Intrarater and Inter-rater Reliability of Maximal Voluntary Neck Muscle Strength Assessment Using a Handheld Dynamometer in Women With Headache and Healthy Women

Ana Paula de Oliveira Carnevalli, PT,[a] Débora Bevilaqua-Grossi, PhD,[a] Ana Izabela Sobral Oliveira, MS,[a] Gabriela Ferreira Carvalho, PhD,[a] César Fernández-De-Las-Peñas, PhD,[b] and Lidiane Lima Florencio, PhD[b]

Abstract

Objective: This study aimed to determine the inter-rater and intrarater reliability, agreement, and minimal detectable change (MDC) of the neck muscle strength test using a handheld dynamometer in healthy women and women with headaches.

Methods: Neck muscle strength in maximal voluntary contraction was measured using the Lafayette Manual Muscle Testing attached to a nonelastic belt in 25 women with migraines and in 25 healthy women. Three repetitions of flexion, extension, and lateral flexion were performed. The tests were performed by 2 examiners on the same day, with a 10-minute interval, and by 1 examiner, with a 1-week interval. The reliability was verified by the intraclass correlation coefficient, the agreement determined by standard error measurement, and the MDC calculated.

Results: The protocol exhibited moderate to excellent intrarater and inter-rater reliabilities in both groups (intraclass correlation coefficient$_{range}$, 0.53-0.90). The standard error measurement ranged from 0.43 to 1.81 and the MDC from 1.49 up to 4.61.

Conclusion: Quantification of neck muscle strength using the handheld dynamometer with an attached nonelastic belt exhibited moderate to excellent intra- and inter-rater reliability in women with and without migraines. Moreover, the standard error measurement and MDC were proven to be useful in the interpretation of data and in guiding clinical decisions. (J Manipulative Physiol Ther 2018;41:621-627)

Key Indexing Terms: *Spine; Cervical Vertebrae; Muscle Strength Dynamometer; Headache*

Introduction

The neck muscles are responsible for proprioception of the head and neck and the maintenance of its posture. Proper control of the 22 pairs of muscles of the cervical spine segment requires different levels of contraction for normal functioning of the head. A relevant parameter in the assessment of cervical spine musculature is the ability to develop force and sustain this force.[1]

Cervical muscle strength is reduced in participants with migraines,[2] tension-type,[3] and cervicogenic headaches.[4] Therefore, assessment of neck muscle strength in clinical practice is necessary to assess individuals with primary or secondary headache disorders. Quantification of neck muscle strength could provide useful clinical information to better define the goals of therapeutic interventions, to observe clinical progress during the therapeutic process, or even to identify the presence of chronic impairments.[5] Therefore, a reliable and accessible method to quantify neck muscle strength is necessary.

Maximal voluntary contraction is 1 of the most common forms of muscle strength assessment to identify deficits in muscle performance.[2,5,6] In general, force production on maximal voluntary contraction measured by customized

[a] Department of Biomechanics, Medicine and Locomotor Apparatus Rehabilitation – Ribeirão Preto Medical School, University of São Paulo, Ribeirão Preto, São Paulo, Brazil.

[b] Department of Physical Therapy, Occupational Therapy, Physical Medicine and Rehabilitation, Universidad Rey Juan Carlos, Alcorcón, Spain.

Corresponding author: Débora Bevilaqua-Grossi, PhD, Department of Biomechanics, Medicine and Locomotor Apparatus Rehabilitation – Ribeirão Preto Medical School, University of São Paulo, Avenida dos Bandeirantes, 3900, Ribeirão Preto, São Paulo, Brazil 14049-900. Tel.: +55 16 3315 4413. (e-mail: *deborabg@fmrp.usp.br*).

Paper submitted August 19, 2017; in revised form November 6, 2017; accepted January 7, 2018.

0161-4754

https://doi.org/10.1016/j.jmpt.2018.01.006

load cells or dynamometer has exhibited good to excellent reliability.[7-11] Available tools to measure muscle strength during a maximal voluntary contraction consist of customized settings that combined load cells with proper stabilization or isokinetic equipment. However, these tools usually require high space and high cost, which makes it difficult to be used in routine clinical practice.

