Clinical Neurophysiology 120 (2009) 947–952

Contents lists available at ScienceDirect

Clinical Neurophysiology

journal homepage: www.elsevier.com/locate/clinph

Inter-rater reliability of motor unit number estimates and quantitative motor unit analysis in the tibialis anterior muscle

S.G. Boe [a,*], B.H. Dalton [a], B. Harwood [a], T.J. Doherty [a,b], C.L. Rice [a,c]

[a] Canadian Centre for Activity and Aging, School of Kinesiology, The University of Western Ontario, Ont., Canada
[b] Department of Clinical Neurological Sciences, Schulich School of Medicine and Dentistry, The University of Western Ontario, Ont., Canada
[c] Department of Anatomy and Cell Biology, The University of Western Ontario, Ont., Canada

ARTICLE INFO

Article history:
Accepted 24 February 2009
Available online 16 April 2009

Keywords:
Electromyography
Motor unit
Motor unit number estimation
Quantitative EMG
Reliability

ABSTRACT

Objective: To establish the inter-rater reliability of decomposition-based quantitative electromyography (DQEMG) derived motor unit number estimates (MUNEs) and quantitative motor unit (MU) analysis.
Methods: Using DQEMG, two examiners independently obtained a sample of needle and surface-detected motor unit potentials (MUPs) from the tibialis anterior muscle from 10 subjects. Coupled with a maximal M wave, surface-detected MUPs were used to derive a MUNE for each subject and each examiner. Additionally, size-related parameters of the individual MUs were obtained following quantitative MUP analysis.
Results: Test–retest MUNE values were similar with high reliability observed between examiners (ICC = 0.87). Additionally, MUNE variability from test–retest as quantified by a 95% confidence interval was relatively low (±28 MUs). Lastly, quantitative data pertaining to MU size, complexity and firing rate were similar between examiners.
Conclusion: MUNEs and quantitative MU data can be obtained with high reliability by two independent examiners using DQEMG.
Significance: Establishing the inter-rater reliability of MUNEs and quantitative MU analysis using DQEMG is central to the clinical applicability of the technique. In addition to assessing response to treatments over time, multiple clinicians may be involved in the longitudinal assessment of the MU pool of individuals with disorders of the central or peripheral nervous system.

1. Introduction

The ability to reliably obtain data from serial studies both within (intra-rater) and across (inter-rater) examiners is a fundamental attribute of any clinical test or diagnostic tool. The rationale underlying this premise is that given a certain degree of reliability, changes observed from test–retest may be attributed to some process, for example progression of disease, other than the variability associated with the test or diagnostic tool employed. One tool, decomposition-based quantitative electromyography (DQEMG), has been developed as a means of performing a quantitative electrodiagnostic evaluation that provides clinically useful information pertaining to the health of the neuromuscular system. This information includes data relating to the physiological characteristics and the estimated number of motor units (MU) in a given muscle or muscle group (Doherty and Stashuk, 2003; Boe et al., 2004; Boe et al., 2007). Previously, this technique has provided data representative of pathophysiological changes occurring at the level of the lower motor neuron in patients with amyotrophic lateral sclerosis (Boe et al., 2007) and Charcot-Marie-Tooth disease (Shy et al., 2007), in addition to demonstrating MU remodeling and an associated decrease in MU number in old and very old men (McNeil et al., 2005a). Furthermore, when applied longitudinally, the results of DQEMG analysis may accurately portray the natural history of a given disease process, in addition to evaluating the effectiveness of potential treatments. With these goals in mind, it is important that the results obtained are reliable from test–retest both within and across examiners, as patients undergoing re-assessment may not be examined by the same clinician.

Although the number of studies and subsequently the extent of the literature relating to DQEMG is expanding, few have evaluated reliability (Boe et al., 2004, 2006; Calder et al., 2008) and none has tested inter-rater reliability comprehensively. Previous work in this laboratory found a high degree of intra-rater reliability of DQEMG from a sample of healthy subjects for motor unit number

* Corresponding author. Address: Rm. 11016, University Centre, Toronto Rehabilitation Institute, 550 University Avenue, Toronto, Ont., Canada M5G 2A2. Tel.: +1 416 597 3422x7881; fax: +1 416 597 3031.
E-mail address: boe.shaun@torontorehab.on.ca (S.G. Boe).

