ABSTRACT: Motor unit number estimation (MUNE) is an important electrophysiological technique for quantitative measurement of motor neuron loss. Although commonly used, there is no consensus concerning the optimal procedure for statistical MUNE, particularly regarding several operator-dependent variables. To assess the variables, we analyzed 500 sequential, submaximal compound muscle action potential (CMAP) responses at three or four stimulus intensities in 10 controls and 10 patients with amyotrophic lateral sclerosis (ALS). In both controls and ALS patients, we found that posttest filtering data based on 20% or 25% windows or 2, 2.5, or 3 SD excludes <5% of data. Windows of 10% or 15% excluded <5% of data in controls but not in ALS patients. Excluding data based on ±2 SD, the coefficient of variation for final MUNE was 12% in controls and 6% in ALS patients. Group sizes of 30 or 50 and sample sizes of 300 to 500 sequential CMAP responses per run yielded the lowest coefficient of variation. We propose that statistical MUNE data should be analyzed based on excluding data >2 SD from the mean, because this is operator independent, includes the majority of data, effectively excludes clearly outlying data, such as fasciculations or movement artifact, and has a reasonable coefficient of variation.

Muscle Nerve **29:** 700–706, 2004

PROPOSED MODIFICATION TO DATA ANALYSIS FOR STATISTICAL MOTOR UNIT NUMBER ESTIMATE

TIMOTHY M. MILLER, MD, PhD,[1] ANDREAS M. KOGELNIK, MD, PhD,[2] and RICHARD K. OLNEY, MD[1]

[1] Department of Neurology, University of California San Francisco, San Francisco, California, USA
[2] Department of Medicine, Stanford University, Palo Alto, California, USA

Accepted 8 January 2004

Motor unit number estimation (MUNE) is an important electrophysiological technique for quantitative measurement of motor neuron loss in patients with amyotrophic lateral sclerosis (ALS) and other diseases that affect the lower motor neurons (reviewed elsewhere[3,7,8,10]). Changes in MUNE may be useful for prognosis,[8] provide an estimate of motor axon loss, and shed light on mechanisms of disease progression or stabilization by distinguishing between axonal loss and collateral sprouting of remaining axons. Recently, MUNE results were applied to therapeutic trials in patients with ALS.[4] As a quantitative measure of axonal loss, MUNE may suggest disease stabilization secondary to a therapeutic intervention. This type of application of MUNE demands a high degree of reproducibility in the technique. In addition, for more widespread application, the performance of MUNE by different investigators should be standardized.

The MUNE is determined by dividing the maximum area of the compound muscle action potential (CMAP) by the area of the average surface-recorded motor unit (SMUP) for a given muscle. Several methods have been developed for determining MUNE.[3,4,7–9] Statistical MUNE is one of the two most commonly used techniques, because of its reproducibility and ease of application. In performing statistical MUNE, the SMUP is derived by measuring the variance of multiple submaximal CMAP responses, which generally follow a Poisson distribution. Based on the formula for a Poisson distribution, the variance is related to the mean SMUP. Typically, an SMUP is determined for three to four different stimulus levels. In recording data for statistical MUNE, the currently available software (Viking; Nicolet Biomedical Inc., Madison, WI) asks the investigator for a "window" level, that is, a range of percent maximum CMAP area into which the responses are expected to fall. The program will then reject any data that fall outside that range. The concept of windows

Abbreviations: ALS, amyotrophic lateral sclerosis; CMAP, compound muscle action potential; MUNE, motor unit number estimate; SMUP, surface motor unit potential

Key words: ALS; motor unit; motor unit number estimate; MUNE; statistical

Correspondence to: T. M. Miller, University of California San Diego, Ludwig Institute for Cancer Research, CMM-East 3072, 9500 Gilman Drive, La Jolla, CA 92093-0670, USA; e-mail: timiller@ucsd.edu