Neck strength measurement using a handheld dynamometer (HHD) is a viable alternative option owing to its pragmatic approach for clinicians to determine neck muscle force. In fact, the advantages of the use of an HHD in clinical practice to assess neck muscles are its relatively low cost, small size, portability, and wide range of segments that could also be evaluated. An HHD is considered a valid and reliable muscle strength instrument for properly assessing the strength of muscles of the shoulder, elbow, knee, and ankle.[12] In addition, the HHD has demonstrated excellent inter-rater reliability for cervical spine muscle assessment in rugby athletes,[13] excellent reliability for healthy participants, and moderate reliability among participants with neck pain.[14] On the other hand, a major criticism of the HHD is the possible addition of the examiner's or the participant's stabilization to the force recorded by the HHD.[1]

To the best of the authors' knowledge, no previous study has investigated the reliability of neck muscle strength assessment using an HHD in participants experiencing a headache. Because reliability is a parameter that relates the measurement error to the variability between the participants,[15] reliability reported for healthy participants or using different measurement techniques might not be the same.

The headache subtype chosen to compose our sample was migraine because of the clinical impact of the association between migraine and neck pain as a result of cervical spine muscle disorder. Participants with migraines are 5 times more likely to present neck pain compared with participants without headaches.[16] In fact, the comorbidity of these 2 painful conditions has a negative impact for both because neck pain delays migraine treatment[17] and chronic migraine is associated with a higher risk for presentation of more severe neck disability.[18] In addition, weakness of neck extensors has been associated with chronic migraine.[2]

Therefore, the objectives of the current study were to determine the inter-rater and intrarater reliability and agreement of neck muscle strength test in healthy women and women with migraines using an HHD customized with a nonelastic belt to avoid external influence. Also, we aimed to describe the minimal detectable change (MDC) attributed to the protocol proposed.

Materials and Methods

Study Design

The study was performed between July 2014 and June 2015 with the approval of the Ethics Committee of the Ribeirão Preto School of Medicine, University of São Paulo (process number 16159/2014). All volunteers signed an informed consent form. The sample size was calculated by adopting a 95% confidence interval, expecting an intraclass correlation coefficient (ICC) of 0.80, and accepting a confidence interval of ICC of 0.30, resulting in a minimum sample of 24 individuals.[19] Accordingly, the final sample was composed of 25 healthy women without a headache and 25 with migraine.

Participants

Consecutive participants experiencing migraines, recruited from a university-based headache center, were screened for eligibility criteria. Patients were diagnosed by neurologists, who were experts in headache management, according to the *International Classification of Headache Disorders* criteria, third edition (ICHD3 beta 2013) down to a third-digit level (code 1.1, 1.3).[20] The inclusion criteria were age between 18 and 55 years and a frequency of at least 1 migraine attack per month.

The exclusion criteria were as follows: other primary or secondary headaches, including medication overuse for headaches according to the ICHD3 criteria; history of neck or head trauma; systemic medical disease; fibromyalgia syndrome; or anesthetic blocks within the past 6 months.

Healthy women aged between 18 and 55 years and with no history of headache and no headache attack over the past year were recruited from the local community. The exclusion criteria for the healthy nonheadache group were the same as those for the migraine group. Informed consent was obtained from all patients.

Clinical Data

Clinical outcomes included migraine and neck pain frequency (days per month), the intensity of migraine attack (numeric pain rate scale 0-10), and number of years with migraine or neck pain. Neck-related disability was assessed using the Brazilian version of Neck Disability Index, which classifies the participants as no disability (0-4 points), mild (5-14 points), moderate (15-24 points), severe (25-34 points), or complete disability (35 or more points).[21]

Procedure

The neck muscle strength was measured using an HHD (Lafayette Manual Muscle Testing System, Lafayette Instrument Co, Lafayette, Indiana). The HHD was customized by attaching a nonelastic belt to resist the head movement, rather than the examiner's hand, to avoid any additional force from the examiner stabilization. The examiners underwent a 15-hour training before the start of the study. Neck movements included flexion, extension, and right and left lateral flexion. Rotation was not verified owing to its rotational component that cannot be precisely registered by any load cell. The order of assessment was randomized in

