Method for Counting Motor Units in Mice and Validation Using a Mathematical Model

Lora A. Major, Janka Hegedus, Douglas J. Weber, Tessa Gordon, and Kelvin E. Jones
Centre for Neuroscience, University of Alberta, Edmonton, Canada

Submitted 24 August 2006; accepted in final form 4 December 2006

INTRODUCTION

One of the most distressing neurological symptoms is the weakness or paralysis that results from disease and injury in the neuromuscular system. Motor neuron diseases are particularly insidious; these diseases progressively paralyze the body, have poor prognosis and/or treatment options, and are ultimately fatal. The cellular mechanisms underlying the death of lower motor neurons in diseases such as amyotrophic lateral sclerosis (ALS) and spinal muscular atrophy (SMA) remain a mystery. In the past decade, significant advances in understanding the pathophysiology of these diseases have been made as a result of research using transgenic rodent models (Friggier et al. 2000; Gurney et al. 1994; Ripps et al. 1995). Of particular relevance to this study, the SOD1 mouse models continue to be an invaluable tool both for understanding the mechanisms of neurodegeneration in ALS and for developing more effective treatments (Bruijn et al. 2004; Cleveland and Rothstein 2001; Nirmalanathan and Greensmith 2005).

To complement the use of these experimental models, there is a need for a better quantitative measure of disease progression. The ability to precisely measure the number of functional motor units in a muscle would provide both a measurement of early disease progression and an assessment of interventions designed to slow degeneration and promote reinnervation of the muscle. This report describes a new data analysis technique for a class of electrodagnostic methods called motor unit number estimation (MUNE) introduced by McComas (1971).

The original MUNE method had some limitations; the most significant confounding factor was alternation caused by the statistical variability of axon thresholds (Brown and Milner-Brown 1976; Major and Jones 2005; Stein and Yang 1990). The technique proposed here uses an incremental stimulus-response protocol that is very similar to McComas’ original method, but does not rely on visual discrimination of response increments. Instead, the proposed technique uses the statistical properties of alternation to determine how many stimuli to deliver to minimize alternation as a confounding factor.

Because there is no MUNE method that is universally accepted as the ideal, it is important to calibrate any suggested MUNE protocol using a recognized “gold standard.” Unfortunately there are few definitive lists of the number of motor neurons innervating individual muscles to compare with MUNE methods.

In response to this absence of a gold standard, we recently developed a mathematical model of the neuromuscular system that could be used to simulate the various MUNE methods (Major and Jones 2005). Because the number of motor units in the model is known, the estimate of this number using different MUNE methods can be objectively assessed. The model also serves as a platform for developing and validating new MUNE protocols, as reported here. The incremental twitch subtraction (ITS-MUNE) protocol is simple and uses the two experimental measures common to all MUNE methods: 1) the maximum response of the whole muscle and 2) an estimate of the average response of a single motor unit. Dividing the whole muscle response by the average motor unit response generates an estimate of the number of individual responses that make up the whole. In our method, the responses are measured as isometric twitch force rather than the more commonly used EMG waveforms.

To estimate the twitch force of an individual motor unit, we developed an interactive software program that randomly se-
lects from a set of stimulus evoked, incremental muscle twitch force waveforms. Contemporary MUNE methods that attempt to extract individual motor unit responses (usually EMG) focus heavily on ensuring that selected motor unit potentials are definitely attributable to the activation exactly one motor unit. ITS-MUNE, on the other hand, treats the problem of motor unit alternation less stringently. The assumption is that with a sufficient number of stimuli, all possible combinations of variable motor unit responses will occur. The goal is to determine the number of stimuli that will maximize the number of individual motor unit responses and minimize the inclusion of responses from multiple motor units. Despite some error being introduced by inclusion of one or more inappropriate response increments, the sample will be a statistically accurate estimate of the population as a whole.

