Motor Unit Number Estimation (MUNE) and Quantitative EMG
(Supplements to Clinical Neurophysiology, Vol. 60)
Editors: M.B. Bromberg

Chapter 25

Decision support for QEMG

L.J. Pino[a], D.W. Stashuk[a,*], S.G. Boe[c] and T.J. Doherty[b]

[a]*Department of Systems Design Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada*

[b]*Department of Clinical Neurological Sciences, University Hospital, London Health Sciences Centre, London, ON N6A 5A5, Canada*

[c]*Heart and Stroke Foundation Centre for Stroke Recovery, Sunnybrook Health Sciences Centre, University Centre, Toronto Rehabilitation Institute, Toronto, ON M5G 2A2, Canada*

1. Introduction

Clinicians use qualitative or quantitative analysis of electromyographic (EMG) signals to help detect the presence or absence of neuromuscular disorders. Characterization of muscle based on qualitative visual and auditory analysis of EMG signals is more prone to subjective bias and misinterpretation than if based on quantitative analysis of EMG signals. A recent study authored by Kendall showed that faculty and residents (blind to the underlying diagnosis of radiculopathy), using video recorded needle based examinations, had an overall 46.9% agreement with the actual diagnosis (Kendall and Werner, 2006). In addition, qualitative methods are less able than quantitative methods to provide effective longitudinal comparisons due to the subjective nature of the qualitative EMG examination. However, the large sets of statistics generated by quantitative analysis of EMG signals make it difficult to manually extract a clinically useful characterization. This is an important factor preventing the widespread adoption of quantitative electromyography (QEMG) by clinicians. Augmenting existing QEMG techniques with methods that transform QEMG generated statistics into a concise muscle characterization that is sensitive to small changes caused by a neuromuscular disease would lead to wider clinical use of QEMG techniques and allow clinicians to more precisely measure the level of disease involvement and therefore evaluate treatment effectiveness (Swash, 2002). It is important, however, that the muscle characterizations provided are transparent, such that their rationale are defendable and easily, intuitively understood. This paper describes a novel method for decision support using QEMG statistics which provides a transparent muscle characterization based on Bayesian aggregation of individual MUP characterizations determined using 'pattern discovery' (PD) methods.

There are several different QEMG methods used for analyzing neuromuscular systems, as shown in Fig. 25.1. The focus of this work is on

[*]*Correspondence to*: Dr. Dan Stashuk, Department of Systems Design Engineering, University of Waterloo, 200 University Avenue West, Waterloo, ON N2L 3G1, Canada.
Tel.: +1-(519)-888-4567, ext. 32982;
Fax: +1-(519)-746-4791;
E-mail: stashuk@uwaterloo.ca

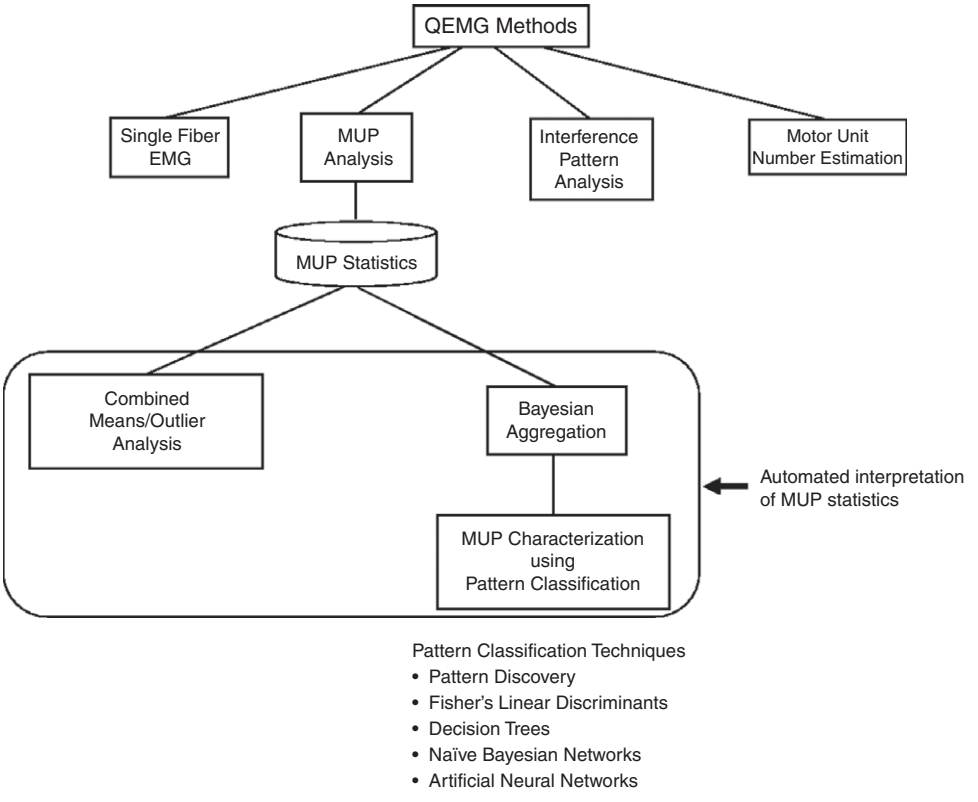

Fig. 25.1 QEMG methods. The focus of this work is on comparing combined means/outlier analysis with pattern discovery based Bayesian aggregation of MUP characterizations.

