsistency in manual loading control (Owens et al., 2007b), the lower reliability of the spinal physiotherapy simulator might be due to using data from a single measurement (Lee and Svensson, 1990). Despite these possibilities, the absence of significant bias in our study corroborated our good agreement in between-session spinal stiffness (global and terminal stiffness) (Table 5). Although speculative, the improved between-day reliability of our MID may be ascribed to specific error minimization strategies (such as provision of standard instructions to the participants and the marking of L3 level for guiding indenter relocation) that should be emphasized in future studies.

In clinical and research situations, the within- and between-day reliability of instrumented spinal stiffness measurements is equally important because an individual's spinal stiffness is usually monitored over time. However, to our knowledge, only two studies reported both the within- and between-day reliability of stiffness-testing devices. Tuttle et al. used a passive movement assessment device to measure cervical spinal stiffness and reported the mean adjusted coefficient of multiple determination (CMD) in within- and between-day comparisons (Table 1) (Tuttle et al., 2008). Likewise, Shirley et al. reported the within- and between-day reliability of the portable stiffness device in measuring spinal stiffness as 0.96, and ranging from 0.83 to 0.85, respectively (Table 1) (Shirley et al., 2002). Although it is inappropriate to compare our ICCs with the CMDs of the former study, the latter study and ours showed no significant difference between the within- and between-day reliability estimates (95% CIs overlapped). Interestingly, our between-day 95% CIs of ICC were narrower than those reported by Shirley et al. (Table 1) (Shirley et al., 2002). Our improvement in precision of CIs might be attributed to our error minimization strategies that were not emphasized in Shirley et al.'s study (Shirley et al., 2002) and the use of heterogeneous participants that increased the values of ICCs (Portney and Watkins, 2008).

4.2. Measurement precision improvement by multiple indentation cycles

Depending on experimental protocols, previous investigators have used the results from a single cycle or an average of 2–5 spinal indentation cycles to calculate the stiffness of a given spinal level (Table 1). Theoretically, averaging a higher number of stiffness values improves measurement precision, however, the spinal stiffness values between successive measurements are affected by both the measurement error and the potential of incomplete viscoelastic recovery following spinal indentation. To minimize the risk of unrecovered viscoelastic change, our participants were given 2-min rest between indentations (Stanton and Kawchuk, 2009). Our results substantiated using an average of three measurements to improve the measurement precision. Although averaging additional measurements may further reduce the SEM, such extra reduction will gradually diminish (Koppenhaver, Parent, Teyhen, Hebert, Fritz, 2009). Since using a mean of three measurements required relatively short data collection duration and yielded excellent reliability, three indentation cycles seemed appropriate for our testing protocol.

Additionally, our methodology yielded better measurement precision than most previous research. Measurement precision of a given stiffness-testing device can be assessed by MDC95. However, it is inappropriate to compare absolute MDC95s among studies because patients' characteristics are different between studies. This problem can be resolved by comparing relative MDC95, which is a ratio between MDC95 and the respective mean spinal stiffness in a study. A smaller relative MDC95 indicates higher measurement precision (Teyhen, 2011). As no reliability studies have calculated

Table 5
Between-day 95% limits of agreement, bias, 95% confidence interval of bias for comparing day1 and day 2 spinal stiffness measurements, and repeatability coefficients of global stiffness and terminal stiffness at day 1 and day 2.

	Mean (SD)	Bias (day 1–day 2)	95% CI for the bias	95% LOA	Repeatability coefficient at day 1	Repeatability coefficient at day 2
Global stiffness (N/mm)	5.99 (0.28)	0.01	−0.10 to 0.12	−0.55 to 0.57	0.53	0.50
Terminal stiffness (N/mm)	5.65 (0.23)	−0.03	−0.12 to 0.06	−0.49 to 0.43	0.46	0.41

SD, standard deviation; CI, confidence interval; LOA, limits of agreement.

MDC95, we calculated the relative MDC95 of the available studies using their reported SEMs, standard deviations, ICCs and mean stiffness values (Table 1). While our within-day relative MDC95s represented 3.51–5.36% of the corresponding mean L3 spinal stiffness, those in other studies represented 3.96–14.28% of their respective spinal stiffness (Latimer et al., 1996a; Edmondston et al., 1998; Shirley et al., 2002). Similarly, our between-day relative MDC95s represented 8.33–8.93% of mean L3 spinal stiffness, whereas those of previous studies represented 25.24–40.01% of their corresponding mean spinal stiffness (Shirley et al., 2002; Owens et al., 2007b). The large relative MDC95 of previous devices might explain the difficulty in finding a relation between LBP symptoms and spinal stiffness, or detecting symptom-related spinal stiffness change in prior research (Owens, DeVocht, Gudavalli, Wilder, Meeker, 2007a).

