SHORT REPORT

ABSTRACT: Low back pain and lumbar radiculopathy are among the most common painful disorders affecting the adult population. This study hypothesizes that there is good correlation between the diagnostic impression of an unblinded electromyographer, using clinical and electromyographic information, and an independent electromyographer, who uses the needle examination only to assess for lumbar radiculopathy. This is a prospective, single-blinded, observational pilot study. The needle examination was electronically recorded, reproduced, and shown to a second examiner, blinded to all clinical data. Diagnostic impressions from both examiners were recorded and evaluated for agreement. Six recorded cases were reviewed by 66 blinded examiners. Overall diagnostic agreement was 46.9% (60.5% faculty level, 28.5% resident level). Logistic regression shows a strong association between training level and agreement on diagnostic impression (odds ratio, 1.9; 95% confidence interval, 1.12–3.22; $P = 0.019$). This study shows that there is fair interrater reliability between faculty-level examiners and poor reliability among resident-level examiners when the needle examination is used to evaluate patients with lumbar radiculopathy.

Muscle Nerve **34:** 238–241, 2006

INTERRATER RELIABILITY OF THE NEEDLE EXAMINATION IN LUMBOSACRAL RADICULOPATHY

RICHARD KENDALL, DO,[1] and ROBERT A. WERNER, MD[2]

[1] Division of Physical Medicine and Rehabilitation, University of Utah Orthopaedics Center, 590 Wakara Way, Salt Lake City, Utah 84108, USA
[2] Department of Physical Medicine and Rehabilitation, University of Michigan, Ann Arbor, Michigan, USA

Accepted 28 February 2006

Over the past 50 years the electrodiagnostic examination has been used in the evaluation of low back pain, but there has not been a single blinded evaluation of the needle examination. An independent (blinded) evaluation of a diagnostic test is a requirement to meet the Class I standards in an evidence-based medicine model. Failure to have an independent evaluation of a diagnostic test would result in the evidence being classified as Class IV (expert opinion).

A review of the literature using the key words electromyography, radiculopathy, reliability, and sensitivity, revealed no blinded electrodiagnostic studies in patients with lumbosacral radiculopathy. Good interrater reliability and validity of a quantified paraspinal mapping technique[2] has been shown, but not in the diagnosis of radiculopathy. The sensitivity of electromyography (EMG) has been reported to range from 49% to 86%[4–9] in subjects with possible to definite radiculopathy, based upon history and physical examination as the gold standard. In other evaluations, EMG findings have demonstrated good correlation with computed tomography, myelography, and surgical findings, with a sensitivity of 80%–100%.[3,4] These studies may have demonstrated the clinical utility of EMG in the diagnosis of lumbar radiculopathy, but they do not meet Class I or II scientific standards for diagnostic tests, because they lack an independent, blinded evaluation. Any unblinded evaluation has the potential for bias, with the history and physical findings influencing interpretation of the electrodiagnostic findings.

There is consensus that the needle examination is the most subjective component of the EMG, yet it is also the most sensitive component of the electrodiagnostic consultation for lumbar radiculopathy. In order to confirm the sensitivity, specificity, and reliability of the needle examination as a diagnostic test, potential biases must be controlled by use of a blinded evaluation. We examined these issues in the present study. In particular, we hypothesized that there is good correlation between the diagnostic impression of an unblinded electromyographer, us-

Abbreviations: EMG, electromyogram; MUAP, motor unit action potential
Key words: electromyography; radiculopathy; reliability; sensitivity
Correspondence to: R. W. Kendall; e-mail: richard.kendall@hsc.utah.edu

Published online 11 April 2006 in Wiley InterScience (www.interscience.wiley.com). DOI 10.1002/mus.20554

ing clinical and electromyographic information, and an independent electromyographer, using only the needle examination to assess for lumbar radiculopathy.

MATERIALS AND METHODS

This study was approved by our local institutional review boards. Patients referred to a university hospital electrodiagnostic laboratory for low back pain or leg pain underwent a standard physical and electrodiagnostic evaluation. The physical examination consisted of sensory testing (both light touch and pinprick with a safety pin), motor examination using the standard American Spinal Injury Association motor examination of key muscles in the lower limbs, determination of stretch reflexes (patellar, Achilles, and medial hamstring) in the lower limbs, and provocative testing (straight leg raise and femoral stretch signs). Clinical diagnostic criteria for a lumbar radiculopathy were based upon the current community standard of back pain and leg pain with paresthesias or numbness in a dermatomal distribution, with or without weakness in that limb.

Electrodiagnostic Examination. All electrodiagnostic examinations were performed using a Teca Synergy device (Oxford Instruments, Pleasantville, New York). The needle examination was digitally recorded through a professional video graphics array (VGA) to a National Television Standards Code (NTSC) converter (Startech Computer Products, Startech.com, Groveport, Ohio) to digital tape using a Sony Video Walkman (Sony Corp., Tokyo, Japan).