techniques that use automated interpretation of quantitative motor unit potential (MUP) statistics for the characterization of a muscle. In the future, other QEMG methods can be combined with these methods as sources of additional information for determining a more comprehensive muscle characterization. Furthermore, as a first effort, feature values for a set of MUPs detected from a single muscle are combined into a muscle characterization naively comprised of just 3 categories: myopathic, normal or neuropathic. These 3 categories provide an initial step towards a useful, robust neuromuscular clinical decision support system. Figure 25.1 shows that there are two methods for automated interpretation of MUP statistics for clinical decision support: means and outlier based analysis and Bayesian aggregation. The means method compares the mean values of standard MUP features for a set of MUPs isolated from a muscle to the distributions of corresponding MUP feature values obtained from a population of control muscles. Using mean values from the control muscles plus or minus two or three standard deviations, thresholds are determined and used to assess whether the muscle under test is "normal" or not (Stewart et al., 1989). The outlier method uses MUP feature values obtained from a population of control muscles to define outlier feature values. The number of outliers in the set of values of standard MUP features for a set of MUPs isolated from a muscle is used to assess whether the muscle is "normal" or not (Stålberg et al., 1994). Recent work has used each of these methods and their combination

(Podnar, 2004). This work highlighted the variability of the results obtained, with either method or their combination, depending on the number of features considered, the number of standard deviations used to define a threshold, and the number and definition of the outliers used to establish abnormality. Bayesian aggregation was first introduced by Pfeiffer (1999) as a method to characterize muscles by combining for each MUP detected their conditional probabilities of being detected in a muscle of a specific category (i.e., myopathic, normal or neuropathic) to arrive at an overall muscle characterization. Analysis of individual MUP feature values lacks sufficient information to accurately characterize a muscle system, so Bayesian aggregation provides a statistically robust method for combining the feature values of several MUPs acquired from a muscle. Pfeiffer used Fisher's linear discriminant analysis (LDA) to determine the conditional probabilities of each MUP (i.e., the MUP characterization); however, any pattern classification method capable of estimating the conditional probability of a category given an observation can be used for Bayesian aggregation. MUP characterization is the estimation of the conditional probabilities (one for each category) of a MUP being detected from a muscle with a given category of disorder. Some other pattern classification based techniques capable of estimating conditional probabilities are shown at the bottom right of Fig. 25.1.

The ability of the pattern classification techniques, shown at the bottom right of Fig. 25.1, to provide MUP characterizations was analyzed (Pino et al., 2007) against the following criteria: accuracy, transparency (ability of the system to explain its conclusions), ability to provide a numeric value for MUP characterization, and ability to handle multiple features simultaneously. Artificial neural networks (ANN) were the only MUP characterization method, mentioned at the bottom right of Fig. 25.1, not considered in that analysis because ANNs lack transparency. All of the other methods were comparably accurate; however, PD showed greater transparency than the other methods because PD could reveal which subsets of features either supported or refuted each element of a MUP characterization using a weight of evidence (WOE) measure.

In this work, the accuracy of the means and outlier based techniques versus Bayesian methods for muscle characterization were evaluated — the circled part of Fig. 25.1 labeled "automated interpretation of MUP statistics". The estimated conditional probabilities (i.e., MUP characterizations) used in Bayesian aggregation were calculated using PD. The Methods section describes Bayesian aggregation using PD methods as well as the means and outlier based techniques. With regard to the PD methods, it introduces a new measure called "compound rule conditional probability" that is a transformation of the WOE provided by PD into a conditional probability so that a set of MUPs can be combined using Bayesian aggregation into a muscle characterization. The Results section reports the sensitivities, specificities, and accuracies for PD-based Bayesian muscle characterization and the means and outlier based techniques. It is followed by the Discussion and Conclusions and Future Work sections.

2. Methods

This section describes the Bayesian aggregation method for muscle characterization used in this work. The new PD methods for calculating MUP characterizations are described as well as the means and outlier based methods that were implemented for comparison.

2.1. Bayesian muscle characterization

Pfeiffer applied Bayesian aggregation of LDA based characterizations of a set of N MUPs detected from a muscle to determine its characterization (Pfeiffer, 1999). A set of category labels $Y = \{y_1, \ldots, y_k, \ldots, y_K\}$ contains K labels and $1 \leq k \leq K$. Characterization of a MUP will produce K conditional probabilities, one for each category. Each MUP conditional probability

$P(y_k|MUP)$ measures the probability of category y_k given the detected MUP. The conditional probability that a muscle belongs to category y_k given a set of N MUP measurements (i.e., $P_n(mus = y_k|\{MUP_N\})$), can be estimated using Eqn 1. The numerator is the product of all of the individual MUP characterizations for category y_k and the denominator is the sum of the products – one product for each category. Eqn 1 is mathematically identical to the Bayesian aggregation used by Pfeiffer (1999). The form of Eqn 1 emphasizes that the order in which the N MUPs of the acquired set are used for Bayesian aggregation does not change the outcome because multiplication is commutative. The prior probabilities used in this work were set to be equal for each category. For instance, for a two-category characterization the prior probabilities were set at 0.5 for each category to determine a muscle characterization that is based only on the electrophysiological evidence provided by the detected MUPs. Figure 25.2 diagrams how Eqn 1 works.

$$P_n(mus = y_k|\{MUP_N\}) = \frac{\prod_i^N P(y_k|MUP_i)}{\sum_{j=1}^{K}\left(\prod_i^N P\left(y_j|MUP_i\right)\right)} \quad (1)$$