Published online 24 March 2004 in Wiley InterScience (www.interscience.wiley.com). DOI 10.1002/mus.20028

was introduced in the second beta version to exclude clearly outlying data, such as fasciculations that interfered with reproducible MUNE results during the Regeneron CNTF study (personal experience, R. K. O.). In normal patients, a 10% window appears to be a reasonable solution to the issue.[6] However, in ALS patients, the generally accepted technique is to expand the window size if >50% of the data fall outside the 10% window. The >50% determination is made by the investigator's estimation, because the current program does not quantify the number of data points rejected. In addition, the degree to which the windows should be expanded has not been standardized. Thus, a consensus has not been reached concerning the optimal procedure for statistical MUNE, particularly regarding several operator-dependent variables such as methods for setting and expanding the recording window size, which clearly affect the results, especially in mildly weak patients.

To address the issue of how best to analyze statistical MUNE data, we have used the MUNE-500 program,[5] a 2002 beta version of statistical MUNE that permits recording of responses to 500 consecutive stimuli for posttest analysis. We have analyzed statistical MUNE data from 10 mild to moderately weak ALS patients and 10 healthy volunteers. The ALS patients were selected for mild to moderate weakness because these patients are the most difficult in whom to perform statistical MUNE.[6] We have calculated SMUP and MUNE values for patients and controls based on (1) excluding outlying data by setting various window sizes or excluding data based on multiples of standard deviation from the mean; (2) varying the group size used for analysis; and (3) varying the sample size of data points for each run. We find that data from healthy volunteers are relatively unaffected by these various permutations, unlike data from ALS patients. Based on this analysis, we offer suggestions for how the performance of statistical MUNE may be improved in an unbiased manner.

METHODS

Subjects. Ten healthy subjects with an average age of 42 years (range, 22–72 years) and 10 ALS patients with an average age of 57 years (range, 42–78 years) were studied. All subjects gave written informed consent for the studies, which were approved by our Institutional Review Board. The patients had probable, laboratory-supported probable, or definite ALS according to the criteria of the World Federation of Neurology.[2] All controls had normal strength and sensation. Strength of abductor digiti minimi in patients ranged from 4 to 5 on the Medical Research Council scale. Ulnar neuropathy was excluded by screening ulnar sensory and motor nerve conduction studies, including across-elbow studies.

Electrophysiological Techniques. All studies were performed on the same abductor digiti minimi muscle for any given subject, using the MUNE-500 program[5] on the Viking IV system (Nicolet Biomedical Inc.). Surface recordings of maximal and submaximal ulnar CMAPs were obtained by stimulating with taped electrodes 6–7 cm proximal to the active recording electrode. The fifth digit was taped to the rest of the hand to minimize movement. All 10 controls and 7 of 10 ALS patients were available for repeat testing. Test–retest studies were performed from 10 min to 3 weeks after the initial recording, except in two control patients in whom repeat testing was performed 6 weeks later. If studies were performed on the same day, the electrodes and markings were removed and replaced to simulate retesting on a different day, which is how the technique is most often applied. There were no changes in strength by history or examination upon repeat testing. To minimize baseline fluctuations and artifacts, such as from movement, patients were asked to recline and relax on a bed with eyes closed and without speaking. In early pilot trials, this approach was markedly more effective at reducing noise than were attempts to restrain the subject's arm and hand. In some cases, the stimulus intensity level was minimally adjusted during the run to offset a drift in the responses. No subject needed to end the test prematurely because of discomfort. The test required 30–45 minutes to complete.

For each MUNE, 500 consecutive, submaximal CMAP responses were recorded at four stimulation levels. The midpoint of responses for the four stimulation levels chosen was approximately 15%, 30%, 45%, and 60%. Stimulation rate was 2 Hz. Occasionally in a patient with ALS, only three stimulation levels were recorded, because a gap ≥10% occurred at one stimulus intensity. All 500 responses at any given stimulus level were collected and analyzed. If any gaps >10% were identified by the initial scan, 50 to 100 responses were recorded in that range to verify the absence of motor units. As is customary, gaps >10% were counted as one toward the MUNE, and percent range was subtracted from the maximum CMAP before calculating final MUNE values.[1] In addition, when gaps ≥10% were present, stimulus intensity was set to record responses for analysis that were on one side or the other of the gap. Gaps