4.3. Asymptomatic and symptomatic participants

Our results not only revealed excellent within- and between-day reliability of spinal stiffness measurements in the two groups but also found no significant difference in the estimates between these groups across all conditions. Since LBP may induce reflexive paraspinal muscle responses during indentation (Johansson and Sojka, 1991), it may increase the variability of spinal stiffness measurements in +LBP. However our findings showed reliable spinal stiffness measurements regardless of the participant's LBP status. The high reliability of stiffness measurement in +LBP might be partly due to our target indentation load (60 N), which likely reduced pain-related responses. The high reliability also implied that this indentation protocol was suitable for testing individuals with and without LBP. Taken together, the consistent within- and between-day spinal stiffness values in +LBP suggests that the indentation process itself is unlikely to generate any therapeutic effect comparable to that of manual therapy.

Although this study was not aimed to compare spinal stiffness values between +LBP and −LBP, it was noted that the mean spinal stiffness values of these groups were not significantly different from one another given their 95% CIs overlapped (Table 4). This finding was not unexpected because many factors (such as age, gender and body mass index) might affect the absolute spinal stiffness values (Wong, 2012). As such, direct comparison of absolute mean spinal stiffness values between the two groups may be misleading.

4.4. Limitations

Since our examiner operated a stiffness-testing device in previous research, our findings cannot be generalized to inexperienced examiners. In addition, these results may not be generalizable to manual assessment of stiffness or to mechanical assessments of spinal stiffness if the applied load is not equal to 60 N. Further, given the resource constraints, the sample size was estimated based on the statistical power of 0.7 instead of the conventional power of 0.8. Although the insignificant difference in the reliability of spinal stiffness measurements between −LBP and +LBP should be interpreted with caution, both groups demonstrated excellent reliability estimates (ICC values between 0.98 and 0.99) with narrow 95% CIs in all testing conditions. Therefore, our findings were unlikely to arise from false negatives. Additionally, our results cannot be generalized to inter-rater reliability of spinal stiffness measurements. However, the variability of inter-rater measurement would most likely be low given the standardized stiffness measurement procedure. Finally, since this study recruited a convenience sample of volunteers with and without LBP, our results may not be generalizable to patients with LBP seeking care from health professionals.

5. Conclusion

Using the average of three spinal stiffness tests, our MID measurements demonstrated excellent within- and between-day reliability in measuring human spinal stiffness regardless of their LBP status. The precision of spinal stiffness measurements was increased by averaging three measurements in the current protocol. The smaller relative MDC95 for our spinal stiffness measurements suggests that our procedures may detect stiffness changes in clinical settings.

Acknowledgments

Mr. Wong is supported by the Alberta Innovates-Health Solutions Graduate Studentship. Dr. Kawchuk is supported by the Canadian Research Chair Program.

References

Abbott JH, Flynn TW, Fritz JM, Hing WA, Reid D, Whitman JM. Manual physical assessment of spinal segmental motion: intent and validity. Manual Therapy 2009;14(1):36–44.
Binkley J, Stratford PW, Gill C. Interrater reliability of lumbar accessory motion mobility testing. Physical Therapy 1995;75(9):786–92.
Bland JM, Altman DG. Measuring agreement in method comparison studies. *Statistical* Methods in Medical Research 1999;8(2):135–60.
Edmondston SJ, Allison GT, Gregg CD, Purden SM, Svansson GR, Watson AE. Effect of position on the posteroanterior stiffness of the lumbar spine. Manual Therapy 1998;3(1):21–6.
Farrar JT, Berlin JA, Strom BL. Clinically important changes in acute pain outcome measures: a validation study. Journal of Pain and Symptom Management 2003; 25(5):406–11.
Fritz JM, Irrgang JJ. A comparison of a modified Oswestry low back pain disability questionnaire and the Quebec back pain disability scale. Physical Therapy 2001; 81(2):776–88.
Fritz JM, Whitman JM, Childs JD. Lumbar spine segmental mobility assessment: an examination of validity for determining intervention strategies in patients with low back pain. Archives of Physical Medicine and Rehabilitation 2005;86(9): 1745–52.
Fritz JM, Koppenhaver SL, Kawchuk GN, Teyhen DS, Hebert JJ, Childs JD. Preliminary investigation of the mechanisms underlying the effects of manipulation: exploration of a multi-variate model including spinal stiffness, multifidus recruitment, and clinical findings. Spine 2011;36(21):1772–81.
Hu Y, Wong YL, Lu WW, Kawchuk GN. Creation of an asymmetrical gradient of back muscle activity and spinal stiffness during asymmetrical hip extension. Clinical Biomech*anics* 2009 2009;24(10):799–806.
Johansson H, Sojka P. Pathophysiological mechanisms involved in genesis and spread of muscular tension in occupational muscle pain and in chronic musculoskeletal pain syndromes: a hypothesis. Medical Hypotheses 1991; 35(3):196–203.
Kawchuk GN, Kaigle AM, Holm SH, Fauvel OR, Ekstrom L, Hansson T. The diagnostic performance of vertebral displacement measurements derived from ultrasonic indentation in an in vivo model of degenerative disc disease. Spine 2001; 26(12):1348–55.
Koppenhaver SL, Parent EC, Teyhen DS, Hebert JJ, Fritz JM. The effect of averaging multiple trials on measurement error during ultrasound imaging of transversus abdominis and lumbar multifidus muscles in individuals with low back pain. Journal of Orthopaedic & Sports Physical Therapy 2009;39(8):604–11.
Kumar S, Stoll S. Device, protocol and measurement of regional spinal stiffness. Journal of Electromyography and Kinesiology 2011;21(3):458–65.
Latimer J, Goodsell MM, Lee M, Maher CG, Wilkinson BN, Moran CC. Evaluation of a new device for measuring responses to posteroanterior forces in a patient population, part 1: reliability testing. Physical Therapy 1996a;79(2):158–65.
Latimer J, Lee M, Adams R, Moran CM. An investigation of the relationship between low back pain and lumbar posteroanterior stiffness. Journal of Manipulative and Physiological Therapeutics 1996b;19(9):587–91.
Latimer J, Lee M, Goodsell M, Maher C, Wilkinson B, Adams Ra. Instrumented measurement of spinal stiffness. Manual Therapy 1996c;1(4):204–9.
Lee M, Svensson NL. Measurement of stiffness during simulated spinal physiotherapy. Clinical Physics and Physiological Measurement 1990;11(3):201–7.
Lee R, Evans J. Load-displacement-time characteristics of the spine under posteroanterior mobilisation. Australian Journal of Physiotherapy 1992;38(2):115–23.
Maher C, Adams R. Reliability of pain and stiffness assessments in clinical manual lumbar spine examination. Physical Therapy 1994;74(9):801–9.

