Journal of Pain & Relief

Inter-rater Reliability of Needle Electromyographic Findings in Patients with Sciatica

Nathan J. Savage[1]*, Julie M. Fritz[2], Richard P. Nielsen[3] and Jeffrey Fraser[4]

[1]Clinical Faculty, Department of Physical Therapy, University of Utah, Salt Lake City, Utah, USA
[2]Faculty, Department of Physical Therapy, University of Utah, Salt Lake City, Utah, USA
[3]President, Rocky Mountain University of Health Professions, Provo, Utah, USA
[4]Director of Electrophysiology and Intraoperative Monitoring, Human Performance and Rehabilitation Centers, Columbus, GA, USA

Abstract

Purpose: Investigate inter-rater reliability of needle electromyographic findings among experienced physical therapist electromyographers.

Methods: Masked review of 24 electromyographic recordings from patients with sciatica referred to physical therapy was undertaken. An examiner unmasked to patient history and physical examination findings digitally recorded and stored insertional and resting electromyographic activity as de-identified digital audio-video files to be analyzed by 2 masked examiners. Examiners provided ratings for individual muscles and overall electrodiagnostic impression. Agreement was assessed using Cohen's kappa (κ) statistics.

Results: Examiner agreement for insertional and resting electromyographic activity for all muscles combined was substantial (κ ≥ 0.68, 95% CI: 0.50 to 0.89; P≤.001), ranging from fair (κ=0.33, 95% CI: -0.25 to 1.0; P>.05) to perfect (κ=1.0, 95% CI: 1.0 to 1.0; P≤.001) for individual muscles examined. Pairwise examiner comparisons revealed moderate (κ=0.43, 95% CI: 0.11 to 0.76; P=.01) to substantial (κ=0.75, 95% CI: 0.48 to 1.0; P<.0001) agreement for the final electrodiagnostic impression and fair (κw=0.31, 95% CI: 0.12 to 0.50; P=0.004) to substantial (κw=0.62, 95% CI: 0.37 to 0.87; P<.0001) agreement for the overall electrodiagnostic impression.

Conclusions: Needle electromyographic activity can be reliably assessed among experienced physical therapist electromyographers in patients with sciatica referred to physical therapy.

Keywords: Agreement; Electromyography; Inter-rater reliability; Radiculopathy; Sciatica

Introduction

Electrodiagnostic (EDX) testing, consisting of needle electromyography (EMG) and nerve conduction studies, is used to evaluate the integrity of the neuromuscular system, including upper and lower motor neurons, the neuromuscular junction, and skeletal muscle [1-5]. Conducted as an extension of the clinical examination, EDX testing is the primary method used to objectively measure and document pathological changes or injury to the neuromuscular system, including proximally located spinal nerve roots [2-5]. Clinicians employ EDX testing to evaluate patients with sciatica [2-4,6] with particular emphasis on the results of the needle EMG examination which has high diagnostic specificity in these patients [2,4,5,7].

Although research has demonstrated the utility of needle EMG for evaluating patients with sciatica [7-9] the lack of examiner masking to the results of a patient's history and physical examination in studies utilizing needle EMG has been identified as a potential source of bias, which may weaken the evidence that needle EMG is a valid diagnostic tool [10]. Recent studies have demonstrated that masking in EDX research can be successfully employed in patients with lumbar spinal stenosis [10,11] as well as lumbosacral radiculopathy (Chouteau et al.) in order to validate the results of the needle EMG examination [12]. Chouteau, et al., Investigating inter-rater reliability between a single unmasked examiner and 2 masked examiners in patients with sciatica, found near perfect agreement for the dichotomized final EDX impression (i.e., evidence of radiculopathy or no evidence of radiculopathy) with Cohen's kappa (κ) values exceeding 0.90. Additionally, the authors found substantial agreement (κ>0.60) for insertional and resting EMG activity of most individual muscles examined. Examiners were board-certified by the American Board of Electrodiagnostic Medicine and practiced together in the same facility where all study-related patients underwent EDX testing. In a related

investigation, Kendall and Werner [13] compared the inter-rater reliability among 66 masked examiners, consisting of both faculty and resident examiners, in patients with sciatica. Examiners analyzed insertional, resting, and volitional EMG activity from 6 recorded cases. The authors found a composite agreement of 47% for the diagnostic impression, consisting of 61% agreement among faculty examiners and 29% agreement among resident examiners. However, these values were not corrected for chance agreement using a Cohen's κ or related statistic [14].