To validate the ITS-MUNE protocol, we performed a blind comparison using both naıêve and expert analysts. The results showed that the protocol is both valid and precise with a high level of reproducibility. We conclude that the ITS-MUNE protocol is a useful adjunct to studies of the pathophysiology of muscle denervation in rodent models and will be a valuable tool for assessing the effectiveness of medical intervention.

M E T H O D S

In this paper, we present empirical data, acquired from 14 SOD1 WT (carrying the normal allele of the human SOD1 gene, i.e., not mutant mice) and 17 wild-type mice of both sexes (Jackson Labs). The animal protocol was approved by the Animal Welfare Committee at the University of Alberta and was in accordance with the Canadian Council on Animal Care guidelines. The experimental methodology will be described, followed by a description of the mathematical model and finally a description of the ITS-MUNE software.

Surgical preparation

Mice were anesthetized with an intraperitoneal injection of a cocktail made up of ketamine (100 mg/ml), acepromazine (Atravet, 10 mg/ml), and 0.9% sodium chloride, at a dosage of 17.5 ml/kg body weight. At regular time intervals, additional anesthetic was administered intraperitoneally to maintain anesthesia. The body temperature of the mice was maintained using a heat lamp. Four functionally distinct hindlimb muscles were examined: three fast-twitch muscles, the tibialis anterior (TA, n = 31 muscles), extensor digitorum longus (EDL, n = 24), and gastrocnemius (GAS, n = 23) and the slow-twitch soleus muscle (SOL, n = 20). The TA, EDL, GAS, and SOL muscle tendons were isolated and individually attached with 4–0 silk to a strain gauge (Kulite model KH-102) to record the isometric force produced by each muscle in response to stimulation of the sciatic nerve. Tendons of the plantaris muscle were cut. Two silver wire electrodes were placed alongside the sciatic nerve, and stimulation was applied to the nerve to evoke isometric twitch contractions. The electrodes were sutured into place to prevent dislocation during stimulation. Hindlimbs of anesthetized mice were immobilized by clamping the knees and the ankles, while being careful not to interfere with the blood supply to the muscles.

Experimental protocol

Stimulus pulses (duration = 100 μs, rate = 0.5/s) were applied to the sciatic nerve, and the amplitude of the pulse was manually adjusted over a range from 0 to 10 V. The stimulus amplitude was increased while monitoring the isometric twitch force on an oscilloscope for discrete increments in peak force. With each increment in force, the stimulus amplitude was held constant for 8–10 stimuli and increased manually to elicit a larger twitch. This process was repeated until the peak force was approximately two thirds of maximum, with the total number of stimuli ranging from 100 to 300. The analog force data were digitized at 2 kHz using Axoscope (Version 8.0, Axon Instruments).

Mathematical model

The full model is a combination of 1) a stochastic model of the response of motor axons to extracellular stimuli of different amplitudes; 2) a deterministic model of motor unit twitch force; and 3) the addition of noise to the final force output to match the experimental data. The model was developed in MATLAB v.7.0, and simulations were run on PCs under Windows XP operating system with a time step of 0.5 ms (2 kHz).

AXON THRESHOLD TO ELECTRICAL STIMULATION. Axons respond to electrical stimulation in an all-or-none fashion, but the threshold is probabilistic rather than a finite value. The probability of generating an action potential increases from 0 to 100% as the stimulus amplitude is increased. Between the subthreshold and suprathreshold stimulus amplitudes, the probability of generating an action potential increases sigmoidally and is well fitted by a cumulative Gaussian function (Bergmans 1970; Brown and Milner-Brown 1976; Hales et al. 2004; Milner-Brown and Brown 1976). For our purposes, the mean axon threshold is defined as the stimulus amplitude that generates an action potential 50% of the time.