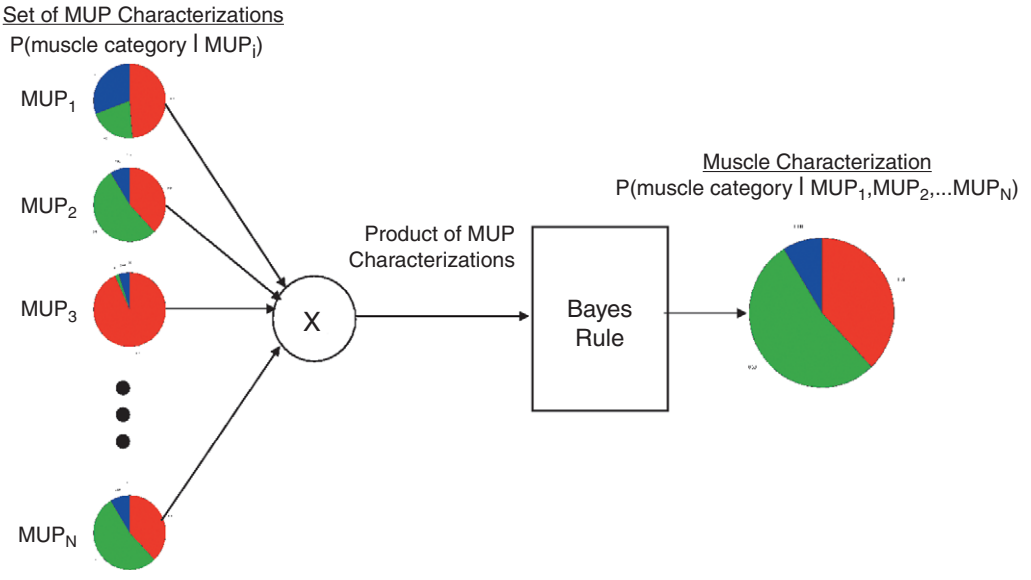

Fig. 25.2 Bayesian aggregation: MUP characterizations to muscle characterization. The pie charts on the left-hand side show the characterizations of MUPs detected from the same muscle. Each category is represented by a color, red is myopathic, green is normal and blue is neuropathic. The proportion of a pie is the probability of being detected from a muscle from that given category, e.g., the upper left pie shows an MUP that has approximately 50% probability of being detected from a myopathic muscle, a 20% probability of being detected from a normal muscle and a 30% probability of being detected from a neuropathic muscle. The product of the probabilities of the MUPs for each category is calculated and the probability of a muscle belonging to a category given the set of MUP characterizations is determined using Eqn 1. The larger pie chart on the right-hand side shows the muscle characterization given the set of MUPs, approximately 38% myopathic, 53% normal and 9% neuropathic in this example.

where:

mus = y_k represents that the muscle is of category y_k;

$\{MUP_N\}$ represents the set of N MUPs sampled from the muscle *mus;*

MUP_i represents the i^{th} motor unit potential from the set $\{MUP_N\}$.

Bayesian aggregation combines the MUP characterizations of a set of MUPs detected from a single muscle into its characterization. Each MUP characterization is expressed as a conditional probability of category y_k (i.e., myopathic, normal, or neuropathic) given the detected MUP. In this work, MUP characterizations were calculated using a new PD* based method.

2.2. Pattern discovery based MUP characterization

PD uses information–theory based statistical inference for the detection of significant patterns. Wong and Wang (Wong and Wang, 1995, 1997, 2003; Wang, 1997) describe the use of information theory–based PD for classification, which involves the use of novel, statistically based analysis of training data. These methods extract information about normal and diseased muscle by defining the occurrences of MUP feature values within specified ranges as discrete events to extract patterns of feature values (events) that can be used to estimate the conditional probabilities of detecting a specific MUP in a myopathic, normal or neuropathic muscle (Pino et al., 2007). These MUP characterizations are transparent, because they are based on statistically valid patterns of QEMG feature values (events) that can be described linguistically (i.e., high amplitude, long duration, moderate number of phases). PD-based characterization is composed of three parts: discovery of patterns, rule selection, and characterization. A specific first order event occurs when a MUP feature has a value within a defined discrete interval. Combinations of several MUP feature values occurring in a single MUP represent a high order event. The order of the event is equal to the number of feature values considered. Events may also include the category of muscle (i.e., myopathic, normal or neuropathic) from which a MUP was detected. A pattern is an event which occurs statistically significantly more often than expected assuming random occurrence. Patterns that include a category label can be used as rules for characterization (i.e., these are combinations of MUP feature values and a muscle category that can be used to predict the muscle category). Given that X_l^r is an r^{th} order pattern that has r discrete MUP feature values for which a rule associating this pattern to the muscle categories exists, and $\mathbf{X}$ is a vector containing the feature values of an observed MUP, the discriminatory power of the pattern X_l^r can be determined by its weight of evidence (WOE). With respect to category y_k, the WOE of a MUP feature vector $\mathbf{X}$ containing a pattern X_l^r is the logarithm of the odds, calculated based on the training data, of $\mathbf{X}$ containing X_l^r given the MUP was detected in a muscle of category y_k versus $\mathbf{X}$ containing X_l^r given the MUP was detected in a muscle of category other than y_k.

$$WOE = \log_2 \frac{P\left(X_l^r | MUP = y_k\right)}{P\left(X_l^r | MUP \neq y_k\right)} \tag{2}$$

where:

$MUP = y_k$ represents that the MUP was detected in a muscle of category y_k;

$MUP \neq y_k$ represents that the MUP was detected in a muscle *not* of category y_k;

X_l^r represents an r^{th} order pattern, l is an index indicating the l^{th}, r^{th} order pattern.

P(X|Y) represents the conditional probability of X occurring given that Y has occurred.

A pattern that is *always* associated with a particular category will produce a WOE of $+\infty$ and a pattern that is *never* associated with a particular category will produce a WOE of $-\infty$.

Patterns associated with multiple categories will produce a WOE that is in between these two extremes. Patterns with WOE values near zero neither support nor refute a specific category. Further details of the PD method used for MUP characterization are available (Pino et al., 2007).