Given the routine use of EDX testing to evaluate patients with suspected nerve root injuries, it is surprising that so few studies have investigated the inter-rater reliability of needle EMG as a diagnostic test [12,13]. Furthermore, no published studies have investigated the inter-rater reliability of EDX testing among physical therapist electromyographers or among patients referred to physical therapy. Research demonstrating the inter-rater reliability of EDX testing as a diagnostic tool in a variety of settings is essential in order to establish the validity of EDX testing in patients with sciatica. The purpose of this investigation was to determine the inter-rater reliability of EDX test findings among experienced physical therapist electromyographers in patients with sciatica referred to physical therapy.

*Corresponding author: Nathan J. Savage, Total Rehab, Inc. 5957 South Fashion Point Drive, Suite 102 South Ogden, Utah 84403, USA, Tel: (801) 827-0200; Fax: (801) 827-0201; E-mail: nathan@totalrehabclinics.com

Methods

Patients

Patients with sciatica participating in a randomized clinical trial [15], comparing different physical therapy interventions were recruited for this investigation. Patients consenting to undergo EDX testing were grouped and analyzed according to the presence or absence of radiculopathy to evaluate the prognostic value of this finding. Digital needle EMG recording were assessed on the final 24 patients that met the inclusion criteria for the randomized trial (Table 1) and consented to undergo EDX testing. Patient demographics and clinical characteristics are found in Table 2. Institutional Review Board approvals were obtained from the University of Utah and Intermountain Healthcare (Salt Lake City, Utah) for this study.

Study procedures

Patients were recruited from physician and outpatient physical therapy clinics from March 2011 to February 2012. Eligible patients provided a separate written informed consent to undergo EDX testing. EDX testing was performed by a single examiner (N.J.S.) unmasked to the patient's medical history, clinical examination findings, and results of the complete EDX testing including assessment of peripheral nerve conduction and volitional EMG findings. The unmasked examiner is a licensed physical therapist with 14 years of clinical experience, who is recognized as a board-certified specialist in orthopaedics and clinical electrophysiology by the American Board of Physical Therapy Specialties and has completed more than 1,000 EDX examinations over the past 7 years.

Electrodiagnostic testing

All EDX tests were performed using a Cadwell Sierra Wave system (Cadwell Laboratories, Kennewick, WA). Patients underwent standardized peripheral sensory and motor nerve conduction studies including F waves in order to rule out other conditions and/or comorbidities such as lumbosacral plexopathy or generalized polyneuropathy [2,4]. Needle EMG testing was performed on a standardized set of 5 limb muscles and the lumbar paraspinals with a disposable 50-millimeter monopolar needle electrode. The muscles selected for examination have been demonstrated to identify 98-100% of EMG-confirmable radiculopathies and include the lumbar paraspinals (L1-5), anterior tibialis (L4-5), medial gastrocnemius (S1-2), posterior tibialis (L5-S1), vastus medialis (L2-4), and biceps femoris short-head (L5-S2) [9]. Additional muscles were tested as needed in order to clarify the overall EDX impression, which occurred when radiculopathy was strongly suggestive based on abnormal EMG findings isolated to a single muscle during the standardized EMG examination and the examiner chose to investigate additional muscles to confirm or rule out the presence of radiculopathy (Table 3). Limb muscles were analyzed at rest and during volitional contraction. The lumbar paraspinal muscles were analyzed at rest only and typically involved 2 needle insertions at or near the L4 and L2 spinous processes with repeated insertions (typically 4-5 each direction) angled to include all of the lumbar spinal nerve roots (L1-5).

Insertional and resting EMG activity was assessed with a gain of 100-200 microvolts per division and a sweep speed of 10 milliseconds per division. Each insertion site included 12-20 separate insertions using the "corners of the block" method. The needle EMG examination was digitally recorded and stored using the Cadwell Sierra Wave "Reel Time" EMG software application as audio-video files. The exported needle EMG recordings had a video rate of 30 frames per second and an audio rate of 512 kilobits per second. These settings enabled the masked examiners to visualize the needle EMG recordings with essentially the same audio and video resolution as the live waveforms observed by the unmasked examiner.