The electrodiagnostic examination consisted of lower-extremity motor and sensory nerve conduction studies and a standardized seven-muscle needle examination, including the lumbar paraspinals. A 50-mm disposable, concentric needle was used to examine the vastus medialis, tibialis anterior, extensor hallucis longus, medial gastrocnemius, tensor fascia lata, gluteus maximus, and low lumbar paraspinals. Spontaneous activity was assessed with a gain of 50 μV, sweep speed of 10 ms/division, and filter settings of 10 Hz to 10 kHz. Voluntary activity was assessed with a gain of 200–500 mV, sweep speed of 10 ms/division, and filter settings of 10 Hz to 10 kHz. The recorded needle examination was edited with Adobe Premiere 6.0 (Adobe Systems, Inc., San Jose, California) to remove extraneous time between muscles and affix labels to identify the muscle and gain. The examination was then saved in motion pictures expert group (mpeg-2) format on a compact disk for viewing by the independent, blinded examiners. The gold standard used for the diagnosis of radiculopathy by needle examination was abnormal insertional activity, spontaneous activity, or motor unit action potential (MUAP) changes in duration, phasicity, amplitude, or recruitment in two or more muscles from the same root level and different peripheral nerves on needle EMG as determined by the principal investigator (R.K.), with normal findings in root levels above and below the affected root.[10]

Subjects. A total of 31 examiners (21 faculty, 10 residents) independently reviewed one or more of six recorded cases for a total of 66 blinded evaluations. Each blinded examiner viewed the recorded needle examination, and was asked to score the spontaneous activity on a 0–4+ scale and MUAP amplitude, duration, phasicity, and recruitment as normal, increased, or decreased. After viewing the entire recorded examination, the examiners were then asked to select a final electrodiagnostic impression from a list of nine choices: normal study; L2, L3, L4, L5, or S1 radiculopathy; lumbar polyradiculopathy; lumbar plexopathy; and peripheral neuropathy. Each examiner was also asked to rate their level of confidence in establishing the diagnosis on a 10-point scale from 1 (unsure) to 10 (very confident).

Statistical Analysis. Statistical analysis was done using SPSS software (SPSS, Inc., Chicago, Illinois). Percent agreement between examiners' diagnostic impression was calculated for all cases and each individual case and then stratified by training level. Logistic regression was used to evaluate an association between training level and agreement of diagnosis. A simple *t*-test was done to evaluate each examiner's confidence of diagnosis, stratifying for correct vs. incorrect responses. The percent agreement on the presence or absence of abnormal spontaneous activity was calculated on an individual muscle basis, as well as on complete agreement within each case. Complete agreement was defined as concordant when all assessments of the spontaneous activity for a given case were interpreted the same among blinded and unblinded examiners. The overall sensitivity of EMG and specificity were also calculated for the presence or absence of a radiculopathy as well as for the specific level of radicular involvement.

RESULTS

Diagnosis. The overall level of agreement between the original (unblinded) examiner and the blinded

Table 1. Summary of percent agreement with clinical gold standard stratified by case and level of training.

Case	Faculty agreement (%)	Resident agreement (%)	All examiners agreement (%)	Mean certainty	Clinical diagnosis
All	60.5	28.5	46.9	5.9	
1	50.0	0.0	33.0	3.0	Normal subject
2	60.0	28.5	41.6	4.0	Lumbar polyradiculopathy
3	50.0	40.0	41.6	6.3	Normal subject
4	62.5	20.0	46.1	6.9	L5 radiculopathy
5	66.6	20.0	45.4	5.2	L5 radiculopathy
6	60.0	0.0	60.0	7.3	Normal subject

evaluators was 46.9% when determining the exact level of the radiculopathy, being higher for faculty-level than resident-level examiners (Table 1). The percent agreement on the final diagnosis among faculty and resident-level examiners is also shown in Table 1. Logistic regression showed faculty-level examiners were twice as likely to agree on the final diagnosis as resident-level examiners (odds ratio, 1.9; 95% confidence interval, 1.12–3.22; $P = 0.019$). The examiners who chose the correct diagnosis were more confident in their diagnostic decision than those who chose the incorrect diagnosis (mean 7.2 certainty with correct diagnosis vs. 4.8 certainty with incorrect diagnosis, $P = 0.0004$). However, regarding level of agreement for identifying the presence of a radiculopathy at any level, sensitivity was 79.4% (27 of 34) and specificity 44.4% (12 of 27) among all examiners, with residents having values of 81.2% and 12.5%, and faculty of 77.7% and 57.8%, respectively.

Spontaneous Activity. Agreement on the presence or absence of abnormal spontaneous activity (positive sharp waves and fibrillation potentials) was 53% (57.5% faculty, 47% residents) when all muscles in the screen were taken into account. On an individual muscle basis, agreement on spontaneous activity ranged from 62% in the paraspinals to 92% in the vastus medialis overall, with faculty examiners ranging from 63% agreement in the paraspinals to 97% agreement in the vastus medialis, and residents from 57% in the tensor fascia lata to 96% in the medial gastrocnemius. There was good correlation (Spearman's rho = 0.655, $P < 0.01$) among blinded and unblinded electromyographers regarding the presence or absence of spontaneous activity in individual muscles. Among the 66 blinded evaluations, 33.3% reached the correct diagnosis and accurately recorded the spontaneous activity, 13.6% reached the correct diagnosis but did not agree on the spontaneous activity, 19.7% had an incorrect diagnosis but accurately noted the spontaneous activity, and 33.3% had an incorrect diagnosis and did not accurately assess the spontaneous activity.