Compound rule conditional probability: an r^{th} order pattern for which a rule associating this pattern to the muscle categories exists and where r is less than the total number of features is called a component rule. The union of disjoint component rules discovered in an observation is called a compound rule and is denoted by x_k^*. The weights of evidence that PD produces with respect to each muscle category y_k need to be expressed as conditional probabilities to obtain a MUP characterization. The compound rule conditional probability (CRCP) of $\mathrm{MUP_i}$ with respect to muscle category y_k is the conditional probability that MUP_i was detected from a muscle of category y_k given the occurrence of the compound rule x_k^* and is expressed as

$$\mathrm{P}\left(MUP_i = y_k | x_k^*\right) = \frac{1}{\left[\left(2^{-WOE}\right) \cdot \left(\frac{1 - P_0(y_k)}{P_0(y_k)}\right)\right] + 1} \tag{3}$$

where:

$MUP_i = y_k$ represents that MUP_i was detected from a muscle of category y_k;

x_k^* represents the compound rule of MUP feature vector $\mathbf{X}$ associated with category y_k;

WOE represents the weight of evidence of the compound rule x_k^* supporting or refuting category y_k;

$P_0(y_k)$ represents the prior probability of muscle category y_k.

The derivation of (3) is provided in the Appendix.

The conditional probability of category y_k given the detected MUP_i can be estimated using the following normalization so that the conditional probabilities across all categories sum to one:

$$P(y_k | MUP_i) = \frac{\mathrm{P}\left(MUP_i = y_k | x_k^*\right)}{\sum_{j=1}^{K} \mathrm{P}\left(MUP_i = y_j | x_k^*\right)} \tag{4}$$

These conditional MUP probabilities can then be combined using Bayesian aggregation (as shown by Eqn 1) to produce a muscle characterization.

2.2. Combined means and outlier based muscle characterization

Means based muscle characterization was developed and evaluated by Stewart et al. (1989). It is a technique where sets of MUP data, each including MUPs from a sample of 15 or more representative motor units collected from a specific muscle, from a population of control reference subjects are analyzed. Let **F** represent the set of MUP features used to represent the acquired MUPs. For each MUP feature in the set **F**, the mean for each set of MUP data (i.e., for the MUPs acquired from each specific muscle) as well as the standard deviation of the mean values (SDm) across the population of control muscles is calculated. Normative limits for the mean of a MUP feature of the set of MUPs detected in the same muscle are defined as the features overall mean ± 2 SDm. Muscle characterization is then determined using a set of at least 15 MUPs, detected from the muscle under test and representative of the motor units sampled, by calculating the mean of each MUP feature in **F** and comparing them to the normative limits calculated from the population of matched control muscles. Muscles with one or more mean feature value below its normative limit are classified as myopathic and those with one or more mean feature value above its normative limit are classified as neuropathic.

Outlier based muscle characterization was developed and evaluated by Stålberg et al. (1994). It is a method of muscle characterization which also uses a set of MUP features, such as **F** defined

above, but it is based on counting outlying MUP feature values. The thresholds for defining an outlier for a MUP feature are determined by analyzing sets of MUP data, each including MUPs from a sample of 20 or more representative motor units collected from a specific muscle, from a population of control reference subjects. For each MUP feature in the set **F** and for each set of MUP data collected from an individual control muscle, a set of feature values comprised of data from the first 20 MUPs detected is sorted in ascending order. The third lowest and the third highest value of the feature in each individual control muscle feature set is then placed in a set of low feature value outliers and a set of high feature value outliers for that feature, respectively. For each feature then, the lower outlier limit is the 5^{th} percentile of the set of its low feature value outliers and the higher outlier limit is the 95^{th} percentile of the set of its high feature value outliers. A muscle that has three or more outliers for the same feature in the set **F** that are all above or below the normative limit is considered neuropathic or myopathic respectively.

The means and outlier based methods both differ slightly when considering three rather than two muscle categories. For characterizing two muscle categories such as normal versus neuropathic, features indicative of complexity can be included in **F**. For characterizing three muscle categories (i.e., myopathic versus normal or neuropathic) phases and turns or other features indicative of MUP complexity are not included in **F** because more complex MUPs can be indicative of either myopathic or neuropathic disorders.

The means and outlier based methods can be used together to characterize a muscle and in this work is called the combined means and outlier (CMO) based method.

2.3. MUP data

The performance of the afore-mentioned methods was compared using clinical MUP data detected from two categories of muscle (normal and neuropathic) and simulated MUP data detected from three muscle categories (myopathic, normal and neuropathic).

2.3.1. Clinical MUP data

The clinical EMG data were decomposed into MUP templates using decomposition based QEMG (dQEMG; Stashuk, 1999). dQEMG typically finds 51 isolated MUPs produced by a single motor unit, aligns them, and uses a median-trimmed average to form the MUP template. A disposable concentric needle electrode (Model N53153; Teca Corp., Hawthorne, NY, USA) was used to acquire intramuscular signals using dQEMG on a Neuroscan Comperio (Neuroscan Medical Systems, El Paso, TX, USA) with a bandpass of 10 Hz–10 kHz at a sampling rate of 31.2 kHz as previously described (Boe et al., 2004, 2005, 2006). These intramuscular signals were acquired during 30 s voluntary isometric contractions performed at between 10% and 20% of each individual subjects' maximal voluntary contraction (MVC).

MUPs were sampled from the biceps brachii and first dorsal interosseous muscles of control subjects and patients with a neuropathic disorder. In total, 1649 MUPs were sampled from 16 healthy control subjects (aged 27 ± 4 years) and 427 MUPs were sampled from 14 patients, including 9 patients (aged 52 ± 12 years) with clinically probable or definite amyotrophic lateral sclerosis as defined by the revised El Escorial criteria (Brooks et al., 2000) and 5 patients (aged 37 ± 11 years) with Charcot–Marie–Tooth disease type X confirmed via genetic testing.