Pathological findings at rest indicative of axonal loss included presence of increased insertional activity, positive waves, fibrillation potentials, and/or complex repetitive discharges. Pathological findings during volitional contraction indicative of axonal loss included presence of neuropathic motor unit potentials (eg: increased polyphasia, increased duration, increased amplitude) and/or abnormal motor unit recruitment (eg: mincreased or decreased).

Evidence of radiculopathy was defined by the presence of at least 1 of the following: 1) pathological findings at rest or during volitional contraction indicative of axonal loss in at least 2 muscles (including the lumbar paraspinal muscles) sharing a common nerve root but from different peripheral nerves, or 2) findings isolated to

Inclusion Criteria	Exclusion Criteria
Chief complaint of pain and/or paresthesia in low back with symptoms extending distal to gluteal fold within previous 24 hours.	Known serious spinal pathology or suspicion of serious pathology based on red flags noted in general medical screening.
Modified Oswestry score ≥20%.	Evidence of CNS involvement including presence of pathological reflexes in physical examination.
Age at least 18 and less than 60 years.	
At least 1 of the following signs of nerve root compression: Positive SLR or crossed SLR test.	Patient report of complete absence of LBP and leg symptoms when seated.
Sensory deficit in symptomatic limb.	Recent surgery (<6 months) to low back including fusion of low back or pelvis.
Diminished myotomal strength in symptomatic limb.	Recent (<2 weeks) epidural steroid injection for LBP and/or leg pain.
Diminished muscle stretch reflex in symptomatic limb.	Current pregnancy.
	Known inability to comply with the treatment schedule.

Abbreviations: CNS, central nervous system; LBP, low back pain; SLR, straight leg raise.

Table 1: Patient selection criteria.

Patient characteristics (n=24)	
Age (years)	39.8 ± 13.1
Gender	
women (%)	10 (42%)
men (%)	14 (58%)
BMI (kg/m²)	27.9 ± 5.5
Smoker (%)	2 (8%)
Average LBP baseline (0-10)	4.8 ± 2.1
Average leg pain baseline (0-10)	4.7 ± 2.5
Oswestry score baseline (0-100)	40.6 ± 14.6
RMDQ score baseline (0-24)	11.8 ± 5.4
Duration current episode (weeks)	27.5 ± 67.9

Table 2: Patient demographic and clinical characteristics.

Muscle	Number of patients (%)
Lumbar paraspinals	23 (96%)
Anterior tibialis	24 (100%)
Medial gastrocnemius	24 (100%)
Lateral gastrocnemius	4 (17%)
Posterior tibialis	24 (100%)
Extensor hallucis longus	7 (29%)
Vastus medialis	24 (100%)
Biceps femoris short-head	24 (100%)

Table 3: Individual muscles sampled during needle EMG examination.

the lumbar paraspinal muscles when they could be reliably examined [1,9]. Patients were classified as having clear, possible, or no evidence of radiculopathy. For analytic purposes, a final EDX impression was given for each patient by dichotomizing patients as having evidence of radiculopathy or not based on the unmasked examiner's classification.

Masked review and validation

All recordings were independently reviewed by 2 masked examiners whose assessment of the insertional and resting EMG activity for the individual muscles tested as well as their overall EDX impression were recorded on a standardized examiner form (Figure 1). Masked examiner A is a licensed physical therapist who is recognized as a board-certified specialist in clinical electrophysiology by the American Board of Physical Therapy Specialties and has completed more than 35,000 EDX examinations over the past 30 years. Masked examiner B is a licensed physical therapist who is recognized as a board-certified specialist in clinical electrophysiology by the American Board of Physical Therapy Specialties and has completed more than 20,000 EDX examinations over the past 20 years.

The needle EMG recordings were de-identified, removing all patient-specific information, with only the gain and sweep speed settings visible along with the name of the individual muscle examined. The de-identified needle EMG recordings were edited in Windows Movie Maker software in order to generate case-specific files and label the individual muscles examined. Each masked examiner was provided an electronic copy of the 24 needle EMG recordings for analysis.

The masked examiners were instructed to complete the standardized examiner form by analyzing insertional and resting EMG activity for the individual muscles examined in each of the 24 recordings provided. They were informed that the individual muscles on the digital recording and on the standardized examiner form appeared in the same order. Each masked examiner was provided with the definition for the presence of radiculopathy mentioned earlier (see Electrodiagnostic Testing section) [9]. No specific instructions or guidance were provided to the masked examiners for the interpretation of insertional or resting EMG activity. The procedures used in this investigation did not follow any specific needle EMG examination protocol or evaluation technique such as lumbar paraspinal mapping [10,11].