DISCUSSION

We found a fair correlation among faculty-level examiners and poor correlation among resident-level examiners in a blinded evaluation of the needle examination. The overall 60% agreement on diagnosis is similar to the sensitivity of the needle examination in identifying lumbar radiculopathy published in prior studies using a clinical or surgical gold standard.[4–9] The sensitivity of the needle EMG improves dramatically to 79% when used to identify a patient with a radiculopathy, without specifying a particular root level.

Our study demonstrates that extensive training is necessary for electromyographers, not only in the recognition of abnormal spontaneous activity and in MUAP analysis, but in clinical application of these data. Studies performed by residents with only a few months of training are less reliable.

When looking at the data regarding abnormal spontaneous activity, we found that there was only 53% agreement on the presence or absence of abnormal spontaneous activity in all muscles studied during a standardized limb evaluation. There was good correlation between examiners on the recognition of abnormal spontaneous activity on a muscle-by-muscle basis, although recognition by residents was less robust. The lumbar paraspinal muscles had the worst agreement among all muscles across all cases (63% faculty, 61% resident, and 62% overall). This may have contributed to the low correlation of agreement on all muscles. The high correlation for recognition of abnormal spontaneous activity on a muscle-by-muscle basis and low agreement on all abnormal muscles can be explained in a number of

ways. First, one third (22 of 66) of the examiners interpreted all the data correctly and agreed on the diagnosis. An additional 9 examiners had the correct diagnosis, despite some disagreement on the presence or absence of abnormal spontaneous activity among the muscles tested, which means that they either disregarded some findings or considered them noncontributory because of the remaining data. Additionally, 13 blinded examiners incorrectly interpreted the abnormal spontaneous activity, used additional MUAP criteria for final diagnosis, or believed there was insufficient data for diagnosis and "guessed" at a diagnosis of either one-level radiculopathy or polyradiculopathy.

We believe that a seven-muscle screen was representative of the community standard for the needle examination, and Dillingham et al.[1] demonstrated that such a screen including paraspinal muscles had recognition rates of 99%–100% of identifiable lumbar radiculopathies. Additionally, we believe that the absence of nerve conduction data (H-reflexes, F-waves, compound muscle action potentials, and sensory nerve action potentials) for the blinded examiners was not a significant factor in overall diagnosis as these tests are typically used to exclude other disorders rather than confirm a diagnosis of radiculopathy.[10]

Although our study represents the standard of care for an EMG study, it has some limitations. While performing the study, an examiner normally has the ability to study given muscles repeatedly or more thoroughly (greater than four quadrants) if abnormal findings are thought to exist. Again, a trigger and delay line are now available on modern machines, and their use would allow for more thorough evaluations of waveforms and recruitment than was possible in this study.

This study was presented in part at the annual scientific meeting of the American Association of Neuromuscular and Electrodiagnostic Medicine, November 2004, Savannah, Georgia. This study was supported by a grant from the American Association of Neuromuscular and Electrodiagnostic Medicine Foundation for Research and Education.

REFERENCES

1. Dillingham TR, Lauder TD, Andary M, Kumar S, Pezzin LE, Stephens RT, et al. Identifying lumbosacral radiculopathies an optimal electromyographic screen. Am J Phys Med Rehabil 2000;79:496–503.
2. Haig AJ, Talley C, Grobler LJ, LeBreck DB. Paraspinal mapping: quantified needle electromyography in lumbar radiculopathy. Muscle Nerve 1993;16:477–484.
3. Haldeman S, Shouka M, Robboy S. Computed tomography, electrodiagnostic and clinical findings in chronic workers' compensation patients with back and leg pain. Spine 1988; 13:345–350.
4. Khatri BO, Baruah J, McQuillen MP. Correlation of electromyography with computed tomography in evaluation of low back pain. Arch Neurol 1984;41:594–597.
5. Kuruoglu R, Oh SJ, Thompson B. Clinical and electromyographic correlations of lumbosacral radiculopathy. Muscle Nerve 1994;17:250–251.
6. Nardin RA, Patel MR, Gudas TF, Rutkove SB, Raynor EM. Electromyography and magnetic resonance imaging in the evaluation of radiculopathy. Muscle Nerve 1999;22:151–155.
7. Schoedinger GR III. Correlation of standard diagnostic studies with surgically proven lumbar disc rupture. South Med J 1987;80:44–46.
8. Tonzola RF, Ackil AA, Shahani BT, Young RR. Usefulness of electrophysiological studies in the diagnosis of lumbosacral root disease. Ann Neurol 1981;9:305–308.
9. Weber F, Albert U. Electrodiagnostic examination of lumbosacral radiculopathies. Electromyogr Clin Neurophysiol 2000; 40:231–236.
10. Wilbourn AJ, Aminoff MJ. AAEM Minimonograph 32: The electrodiagnostic examination in patients with radiculopathies. Muscle Nerve 1998;21:1612–1631.