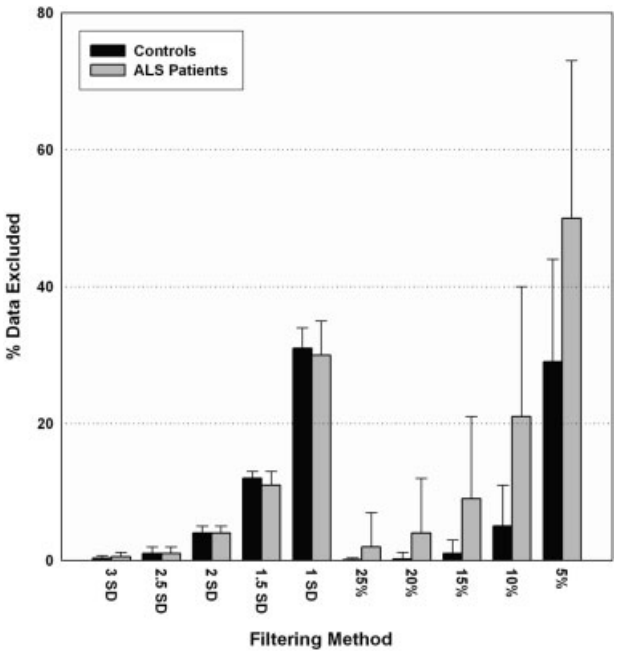

FIGURE 1. Percentage of data excluded by different methods of filtering. The percentage of data excluded by various methods of filtering was determined from 3 to 4 runs of 500 data points for 10 ALS patients and 10 controls. (Error bars represent standard deviations.)

occurred in only 3 of 10 ALS patients, as might have been expected, because the ALS patients were selected to have only mild to moderate weakness.

The runs of 500 data points were divided into 10 groups of 50 for analysis (Fig. 1; Table 1). As detailed in Results, the group size and number of data points varied in subsequent analyses. Each group was analyzed based on excluding data based on 5%, 10%, 15%, 20%, and 25% windows, or ± 1, 1.5, 2, 2.5, or 3 standard deviations (SD). For each data set, SMUP was calculated according to the following equation: mean SMUP = variance/(mean CMAP area − minimum CMAP area). Variance = sum $(x_n - x_{mean})^2/n - 1$, where x is the individually recorded CMAP areas, x_{mean} is the mean submaximal CMAP area, and n is the number of CMAP responses. For each stimulus level, SMUP was determined by taking the average of the values for the 10 groups of 50.

Final SMUP values were based on either an average or a weighted average. For the weighted average, the mean SMUP amplitude for each stimulus level was multiplied by the number of motor units estimated at each level, the products added, and the result divided by the sum of the motor units determined at each level.[10] For the simple average, final MUNE values were calculated by dividing the maximum CMAP area by the simple arithmetic average SMUP.

RESULTS

To begin to understand the effects of different methods of excluding outlying data collected for statistical MUNE, we first analyzed the average percentage of data excluded for each MUNE determination consisting of three or four runs of 500 data points for control subjects and ALS patients. The parameters used to exclude data are indicated in Methods. For control subjects, 80 runs from 20 independent trials were analyzed. For ALS patients, 59 runs from 17 independent trials were analyzed. The results are presented in Figure 1. For control subjects, ± 3 SD, 2.5 SD, or 2 SD and 10% or larger windows excluded on average <5% of the data. This is not surprising given the normal range of submaximal CMAP response of 5–10% in control subjects. In ALS patients, ± 3 SD, 2.5 SD, or 2 SD and 20% or 25% windows excluded, on average <5% of the data (Fig. 1), although there was a wide variation, with even 25% windows excluding as much as 21% of the data in some runs (data not shown). The increased exclusion of data by smaller windows sizes in ALS patients is expected given the wider range of submaximal CMAP responses of 5 to 45%. Thus, in both ALS patients and control subjects, ± 3 SD, 2.5 SD, or 2 SD