1388-2457/$36.00
doi:10.1016/j.clinph.2009.02.168

estimates (MUNE) derived for the thenar muscle group (Boe et al., 2004). In an extension of this work, we further established the intra-rater reliability of both MUNEs and quantitative MU analysis obtained using DQEMG from the biceps-brachii and first dorsal interosseous muscles (Boe et al., 2006). In this study, the variability associated with the test–retest MUNE values was analyzed in an effort to ascertain the range of values expected on retest. This in turn would allow clinicians to differentiate between methodological variability and the effect of either an underlying pathological process or its treatment when using MUNE as an outcome measure in longitudinal studies.

One recent study reported high inter-rater reliability of the data analysis component of DQEMG methodology when applied to MUs of the extensor carpi radialis muscle (Calder et al., 2008). However, due to the importance of establishing the reliability of DQEMG as a clinical tool, coupled with the lack of studies comprehensively examining the inter-rater reliability of both the data collection and analysis components of the technique, further studies are needed. Thus, the purpose here was to investigate the within-subject (test–retest) reliability of quantitative MU analysis and MUNEs in the tibialis anterior (TA) muscle obtained using DQEMG performed by two independent examiners.

2. Methods

2.1. Subjects

Ten healthy male subjects aged 20–30 years (25 ± 3 years) recruited from the university environment, volunteered to participate in the study. All subjects provided written, informed consent and our institutional ethics review board approved the study.

2.2. Inter-rater testing

The initial test and subsequent retest was performed by two different examiners on the same day for all subjects with a minimum of 1 hour separating the data collection sessions. Both examiners were new users of DQEMG who were trained with regard to data collection and analysis by an individual experienced with the technique. Testing order was pseudo-randomized between the examiners, in that each performed the initial test in five out of the 10 subjects. The location of the active surface electrode for the initial test was marked on the skin by the first examiner (outlined below) and, following re-measurement to ensure accuracy, this position was utilized by the second examiner for placement of the active electrode. Data analyses were performed independently by each examiner following the completion of data collection for an individual subject. Lastly, examiners remained blind to the results of the opposing examiner until the completion of data collection and analysis for all subjects.

2.3. Torque measurement

The protocol used to measure the torque output of the dorsiflexors has been previously reported (McNeil et al., 2005a,b). Subjects were seated in a custom-made force dynamometer with their right ankle positioned in 30° of plantar flexion, and an angle of 90° at both the hip and knee joints. A C-clamp positioned ~ 6 cm proximal to the right knee joint provided a compressive force on the distal thigh to minimize hip flexion during the dorsiflexion contractions. Velcro straps across the toes and the dorsum of the foot secured the limb to the dynamometer footplate. Torque data were converted to digital format by a 12-bit A/D converter (CED model 1401 Plus, Cambridge Electronic Design, Cambridge, UK) and sampled on-line at 500 Hz using a commercially available software package (Spike 2 v. 4.13; Cambridge Electronic Design, Cambridge, UK). Although we have controlled the torque output using a customized dynamometer in this "idealized" setting, the use of the root mean square of the surface EMG signal can provide a similar degree of feedback and control of contraction intensity in a clinical setting in the absence of a dynamometer (Boe et al., 2008).

2.4. Electromyographic data collection

Intramuscular signals were obtained with a commercially available 25 mm, 30 gauge, disposable, concentric needle electrode (Model N53153; Teca Corp., Hawthorne, NY) using a bandpass setting of 10 Hz to 10 kHz and DQEMG software on a Neuroscan Comperio (Neuroscan Medical Systems, El Paso, TX)(Doherty and Stashuk, 2003; Boe et al., 2004). Using self-adhering electrodes (Kendall-LTP, Chicopee, MA) cut into strips (1 × 3 cm), surface signals were obtained using a bandpass setting of 5 Hz to 5 kHz. The active electrode was placed 7 cm distal to the tibial tuberosity and 2 cm lateral to the anterior tibial border with the reference electrode placed over the distal tendon of the TA just proximal to the ankle joint line. A full-sized electrode (2 × 3 cm) placed on the patella served as a ground.