2.3.2. Simulated data

To help examine the relationship between level of involvement and MUP and muscle characterization performance, EMG signals were simulated using a physiologically based model (Hamilton-Wright and Stashuk, 2005), extended to allow simulation of the affects of neuropathic and myopathic disorders (Hamilton-Wright et al., 2003). To simulate a neuropathy, motor units are reorganized progressing from random motor neuron

death to random reinnervation of orphaned fibers by nearby surviving motor neurons. To simulate an inflammatory myopathy, a small percentage of randomly selected healthy muscle fibers are "infected". The majority of infected fibers is atrophied by a small fraction, and a small percentage of fibres is hypertrophied by a small fraction. This process is iterated by infecting additional fibers, similarly atrophying and hypertrophying the newly infected fibers as well as further atrophying and hypertrophying the previously infected fibers until the prescribed level of involvement is reached. A fiber is considered non-functioning (i.e., dead) when its diameter is below a critical threshold.

The physiologically based model simulates the recruitment of motor units necessary to bring the level of force produced by a diseased muscle up to a prescribed percentage of the MVC force of a healthy muscle. EMG signals detected using a concentric needle at various intramuscular positions of several different muscles during approximately 7–10% MVC were simulated, then decomposed and MUP templates were calculated. This method mimics the completion of several EMG studies across different individuals and includes background MUP interference and noise typical of that present in clinical studies. A comparison completed by Hamilton-Wright and Stashuk (2005) of decomposed MUPs detected from simulated healthy muscles versus real healthy muscles shows good correspondence.

In total, 500 MUPs were extracted from simulated EMG signals of normal muscle, 500 from myopathic muscle and 500 from neuropathic muscle. The myopathic MUPs were simulated to originate from muscles with 25%, 50% and 75% muscle fiber loss. The neuropathic MUPs were simulated to originate from muscles with 25%, 50% and 75% motor unit loss. The number of MUPs from each level of involvement was approximately the same. Each MUP was labeled as being detected from either a normal, myopathic or neuropathic muscle. This broad categorization allows comparison with the literature that uses these labels (Stewart et al., 1989; Pattichis et al., 1995; Pfeiffer, 1995; Pfeiffer and Kunze, 1995).

2.4. Feature extraction

Values for the following MUP features, measured from a MUP template, were input to the different methods: amplitude, duration, phases, turns and area-to-amplitude ratio (AAR), otherwise known as thickness. For the CMO techniques, phases and turns were included for the 2-category characterization of the clinical data and were not included for the 3-category characterization of simulated data. All features were available for 2- and 3-category testing, using Bayesian aggregation based on PD MUP characterizations. The CMO methods analyzed continuous data. For PD-based MUP characterization, MUP data were quantized into 3 intervals (e.g., low, medium or high), derived using the training data (Pino et al., 2007). Different combinations of features were selected to determine how the performance of the methods varied with different feature sets.

2.5. Muscle characterization performance

For the CMO method, half of the control subject data were used to establish normative limits and the other half of the control subject data were used to test for specificity. All of the patient data were used to test the sensitivity of the CMO method. The clinical data were used to establish normative limits for testing of the clinical data and the simulated data were used to establish normative limits for testing of the simulated data.

The following method was used for training and testing of Bayesian aggregation. The patient and control MUP data were divided into 8 sets. Eight iterations of testing were performed where one set of data from both the controls and patients was withheld for testing and the other 7 sets from both the controls and patients were used for training. PD was initially used to

calculate the conditional probabilities for each category given a detected MUP. These conditional probabilities for the set of MUPs detected from the test muscle were then aggregated using Eqn 1 to determine the conditional probabilities for each of the muscle categories (i.e., to characterize the muscle). To assess the accuracy, specificity and sensitivity of the muscle characterizations, a muscle was labeled as normal or neuropathic based on the higher of the two conditional probabilities for the clinical data or the muscle was labeled as myopathic, normal or neuropathic based on the highest of the three conditional probabilities for the simulated data and the assigned muscle label was then compared to its actual category. The number of true negatives and true positives from each test were accumulated to arrive at the overall specificity and sensitivity results for Bayesian muscle characterization.

In this work, accuracy was defined as the average of sensitivity and specificity. The traditional definition of accuracy [*(true negatives + true positives)/all muscles tested*] was not used because it is biased towards the category that has the largest number of test muscles to be characterized. The traditional accuracy measure would be skewed by the unequal proportion of subjects to patients. For instance, the traditional accuracy measure would underweigh any results for the patients because there were fewer patients than controls.

For the clinical data, sensitivity was defined as the total number of muscles characterized as neuropathic divided by the total number of "true" neuropathic muscles. Specificity was defined as the total number of muscles characterized as normal divided by the total number of "true" normal muscles. The term sensitivity–specificity deviation (SSD) was defined as:

$$SSD = \sqrt{(A - Sens)^2 + (A - Spec)^2/2} \qquad (5)$$

where A is accuracy – the mean of specificity and sensitivity; *Sens* is sensitivity; and *Spec* is specificity.

SSD was used to determine how well a characterization method maximized both specificity and sensitivity. Because the simulated data had three categories, accuracy was the mean of the accuracy of each category, i.e., A_{y_k} is the accuracy for category y_k – the number of observations of category y_k that were classified as y_k divided by the total number of observations with "true" category y_k. The SSD for three categories was then:

$$SSD = \sqrt{(A - A_{myo})^2 + (A - A_{norm})^2 + (A - A_{neur})^2/3} \qquad (6)$$

where A_{myo}, A_{norm} and A_{neur} are the accuracies for the myopathic, normal and neuropathic categories respectively;

A is the overall accuracy – the mean of A_{myo}, A_{norm} and A_{neur}.

3. Results

Table 25.1 shows muscle characterization results for clinical data using the CMO method. The feature set with the highest accuracy (61.5%) for the combined method was amplitude, duration and phases. Accuracy values for the combined method ranged from 51.9% to 61.5% with an SD of 3.7% across the different sets of features used for analysis. The SSD varied considerably from 3.8% to 17.3% depending on the feature set used.