On the standardized examiner form, insertional EMG activity was rated as decreased, increased, or normal if left blank. Resting EMG activity, which included evaluating for the presence of fibrillation potentials, positive waves, complex repetitive discharges, or other neuropathic findings, was rated as present or normal if left blank. The author chose a dichotomous scale for rating resting EMG activity as opposed to the commonly used graduated, semi-quantitative scale (ie, rating the relative number of fibrillation potentials and/or positive waves recorded as 0, +1, +2, +3, +4) to define the presence of resting EMG activity because the amount of abnormal EMG activity was not of primary concern, rather the existence and location of abnormal resting EMG activity in order to identify the presence of nerve root injury [4].

The following system was used for scoring the insertional and resting EMG activity of individual muscles examined: normal insertional and resting EMG activity=0; normal or increased insertional EMG activity with the presence of sustained abnormal resting EMG activity=1. Space was provided on the standardized examiner form for comments by the masked examiners.

Examiners provided an overall EDX impression for each patient including the involved nerve root(s) when a radiculopathy was deemed present. Patients were classified as having evidence of radiculopathy, possible evidence of radiculopathy, or no evidence of radiculopathy on the standardized examiner form. All patients were ultimately given a final EDX impression by dichotomizing them into those with evidence of radiculopathy and those with no evidence of radiculopathy. This was accomplished by combining patients with evidence of radiculopathy and possible evidence of radiculopathy into one group and comparing them to patients with no evidence of radiculopathy.

Since the masked examiners only had access to the insertional and resting EMG activity portions of the needle EMG examination, they were unable to comment on other EDX possibilities such as mononeuropathy, polyneuropathy, plexopathy, or myopathy.

Statistical analysis

PASW Statistics for Windows, Version 18.0 (SPSS Inc., Chicago, IL) was used to compute inter-rater reliability statistics by comparing examiners in a pairwise fashion (UnMask:MaskA, UnMask:MaskB, and MaskA:MaskB). Cohen's κ statistic was calculated for insertional and resting EMG activity of individual muscles examined as well as the final EDX impression. For the overall EDX impression, because the categories are ordered, a linear weighted kappa (κw) statistic was calculated (http://www.vassarstats.net/kappa.html) [16]. This was done because patients categorized as having possible evidence of

Figure 1: Standardized examiner form.

radiculopathy are more closely related to patients categorized as having clear evidence or no evidence of radiculopathy than either of those categories relate to one another [14,16]. Strength of agreement was based on the following scale of κ values: <0=Poor agreement; 0.01-0.20=Slight agreement; 0.21-0.40=Fair agreement; 0.41-0.60=Moderate agreement; 0.61-0.80=Substantial agreement; 0.81-1.00=Almost perfect agreement [14].

EDX sensitivity and specificity values were calculated comparing the unmasked examiner to the masked examiners in a pairwise fashion (UnMask:MaskA and UnMask:MaskB). The unmasked examiner's final EDX impression – which included knowledge of the patient's history, physical examination, and complete EDX test results – served as the gold standard for all calculations. None of the patients in this investigation were found to have comorbid conditions (eg, mononeuropathy, generalized polyneuropathy, lumbosacral plexopathy, or myopathy) based on their complete EDX testing results, which included peripheral nerve conduction including F waves and analysis of volitional needle EMG findings.

A secondary analysis was performed in which the overall and final EDX impressions were determined based on the raw assessment of insertional and resting EMG activity provided by the masked examiners on the standardized examiner form. The secondary analysis compares the overall EDX impression provided by the masked examiners to a forced classification of patients based strictly on the ratings of insertional and resting EMG activity provided by the masked examiners. The purpose of the secondary analysis was to determine if the definition of radiculopathy provided to the masked examiners prior to the study was consistently followed.

Preliminary power analysis revealed that 24 needle EMG recordings would provide 90% power to detect substantial agreement (κ>0.60)

between examiners using a one-tailed test of statistical significance at an alpha level of 0.05 assuming the null is κ=0 [14].