Fig 1. *Neck strength measurement proposed in flexion (A), extension (B), and lateral flexion (C) using a handheld dynamometer.*

the first test, and all subsequent tests were performed using the same sequence. A nonelastic belt was also fastened at the level of the T3 vertebra and the superior iliac spine to avoid compensations. The neck flexor strength was measured with the participants in a supine position and with the head and neck in a neutral position, the knee flexed, and the arms alongside the body (Fig 1A). The extension strength was measured with the participants in a prone position, with the arms alongside the body (Fig 1B). To measure neck lateral flexor strength, the participants were side-lying, with an adjustable pillow to maintain the head and neck in a neutral position (Fig 1C). All participants performed 3 repetitions of flexion, extension, and lateral flexion, using maximum voluntary effort with standard verbal encouragement, with a 15-second interval between trials and a 2-minute interval between each movement direction.

For inter-rater reliability and agreement, participants were assessed on the same day by 2 different examiners with a 10-minute interval between assessments. The participant was removed from the test position and was positioned again by the second examiner. To determine the intrarater reliability and agreement, participants were assessed by the same examiner with a 1-week interval in the same place and day period of the first test. The examiners were not blinded to the diagnostic condition because they should ask if migraineurs presented any sign or symptom during the test. The participants with migraines were pain-free at the time of examination. If a migraine attack suddenly occurred during the test, the test was interrupted and the volunteer was excluded from the analysis.

All procedures followed were performed in compliance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1975, as revised in 2008. The study was approved by the Committee of Ethics in Research of the Ribeirão Preto Medical School of the University of São Paulo (process HCRP no. 16159/2014).

Statistical Analysis

Descriptive analysis was used to characterize the sample, including age, neck pain features, neck pain-related disability, and migraine features. The reliability and agreement were calculated for both parameters: higher value among the 3 trials and the mean of the values. It aimed to verify whether the properties measured remain similar under conditions in which reference values are different to advise which value would be the most reliable.

The ICC was calculated to determine the intrarater and inter-rater reliabilities.[22,23] The $ICC_{2,3}$ was calculated using the mean among the 3 measures, and $ICC_{2,1}$ was calculated using the peak for each movement direction assessed using the Statistical Package for the Social Sciences version 16.0. The reliability was classified as poor (≤0.40), moderate (0.41-0.60), substantial (0.61-0.80), or excellent (≥0.81).[24]

Intra- and inter-rater agreements were verified by the standard error of measurement (SEM).[15,25] The SEM was calculated on Microsoft Excel using the following formula: $SEM = SD \times \sqrt{(1-ICC)}$. To facilitate clinical interpretation of SEM, the proportion that it represents in percentage of the mean values found for each group was calculated. In addition, MDC was calculated based on values of the 1-week interval test–retest according to the following formula: $MDC = 1.96 \times (\sqrt{2}) \times SEM$. The MDC represents the minimal change that should be considered to guarantee a real change when applying the proposed assessment.[25]

Results

A total of 50 women participated in the study: 25 with migraine (mean age: 27 ± 10 years) and 25 without headache (mean age: 27 ± 9 years). Table 1 shows the clinical characteristics of both groups.

The migraine group assessment exhibited intrarater reliability ranging from substantial to excellent for an average value (ICC_{range}, 0.72-0.84) and moderate to substantial for a peak value (ICC_{range}, 0.53-0.75). The inter-rater reliability was substantial to excellent for an average value (ICC_{range}, 0.79-0.90) and ranking from substantial for a peak value (ICC_{range}, 0.63-0.78). The SEM, MDC, and detailed ICC values for the migraine group are presented in Table 2.