2.5. Experimental protocol

With the subject secured in the dynamometer, a maximum M wave was elicited via supramaximal stimulation of the common peroneal nerve, posterior to the fibular head. The stimulation intensity was increased incrementally until the size of the M wave (based on negative-peak amplitude) reached a plateau. To ensure a supramaximal response the stimulation intensity was then increased a further 15%. Markers indicating negative onset, negative peak, negative-peak duration, and positive peak of the M wave were automatically positioned and following a visual check of the markers (and manual adjustments if required), the negative-peak amplitude of the M wave was automatically calculated and stored.

Next, subjects performed a series of 3, 4-s isometric maximal voluntary contractions (MVC) of the dorsiflexors, with each effort separated by 3 min of rest. Subjects were aided in these maximal contractions by visual feedback in the form of their torque output displayed on a computer monitor placed in front of the subject and via strong verbal encouragement from the examiner. The peak torque of these maximal contractions was determined and a number corresponding to 25% of this value established. A target line equivalent to this 25% value was positioned on the computer monitor, allowing for subsequent, sub-maximal contractions to be performed at a torque output consistent with this value, ensuring that all subjects were voluntarily contracting at a similar percentage of their MVC (see results for force data). With the 25% target line established, the concentric needle electrode was inserted into the TA muscle 5–10 mm proximal to the active surface electrode. Subjects were then asked to minimally contract the muscle isometrically in order to allow the examiner to locate an optimal needle position for data acquisition. This was achieved by adjusting the position of the needle to minimize the rise time of the MUPs of the first 2–3 recruited MUs. With the needle maintained manually in a stable position by the examiner, the subject was then instructed to increase the contraction force to the desired percent of MVC. If upon reaching this 25% target line the signal was of poor quality based on visual inspection, the subject was asked to relax, the needle was repositioned and the process repeated until adequate signal quality was obtained. Subjects were instructed to maintain stable contractions at the target intensity (25%) for 30 s

of data collection. Contractions were performed until a minimum of 20 MUP trains were obtained from superficial, intermediate and deep needle detection sites collectively. Rest periods of at least 1 min were provided between contractions; during this time, the operator repositioned the needle in order to ensure sampling from different MUs. If more than three contractions were required to obtain the minimum number of MUPs, a second needle insertion at a different site was undertaken.

2.6. Data reduction and analysis

The DQEMG methodology and associated algorithms have been described previously (Stashuk, 1999, 2001; Doherty and Stashuk, 2003). Briefly, DQEMG decomposes the composite EMG signal detected via a needle electrode into its constituent MUP trains using shape and temporal information related to the individual MUP discharges in addition to MU firing time statistics. Using these needle-detected MUPs as triggers for spike-triggered averaging, a component of DQEMG, decomposition-based spike-triggered averaging, provides a sample of MUPs detected via surface electrodes (S-MUPs) which are representative of the sizes of the MUs in the underlying muscle of interest. Parameters associated with the size of these needle (peak–peak voltage, duration, area–amplitude ratio (AAR)) and surface-detected (negative-peak amplitude) MUPs in turn provide information about the physiological characteristics of the underlying MU pool (Boe et al., 2004, 2005, 2007). A statistically significant sample of these S-MUPs ($\geqslant 20$) is then averaged to determine the mean S-MUP size. Using the same electrodes employed to detect the surface EMG signal, a maximum M wave is obtained, and a MUNE is derived by dividing a size-related parameter of the mean S-MUP (i.e., negative-peak amplitude) into the corresponding size parameter of the maximum M wave.

Following needle-detected signal decomposition and analysis, the MUP trains and needle and surface-detected MUPs were reviewed with regard to their acceptability based on criteria outlined by previous reports (Doherty and Stashuk, 2003; Boe et al., 2004, 2005). Briefly, MUP trains accepted for further analysis included: (1) a consistent firing rate versus time plot, based on visual inspection; (2) a physiological firing rate quantified by a Gaussian-shaped main peak of the interdischarge interval histogram with an associated coefficient of variation of less than 0.3 (Fuglevand et al., 1993; Stashuk, 1999); and (3) a minimum of 51 detected potentials. The needle and surface-detected MUP waveforms were then visually checked to ensure that the onset, end and peak markers were accurate (and repositioned manually if necessary). Lastly, the S-MUP waveform onset was required to occur within 10 ms of the needle-detected MUP waveform onset to be included in the analysis. Motor unit potential trains and needle and surface-detected MUPs that failed to meet all the inclusion criteria were excluded from further analysis.