Table 25.2 shows muscle characterization results for the clinical data using PD-based Bayesian muscle characterization. Accuracy ranged from 80.8% to 84.4% (SD 1.6%). The feature set that had the highest accuracy (84.4%) was amplitude, duration, and phases with no further improvement observed with the addition of thickness or turns. Except for the amplitude and duration feature set, the SSD using PD-based Bayesian muscle characterization for all other feature sets was less than 3%. Regardless of the feature set used, PD-based Bayesian muscle characterization had about a 20% improvement in accuracy compared to the CMO method. The SD in accuracy across the feature sets for PD-based Bayesian muscle

TABLE 25.1

COMBINED MEANS AND OUTLIER BASED MUSCLE CHARACTERIZATION – CLINICAL DATA RESULTS

Sensitivity (%)	Specificity (%)	Accuracy (%)	SSD (%)	Features
61.5	53.8	57.7	3.8	Amp/Dur
69.2	53.8	61.5	7.7	Amp/Dur/Phs
69.2	38.5	53.8	15.4	Amp/Dur/Phs/Trns
61.5	50.0	55.8	5.8	Amp/Dur/AAR
69.2	34.6	51.9	17.3	Amp/Dur/AAR/Phs/Trns

TABLE 25.2

PD-BASED BAYESIAN MUSCLE CHARACTERIZATION – CLINICAL DATA RESULTS

Sensitivity (%)	Specificity (%)	Accuracy (%)	SSD (%)	Features
72.7	88.9	80.8	8.1	Amp/Dur
81.8	87.0	84.4	2.6	Amp/Dur/Phs
81.8	81.5	81.6	0.2	Amp/Dur/Phs/Trns
81.8	83.5	82.6	0.8	Amp/Dur/AAR
81.8	87.0	84.4	2.6	Amp/Dur/AAR/Phs/Trns

characterization remained at about 2%, almost half the SD achieved for the CMO method. The SSD using PD-based Bayesian muscle characterization was smaller than that obtained using the CMO method.

Tables 25.3 and 25.4 show results based on simulated data using the CMO and PD-based Bayesian muscle characterization method, respectively. Table 25.3 does not include phases and turns as features since the CMO method does not consider them when analyzing three categories. PD-based Bayesian muscle characterization was at least 10% more accurate compared to the CMO method. In general, increasing the number of features used for PD-based Bayesian muscle characterization resulted in marginal increases in accuracy. This is in contrast to the CMO method whose accuracy decreased as the number of features used increased because of poorer sensitivity for detecting myopathic muscles. Across the various feature sets, the SSD values obtained using PD-based Bayesian muscle characterization were similar.

4. Discussion

Muscle characterization of individual and groups of muscles is an important step towards developing an understanding of the presence and/or extent of involvement of neuromuscular disorders. As described in this work, both the Bayesian and CMO methods for muscle characterization are based on analyzing quantitative MUP data extracted from sets of 15 or more representative MUPs acquired from a muscle of interest. The muscle characterization is then determined by comparing the test data with reference data in some way. The CMO method uses parametric or non-parametric methods to defined thresholds based only on control reference (i.e., "normal") data to make a binary decision. In contrast,

TABLE 25.3

COMBINED MEANS AND OUTLIER BASED MUSCLE CHARACTERIZATION – SIMULATED DATA RESULTS

Myo accuracy (%)	Norm accuracy (%)	Neuro accuracy (%)	Overall accuracy (%)	SSD (%)	Features
92.3	58.3	80.8	77.1	14.1	Amp/Dur
88.5	58.3	80.8	75.9	12.8	Amp/Dur/AAR

TABLE 25.4

PD-BASED BAYESIAN MUSCLE CHARACTERIZATION – SIMULATED DATA RESULTS

Myo accuracy (%)	Norm accuracy (%)	Neuro accuracy (%)	Overall accuracy (%)	SSD (%)	Features
96.2	88.0	80.8	88.3	6.3	Amp/Dur
96.2	92.0	80.8	89.6	6.5	Amp/Dur/Phs
96.2	84.0	88.5	89.5	5.0	Amp/Dur/Phs/Trns
88.5	84.0	84.6	85.7	2.0	Amp/Dur/AAR
96.2	92.0	88.5	92.2	3.1	Amp/Dur/AAR/Phs/Trns

the Bayesian method uses control and disease category data to allow the test data to be used to estimate individual MUP conditional probabilities, which in turn are aggregated to estimate the muscle level conditional probabilities used for characterization. The results of the analysis of the data used in this study highlight the advantages of Bayesian muscle characterization compared to the means and outlier based method in terms of accuracy, variance of accuracy across different feature sets and the ability to provide conditional probabilities.

The high muscle characterization accuracy achieved using PD-based Bayesian aggregation demonstrates the advantages of characterizing not only control reference data but also disease category data and of using Bayesian probability methods. It also suggests that PD provides reasonably accurate estimates of the MUP conditional probabilities as previously demonstrated by Pino et al. (2007). In addition, PD-based Bayesian muscle characterization achieved substantially lower SSD than the CMO method. PD-based Bayesian characterization seems to automatically maximize both sensitivity and specificity partially because of the robust nature of the Bayesian methods and partially because PD does not transform the features it analyzes. With unknown "costs" for false negative and false positive muscle characterizations a characterization method that maximizes both sensitivity and specificity is favored. While the results obtained for the means and outlier based method on these data cannot be directly compared with the analysis of data acquired from different patients by others because of differences in the level of disease involvement, it is not expected that CMO performance relative to the Bayesian method would change significantly with different data. Likewise, it is expected that PD-based Bayesian muscle characterization would achieve similar results to those obtained using LDA-based

MUP characterization, except with a lower SSD, if applied to the data used by Pfeiffer (1999).