Table 1. Percentage difference in final MUNE values without filtering and with different methods of filtering.*

Filter	3 SD	2.5 SD	2 SD	1.5 SD	1 SD	5% Window	10% Window	15% Window	20% Window	25% Window
Control subjects										
Mean ± SD	2 ± 2	2 ± 2	3 ± 2	11 ± 4	37 ± 5	8 ± 5	1.5 ± 2	1.5 ± 2	0.5 ± 1	0.2 ± 0.4
Range	0 to 8	0 to 7	1 to 8	9 to 21	26 to 44	2 to 17	0 to 5	0 to 5	0 to 3	0 to 1
ALS Patients										
Mean ± SD	5 ± 9	6 ± 7	11 ± 14	19 ± 16	35 ± 19	30 ± 18	14 ± 18	10 ± 9	6 ± 8	8 ± 11
Range	0 to 30	0 to 21	1 to 47	1 to 61	1 to 78	6 to 53	0 to 55	2 to 29	0 to 26	0 to 30

**Each method of filtering increases the MUNE value by the percentage indicated. In 10 ALS patients and 10 controls, final MUNE values based on various methods of filtering were compared (percentage difference) to final MUNE values determined without any filtering.*

or 20% or 25% filtering is optimal to exclude a minimum amount of data.

We next addressed whether excluding data based on the various parameters affects final MUNE values. The MUNE values were calculated based on a nonweighted, average SMUP area for each trial. The values were compared with the MUNE calculated based on analyzing all 500 responses. The results are presented in Table 1. For control subjects, analyzing all data, or based on ± 3 SD, 2.5 SD, or 2 SD or ≥10% windows had relatively little effect on final MUNE value. Windows of 1.5 SD, 1 SD, or 5% resulted in an 11%, 37%, or 8% difference, respectively. In contrast, in ALS patients, all methods of excluding data resulted in greater than 5% change in final MUNE values, with a wide range of differences from 0% to 21% or greater in all groups.

Given the lack of a gold standard against which to compare these various values, it is not clear which set of values is appropriate for the ALS patients. However, the results and previous experience in studying ALS patients do provide information on which technique to choose for excluding data. Based on operator independence, the ability to exclude clearly outlying data such as fasciculations or movement artifact, and inclusion of the majority of the data, we decided to analyze further windows of ±2 SD.

Our initial analysis was based on dividing 500 data points into 10 groups of 50 for analysis, but the optimum group number and the number of data points needed per run are unknown. Given that the reproducibility of the technique is of prime importance in analyzing different methods of data analysis,

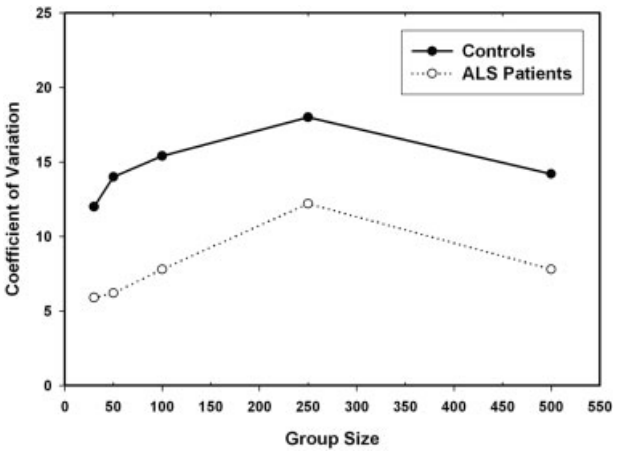

FIGURE 2. Effect of group size on coefficient of variation. Using different group sizes to divide the data in each run before analysis, coefficients of variation were determined for two separate MUNEs for 7 ALS patients and 10 controls. Filter used was ± 2 SD.