2.7. Statistics

Prior to statistical analyses all data were tested for normality and those that were outside the normal distribution were analyzed using non-parametric statistics. Thus, to investigate test–retest differences between individual subjects, either a standard pairwise *t* test or the Wilcoxon ranked sums test was employed (SPSS V. 16.0, Chicago, IL). An a priori alpha level of $p < 0.10$ was utilized to denote significance in an effort to provide a more conservative evaluation of reliability, as our intent was to demonstrate similarity (non-significance) between the variables of interest. Additionally, relative test–retest reliability of the MUNE values and quantitative MUP parameters were assessed using a two-way random, single measure intra-class correlation coefficient (ICC, GraphPad Prism 4; GraphPad Software v. 4.02, San Diego, CA). Lastly, using methods employed in previous studies of MUNE reliability (Bromberg, 1993; Simmons et al., 2001; Boe et al., 2006), the mean percent difference was calculated (absolute difference between the test and retest MUNE values divided by their mean) in addition to determining 95% confidence intervals for the predicted true score (see Eq. (1)), of each individual subject's MUNE using the standard error of measurement (Pedhazur and Pedhazur Schmelkin, 1991; Boe et al., 2006). This method, which incorporates the variability of the sample population using the resulting correlation coefficient and observed test values, is useful in quantifying the amount of error associated with the methodology when interpreting an individual subjects 'true' test performance.

$$T = (1 - r_{xx})M + r_{xx}X \quad (1)$$

where *T*, predicted true score; r_{xx}, correlation coefficient of the observed test–retest values; *M*, mean of the observed test values; and *X*, subjects observed test value

3. Results

3.1. Torque data

Maximal voluntary contraction values were similar between the test and retest ($z = 0.46$, $p > 0.10$), with values (mean ± standard deviation) of 43.5 ± 8.2 (test) and 43.1 ± 6.8 Nm (retest). Throughout the sub-maximal, 30-s isometric contractions (target torque of 25% MVC), subjects were able to generate a similar torque output between tests ($z = 0.46$, $p > 0.10$), producing an average of 24.5 ± 0.8 and 24.6 ± 0.6% of their MVC or, in absolute values, 10.7 ± 2.1 and 10.6 ± 1.8 Nm for the test and retest, respectively.

3.2. Needle-detected MUP size, complexity and firing rate

Following data reduction and analysis, an average of 26 ± 3 (range = 20–30) and 27 ± 4 (range = 22–33) MUs were sampled for each individual subject for the test and retest, respectively, from an average of 4 ± 1 (range = 3–5; test) and 5 ± 2 (range = 3–8; retest) contractions. Needle-detected MUP size parameters were similar between tests, with no significant differences observed for needle-detected MUP AAR ($z = 0.15$, $p > 0.10$) and amplitude [peak–peak voltage ($z = 0.97$, $p > 0.10$); Table 1], however a significant difference was found for MUP duration [$t(9) = 2.69$, $p < 0.10$]. With regard to MUP complexity, the number of turns was similar ($z = 1.27$, $p > 0.10$) between the initial test and retest sessions, while a significant difference [$t(9) = 1.85$, $p < 0.10$] was detected for the number of phases (Table 1). Lastly, mean MU firing rates were similar between tests ($z = 0.46$, $p < 0.10$; Table 1). Individual subjects' needle-detected MUP parameters from test–retest are summarized in Table 1. Additional reliability analysis using the ICC revealed low-moderate reliability for the majority of the quantitative MUP parameters, including needle-detected MUP peak–peak voltage (0.24), duration (−0.10), AAR (0.40), number of turns (0.15) and phases (0.03) and lastly firing rate (0.65). For MUP duration, a negative value was produced due to the magnitude of the error variance, a finding consistent with a previous study examining reliability using DQEMG (Calder et al., 2008).

3.3. Surface-detected MUP size, maximum M wave and MUNE

Surface-detected MUP size based on negative-peak amplitude was similar between subjects for the test and retest ($z = 0.97$, $p > 0.10$; Table 2). Similar to S-MUP size, no significant differences were observed between subjects for the test and retest

Table 1
Quantitative needle-detected MUP parameter values for the test and retest.

