Reliability of a Measurement of Neck Flexor Muscle Endurance
Kevin D Harris, Darren M Heer, Tanja C Roy, Diane M Santos, Julie M Whitman and Robert S Wainner
PHYS THER. 2005; 85:1349-1355.

The online version of this article, along with updated information and services, can be found online at: http://ptjournal.apta.org/content/85/12/1349

Collections
This article, along with others on similar topics, appears in the following collection(s):
Injuries and Conditions: Neck Tests and Measurements

e-Letters
To submit an e-Letter on this article, click here or click on "Submit a response" in the right-hand menu under "Responses" in the online version of this article.

E-mail alerts
Sign up here to receive free e-mail alerts
Reliability of a Measurement of Neck Flexor Muscle Endurance

Background and Purpose. Neck flexor muscle endurance has been negatively correlated with cervical pain and dysfunction. The purposes of this study were to determine rater reliability in subjects both with and without neck pain and to determine whether there was a difference in neck flexor muscle endurance between the 2 groups. Subjects. Forty-one subjects with and without neck pain were enrolled in this repeated-measures reliability study. Methods. Two raters used an isometric neck retraction test to assess neck flexor muscle endurance for all subjects during an initial session, and subjects without neck pain returned for testing 1 week later. Results. For the group without neck pain, intrarater reliability was good to excellent (intraclass correlation coefficient [ICC(3,1)] = .82–.91), and interrater reliability was moderate to good (ICC[2,1] = .67–.78). The associated standard error of measurement (SEM) ranged from 8.0 to 11.0 seconds and from 12.6 to 15.3 seconds, respectively. For the group with neck pain, interrater reliability was moderate (ICC[2,1] = .67, SEM = 11.5). Neck flexor muscle endurance test results for the group without neck pain (X = 38.95 seconds, SD = 26.4) and the group with neck pain (X = 24.1 seconds, SD = 12.8) were significantly different. Discussion and Conclusion. Reliability coefficients differed between the 2 groups and ranged from moderate to excellent and improved after the first test session. The interrater reliability of data obtained with the neck flexor muscle endurance test in people with neck pain must be improved in order for clinicians to distinguish a clinically meaningful change from measurement error. Neck flexor muscle endurance was both statistically and clinically greater for subjects without neck pain than for those with neck pain. [Harris KD, Heer DM, Roy TC, et al. Reliability of a measurement of neck flexor muscle endurance. Phys Ther. 2005;85:1349–1355.]

Key Words: Cervical impairment, Cervical pain, Chin tuck, Deep neck flexors.

Kevin D Harris, Darren M Heer, Tanja C Roy, Diane M Santos, Julie M Whitman, Robert S Wainner
Cervical spine pain affects nearly 50% of the population at some time in their lives. It is often associated with a variety of complications, such as headache and loss of range of motion in the cervical spine. In addition, cervical impairment (defined as any cervical pain, loss of range of motion, or decreased strength [defined as the force-generating capacity of a muscle] or endurance of the neck musculature, including, but not limited to, the sternocleidomastoid [SCM], trapezius, and deep neck flexor [DNF] muscles) has been found in up to 70% of people with headaches. Previous research investigating cervical impairment suggests that DNF muscle weakness might be a causative or contributory factor in the pathogenesis of head and neck pain.

The DNF muscles are small stabilizing muscles located on the anterior and anterolateral surfaces of the cervical spine deep to the SCM muscle. Although the muscles comprising the DNFs are not clearly defined, the longus capitis, longus colli, rectus capitis anterior, and occasionally the rectus capitis lateralis muscles have been cited in the literature. The location of the DNFs suggest that they potentially play an important role in stabilizing the cervical spine. Mayoux-Benhamou reported that the longus colli and dorsal neck muscles form a sleeve that stabilizes the cervical spine in all positions against the effects of gravity. It is theorized that when muscle performance is impaired, the balance between the stabilizers on the posterior aspect of the neck and the DNFs will be disrupted, resulting in loss of proper alignment and posture, which is then likely to contribute to cervical impairment.

Assessment of DNF muscle performance properties is arguably important given the association between DNF muscle torque and endurance deficits and patient complaints of head and neck pain. In addition, the results of one randomized controlled trial of patients with neck and headache complaints showed that a group of patients who received endurance exercises, attempting to target the DNF musculature as part of a multimodal intervention, experienced a significant reduction in neck pain and headache frequency as compared with a control (no intervention) group.