Table 1. *Clinical Characteristics of Nonheadache and Migraine Groups*

Sample Characteristics	Healthy People (n = 25)	Migraine (n = 25)
Age (y)	27 (9)	27 (10)
Migraine frequency (d/mo)	-	9.1 (6.3)
Migraine intensity (NPRS, 0-10)	-	7.7 (1.7)
Neck pain n (%)	8 (32)	21 (84)
Frequency of neck pain (d/mo)	8.4 (7.8)	13.2 (8.7)
Neck pain intensity (NPRS, 0-10)	4.7 (0.9)	5.2 (2.3)
NDI classification n (%)		
None	22 (88)	5 (20)
Mild	3 (12)	17 (68)
Moderate	-	2 (8)
Severe	-	1 (4)
Complete	-	-

NDI, Neck Disability Index; *NPRS*, Numeric Pain Rating Scale.

The nonheadache group presented substantial to excellent intrarater reliability for an average value (ICC_{range}, 0.79-0.90), and ranking from substantial to excellent for a peak value (ICC_{range}, 0.63-0.82) was considered. The inter-rater reliability of this group was substantial to excellent for the average (ICC_{range}, 0.78-0.86) and moderate to substantial when the peak of 3 repetitions was used (ICC_{range}, 0.57-0.72). The SEM, MDC, and detailed ICC values for the healthy group are presented in Table 3.

Table 4 shows the mean force registered on each direction for both groups. These data are accompanied by their respective SEMs and the percentage representing the magnitude of the SEM. The SEM (kilogram force) ranged from 10% to 21% and from 8% to 19% for the migraine and the healthy nonheadache groups, respectively.

Discussion

The present study predominantly reported substantial to excellent intra- and inter-rater reliabilities for neck muscle strength measurement using a customized HHD in women with and without migraines. Two exceptions were observed in the peak values: the migraine group presented moderate intrarater reliability at left lateral flexion and flexion, and the nonheadache group presented moderate inter-rater reliability at extension.

The finding that the reliability observed in the current study is higher than that reported by Shahid et al,[14] who tested another brand of HHD, is interesting to note. In their experiment, the participants were required to stabilize their head in a neutral position while the examiner produced a force, and the value needed to change this neutral position was registered. These authors reported moderate to excellent reliability for individuals without neck pain and weak to excellent reliability for participants with neck pain.[14] On the other hand, our findings of reliability are similar to those reported by Versteegh et al,[26] who tested a self-generated resistance with an HHD to verify the strength of the cervical muscles in healthy participants. The reliability of cervical muscle strength assessment has been hypothesized to be better when participants are asked to produce force against a resistance than when they have to maintain a standard head posture.

In the current study, we provided the SEM and MDC that are expressed in the same scale of the measure, that is, kilogram force, which makes it easier for interpretation in the clinical setting. The SEM represents the absolute amount of measurement error, and MDC represents the minimum value needed for the evaluation proposed to identify a change. Accordingly, only the differences that overcome these values can be considered a real change, with a 95% level of confidence. In such cases, both scores would be valuable to decide whether or not the proposed examination procedure would fit within research and clinical practice, depending on between-group differences where the SEM should be considered or the amount expected to be improved when MDC should be known.[15,25] The SEM of both groups was observed to range from 0.54 to 1.66 for intrarater agreement and from 0.43 to 1.81 for inter-rater agreement.

Table 2. *Inter-rater and Intrarater Reliability of Neck Muscle Strength for the Migraine Group (n = 25)*

	Mean					Peak				
	Intrarater			Inter-rater		Intrarater			Inter-rater	
	ICC (95% CI)	SEM	MDC	ICC (95% CI)	SEM	ICC (95% CI)	SEM	MDC	ICC (95% CI)	SEM
Flexion	0.77 (0.48-0.91)	0.54	1.49	0.85 (0.65-0.93)	0.47	0.60 (0.27-0.80)	0.83	2.31	0.64 (0.34-0.82)	0.81
Extension	0.76 (0.54-0.89)	1.50	4.16	0.90 (0.78-0.96)	0.84	0.69 (0.41-0.85)	1.66	4.61	0.78 (0.56-0.90)	1.34
RLF	0.84 (0.63-0.93)	0.82	2.26	0.90 (0.78-0.96)	0.61	0.75 (0.51-0.88)	1.03	2.87	0.72 (0.46-0.86)	1.09
LLF	0.72 (0.39-0.88)	1.04	2.88	0.79 (0.52-0.91)	0.82	0.53 (0.18-0.76)	1.46	4.06	0.63 (0.32-0.82)	1.16

CI, confidence interval; *ICC*, intraclass correlation coefficient; *LLF*, left lateral-flexion; *MDC*, minimal detectable change; *RLF*, right lateral-flexion; *SEM*, standard error of measurement.