Parameter	Peak–peak voltage (μV)		Duration (ms)		AAR		Number of Phases		Number of Turns		Firing Rate (Hz)	
Subject	Rater 1	Rater 2	Rater 1	Rater 2	Rater 1	Rater 2	Rater 1	Rater 2	Rater 1	Rater 2	Rater 1	Rater 2
1	923.8	594.1	7.3	7.9	1.5	1.3	2	3	3	4	12.3	11.5
2	1140.6	711.3	7.9	6.4	1.3	1.3	3	3	3	3	13.6	11.3
3	762.2	762.4	6.1	7.3	1.1	1.2	3	3	4	4	14.8	14.5
4	625.9	597.4	6.1	9.8	1.3	1.4	2	3	2	4	12.2	12.4
5	837.6	739.9	5.6	7.7	1.1	1.2	3	2	3	3	12.9	13.3
6	374.3	483.6	5.3	6.8	0.9	1.3	3	2	4	3	14.2	14.6
7	463.5	690.2	6.8	9.7	1.4	1.6	3	3	5	4	12.6	12.8
8	841.7	550.1	7.5	7.7	1.4	1.1	3	3	3	4	14.6	14.1
9	744.3	928.8	6.6	6.8	1.6	1.2	2	3	3	3	12.9	13.9
10	720.5	571.5	7.3	8.8	1.3	1.5	2	3	3	4	13.0	12.8
Mean	743.4	662.9	6.7	7.9	1.3	1.3	3	3	3	4	13.3	13.1
SD	220.6	129.9	0.8	1.2	0.2	0.2	0.4	0.3	0.6	0.4	0.9	1.2

Table 2
Test–retest values for the S-MUP, maximal M Wave and MUNE.

	Surface-detected MUP		M Wave		MUNE		
Parameter	Negative-peak amplitude (μV)		Negative-peak amplitude (mV)				
Subject	Rater 1	Rater 2	Rater 1	Rater 2	Rater 1	Rater 2	Predicted ± CI
1	52.5	52.1	7.5	6.8	214	184	207 ± 28
2	216.3	285.5	7.7	8.0	95	90	97 ± 28
3	131.0	105.5	6.5	6.7	71	93	74 ± 28
4	65.5	48.8	8.7	7.0	184	170	179 ± 28
5	107.7	92.4	5.7	6.4	68	98	72 ± 28
6	77.6	76.8	8.1	8.2	117	146	117 ± 28
7	87.9	68.2	5.9	5.6	79	123	82 ± 28
8	84.0	89.7	9.6	10.0	174	164	170 ± 28
9	71.6	63.1	3.8	4.0	70	89	74 ± 28
10	87.9	100.7	6.8	6.6	104	100	105 ± 28
Mean	98.2	98.3	7.0	6.9	118	126	118 ± 28
SD	46.9	68.6	1.7	1.6	54	37	50

for maximum M wave negative-peak amplitude [$t(9) = 0.46$, $p > 0.10$; Table 2]. Corresponding to the similarities observed for both S-MUP and M wave size, the negative-peak amplitude MUNE values obtained from the test and retest sessions were comparable [$t(9) = 1.08$, $p > 0.10$; Table 2]. Subsequent analyses to determine the relative reliability of individual subject's MUNE and S-MUP negative-peak amplitude values from test–retest revealed a high degree of reliability, based on respective ICC values of 0.87 and 0.90. Lastly, analyses to determine the range of methodological variability expected from test–retest revealed a 95% confidence interval about an individual subjects predicted true score (MUNE) of ±28 MUs (Table 2, Fig. 1).

4. Discussion

Consistent with, and extending the observations from a previous study examining the *intra*-rater reliability of DQEMG (Boe et al., 2006), the current results demonstrate that quantitative data pertaining to both the estimated number and physiological characteristics of MUs may be obtained reliably using DQEMG within individual subjects between two different examiners. The significance of this finding is that these results encompass the *inter*-rater reliability of both the data collection and analysis components of DQEMG, as opposed to data analysis alone (Calder et al., 2008).

In light of the biological variability observed previously for both MUNE and quantitative MU analysis (Boe et al., 2004, 2006, 2007), the ability to obtain reliable results within each individual subject (as opposed to across group means) is central to the clinical applicability of this technique. Of equal importance is the ability to reliably obtain data between different examiners using DQEMG.