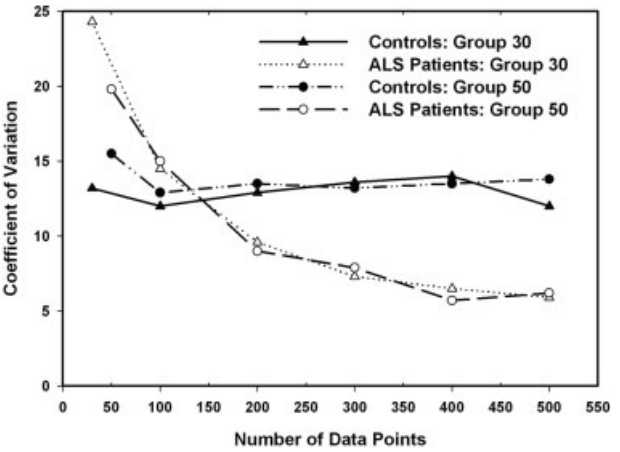

FIGURE 3. Effect of number of data points on coefficient of variation. Using different sample sizes for each run, coefficients of variation were determined for two separate MUNEs for 7 ALS patients and 10 controls. Group sizes of 30 or 50 and a filter of ± 2 SD were used for these calculations.

we analyzed the variables with respect to the coefficient of variation. Seven of 10 ALS patients and all 10 controls were available for repeat testing, and those trials were used to calculate the coefficient of variation. We first analyzed all 500 data points per run with group sizes of 30, 50, 100, 250, and 500. In both ALS patients and controls (Fig. 2), there was a trend toward decreased coefficients of variation with smaller group sizes of 30 or 50. Although analysis of all the data points at one time, that is, a group of 500, yielded a reasonable coefficient of variation in both ALS patients and controls, this necessitates including data for analysis even in cases of known error, for example, drift in the data. Group sizes of 30 and 50 were therefore selected for further analysis.

Figure 3 shows the effect of limiting the number of data points per run on the coefficient of variation for final MUNE values, using group sizes of both 30 and 50. It is surprising that, in controls, there is little effect on coefficient of variation even when using only 30 or 50 data points. For ALS patients, increasing the number of data points per run yields a lower coefficient of variation for final MUNE values, although sample sizes of 400 and 500 were only marginally better than 300 data points.

In Table 2, we summarize the results of applying the optimized settings to our data set. Using 500 data points per run, divided into groups of 30 and filtered using ± 2 SD, we calculated both weighted and nonweighted SMUP amplitudes, MUNE values, and coefficients of variation. The MUNE values ranged from 74 to 137 (average value, 101) for control subjects, and from 10 to 102 for ALS patients for the nonweighted MUNE calculation; coefficients of vari-

Table 2. Comparison of weighted and nonweighted MUNE calculation with a 2-SD filter.*

	Nonweighted			Weighted		
	SMUP (μV)	MUNE	Coefficient of variation	SMUP (μV)	MUNE	Coefficient of variation
Control Subjects						
Mean ± SD	119 ± 46	101 ± 19	12 ± 9	117 ± 46	103 ± 19	14 ± 11
Range	68 to 218	74 to 137	0 to 25	67 to 219	74 to 138	0 to 38
ALS Patients						
Mean ± SD	173 ± 108	52 ± 34	6 ± 7	173 ± 108	55 ± 35	6 ± 6
Range	68 to 426	10 to 102	2 to 20	68 to 426	10 to 105	1 to 20

**For 10 ALS patients and 10 controls, SMUP size and MUNE were determined. Coefficients of variation were based on 7 ALS patients and 10 controls. Calculations were based on ± 2 SD filter, group size of 30, and sample size of 500 data points per run.*

ation were 12 and 6, respectively. The difference between weighted and nonweighted MUNE calculations was small, averaging 2% (range, 0 to 8%) for controls and 6% (range 0 to 16%) for ALS patients. Final values were nearly identical using a group size of 50 instead of 30.