However, one difficulty in assessing performance of the DNFs is isolating their contribution from that of the more superficial SCM and anterior scalene muscles. Current methods of manual muscle testing are inadequate to isolate differences between the SCM and deep neck flexor muscle groups, both of which work in concert to produce a cervical flexion moment. Although the SCM muscle makes the largest contribution of all the neck flexors to performing neck flexion, electromyographic (EMG) studies provide evidence that DNFs are active in this position and that they remain contracted in static and small phasic flexion contractions. Falla et al recently demonstrated that the DNFs are increasingly active during cranio cervical flexion (chin tuck). With EMG electrodes placed on the posterior oropharyngeal wall, they were able to record

KD Harris, PT, MPT, CSCS; DM Heer, PT, PTA, MPT, OCS; TC Roy, PT, MPT; and DM Santos, PT, MPT, OCS, were students in the US Army–Baylor Graduate Program in Physical Therapy, San Antonio, Tex, at the time of this study and completed the project in partial fulfillment of the Master in Physical Therapy degree requirements. Address all correspondence to: Captain Kevin Harris, Group Physical Therapist, 10th Special Forces Group (Airborne), Troop Medical Clinic #10, 7490 Sutherland Cir, Ft Carson, CO 80913 (USA) (kevin.d.harris@us.army.mil).

JM Whitman, PT, DSc, OCS, FAAOMPT, is Assistant Professor, Regis University, Denver, Colo, and US Army–Baylor University Postprofessional Doctoral Program in Orthopaedic and Manual Physical Therapy.

RS Wainner, PT, PhD, OCS, ECS, FAAOMPT, is Assistant Professor, Faculty, and Project Advisor, US Army–Baylor Graduate Program in Physical Therapy.

Mr Harris, Ms Roy, and Dr Whitman provided concept/idea/research design. All authors provided writing. Mr Heer, Ms Santos, and Dr Wainner provided data collection. Mr Harris, Ms Roy, Ms Santos, Dr Whitman, and Dr Wainner provided data analysis. Mr Harris, Dr Whitman, and Dr Wainner provided project management. Mr Harris, Mr Heer, Ms Roy, and Dr Whitman provided subjects. Dr Wainner provided clerical support. Dr Whitman and Dr Wainner provided facilities/equipment, institutional liaisons, and consultation (including review of manuscript before submission).

This study was approved by the Institutional Review Board of Wilford Hall USAF Medical Center (WHMC).

This research, in part, was given as a poster presentation at the Annual Conference of the American Academy of Orthopaedic Manual Physical Therapists; October 3–6, 2003; Orlando, Fla.

The opinions or assertions contained herein are the private views of the authors and are not to be construed as official or as reflecting the views of the Department of the Army, the Department of the Air Force, or the Department of Defense.

This article was received February 27, 2005, and was accepted June 30, 2005.
increasing levels of EMG signals of the longus colli and longus capitis muscles with increasing levels of cranio-cervical flexion while the levels of EMG signals of the SCM and anterior scalene muscles, although somewhat active, remained relatively stable. Therefore, a test of neck flexor muscle performance that utilizes a measure of endurance and includes a chin-tuck position should maximize the contributions of the DNF muscles.8,16

Unfortunately, physical therapists are currently limited by the lack of a clinically efficient noninstrumented neck flexor muscle endurance test that yields valid and reliable data and that attempts to maximize the contributions of the DNF muscles. To our knowledge, Placzeck et al,4 Grimmer,9 and Blizzard et al17 are the only researchers who used a noninstrumented assessment technique that could be easily performed in any clinical setting. Unfortunately, Placzeck et al tested only women, and no reliability measurements were obtained. Blizzard et al17 conducted only intrarater reliability testing and did not separate subjects with headaches from those without headaches in their calculations of reliability. Grimmer9 tested only subjects without neck pain and only performed intrarater reliability testing. Although instrumented tests that yield valid and reliable data have been cited in the literature, they are seldom used clinically due to cost and time constraints.18 In addition, none have reported the average hold time for a test of neck flexor muscle endurance.

The primary purpose of this study was to determine the interrater and intrarater reliability of measurements obtained using a simple clinical test of neck flexor muscle endurance in subjects without neck pain and the intrarater reliability of measurements obtained with the test in subjects with neck pain. The secondary purpose was to determine whether neck flexor muscle endurance was less in subjects with neck pain than in those without neck pain.