As highlighted above, a considerable degree of between-subject variability has been observed in the MUNE values obtained for several muscles (Boe et al., 2004, 2006, 2007) including the TA in the present study. In light of the appreciably lower degree of within-subject variability observed from test–retest in this and a previous DQEMG MUNE study (Boe et al., 2006), it is likely that much of the between-subject variability is attributable to differences in MU number between individual subjects. This is supported by the large number of MUNE studies using a variety of techniques that have shown a level of between-subject variability that is similar to, or greater than that observed presently (McComas et al., 1971; Sica et al., 1974; Brown et al., 1988; Doherty and Brown, 1993; Shefner et al., 1999; Boe et al., 2004; McNeil et al., 2005a; Boe et al., 2007). Further, evidence of between-subject differences in the numbers of MUs has been demonstrated in human cadaveric studies that have examined the number of motor neurons in the ventral horn of the lumbar cord in addition to counting the number of efferent axons supplying various muscles in humans (Feinstein et al., 1955; Blevins, 1967; English and Blevins, 1969; de Carvalho, 1976; Kawamura et al., 1977a; Kawamura et al., 1977b; Carvalho et al., 1988). While limited in number, these cadaveric studies parallel the variability reported for MU numbers estimated using electromyography.

Similar to other quantitative EMG-based diagnostic tools, the results of DQEMG analysis exhibit a degree of variability from test–retest within individual subjects for both MUNE and the quantitative MU analysis; with MUNE quantified as the mean percent difference between test and retest in addition to the application of 95% confidence intervals (see methods for details)(Bromberg, 1993; Simmons et al., 2001; Boe et al., 2006). While the mean percent difference cal-

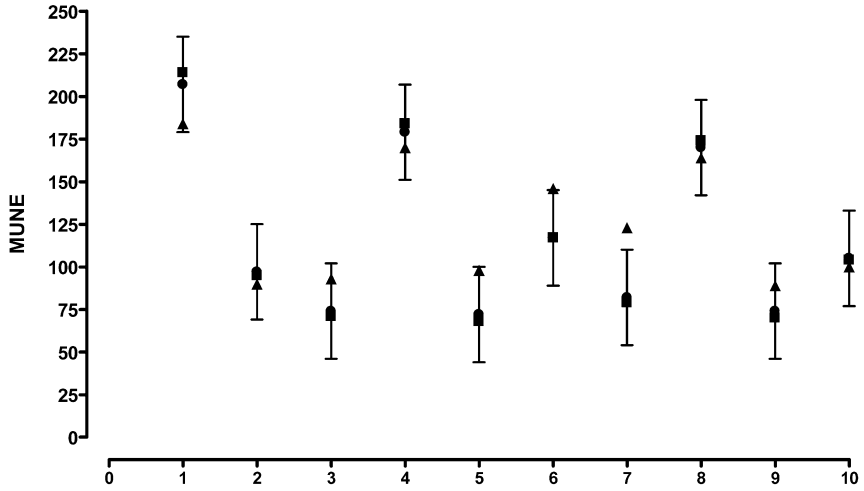

Fig. 1. Individual subjects test (square), retest (triangle) and predicted (circle) MUNE values. Bars represent the range of variability expected from test–retest of the 95% confidence interval about the predicted true score (circle) of ±28 MUs. With a single exception (subject seven), all subjects retest MUNE values were within the projected range.

culated in the present study (19.1%) is higher than those observed using DQEMG in the first dorsal interosseous (7.9%) and biceps-brachii (16.1%) muscles, it is comparable to those reported for other MUNE techniques (see Boe et al., 2006, Table 3 for review) (Bromberg, 1993; Felice, 1995; Shefner et al., 1999; Boe et al., 2006).

The 95% confidence interval calculated for the test–retest TA MUNE values and centered about each subject's predicted true score was ±28 MUs (Table 2), a value considerably lower than the ±41 MUs calculated previously in the first dorsal interosseous and biceps-brachii muscles (Boe et al., 2006). The significance of this confidence interval value corresponds to the interpretation of follow-up electrodiagnostic studies of individual subjects using MUNE as an outcome measure. Specifically, the ±28 MUs represent the magnitude of change in MU number that would have to occur in an individual subject in order to conclude (with 95% confidence) that the change was due to a process other than the variability associated with the technique itself (Table 2, Fig. 1). This 95% confidence interval (i.e., ±28 MUs) is related to each individual subject's predicted true score, not the range of MUNE values observed across the entire sample of subjects (Table 2). Consequently, the calculation of the predicted true score ± a 95% confidence interval may be especially useful in monitoring individual subjects at repeat assessments. With the exception of subject seven, whose retest MUNE value was outside of the expected range (i.e., retest MUNE value > predicted true score ±28 MUs), all other subjects had retest values within the projected range (Fig. 1).