To model the threshold of a single axon, we used a Gaussian probability distribution function

$$
\Theta(x) = \frac{1}{\sigma_i \sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{(s-\mu_i)^2}{2\sigma_i^2}\right) ds
$$

(1)

where $s$ is the stimulus amplitude, $\mu_i$ is the mean threshold of the $i$th axon, $\sigma_i$ is the SD of the $i$th axon threshold, and all parameters have the same units. For each axon, the SD was calculated as

$$
\sigma_i = \frac{1.65 \times \mu_i}{100}
$$

(2)

where the relative spread of 1.65% was measured from normal human motor axons (Hales et al. 2004). To simulate the thresholds of a population of axons, the mean thresholds, $\mu_i$, were randomly chosen from a normal distribution. We determined the range of $\mu_i$ in mouse experiments to be a factor of 2.3 across the population of motor nerve axons.

This was determined by measuring the range of stimulus amplitudes in the stimulus-response curve from subthreshold to whole muscle force. The combination of sigmoidal thresholds, a relative spread of 1.65%, and mean thresholds for the population having a normal distribution over a limited range resulted in overlap of the threshold probability curves. This overlap results in stimulus-to-stimulus variation of the axons excited with a constant stimulus (Brown and Milner-Brown 1976; Milner-Brown and Brown 1976).

There is no evidence for a systematic relationship between threshold and motor axon diameter using percutaneous stimuli in human nerves (Doherty and Brown 1993; Doherty and Stashuk 2003). Therefore in the model, axon thresholds were randomly associated with motor unit twitch forces. This resulted in random recruitment of motor units with respect to twitch force as stimulus intensity was increased.

MOTOR UNIT TWITCH FORCE. When an action potential was generated in an axon, the associated motor unit generated a twitch force described by

$$
f(t) = \frac{P_i}{T_i} \times t \times e^{-t/T_i}
$$

(3)

where $t$ is the time (ms), $P_i$ is the peak amplitude, and $T_i$ is the contraction time (ms) of the $i$th motor unit in a population of size $N$.
The values for peak twitch force and contraction time in the population are given by

\[ P_i = e^{\frac{\log(P_{ratio})}{N_i}} \]

and

\[ T_i = T_{ratio} \left( \frac{1}{P_i} \right)^{\frac{1}{N_i}} \]

respectively, where \( P_{ratio} \) is the scaling of smallest to largest motor unit twitch amplitudes and \( c \) is a coefficient for scaling twitch contraction times. \( P_{ratio} \) was determined experimentally. The equation \( c = \log_{P_{ratio}}(P_{ratio}) \) determines the value of the coefficient, where \( T_{ratio} \) is the ratio of slowest to fastest contraction times \( (T_{1,2} / T_{min}) \). This factor was estimated from the mice experiments to be 23 ms to 17 ms and was used for all simulations regardless of the muscle being simulated. A sample of single motor unit twitches generated using Eqs. 3–5 are shown in Fig. 1A. The total force output of a simulated muscle was the sum of the forces of the motor units activated by a stimulus pulse.

SIMULATING EXPERIMENTAL NOISE. The force data recorded from the mouse experiments were contaminated by small fluctuations originating from mechanical perturbations of the experimental platform, electromagnetic interference, and digitization at different amplifier gains. To faithfully simulate realistic force data, we added noise to the simulated twitches, because the noisiness of recorded waveforms impacts the reliability of the proposed MUNE method. The appropriate spectrum of random noise was estimated by analyzing several examples of force recordings from the mice. This iterative process of matching the power spectrum of simulated noise to samples of actual baseline force recordings led to the addition of a random signal of zero mean and 0.28 mN SD within the frequency band of 0 to 100 Hz. This noise has been added to the twitches in Fig. 1B to provide a visual impression of the relative amount of noise in our experimental and simulated data.

MUNE analysis method

The basic principle underlying all MUNE techniques is division of the total muscle response by an estimate of the average motor unit response. The difference between the different MUNE techniques is the method used to estimate the average single motor unit response. In our method, 10–20 responses are selected randomly from the 100–300 incremental force waveforms recorded in an experiment (or simulated with the model, for the purpose of validation in this study).