The current statistical MUNE program stops collecting further data for a given run when the standard error of the mean (SEM) of SMUPs is less than 10%. We tested the approach in our data set (Table 3). Using group sizes of either 30 or 50, we sequentially calculated SMUP values for each group. Each SMUP value for a group was added to the previous SMUP values, and the standard error of the mean was calculated. When the SEM was <10%, no further groups were analyzed and the mean SMUP was calculated for the run. We used a minimum of four groups to insure adequate sample size. Because we collected 500 data points per run, the maximum number of groups was 10 for group sizes of 50 and 16 for group sizes of 30. Using group sizes of 50, we found that, on average, six groups were needed before the SEM reached less than 10%. For a group size of 30, seven and nine groups were needed in controls and ALS patients, respectively, to reach a standard error of less than 10%. Compared with using all 500 data points in each run (Table 2), SMUP and MUNE values were similar, whereas coefficients of variation were slightly increased.

DISCUSSION

We have analyzed the effects of various methods of excluding data collected for statistical MUNE. In control subjects, we find relatively little data excluded and only small differences in final MUNE values using ± 2 SD or 10%, 15%, 20%, 25% windows. In contrast, in ALS patients, 10% and 15% windows excluded on average 21% and 9% of data, respectively, with 50% to 70% of data being excluded in some runs; 20% or 25% windows included most of the data in the majority of runs. The use of ± 2 SD excluded about 4% of data in both ALS patients and control subjects. The data are consistent with the findings of Henderson and colleagues,[5] who recently demonstrated that SMUP sizes varied little in control patients when data was excluded based on 10% windows or ± 2 SD. Our study differs from that of Henderson and colleagues in that we examined the effects of different analysis windows not

Table 3. Results of stopping data collection after standard error <10%.*

Group size		Number of groups	SMUP (μV)	MUNE	Coefficient of variation
Control subjects					
50	Mean ± SD	6 ± 2	114 ± 48	107 ± 24	14 ± 12
30	Mean ± SD	7 ± 4	117 ± 46	104 ± 24	15 ± 11
ALS Patients					
50	Mean ± SD	6 ± 3	171 ± 100	58 ± 31	9 ± 9
30	Mean ± SD	9 ± 5	173 ± 101	57 ± 31	9 ± 8

**Using group sizes of either 30 or 50, SMUP values for each group were calculated sequentially. Each SMUP value for a group was added to the previous SMUP values and the SEM was calculated. When the SEM was <10%, no further groups were analyzed and the mean SMUP was calculated for the run. A minimum of four groups was used to insure adequate sample size. The average number of groups used and the resultant SMUP and MUNE were based on 10 ALS patients and 10 controls. Coefficients of variation are based on 7 ALS patients and 10 controls.*

only on SMUP size but also on MUNE and in that we propose a new approach for performing statistical MUNE.

From the current data and previous experience with studying ALS patients, we conclude that limits of 2 SD may be the most appropriate way to exclude outlying data, based on the following. First, there is a general consensus that some type of filter is needed. With the initial beta version of the statistical MUNE program, investigators found that data collection in ALS patients was significantly hampered by fasciculations or artifacts, such as movement during data acquisition. Windows were introduced to exclude these responses. Second, there is no consensus as to how far to increase size when >50% of data falls outside the typically set 10% window and there is no easy, reproducible way to recognize when "too much" data has been excluded by the 10% window. Setting windows based on ± 2 SD eliminates this operator-dependent variable while still excluding outlying data. Third, using limits of 2 SD, by definition, includes the vast majority of the responses.

We have tested our hypothesis that ± 2 SD is a reasonable approach to analyzing MUNE data in both ALS patients and normal subjects. The calculated MUNE and SMUP are consistent with previously published data for the statistical MUNE. In addition, the coefficient of variation of 6% in ALS patients and 12% in controls falls within the range of reproducibility previously established for this technique and suggests that this will be a robust technique to follow ALS longitudinally. Our proposed approach differs from the current program for the sequencing of tasks and for objective method of selecting data for inclusion in the analysis. With the current program, the operator first sets recording windows through which data are selected for analysis, and the program then calculates mean SMUP size based on the variance of the included data. In contrast, we propose, first, that all responses are recorded and, second, that statistical analysis sets windows at ± 2 SD for inclusion of data and then calculates mean SMUP size at that stimulus level.