With the exception of duration and the number of phases related to the needle-detected MUPs, all parameters associated with the quantitative MU analysis demonstrated no significant differences between test and retest (Tables 1 and 2). The inter-rater reliability reported in this study compares favorably to a previous reliability study which reported a mean difference percentage of 44.2% for biceps-brachii S-MUP amplitude (Bromberg, 1993), and is consistent with our previous work on the first dorsal interosseous (5.9%) and biceps-brachii (7.8%) muscles (Boe et al., 2006). An additional *intra*-rater study using DQEMG reported ICC values for a number of quantitative MUP parameters similar to those determined in the present *inter*-rater study, including S-MUP negative-peak amplitude (ICC values of 0.90 and 0.90; Calder et al., 2008 and present study, respectively) and needle-detected MUP peak–peak voltage (ICC values of 0.37 and 0.24; Calder et al., 2008 and present study, respectively). Consistent with this previous study (Calder et al., 2008), we also observed lower ICC values for several needle-detected MUP parameters including discharge rate (0.65), AAR (0.40) and duration (−0.10). These lower values may be attributed to the ICC calculation itself, as the ICC is influenced by the variance of the parameter in the population being assessed (Bartko, 1966; Shrout and Fleiss, 1979; Laschinger, 1992). For example, when applied to a variable that displays minimal variation across the sample, small differences among raters are magnified, thus resulting in a lower ICC value. An alternate interpretation of these results which should be considered is the possibility that these parameters cannot be obtained reliably within the context of this and the previous study (Calder et al., 2008).

In light of the results obtained in this study, there are some limitations which should be considered. First, this study had a relatively small number of subjects. With more subjects we would likely have observed a smaller confidence interval associated with the MUNE values. The reduction in confidence interval size would in-turn decrease the magnitude of change in the MUNE (from test–retest) required to infer that the change is due to a process rather than the variability associated with the technique. Second, this study did not examine patients with neuromuscular disorders; as such, future work should examine the inter-rater reliability of DQEMG in clinical populations in a manner similar to those utilized in the current study to ensure the clinical utility of the technique.

In examining the inter-rater reliability of DQEMG derived MUNEs and quantitative MU analyses, our results indicate that although there is a degree of methodological variability present, this clinically useful information can be obtained reliably from two different examiners. Furthermore, in keeping with a previous study (Boe et al., 2006), we have quantified this variability from a MUNE perspective in the form of a 95% confidence interval, thus providing the potential ability to identify changes in MU number that result from either an underlying disease process or its treatment.

References

Bartko JJ. The intraclass correlation coefficient as a measure of reliability. Psychol Rep 1966;19:3–11.

Blevins CE. Innervation patterns of the human stapedius muscle. Arch Otolaryngol 1967;86:136–42.

Boe SG, Stashuk DW, Doherty TJ. Motor unit number estimation by decomposition-enhanced spike-triggered averaging: control data, test–retest reliability, and contractile level effects. Muscle Nerve 2004;29:693–9.

Boe SG, Stashuk DW, Brown WF, Doherty TJ. Decomposition-based quantitative electromyography: effect of force on motor unit potentials and motor unit number estimates. Muscle Nerve 2005;31:365–73.

Boe SG, Stashuk DW, Doherty TJ. Within-subject reliability of motor unit number estimates and quantitative motor unit analysis in a distal and proximal upper limb muscle. Clin Neurophysiol 2006;117:596–603.

Boe SG, Stashuk DW, Doherty TJ. Motor unit number estimates and quantitative motor unit analysis in healthy subjects and patients with amyotrophic lateral sclerosis. Muscle Nerve 2007;36:62–70.

Boe SG, Rice CL, Doherty TJ. Estimating contraction level using root mean square amplitude in control subjects and patients with neuromuscular disorders. Arch Phys Med Rehabil 2008;89:711–8.