Table 3. *Inter-rater and Intrarater Reliability of Neck Muscle Strength for the Healthy Nonheadache Group (n = 25)*

	Mean					Peak				
	Intrarater			Inter-rater		Intrarater			Inter-rater	
	ICC (95% CI)	SEM	MDC	ICC (95% CI)	SEM	ICC (95% CI)	SEM	MDC	ICC (95% CI)	SEM
Flexion	0.79 (0.52-0.91)	0.59	1.62	0.82 (0.64-0.93)	0.43	0.65 (0.29-0.84)	0.84	2.32	0.68 (0.39-0.84)	0.67
Extension	0.90 (0.78-0.96)	0.86	2.38	0.78 (0.51-0.90)	1.21	0.82 (0.64-0.92)	1.22	3.39	0.57 (0.23-0.78)	1.81
RLF	0.83 (0.56-0.93)	0.86	2.40	0.86 (0.69-0.94)	0.67	0.63 (0.29-0.82)	1.35	3.74	0.72 (0.47-0.87)	1.01
LLF	0.8 (0.58-0.92)	0.87	2.42	0.84 (0.64-0.93)	0.78	0.67 (0.55-0.91)	1.19	3.29	0.60 (0.28-0.89)	1.26

CI, confidence interval; *ICC*, intraclass correlation coefficient; *LLF*, left lateral-flexion; *MDC*, minimal detectable change; *RLF*, right lateral-flexion; *SEM*, standard error of measurement.

In general, lower SEM and MDC values and higher ICC values were obtained when the average value of the 3 trials was used for analysis rather than the peak value. Therefore, we suggest using this variable in clinical practice. The advantage of the proposed assessment is that a simple adaptation of the HHD with a nonelastic belt associated with proper patient's stabilization can be used to easily replicate our method in the clinical setting.

An important point to consider is the external validity of these measurement properties that would be replicable when neck muscle strength is assessed, as described herein, and the sample included presents similar characteristics than the sample included in our study. The generalization of the reliability is better represented by the confidence interval reported for each ICC, that is, the range of possible ICC values in the population with a 95% of confidence. These ranges indicate that the reliability in general population may be lower or higher than the actual value estimated to our sample. Moreover, the satisfactory reliability demonstrated here does not guarantee the validity of the measurement. Despite the validity of the HHD demonstrated for other segments,[12] to the best of our knowledge, there is no report on its validity for assessment of neck muscle strength.

The present study is the first to assess the reliability of an HHD to measure the neck muscle strength of participants with headaches. The applicability of the protocol and the reference parameters regarding neck muscle strength assessment provided to subsidize proper clinical practice and research in the field were identified as the strengths of the study. The current study agrees with quality criteria recommendation guidelines.[15,22,23]

Limitations

First, the assessment proposed did not include cervical rotation because of the characteristics of the HHD that is not able to verify the complete vector of this movement. However, this finding is also a limitation of all load cells used for neck muscle strength assessment. Second, the same psychometric properties cannot be assumed to be found in men with or without migraine and in different age groups. Future studies should include men with migraines and also different headache disorders to compare the reliability of cervical muscle strength assessment.