Method

Subjects

Twenty subjects with neck pain (mean age = 38 years, SD = 10) and 20 subjects without neck pain (mean age = 33 years, SD = 8) completed the study. A total of 41 subjects with and without neck pain (25 female, 16 male; mean age = 36.04 years) were enrolled in this study. The inclusion criterion for the subjects with neck pain was a primary complaint of neck pain. Inclusion criteria for subjects without neck pain were: (1) no complaints of current neck pain, (2) no symptomatic cervical joints upon cervical spine palpation examination, and (3) no reports of neck or thoracic region pain or headaches in the last year resulting in limitations of daily activities. Exclusion criteria for both groups were: (1) a history of spinal surgery to the thoracic or cervical spine, (2) known cervical abnormalities, (3) known systemic, muscular, or connective tissue disorders, (4) a history of cancer, and (5) a history of significant injury to the neck or upper thoracic spine. Additional exclusion criteria for subjects in the neck pain group included: difficulty falling asleep at night due to neck symptoms, diminished motor function of 2 or more muscles of the same nerve root level, an absent upper-extremity muscle stretch reflex, or a positive Valsalva maneuver. Subjects also were excluded if they had a disk herniation on magnetic resonance imaging or a positive electromyogram consistent with a radiculopathy. All participating subjects signed an informed consent document following a detailed explanation by an investigator.

Neck Flexor Muscle Endurance Test

The neck flexor muscle endurance test used in this study was performed in a supine, hook-lying position and was operationally defined as follows: with the chin maximally retracted and maintained isometrically, the subject lifted the head and neck until the head was approximately 2.5 cm (1 in) above the plinth while keeping the chin retracted to the chest (Fig. 1). Once in position, a line
was drawn across 2 approximated skin folds along the subject’s neck, and the rater placed his or her left hand on the table just below the occipital bone of the subject’s head (Fig. 2). Verbal commands (ie, “Tuck your chin” or “Hold your head up”) were given when either the line edges began to separate or the subject’s head touched the rater’s left hand. The test was terminated if the edges of the lines no longer approximated each other due to loss of chin tuck or the subject’s head touched the rater’s hand for more than 1 second. To our knowledge, this operational definition has not been used before.

**Testing Procedure**

Two raters trained in the application of the neck flexor endurance test performed all testing. Each rater alternated being the first rater. Each subject was positioned by the first rater and provided with verbal instructions and pictures depicting proper performance of the test (Fig. 1). The rater then guided the subject through the motions required for performance of the test 2 times. Following the instructions and guidance through the test procedure, the subject was asked to perform the test as operationally defined. A stopwatch was used to measure the length of time (in seconds) that the subject could hold the test position. Measurements were recorded to the nearest second.

Following conclusion of the test, the first rater removed the previously drawn line across the approximated skin folds and instructed the subject to rest for 10 minutes. During this time, the subject was instructed to remain supine and turn his or her neck from side to side through a pain-free range of motion at least 3 times without raising the head from the table. The first rater then departed the room, and the second rater entered and repeated the same procedure. Subjects with neck pain were considered to have completed the study after the second measurement. Subjects without neck pain returned 1 week later and were retested by both examiners in the exact same manner previously described, although the examiners alternated rater position. Each test session lasted approximately 20 minutes.

**Data Analysis**

For subjects without neck pain, intraclass correlation coefficients (ICC[2,1]) for intrarater reliability of the neck flexor muscle endurance test measurements were calculated using the scores from both the first and second test sessions. In addition, ICC(3,1) values for intrarater reliability of the test measurements over a 1-week period were calculated separately for each rater using the measurements obtained in sessions 1 and 2. For subjects with neck pain, ICC(2,1) values for interrater reliability of the neck flexor muscle endurance test measurements were calculated using the scores from raters 1 and 2 for the first test session. Reliability coefficients were interpreted according to the following modification of the criteria proposed by Portney and Watkins\(^\text{19}\): poor \((r<.50)\), moderate \((.50< r<.75)\), good \((.75< r<.90)\), and excellent \((r> .90)\). The standard error of measurement (SEM) was calculated for all reliability coefficients and is a facet of the reliability of the measurement. The interpretation of SEM is dependant on the type of reliability coefficient that is used in its computation. For example, if the estimate is based on test-retest reliability (ICC[3,1]), then the SEM is indicative of the range of scores that can be expected on retesting. If rater reliability (ICC[2,1]) is used, the SEM reflects the extent of expected error in different raters’ scores. To determine whether fatigue was a confounding factor, a paired-samples t test \((\alpha<.05)\) was used to compare each subject’s first and second measurement times within the same session. Because neck flexor muscle endurance test measurements of subjects with neck pain demonstrated limited variability, which compromises the validity of the reliability coefficients, the coefficient of variance of method error (CVme) also was calculated.\(^\text{19}\) Method error is a measure of the discrepancy between 2 sets of
repeated scores, or their difference score. It is converted to a percentage using the coefficient of variation to produce the CVme for comparison relative to the size of the mean differences. A one-tailed independent t test (α<.05) was used to determine whether there was a difference in neck flexor muscle endurance test measurement times between subjects with and without neck pain for the first test session. We used SPSS 10.1* and Excel 2000† to perform all statistical analyses.