Bromberg MB. Motor unit estimation: reproducibility of the spike-triggered averaging technique in normal and ALS subjects. Muscle Nerve 1993;16:466–71.

Brown WF, Strong MJ, Snow R. Methods for estimating numbers of motor units in biceps-brachialis muscles and losses of motor units with aging. Muscle Nerve 1988;11:423–32.

Calder KM, Agnew MJ, Stashuk DW, McLean L. Reliability of quantitative EMG analysis of the extensor carpi radialis muscle. J Neurosci Methods 2008;168:483–93.

Carvalho VC, Cintra AI, Santo Neto H. Motor units and functional anatomy of the human musculus opponens digiti minimi. Anat Anz 1988;165:65–8.

de Carvalho VC. Study of motor units and arrangement of myons of human musculus plantaris. Acta Anat (Basel) 1976;96:444–8.

Doherty TJ, Brown WF. The estimated numbers and relative sizes of thenar motor units as selected by multiple point stimulation in young and older adults. Muscle Nerve 1993;16:355–66.

Doherty TJ, Stashuk DW. Decomposition-based quantitative electromyography: methods and initial normative data in five muscles. Muscle Nerve 2003;28:204–11.

English DT, Blevins CE. Motor units of laryngeal muscles. Arch Otolaryngol 1969;89:778–84.

Feinstein B, Lindegard E, Nyman E. Morphological studies of human motor units in normal human muscles. Acta Anat 1955;2:127.

Felice KJ. Thenar motor unit number estimates using the multiple point stimulation technique: reproducibility studies in ALS patients and normal subjects. Muscle Nerve 1995;18:1412–6.

Fuglevand AJ, Winter DA, Patla AE. Models of recruitment and rate coding organization in motor-unit pools. J Neurophysiol 1993;70:2470–88.

Kawamura Y, O'Brien P, Okazaki H, Dyck PJ. Lumbar motoneurons of man II: the number and diameter distribution of large- and intermediate-diameter cytons in motoneuron columns of spinal cord of man. J Neuropathol Exp Neurol 1977a;36:861–70.

Kawamura Y, Okazaki H, O'Brien PC, Dych PJ. Lumbar motoneurons of man: (I) number and diameter histogram of alpha and gamma axons of ventral root. J Neuropathol Exp Neurol 1977b;36:853–60.

Laschinger HK. Intraclass correlations as estimates of interrater reliability in nursing research. West J Nurs Res 1992;14:246–51.

McComas AJ, Fawcett PR, Campbell MJ, Sica RE. Electrophysiological estimation of the number of motor units within a human muscle. J Neurol Neurosurg Psychiatry 1971;34:121–31.

McNeil CJ, Doherty TJ, Stashuk DW, Rice CL. Motor unit number estimates in the tibialis anterior muscle of young, old, and very old men. Muscle Nerve 2005a;31:461–7.

McNeil CJ, Doherty TJ, Stashuk DW, Rice CL. The effect of contraction intensity on motor unit number estimates of the tibialis anterior. Clin Neurophysiol 2005b;116:1342–7.

Pedhazur E, Pedhazur Schmelkin L. Measurement design and analysis: an integrated approach. Reliability. Hillsdale, NJ: Lawrence Erlbaum; 1991.

Shefner JM, Jillapalli D, Bradshaw DY. Reducing intersubject variability in motor unit number estimation. Muscle Nerve 1999;22:1457–60.

Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979;86:420–8.

Shy ME, Siskind C, Swan ER, Krajewski KM, Doherty T, Fuerst DR, et al. CMT1X phenotypes represent loss of GJB1 gene function. Neurology 2007;68:849–55.

Sica REP, McComas AJ, Upton ARM. Motor unit estimation in small muscles of the hand. J Neurol Neurosurg Psychiatry 1974;37:55.

Simmons Z, Epstein DK, Borg B, Mauger DT, Kothari MJ, Shefner JM. Reproducibility of motor unit number estimation in individual subjects. Muscle Nerve 2001;24:467–73.

Stashuk D. EMG signal decomposition: how can it be accomplished and used? J Electromyogr Kinesiol 2001;11:151–73.

Stashuk DW. Decomposition and quantitative analysis of clinical electromyographic signals. Med Eng Phys 1999;21:389–404.