Conclusion

This study found that quantification of cervical muscle strength (maximal voluntary contraction) using the HHD Lafayette Manual Muscle Testing System with an attached nonelastic belt exhibited moderate to excellent intra- and

Table 4. *Neck Strength Measure and the Proportion That Standard Error Measurement Represents From Nonheadache (n = 25) and Migraine (n = 25) Groups*

		Mean					Peak				
		Intrarater			Inter-rater		Intrarater			Inter-rater	
		Group Mean (kgf)	SEM (kgf)	%	SEM (kgf)	%	Group Mean (kgf)	SEM (kgf)	%	SEM (kgf)	%
Flexion	Nonheadache	3.84	0.59	15	0.43	11	4.33	0.84	19	0.63	14
	Migraine	3.61	0.54	15	0.47	13	4.19	0.83	20	0.81	19
Extension	Nonheadache	10.20	0.86	8	1.21	12	11.15	1.22	11	1.81	16
	Migraine	8.71	1.50	17	0.84	10	9.80	1.66	17	1.34	14
RLF	Nonheadache	6.79	0.86	13	0.67	10	7.55	1.35	18	1.01	13
	Migraine	6.40	0.82	13	0.61	10	7.17	1.03	14	1.09	15
LLF	Nonheadache	6.84	0.87	13	0.78	11	7.58	1.19	16	1.26	17
	Migraine	6.27	1.04	17	0.82	13	6.93	1.46	21	1.16	17

LLF, left lateral-flexion; *RLF*, right lateral-flexion; *SEM*, standard error of measurement.

inter-rater reliability in women with and without migraines. This can be considered a reliable and easily applicable tool for clinical settings and research. Moreover, SEM and MDC were proven to be useful in the interpretation of data and to guide clinical decisions.

Funding Sources and Conflicts of Interest

This work was supported by São Paulo Research Foundation by a grant given for the first author to develop the research related to the current manuscript (grant number: 2015/15854-0). No conflicts of interest were reported for this study.

Contributorship Information

Concept development (provided idea for the research): D.B.G., L.L.F.
Design (planned the methods to generate the results): D.B.-G., G.F.C., L.L.F.
Supervision (provided oversight, responsible for organization and implementation, writing of the manuscript): A.P.d.O.C., L.L.F.
Data collection/processing (responsible for experiments, patient management, organization, or reporting data): A.P.d.O.C., A.I.S.O., L.L.F.
Analysis/interpretation (responsible for statistical analysis, evaluation, and presentation of the results): A.P.d.O.C., D.B.-G., G.F.C., C.F.-D.-L.-P., L.L.F.
Literature search (performed the literature search): A.P.d.O.C.
Writing (responsible for writing a substantive part of the manuscript): A.P.d.O.C., L.L.F.
Critical review (revised manuscript for intellectual content, this does not relate to spelling and grammar checking): D.B.-G., A.I.S.O., G.F.C., C.F.-D.-L.-P.

Practical Applications

- Neck strength measurements using an HHD exhibit predominantly excellent reliability.
- Reported error measurements guide the interpretation of clinical and research data.
- A reliable and easily applicable tool for neck muscle assessment was provided.

References

1. Dvir Z, Prushanky T. Cervical muscles strength testing: methods and clinical implications. *J Manipulative Physiol Ther.* 2008;31(7):518-524.
2. Florencio LL, de Oliveira AS, Carvalho GF, et al. Cervical muscle strength and muscle coactivation during isometric contractions in patients with migraine: a cross-sectional study. *Headache.* 2015;55(10):1312-1322.
3. Oksanen A, Pöyhönen T, Ylinen JJ, et al. Force production and EMG activity of neck muscles in adolescent headache. *Disabil Rehabil.* 2008;30(3):231-239.
4. Jull G, Amiri M, Bullock-Saxton J, Darnell R, Lander C. Cervical musculoskeletal impairment in frequent intermittent headache. Part 1: Subjects with single headaches. *Cephalalgia.* 2007;27(7):793-802.