Results

Analysis of insertional and resting EMG activity

Agreement among examiners for insertional and resting EMG activity for individual muscles combined was substantial (κ ≥0.68, 95% CI: 0.50 to 0.89; P ≤ .001) across all pairwise comparisons. The level of agreement for individual muscles examined ranged from fair (κ=0.33, 95% CI: -0.25 to 1.0; P>.05) to perfect (κ=1.0, 95% CI: 1.0 to 1.0; P≤.001) across all pairwise examiner comparisons with the biceps femoris short-head and medial gastrocneumius muscles having the lowest levels of agreement and the vastus medialis muscle having the highest level of agreement (Table 4). A summary of each examiners raw EMG assessment of insertional and resting EMG activity and the overall EDX impression are found in Table 5.

Analysis of the final and overall electrodiagnostic impressions

The level of agreement among the electromyographers for the final EDX impression ranged from moderate to substantial. Agreement between the unmasked examiner and masked examiner A was substantial with a κ value of 0.75 (95% CI: 0.48 to 1.0; P<0.0001). The level of agreement between the unmasked examiner and masked examiner B was moderate with a κ value of 0.53 (95% CI: 0.24 to 0.81; P=0.002). The level of agreement between masked examiner A and masked examiner B was moderate with a κ value of 0.43 (95% CI: 0.11 to 0.76; P=0.010) (Table 6).

The raw level of agreement among the electromyographers for the overall EDX impression ranged from fair to substantial. Agreement between the unmasked examiner and masked examiner A was

Muscle	Unmasked examiner *vs.* Masked examiner A	Unmasked examiner *vs.* Masked examiner B	Masked examiner A *vs.* Masked examiner B
Paraspinals	0.62† (0.16,1.0)	0.62† (0.16,1.0)	1.0† (1.0,1.0)
Anterior tibialis	0.65† (0.02,1.0)	1.0† (1.0,1.0)	0.65† (0.02,1.0)
Medial gastrocnemius	0.78† (0.50,1.0)	0.78† (0.50,1.0)	0.50* (0.07,0.93)
Posterior tibialis	0.75† (0.43,1.0)	0.60† (0.21,0.99)	0.83† (0.51,1.0)
Vastus medialis	1.0† (1.0,1.0)	1.0† (1.0,1.0)	1.0† (1.0,1.0)
Biceps femoris short-head	0.51* (0.06,0.95)	0.70† (0.32,1.0)	0.33 (-0.25,0.91)
All muscles combined‡	0.72† (0.57,0.87)	0.74† (0.59,0.89)	0.68† (0.50,0.86)

*P<0.05
†P≤0.001
‡Includes lateral gastrocnemius and extensor hallucis longus which had too few cases to analyze individually.

Table 4: Cohen's kappa values (95% CI) for insertional and resting EMG activity of individual muscles tested.

Muscle	Unmasked examiner	Masked examiner A	Masked examiner B
Lumbar paraspinals	0=19, 1=4	0=21, 1=2	0=21, 1=2
Anterior tibialis	0=22, 1=2	0=23, 1=1	0=22, 1=2
Medial gastrocnemius	0=17, 1=7	0=19, 1=5	0=19, 1=5
Lateral gastrocnemius	0=3, 1=1	0=3, 1=1	0=3, 1=1
Posterior tibialis	0=18, 1=6	0=20, 1=4	0=21, 1=3
Extensor hallucis longus	0=2, 1=5	0=2, 1=5	0=4, 1=3
Vastus medialis	0=24, 1=0	0=24, 1=0	0=24, 1=0
Biceps femoris short-head	0=19, 1=5	0=22, 1=2	0=21, 1=3
No radiculopathy	10	11	16
Possible radiculopathy	2	5	7
Radiculopathy	12	8	1

Frequency of insertional and resting EMG activity and overall EDX impression for each examiner.
0, normal insertional and resting EMG activity; 1, and normal or increased insertional EMG activity with the presence of sustained abnormal resting EMG activity
EDX: electrodiagnostic

Table 5: Raw findings for insertional and resting EMG activity and overall EDX impression.

	Unmasked examiner *vs.* Masked examiner A	Unmasked examiner *vs.* Masked examiner B	Masked examiner A *vs.* Masked examiner B
Cohen's κ (95%CI)	0.75 (0.48,1.0)	0.53 (0.24,0.81)	0.43 (0.11,0.76)
One-sided P value	<.0001	.002	.01
EDX: electrodiagnostic; CI: confidence interval.			