Maitland G, Hengeveld E, Banks K, English K. Maitland's vertebral manipulation. 7th ed. Oxford: Butterworth-Heinemann; 2005.
Maughan EF, Lewis JS. Outcome measures in chronic low back pain. European Spine Journal 2010;19(9):1484–94.
Nicholson L, Adams R, Maher C. Reliability of a discrimination measure for judgements of non-biological stiffness. Manual Therapy 1997;2(3):150–6.
Nicholson LL, Adams RD, Maher CG. Manual discrimination capability when only viscosity is varied in viscoelastic stiffness stimuli. Journal of Manipulative and Physiological Therapeutics 2003;26(6):365–73.
Nunnally JC. Psychometric theory. New York: McGraw-Hill; 1978.
Owens EF, DeVocht JW, Gudavalli MR, Wilder DG, Meeker WC. Comparison of posteroanterior spinal stiffness measures to clinical and demographic findings at baseline in patients enrolled in a clinical study of spinal manipulation for low back pain. Journal of Manipulative and Physiological Therapeutics 2007a;30(7):493–500.
Owens EF, DeVocht JW, Wilder DG, Gudavalli MR, Meeker WC. The reliability of a posterior-to-anterior spinal stiffness measuring system in a population of patients with low back pain. Journal of Manipulative and Physiological Therapeutics 2007b;30(2):116–23.
Portney L, Watkins M. Foundations of clinical research: applications to practice. 3rd ed. New Jersey: Prentice Hall; 2008.
Roebroeck ME, Harlaar J, Lankhorst GJ. The application of generalizability theory to reliability assessment: an illustration using isometric force measurements. Physical Therapy 1993;73(6):386–95.
Rosner B. Fundamentals of biostatistics. 5th ed. California: Duxbury; 2000.
Seffinger MA, Najm WI, Mishra SI, Adams A, Dickerson VM, Murphy LS, et al. Reliability of spinal palpation for diagnosis of back and neck pain: a systematic review of the literature. Spine 2004;29(19):E413–25.
Shoukri M, Asyali M, Donner A. Sample size requirements for the design of reliability study: review and new results. Statistical Methods in Medical Research 2004;13(4):251–71.
Shirley D, Ellis E, Lee M. The response of posteroanterior lumbar stiffness to repeated loading. Manual Therapy 2002;7(1):19–25.
Shirley D, Hodges PW, Eriksson AE, Gandevia SC. Spinal stiffness changes throughout the respiratory cycle. Journal of Applied Physiology 2003;95(4):1467–75.
Stanton TR, Kawchuk GN. Reliability of assisted indentation in measuring lumbar spinal stiffness. Manual Therapy 2009;14(2):197–205.
Teyhen DS. Rehabilitative ultrasound imaging for assessment and treatment of musculoskeletal conditions. Manual Therapy 2011;16(1):44–5.
Tuttle N, Barrett R, Laakso L. Postero-anterior movements of the cervical spine: repeatability of force displacement curves. Manual Therapy 2008;13(4):341–8.
Tuttle N. Is it reasonable to use an individual patient's progress after treatment as a guide to ongoing clinical reasoning? Journal of Manipulative and Physiological Therapeutics 2009;32(5):396–403.
Werneke M, Hart DL, Cook D. A descriptive study of the centralization phenomenon. A prospective analysis. Spine 1999;24(7):676–83.
Wong AYL. Understanding low back biomechanics: through the relation between spinal stiffness and back muscle activity. Saarbrücken: LAP Lambert Academic Publishing; 2012.