With regard to how many and what features to use, there were no statistically significant differences in the accuracy of PD-based Bayesian muscle characterization across the different feature sets studied according to paired *t* tests at a significance level of 0.05 and Bayesian muscle characterization can incorporate feature values that measure MUP complexity regardless of the number of categories. In general, the Bayesian method has robust performance across various feature sets and the ability to improve its performance given additional information from additional features. In contrast, the CMO method can be "tuned" to maximize sensitivity and specificity by using different feature sets or changing the width of the normative limits as discussed by Podnar (2004) and in general performance varied considerably with changes in feature set composition as well as the number of features used. Increasing the number of features used for CMO analysis resulted in higher sensitivity but lower specificity, which is similar to the trend reported by Podnar (2004) (79% specificity using area and duration and only 71% when using all features). Furthermore, the CMO method depending on the number of categories considered may not be able to use feature values that measure MUP complexity.

The CMO method cannot easily provide a numerical confidence measure for a muscle characterization. Confidence can be inferred by the CMO method by varying the normative limits and or counting the change in the numbers of features that indicate normality or abnormality. For instance, there is higher confidence in the characterization of a muscle as abnormal with three features with values outside the normative limits of ± 3 SD_m as compared to a muscle that is characterized as abnormal with only one feature with values outside the normative limits of ± 2 SD_m but it is unclear how this can be used to determine a numerical value of confidence. In addition, a goal for any future clinical decision support system is reporting the level of disease involvement. MUPs detected from a muscle at different points in time (e.g., for longitudinal study) over which a neuropathic or myopathic process progresses to greater levels of involvement will in general show increases or decreases in MUP size respectively and both of these categories will in general have increased MUP complexity. It is uncertain how the CMO method can provide quantifiable information that would provide accurate indications for level of involvement useful for longitudinal study of disease progression. Another aspect of the CMO method is the uncertainty of how to deal with conflicts between low and high outliers (i.e., a set of MUPs acquired in the same muscle may have three or more low outliers and three or more high outliers). In contrast, Bayesian muscle characterization can report numeric quantities expressing the confidence in a MUP or muscle characterization. It provides for each muscle category (i.e., myopathic, normal or neuropathic) a conditional probability expressing the probability of the muscle being of that specific category given the set of acquired MUP data. These conditional probabilities can in turn be used to support clinical decisions based on the muscle characterization and can be directly used to track disease progression.

Bayesian muscle characterization requires the characterization of each MUP in the set of MUPs acquired from the muscle of interest. There are a number of methods for estimating the MUP conditional probabilities comprising a MUP characterization. Pino et al. (2007) compared MUP characterizations using LDA and PD. They discovered that each provided valid estimates of the required conditional probabilities. However, PD-based MUP characterization can also provide a transparent explanation by providing the WOE of the sets of features that supported or refuted the characterization (Pino et al., 2007). Therefore PD-based Bayesian muscle characterization also has the advantage of being able to provide an in-depth rationale for its results down to the indi-

vidual MUP feature values of the set of MUPs used to establish the characterization. Therefore, PD-based Bayesian muscle characterization can provide a basis for clinical decision support systems.

5. Conclusions and future work

This work suggests that muscle characterization using PD-based Bayesian aggregation is well suited to form the basis of a neuromuscular clinical decision support system because the results showed higher accuracy and lower SSD than the CMO method. This is most likely because muscle characterization, using PD-based Bayesian aggregation, automatically maximizes both sensitivity and specificity. This leads to the expectation that clinicians characterizing muscles using PD-based Bayesian aggregation will make fewer errors. In addition, previous work demonstrated the transparency of PD-based MUP characterization (Pino et al., 2007), which provides insight into the rationale of the muscle characterizations provided. It also confirmed that the specificity of conventional means and outlier based analysis reduces and its sensitivity increases as more features are used for characterization.

Future work will further examine the conditional probabilities of the muscle characterizations produced by Bayesian aggregation for accuracy and consistency as well as correlation to the level of involvement for both myopathic and neuropathic disorders. Simulated data are useful for examining this correlation since muscles can be simulated with known levels of involvement for myopathic and neuropathic disorders. Ideally, sets of patients with low, medium and high levels of involvement should correlate strongly with the conditional probabilities of the muscle characterizations. Such correlations would warrant an evaluation of how the system affects physician performance compared to current clinical practice.

Summary

For clinicians to use quantitative electromyography (QEMG) to help determine the presence or absence of neuromuscular disease, they must manually interpret an exhaustive set of motor unit potential (MUP) or interference pattern statistics to formulate a clinically useful muscle characterization. A new method is presented for automatically categorizing a set of quantitative electromyographic (EMG) data as characteristic of data acquired from a muscle affected by a myopathic, normal or neuropathic disease process, based on discovering patterns of MUP feature values. From their numbers of occurrence in a set of training data, representative of each muscle category, discovered patterns of MUP feature values are expressed as conditional probabilities of detecting such MUPs in each category of muscle. The conditional probabilities of each MUP in a set of MUPs acquired from an examined muscle are combined using Bayes' rule to estimate conditional probabilities of the examined muscle being of each category type. Using simulated and clinical data, the ability of a "pattern discovery" based Bayesian (PD-based Bayesian) method to correctly categorize sets of test MUP data was compared to conventional methods which use data means and outliers. The simulated data were created by modeling the effects of myopathic and neuropathic diseases using a physiologically based EMG signal simulator. The clinical data was from controls and patients with known neuropathic disorders. PD-based Bayesian muscle characterization had an accuracy of 84.4% compared to 51.9% for the means and outlier based method when using all MUP features considered. PD-based Bayesian methods can accurately characterize a muscle. PD-based Bayesian muscle characterization automatically maximizes both sensitivity and specificity and provides transparent rationalizations for its characterizations. This leads to the expectation that clinicians using PD-based Bayesian muscle characterization will be provided with improved decision support compared to that provided by the status quo means and outlier based methods.