**Results**

One subject without neck pain was dropped from the study due to inability to follow up in the 1-week time period. Descriptive statistics by group, test session, and rater are shown in Table 1. Intrarater reliability ranged from good to excellent (ICC[3,1] = .82–.91) for subjects without neck pain. Intrarater reliability was moderate to good (ICC[2,1] = .67–.78) for subjects without neck pain and was moderate (ICC[2,1] = .67) for subjects with neck pain. All ICCs and associated SEMs, along with CVme values, are listed in Table 2. Neck flexor endurance time was significantly greater in the group without neck pain (X=38.95 seconds, SD=26.4) than in the group with neck pain (X=24.1 seconds, SD=12.8) (P=.013) and is illustrated in Figure 3. There were no differences in subjects’ first and second measurement times within the same session for either the group without neck pain (P=.52) or the group with neck pain (P=.43).

**Discussion**

To our knowledge, this is the first study to assess the interrater reliability of measurements obtained from a non-instrumented clinical test of neck flexor muscle endurance in subjects with neck pain. The intrarater reliability for the group without neck pain was moderate for session 1 and good for session 2. Intrarater reliability for the neck pain group was moderate.

There are several possible reasons for the difference in reliability between the groups. One explanation is that, for the group without neck pain, raters had increased practice associated with a second testing session, which may have improved their ability to accurately judge test termination criteria. Another reason may be that subjects with neck pain had less neck flexor muscle endurance. It has been suggested that measurements obtained from subjects with poor endurance capacity demonstrate greater variability on retest than subjects with better endurance capacity. Yet another possibility to consider is range restriction of the scores for the neck pain group, which can deflate ICC values. However, low between-subjects variability did not appear to affect reliability coefficients based on the low CVme value. If subjects had experienced fatigue due to inadequate recovery, this could have resulted in unstable (lower) retest scores, thereby violating an assumption of reliability. It is not likely that fatigue was a factor, because subjects’ scores did not differ significantly between the first and second test applications within a single test session.

Intrarater reliability of neck flexor muscle endurance measurements ranged from good to excellent when assessed in subjects without neck pain. This finding is consistent with the findings of Grimmer, who performed a similar study and also reported good intrarater reliability. In our study, subjects were tested 1 week apart, whereas subjects in Grimmer’s study were tested 1 month apart, indicating that neck flexor muscle test measurements are stable over a short period of time.

It is important for clinicians to understand the amount of error associated with any measurement of performance obtained from an individual patient. Although

### Table 1.

Descriptive Statistics for Group, Test Session, and Rater (All Measurements in Seconds)

|                     | No Neck Pain | Test Session 1 | | | Test Session 2 | | | Neck Pain | | |  
|---------------------|--------------|----------------|-----------------|-----------------|----------------|----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| X                   | 44           | 37.1           | 45.4            | 33.8            | 25.5            | 20.9            |  
| SD                  | 25.57        | 26.68          | 26.5            | 24.86           | 12.53           | 10.99           |  