5. Strimpakos N. The assessment of the cervical spine. Part 2: Strength and endurance/fatigue. *J Bodyw Mov Ther.* 2011;15(4):417-430.
6. Koning CH, Van den Heuvel SP, Staal JB, Smits-Engelsman BC, Hendriks EJ. Clinimetric evaluation of methods to measure muscle functioning in patients with non-specific neck pain: a systematic review. *BMC Musculoskelet Disord.* 2008;9:142.
7. Strimpakos N, Sakellari V, Gioftsos G, Oldham J. Intratester and intertester reliability of neck isometric dynamometry. *Arch Phys Med Rehabil.* 2004;85(8):1309-1316.
8. Seng KY, Lee Peter VS, Lam PM. Neck muscle strength across the sagital and coronal planes: an isometric study. *Clin Biomech.* 2002;17(7):545-547.
9. Ylinen J, Salo P, Nykänen M, Kautiainen H, Häkkinen A. Decreased isometric neck strength in women with chronic neck pain and the repeatability of neck strength measurements. *Arch Phys Med Rehabil.* 2004;85(8):1303-1308.
10. Cagnie B, Cools A, De Loose V, Cambier D, Danneels L. Differences in isometric neck muscle strength between healthy controls and women with chronic neck pain: the use of a reliable measurement. *Arch Phys Med Rehabil.* 2007;88(11):1441-1445.
11. Chiu TT, Lo SK. Evaluation of cervical range of motion and isometric neck muscle strength: reliability and validity. *Clin Rehabil.* 2002;16(8):851-858.
12. Stark T, Walker B, Phillips JK, Fejer R, Beck R. Hand-held dynamometry correlation with the gold standard isokinetic dynamometry: a systematic review. *PMR.* 2011;3(5):472-479.
13. Geary K, Green BS, Delahunt E. Intrarater reliability of neck strength measurement of rugby union players using a handheld dynamometer. *J Manipulative Physiol Ther.* 2013;36(7):444-449.
14. Shahidi B, Johnson CL, Curran-Everett D, Maluf KS. Reliability and group differences in quantitative cervicothoracic measures among individuals with and without chronic neck pain. *BMC Musculoskelet Disord.* 2012;13(215):1-11.
15. Vet HC, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. *J Clin Epidemiol.* 2006;59(10):1033-1039.
16. Plesh O, Adams SH, Gasky SA. Self-reported comorbid pain in severe headaches or migraines in US national sample. *Headache.* 2012;52(6):946-956.
17. Calhoun AH, Ford S. Double-blind, placebo-controlled, crossover study of early-intervention with sumatriptan 85/naproxen sodium 500 in (truly) episodic migraine: what's neck pain got to do with it? *Postgrad Med.* 2014;126(2):86-90.
18. Florencio LL, Chaves TC, Carvalho GF, et al. Neck pain disability is related to the frequency of migraine attacks: a cross-sectional study. *Headache.* 2014;54(7):1203-1210.
19. Bonnett DG. Sample size requirements for estimating intraclass correlations with desired precision. *Stat Med.* 2002;21(9):1331-1335.
20. Headache Classification Committee of the International Headache Society. The international classification of headache disorders, 3rd ed (beta version). *Cephalalgia.* 2013;33(9):629-808.
21. Cook C, Richardson JK, Braga L, et al. Cross-cultural adaptation and validation of the Brazilian Portuguese version of the Neck Disability Index and Neck Pain and Disability Scale. *Spine.* 2006;31(14):1621-1627.
22. Kottner J, Audige L, Brorson S, Donner A, Gajewski BJ, Hróbjartsson A. Guidelines for reporting reliability and agreement studies (grras) were proposed. *J Clin Epidemiol.* 2011;64(1):96-106.
23. COSMIN. VU University Medical Center. Available at: http://www.cosmin.nl/images/upload/files/Tabel%20met%20definities-new.pdf. Accessed September 29, 2014.
24. Eliasziw M, Young SL, Woodbury MG, Fryday-Field K. Statistical methodology for the concurrent assessment of interrater and intrarater reliability using goniometric measurements as an example. *Phys Ther.* 1994;74(8):777-788.
25. Scholtes VA, Terwee CB, Poolman RW. What makes a measurement instrument valid and reliable? *Injury.* 2011;42(3):236-240.
26. Versteegh T, Beaudet D, Greenbaum M, Hellyer L, Tritton A, Walton D. Evaluating the reliability of a novel neck-strength assessment protocol for healthy adults using self-generated resistance with a hand-held dynamometer. *Physiother Can.* 2015;67(1):58-64.