From each of the randomly selected waveforms, the next largest (or second next largest) trace is subtracted, giving a variety of motor unit response samples. Some of these twitch candidates may be rejected because of nonphysiological characteristics. Any rejected sample is replaced with a new random selection.

ITS-MUNE does not guarantee that every accepted twitch sample is attributable to the activation of one motor unit. It is assumed, however, that by rejecting candidates that are obviously nonphysiological, averaging 10–20 twitches will offer a reasonably valid estimate of the average motor unit twitch.

What follows is a detailed description of the steps involved in this sampling process.

SOFTWARE. A custom computer program—a graphical user interface (GUI) based in MATLAB—called ITS-MUNE was developed to process the force data and can be downloaded from www.uofaweb.ualberta.ca/compneurlab/research.cfm. The major tasks handled by the program are shown in Fig. 2 and include 1) filtering and alignment of the incremental twitch force traces, 2) random selection of twitch responses, 3) subtraction of two rank ordered force traces to generate a candidate motor unit twitch response, 4) automatic and supervised rejection of nonphysiological motor unit twitches, and 5) averaging motor unit twitch samples to estimate mean twitch force and calculate the MUNE.

FILTERING AND ALIGNMENT. In Fig. 2A, selections of the raw data acquired by Axoscope are shown at different levels of amplifier gain. The force data are recorded at various levels of amplification, because of the wide range of incremental forces from TA and GAS muscles and stored in separate digital data files. The full data set from a single muscle, consisting of multiple files, is loaded into the ITS-MUNE software simultaneously. The program also handles loading of MATLAB binary file formats (.mat) from the model simulations.

After file loading, the force data are separated into 300-ms traces and individually low-pass filtered (1st-order Butterworth with 100-Hz cut-off) to remove high-frequency variations that are not relevant to the force signal. The average baseline (from time 0 to force onset) of each trace was removed so that all records were aligned to zero baseline force. The data generated during this step of processing are shown in raster format in Fig. 2B.

The analyst has the option of deleting individual force traces from the collection. The purpose of this facility is to remove waveforms that are clearly corrupted by movement artifact, amplifier saturation, or experimental error. This manual elimination of inappropriate data are the only element of ITS-MUNE processing where the incremental force waveforms must be visually inspected as a group. Because sampling of motor unit twitches is automated, visual identification of motor unit force increments is never required.

RANDOM SELECTION AND ACCEPTANCE OF CANDIDATE TWITCHES. The next stage of the analysis involves choosing a user-defined number of motor unit twitches to extract from the full data set. This number was set to 20 for the muscles that produce large whole muscle forces (i.e., TA and GAS) and 10 or 15 for the smaller force muscles. These numbers were determined based on experience of the expert analysts and a previous report of sample size considerations (Slawnych et al. 1997). The user initiates the random selection of...
force responses and the software selects a random waveform and subtracts the next smaller waveform (Fig. 2B, arrows).

The software rejects some candidate motor unit twitches automatically; the user must reject others manually. A candidate twitch is automatically rejected if 1) it has a region of negative force or 2) if the peak force is smaller than a specified minimum. An example of the negative force criterion is shown in Fig. 2C. The presence of negative force is evaluated within a user-defined window (Fig. 2, C and D, vertical lines). When a negative force region exists within the window, the program automatically discards the candidate twitch and generates a new candidate by subtracting the second next smaller waveform (Fig. 2B, 2nd order twitch; Fig. 2D, resulting subtraction). The minimum peak force criterion is user-defined and was set as 0.5 mN for the data presented here. If the analyst has reason to believe that the muscle under study has motor units smaller than 0.5 mN, this parameter can be adjusted to accept smaller units. The minimum peak force could even be specified as zero, so that no twitch candidate would be automatically rejected, as long as its peak was positive.

If a candidate motor unit twitch is not automatically rejected, the waveform is presented to the user (Fig. 3). The user is required to accept or reject the subtracted trace based on visual inspection. While admittedly a subjective process, in practice, the user is most often faced with a decision between traces that “look” like a twitch (Fig. 3, right) and those that do not (Fig. 3, left). If the candidate twitch is rejected, another subtraction is performed using the next smaller waveform. If this candidate twitch is also rejected, a new incremental muscle force waveform is randomly selected.