Appendix

Derivation of compound pattern conditional probability

where:

$Y = y_k$ represents that observation x is from category y_k.

$Y \neq y_k$ *represent that observation* x is *not* from category y_k.

x^* is a compound rule of an observation x.

WOE = weight of evidence of the compound rule.

$P_0(y_k)$ is the prior probability of category y_k.

$$WOE = \log_2 \frac{P(x^*|Y = y_k)}{P(x^*|Y \neq y_k)}$$

$$WOE = \log_2 \frac{\Pr(x^*, Y = y_k)(1 - P_0(Y = y_k))}{P_0(Y = y_k)((\Pr(x^*) - \Pr(x^*, Y = y_k))}$$

$$\textit{Now letting} \quad \Phi = \frac{1 - P_0(Y = y_k)}{P_0(Y = y_k)}$$

$$= \log_2 \left(\frac{Pr(x^*, Y = y_k)}{(\Pr(x^*) - \Pr(x^*, Y = y_k))} \cdot \Phi \right)$$

$$= \log_2 \left(\frac{\frac{1}{\Pr(x^*)}}{\frac{\Pr(x^*, Y = y_k)}{}} \cdot \Phi \right)$$

$$\textit{but} \quad \Pr(Y = y_k|x^*) = \frac{\Pr(x^*, Y = y_k)}{\Pr(x^*)}, \quad \textit{so}$$

$$WOE = \log_2 \frac{\Phi}{\frac{1}{\Pr(Y = y_k|x^*)} - 1}$$

rearranging and substituting Φ *back in, so* $\therefore$

$$\Pr(Y = y_k|x^*) = \frac{1}{\left[(2^{-WOE}) \cdot \left(\frac{1 - P_0(Y = y_k)}{P_0(Y = y_k)} \right) \right] + 1}$$

References

Boe, S.G., Stashuk, D.W. and Doherty, T.J. (2004) Motor unit number estimation by decomposition-enhanced spike-triggered averaging: control data, test–retest reliability, and contractile level effects. *Muscle Nerve*, 29: 693–699.

Boe, S.G., Stashuk, D.W., Brown, W.F. and Doherty, T.J. (2005) Decomposition-based quantitative electromyography: effect of force on motor unit potentials and motor unit number estimates. *Muscle Nerve*, 31: 365–373.

Boe, S.G., Stashuk, D.W. and Doherty, T.J. (2006) Within-subject reliability of motor unit number estimates and quantitative motor unit analysis in a distal and proximal upper limb muscle. *Clin. Neurophysiol.*, 117: 596–603.

Brooks, B.R., Miller, R.G., Swash, M. and Munsat, T.L. (2000) El Escorial revisited: revised criteria for the diagnosis of amyotrophic lateral sclerosis. *ALS Other Motor Neuron Disord.*, 1: 293–299.

Hamilton-Wright, A. and Stashuk, D.W. (2005) Physiologically based simulation of clinical EMG signals. *IEEE Trans. Biomed. Eng.*, 52: 171–183.

Hamilton-Wright, A., Stashuk, D.W. and Doherty, T.J. (2003) Simulation of disease effects on muscle structure, activation and acquired electromyography. *Muscle Nerve*, 28(S12): S126.

Kendall, R. and Werner, R.A. (2006) Interrater reliability of the needle examination in lumbosacral radiculopathy. *Muscle Nerve*, 34: 238–241.

Pattichis, C.S., Schizas, C.N. and Middleton, L.T. (1995) Neural network models in EMG diagnosis. *IEEE Trans. Biomed. Eng.*, 42: 486–496.

Pfeiffer, G. (1999) The diagnostic power of motor unit potential analysis: an objective bayesian approach. *Muscle Nerve*, 22: 584–591.

Pfeiffer, G. and Kunze, K. (1995) Discriminant classification of motor unit potentials (MUPs) successfully separates neurogenic and myogenic conditions. A comparison of multi- and univariate diagnostic algorithms for MUP analysis. *Electroencephalogr. Clin. Neurophysiol.*, 97: 191–207.

Pino, L.J., Stashuk, D.W., Boe, S.G. and Doherty, T.J. (2007) Motor unit potential characterization using "pattern discovery". *Med. Eng. Phys.*, 30: 563–573

Podnar, S. (2004) Criteria for neuropathic abnormality in quantitative anal sphincter electromyography. *Muscle Nerve*, 30: 596.

Stålberg, E., Bischoff, C. and Falck, B. (1994) Outliers, a way to detect abnormality in quantitative EMG. *Muscle Nerve*, 17: 392–399.

Stashuk, D.W. (1999) Decomposition and quantitative analysis of clinical electromyographic signals. *Med. Eng. Phys.*, 21: 389–404.

Stewart, C.R., Nandedkar, S.D., Massey, J.M., Gilchrist, J.M., Barkhaus, P.E. and Sanders, D.B. (1989) Evaluation of an automatic method of measuring features of motor unit action potentials. *Muscle Nerve*, 12(2): 141–148.

Swash, M. (2002) What does the neurologist expect from clinical neurophysiology? *Muscle Nerve*, 25: S134–S138.

Wang, Y. (1997) *High-Order Pattern Discovery and Analysis of Discrete-Valued Data Sets.* University of Waterloo, Waterloo, ON, PhD Thesis.

Wong, A.K.C. and Wang, Y. (1995) Discovery of high order patterns. *Proc. IEEE Int. Conf. Syst. Man Cybernet.*, 2: 1142–1147.

Wong, A.K.C. and Wang, Y. (1997) High-order pattern discovery from discrete-valued data. *IEEE Trans. Knowl. Data Eng.*, 9: 877–893.

Wong, A.K.C. and Wang, Y. (2003) Pattern discovery: a data driven approach to decision support. *IEEE Trans. Syst. Man Cybernet.*, 33: 114–124.