Table 6: Inter-rater reliability of the final EDX impression.

	Unmasked examiner *vs.* Masked examiner A	Unmasked examiner *vs.* Masked examiner B	Masked examiner A *vs.* Masked examiner B
Weighted κ_w (95%CI)	0.62 (0.37,0.87)	0.31 (0.12,0.50)	0.32 (0.09,0.55)
One-sided P value	<.0001	.004	<.05
EDX: electrodiagnostic; κ_w=linear weighted kappa value			

Table7: Inter-rater reliability of the overall EDX impression.

substantial with a κw value of 0.62 (95% CI: 0.37 to 0.87; P<0.0001). The level of agreement between the unmasked examiner and masked examiner B was fair with a κw value of 0.31 (95% CI: 0.12 to 0.50; P=0.004). The level of agreement between masked examiner A and masked examiner B was fair with a κw value of 0.32 (95% CI: 0.09 to 0.55; P<.05) (Table 7).

The sensitivity and specificity values for the final EDX impression comparing the unmasked examiner and masked examiner A were 0.86 (95% CI: 0.56 to 0.97) and 0.90 (95% CI: 0.54 to 0.99), respectively. These values comparing the unmasked examiner and masked examiner B were 0.57 (95% CI: 0.30 to 0.81) and 1.0 (95% CI: 0.66 to 1.0), respectively.

A secondary analysis was performed in which the overall and final EDX impressions for each patient were categorized based on the raw assessment of insertional and resting EMG activity provided by the masked examiners. This was performed in order to classify patients strictly based upon the definition of radiculopathy provided to each masked examiner prior to beginning the study. The secondary analysis resulted in the level of agreement for the final EDX impression ranging from substantial to almost perfect across all pairwise examiner comparisons. The level of agreement between the unmasked examiner and masked examiner A improved from substantial to almost perfect with a κ value of 0.83 (95% CI: 0.60 to 1.0). The level of agreement between the unmasked examiner and masked examiner B and the level of agreement between masked examiner A and masked examiner B improved from moderate to substantial in both instances, each with κ values of 0.60 (95% CI: 0.29 to 0.91).

Additionally, the secondary analysis resulted in the level of agreement for the overall EDX impression being substantial for all pairwise examiner comparisons. Agreement between the unmasked examiner and masked examiner A remained substantial but the κw value improved to 0.70 (95% CI: 0.48 to 0.92). The level of agreement between the unmasked examiner and masked examiner B and the level of agreement between masked examiner A and masked examiner B improved from fair to substantial in both instances, with κw values of 0.63 (95% CI: 0.35 to 0.91) and 0.66 (95% CI: 0.44 to 0.88), respectively.

Discussion

The results of this investigation demonstrate that needle EMG findings can be reliably assessed among experienced physical therapist electromyographers in patients with sciatica referred to physical therapy. The inter-rater reliability values for assessing insertional and resting EMG activity were substantial for individual muscles examined indicating that needle EMG findings can be used reliably to evaluate for the presence of nerve damage in patients with sciatica. The level of agreement among examiners for the final EDX impression ranged from moderate to substantial – improving to substantial to almost perfect

when findings at rest were strictly classified – supporting the use of needle EMG in patients with sciatica as a diagnostic test.

Overall, the assessment of insertional and resting EMG activity for individual muscles examined was substantial; however, patterns emerged among examiners which may be indicative of individual clinical preferences for analyzing and recording the results of the needle EMG examination. First, the unmasked examiner consistently rated increased insertional EMG activity in conjunction with the presence of abnormal resting EMG activity and rated very few muscles as having decreased insertional EMG activity. Second, masked examiner A rated several muscles as having decreased insertional EMG activity, including rating some muscles with fibrillation potentials and/or positive waves as having both increased and decreased insertional EMG activity (which is plausible in the presence of axonal loss and associated chronic soft tissue changes muscle atrophy). Finally, masked examiner B rated all muscles as having normal insertional EMG activity. These findings clearly indicate that individual examiners not only differ in their assessment of insertional and resting EMG activity but may also place varying degrees of emphasis on the importance of insertional and resting EMG activity in formulating their overall EDX impression.