### Table 2.

Statistical Indexes of Reliability

<table>
<thead>
<tr>
<th></th>
<th>No Neck Pain</th>
<th>Intrarater Reliability</th>
<th>Test Session 1</th>
<th>Test Session 2</th>
<th>Rater A</th>
<th>Rater B</th>
<th>Intrarater Reliability</th>
<th>Neck Pain</th>
<th>Inter rater Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICC</td>
<td>.67</td>
<td>.78</td>
<td>.91</td>
<td>.82</td>
<td>.67</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SEM (s)</td>
<td>15.3</td>
<td>12.6</td>
<td>8.0</td>
<td>11.0</td>
<td>11.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CVme (%)</td>
<td>36.1</td>
<td>25.0</td>
<td>17.6</td>
<td>31</td>
<td>27.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* ICC=intraclass correlation coefficient, SEM=standard error of the mean, CVme=coefficient variation of method error.
* ICC(2,1).
* ICC(3,1).
The ICC provides information about a measure’s ability to differentiate among patients, the SEM expresses measurement error in the same units as the original measurement and is not influenced by variability among patients. Therefore, the SEM may be used to distinguish whether repeated measurements taken from a subject represent true change or measurement error based on the confidence interval used. For example, the SEM associated with the measurements taken in subjects with neck pain was 11.5 seconds. If another therapist repeated this measurement, the second measurement would need to differ from the first measurement by more than 27 seconds to be 90% confident that the difference was due to real patient change and not just measurement error in the 2 measurements (minimum detectable change computed as $1.645 \times \sqrt{2} \times \text{SEM}$). This is a considerable amount of measurement error and lack of precision given that the range of measurements for subjects with neck pain was 58 seconds.

The secondary purpose of this study was to compare neck flexor muscle endurance of subjects with neck pain with that of subjects without neck pain. Our findings were consistent with the results of previous studies, which showed that neck flexor muscle endurance was reduced in subjects with cervicogenic headaches. Whether this relationship is causal in nature remains unknown. Clinicians also should exercise caution when comparing test results from individual patients with our values or those reported by other researchers until normal values for people without neck pain are established.

Although acceptable test reliability was found for subjects without neck pain, such a finding is not of significant clinical utility for therapists who want to use the test in the management of patients with neck pain. Our intention for assessment of intrarater reliability was due to the desire to assess this particular measurement property and the associated clinical implication for therapists when treating patients with neck pain. Future studies of this neck flexor muscle endurance test, we believe, should assess intrarater reliability for subjects with neck pain and perhaps attempt to further objectify and classify those subjects with neck pain upon admission to a study using a measurement tool, such as the Neck Disability Index, instead of simply recruiting subjects based on their report of having neck pain. A test with acceptable intrarater reliability would allow an individual therapist to accurately assess changes in neck flexor muscle endurance over time and is compatible with standard practice patterns in which patients usually are seen by the same therapist at subsequent visits. Additional work also needs to be done with regard to quantifying the contributions of the DNF muscles. Based on previous work, the test position used in this study was thought to maximize the contribution of the DNF muscles. However, we did not measure isolated DNF muscle activity, so their contribution to the endurance test scores is unknown.

We did not use instrumentation in this study because the aim of the study was to develop a simple, cost-effective test that could easily be implemented in the clinic while minimizing clinician and subject burden of a time-consuming instrumented test.

Although the use of standardized equipment, such as handheld dynamometers, may improve test reliability, the development of a noninstrumented test of neck flexor muscle endurance that demonstrates good to excellent reliability in people with neck pain remains a desirable goal. This study represents another step toward that goal. If successful, clinicians will have a measurement tool that may prove useful in the management of patients with neck pain, and researchers will be challenged to determine the diagnostic properties, prognostic ability, and normal values of such a test.

**Conclusions**

For subjects without neck pain, the intrarater reliability of measurements obtained 1 week apart with a clinical neck flexor muscle endurance test was good to excellent. The interrater reliability ranged from moderate to good and may improve over time with practice. For subjects with neck pain, interrater reliability was moderate, and intrarater reliability is unknown. Subjects with neck pain demonstrated clinically and statistically lower neck flexor muscle endurance scores than subjects without neck pain. When using the neck flexor muscle endurance test described in this report in the management of patients with neck pain, measurements should be obtained by a single rater in order to distinguish a clinically meaningful change from measurement error. Future studies are needed to determine intrarater reliability in people with neck pain as well as the diagnostic, prognostic, and evaluative properties of the neck flexor muscle endurance test.
References


Reliability of a Measurement of Neck Flexor Muscle Endurance
Kevin D Harris, Darren M Heer, Tanja C Roy, Diane M Santos, Julie M Whitman and Robert S Wainner
PHYS THER. 2005; 85:1349-1355.