For this group of data, we do not find significant differences in MUNE or coefficient of variation based on weighted versus nonweighted SMUP values. This may be specific to our patient group or may be secondary to collecting the data at four set stimulus levels. In the report by Shefner and colleagues[11] that demonstrated improved reliability with weighted SMUP values, the stimulus levels had been determined by the program rather than examining four predetermined ranges.

We have found that using group sizes of 30 or 50 yields the lowest coefficient of variation. In addition to providing greater reliability, analyzing the data based on this smaller group size may dampen some sources of noise in the data such as small drift in the CMAP responses during recording.[5] Our data suggest that using 300 or more sequential responses yields the lowest coefficient of variation in ALS patients. Recording all 300 responses at one time, rather than dividing them into groups of 30 during the recording, would have several practical advantages, including allowing the operator to more easily recognize changes in the baseline or drift in the data during the recording, and would save some time in data collection. We have also demonstrated the reliability of the current approach of stopping data collection when the SEM for SMUP values for individual groups within a run reaches less than 10%. Overall, SMUP and MUNE values were similar. The coefficients of variation were slightly higher (1–2%) using this method rather than calculating coefficients of variation based on all 500 data points. On average, the number of groups needed to reach SEM <10% corresponded to 200 to 300 data points, similar to the results of the optimal number of data points based on Figure 3. The advantage of using this approach is that, for some patients, testing times would be shortened. We propose analyzing statistical MUNE by excluding data that is beyond 2 SD, because this eliminates the need for the operator to determine window values and includes the majority of the data while still excluding clearly outlying data, such as fasciculations or movement artifact. We have demonstrated that this method is reliable in control subjects and ALS patients and likely applies to other diseases associated with lower motor neuron dysfunction.

Presented at the joint meeting of the American Association of Electrodiagnostic Medicine and International Congress of Clinical Neurophysiology, September 2003, San Francisco, California. The authors thank Dr. C. Lomen-Hoerth for helpful comments regarding the study and manuscript.

REFERENCES

1. Bromberg M. Consensus: motor unit number estimation. Clin Neurophysiol Suppl 2003;55:335–338.
2. Brooks BR, Miller RG, Swash M, Munsat TL. El Escorial revisited: revised criteria for the diagnosis of amyotrophic lateral sclerosis. Amyotroph Lateral Scler Other Motor Neuron Disord 2000;1:293–299.
3. Daube JR. Estimating the number of motor units in a muscle. J Clin Neurophysiol 1995;12:585–594.
4. Gooch CL, Harati Y. Motor unit number estimation, ALS and clinical trials. Amyotroph Lateral Scler Other Motor Neuron Disord 2000;1:71–82.

5. Henderson RD, McClelland R, Daube JR. Effect of changing data collection parameters on statistical motor unit number estimates. Muscle Nerve 2003;27:320–331.
6. Lomen-Hoerth C, Olney RK. Effect of recording window and stimulation variables on the statistical technique of motor unit number estimation. Muscle Nerve 2001;24:1659–1664.
7. Olney RK. Motor unit number estimation (MUNE): pro. Amyotroph Lateral Scler Other Motor Neuron Disord 2002;3 Suppl 1:S91–S92.
8. Olney RK, Lomen-Hoerth C. Motor unit number estimation (MUNE): how may it contribute to the diagnosis of ALS? Amyotroph Lateral Scler Other Motor Neuron Disord 2000;1 Suppl 2:S41–S44.
9. Olney RK, Yuen EC, Engstrom JW. Statistical motor unit number estimation: reproducibility and sources of error in patients with amyotrophic lateral sclerosis. Muscle Nerve 2000;23:193–197.
10. Shefner JM, Gooch CL. Motor unit number estimation in neurologic disease. Adv Neurol 2002;88:33–52.
11. Shefner JM, Jillapalli D, Bradshaw DY. Reducing intersubject variability in motor unit number estimation. Muscle Nerve 1999;22:1457–1460.