In this study, an experienced unmasked electromyographer was in moderate to substantial agreement with 2 experienced masked electromyographers on the final EDX impression in patients with sciatica referred to physical therapy. The level of agreement found in this investigation was not as high as that achieved by the electromyographers in the study by Chouteau et al. [12] employing a similar study design. This may be explained by the fact that the electromyographers in this investigation are geographically separate from one another and have never practiced together, which makes it more likely that they perform and analyze needle EMG examinations in distinctly different ways. While this fact may limit the internal validity of our study, it makes this investigation more pragmatic and improves the generalizability of our findings for electromyographers in clinical practice.

Although the levels of agreement in this study did not reach those of Chouteau et al. [12], it is worth noting that the majority of disagreement in this investigation occurred across a subset of 5 patients that were judged to have radiculopathy in 4/5 cases by the unmasked examiner, judged to have radiculopathy in 5/5 cases by masked examiner A, and judged to have radiculopathy in 0/5 cases by masked examiner B. Outside of that subset of patients, disagreement on the final EDX impression among examiners was found in only 3 other cases. The level of agreement on the final EDX impression between the unmasked examiner and examiner A was found to be substantial with a κ value of 0.75. A less robust level of agreement was found between the unmasked examiner and masked examiner B, as well as between masked examiner A and masked examiner B, with moderate κ values of 0.53 and 0.43, respectively. These values are likely clinically meaningful given the percentage agreement between the unmasked examiner and

masked examiner A was 88% (21/24 cases), 75% (18/24 cases) between the unmasked examiner and masked examiner B, and 71% (17/24 cases) between masked examiner A and masked examiner B [17]. These values are higher than those achieved in the study by Kendall and Werner [13]. Which employed a slightly different research design and data analytic approach than that used in this investigation.

A secondary analysis was performed which classified patients based strictly on analysis of insertional and resting EMG activity as recorded by the masked examiners and following the definition of radiculopathy provided prior to beginning the study. The secondary analysis resulted in significant improvements in the level of agreement among examiners in both the final and overall EDX impressions. This may be explained by the fact that the masked examiners are clinicians who routinely consider a patient's complete history, clinical examination, and EDX testing results in practice when determining if an abnormality such as lumbosacral radiculopathy is present. In this investigation, almost every instances of disagreement on the overall EDX impression involved the masked examiners categorizing observed insertional and resting EMG abnormalities as indicative of a possible radiculopathy, as opposed to presenting clear evidence of radiculopathy. In other words, the level of confidence the masked examiners had for declaring the presence of radiculopathy appeared to be insufficient based upon their assessment of insertional and resting EMG activity alone; this despite the fact that the observed abnormalities ultimately fit the strict definition provided defining the presence of radiculopathy.

The sensitivity and specificity values calculated for this study are consistent with published reports which demonstrate that needle EMG tends to be more specific than sensitive [9]. Specificity was measured to be ≥90% across all pairwise examiner comparisons, ranging from 90% to 100%. Clinically, this makes needle EMG a more reliable EDX test for ruling-in a radiculopathy in the presence of abnormal findings than for ruling-out a radiculopathy in the absence of findings. This is significant in terms of the larger prognostic study because it improves the likelihood that patients were properly classified based on the results of their needle EMG examination. In the larger prognostic study, 19 of 38 (50.0%) patients were classified as having evidence of radiculopathy, a percentage that is consistent with previous research [12,18-20] therefore, the likelihood that patients were misclassified based on incidental, false-positive EMG findings is unlikely given the demonstrated diagnostic specificity in this study.

The case can be made that the findings in this investigation are both pragmatic and generalizable to the clinical setting for a couple of reasons. First, while all examiners are practicing electromyographers, they are geographically separate and have never practiced together. Second, patients included in this study underwent EDX testing in 1 of 8 different physical therapy clinics with diverse environmental factors impacting the fidelity of the EMG recordings in several instances, a fact which was noted by the masked examiners. Despite efforts by the unmasked examiner to correct or minimize the impact of these environmental factors, at times it was difficult to obtain a clean electrical baseline for analyzing insertional, resting, and volitional EMG activity. Obtaining good electrical fidelity for the performance and interpretation of EDX testing is a challenge routinely encountered by electromyographers in clinical practice. The presence of such factors in this investigation strengthens the generalizability of the results. Third, because nearly all EDX testing was performed either prior to or immediately following a scheduled physical therapy treatment session, the constraints of time (as in clinical practice) may have impacted the quality of EMG recordings produced. Comments from the masked examiners noted the rapid nature of needle insertions at times impacted their ability to

properly analyze insertional and resting EMG activity. Despite these challenges, none of which are foreign to clinical practice, an acceptable level of inter-rater reliability was found among experienced physical therapist electromyographers for analyzing needle EMG findings in patients with sciatica.

Conclusions

The results of needle EMG testing in patients with sciatica referred to physical therapy can be reliably assessed by experienced physical therapist electromyographers. This was a more pragmatic study than previously published investigations and the findings can be generalized to electromyographers in clinical practice. The results of this investigation support the use of masking in EDX research to validate the use of needle EMG as a diagnostic test.

Acknowledgements

References

1. Daube JR, Rubin DI (2009) Needle electromyography. Muscle Nerve 39: 244-270.

2. Dumitru D (2001) Electrodiagnostic Medicine: Hanley & Belfus.

3. Katirji B (2007) Electromyography in Clinical Practice: A Case Study Approach (2 ed.): Mosby.

4. Kimura J (2001) Electrodiagnosis in Diseases of Nerve and Muscle: Principles and Practice: Oxford University Press, USA.

5. Preston DC, Shapiro B (2005) Electromyography and Neuromuscular Disorders: Clinical-Electrophysiologic Correlations: Butterworth-Heinemann.

6. Lee JH, Lee SH (2012) Physical examination, magnetic resonance image, and electrodiagnostic study in patients with lumbosacral disc herniation or spinal stenosis. J Rehabil Med 44: 845-850.

7. Coster S, de Bruijn SF, Tavy DL (2010) Diagnostic value of history, physical examination and needle electromyography in diagnosing lumbosacral radiculopathy. J Neurol 257: 332-337.

8. Cho SC, Ferrante MA, Levin KH, Harmon RL, So YT (2010) Utility of electrodiagnostic testing in evaluating patients with lumbosacral radiculopathy: An evidence-based review. Muscle Nerve 42: 276-282.

9. Dillingham TR (2002) Electrodiagnostic approach to patients with suspected radiculopathy. Phys Med Rehabil Clin N Am 13: 567-588.

10. Haig AJ, Yamakawa K, Kendall R, Miner J, Parres CM, et al. (2006) Assessment of the validity of masking in electrodiagnostic research. Am J Phys Med Rehabil 85: 475-481.

11. Haig AJ, Tong HC, Yamakawa KS, Quint DJ, Hoff JT, et al. (2005) The sensitivity and specificity of electrodiagnostic testing for the clinical syndrome of lumbar spinal stenosis. Spine (Phila Pa 1976) 30: 2667-2676.

12. Chouteau WL, Annaswamy TM, Bierner SM, Elliott AC, Figueroa I (2010) Interrater reliability of needle electromyographic findings in lumbar radiculopathy. Am J Phys Med Rehabil 89: 561-569.

13. Kendall R, Werner RA (2006) Interrater reliability of the needle examination in lumbosacral radiculopathy. Muscle Nerve 34: 238-241.

14. Sim J, Wright CC (2005) The kappa statistic in reliability studies: use, interpretation, and sample size requirements. Phys Ther 85: 257-268.

15. Fritz JM, Thackeray A, Childs JD, Brennan GP (2010) A randomized clinical trial of the effectiveness of mechanical traction for sub-groups of patients with low back pain: study methods and rationale. BMC Musculoskelet Disord 11: 81.

16. Lowry R (2013) Kappa as a Measure of Concordance in Categorical Sorting. Retrieved January 12, 2013

17. McHugh ML (2012) Interrater reliability: the kappa statistic. Biochem Med (Zagreb) 22: 276-282.

18. Annaswamy TM, Bierner SM, Chouteau W, Elliott AC (2012) Needle

electromyography predicts outcome after lumbar epidural steroid injection. Muscle Nerve 45: 346-355.

19. Fish DE, Shirazi EP, Pham Q (2008) The use of electromyography to predict functional outcome following transforaminal epidural spinal injections for lumbar radiculopathy. J Pain 9: 64-70.

20. Lauder TD, Dillingham TR, Andary M, Kumar S, Pezzin LE, et al. (2000) Effect of history and exam in predicting electrodiagnostic outcome among patients with suspected lumbosacral radiculopathy. Am J Phys Med Rehabil 79: 60-68.