This process of random software selection and acceptance is repeated until the desired number of motor unit twitch samples is obtained (e.g., 20 in the case of TA or GAS). If the specified number of samples is not obtained, e.g., because of a high rejection rate, it is necessary for the user to reduce the number of samples and restart this process.

CALCULATING MEAN TWITCH FORCE AND MUNE. After completing the process of random selection and acceptance, the set of accepted motor unit twitch waveforms is presented along with the distribution of twitch amplitudes and the original incremental force waveforms (Fig. 4). This figure presents the user with a summary of the sampling process where a number of important features can be qualitatively assessed. First, the random selections of whole muscle force used in the analysis are highlighted in color, and the user can check if samples...
were selected throughout a range of recorded forces (Fig. 4, left). Next, the overlay of the estimated motor unit twitches is examined to compare the sample population at the same scale (Fig. 4, top right). This provides a final opportunity for the user to delete motor unit twitches that have a questionable waveform in comparison to the rest of the samples (e.g., double humped yellow twitch in Fig. 4, top right). Finally, a histogram of peak amplitudes is displayed (Fig. 4, bottom right).

The average motor unit twitch force is computed and divided into the whole muscle force to calculate the motor unit number estimate (Fig. 4, parameter area, top left). The whole muscle peak force amplitude is not part of the experimental dataset because the incremental stimulation paradigm is typically carried out to about two thirds of the maximum to prevent damage to the preparation. The whole muscle force is measured experimentally at the end of the paradigm, and the value noted and entered manually. A diagram of the workflow required for ITS-MUNE analysis is shown in Fig. 5.

**MUNE validation methods**

To assess whether the ITS-MUNE program measures what it is supposed to measure, the number of motor units in a muscle, we analyzed simulated data and compared the estimate to the known number of motor units in the model. Eight files containing simulated data were generated based on the experimental measurements: two sets of data from each of the four muscles (Table 1). Seven different analysts scored the simulated data, four of whom were naïve to the method and its implications. Naïve users were trained by watching a video instruction guide to ensure consistency. The video provided guidelines on choosing the user-defined parameters. For example, “Collect 10 samples for muscles with a whole muscle force <150 mN, 20 samples for those >400 mN, and 15 samples for the remaining muscles.” All seven analysts were blind to the muscle type, the actual number of motor units in the simulated data files, and the results of previous analysts.

We approximated the manual incremental stimulus paradigm by using 30 distinct stimulus amplitudes, evenly spaced between minimum threshold and 67% of maximum, in the simulations. The number of stimuli at each of the 30 stimulus amplitudes varied from 3 to 10, depending on the peak force of the whole muscle force. For each simulated muscle, 90 to 300 stimuli were given, and the order of stimuli was random with respect to amplitude. These parameters were chosen based on a post hoc analysis of the manual experimental data for different muscles.

**Statistical analysis**

Statistical analyses were performed using MATLAB Version 7.2 (R2006a) and its optional STATISTICS toolbox. To quantify the overall
COUNTING MOTOR UNITS IN MICE

TABLE 1. Muscle model parameters

<table>
<thead>
<tr>
<th>Muscle</th>
<th>$N_{\text{min}}$</th>
<th>$N_{\text{max}}$</th>
<th>$P_{\text{max}}$ mN</th>
<th>$P_{\text{ratio}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>TA (n = 31)</td>
<td>55</td>
<td>130</td>
<td>1.2</td>
<td>10.2</td>
</tr>
<tr>
<td>SOL (n = 20)</td>
<td>29</td>
<td>56</td>
<td>1.2</td>
<td>5.31</td>
</tr>
<tr>
<td>GAS (n = 23)</td>
<td>68</td>
<td>141</td>
<td>1.7</td>
<td>10.2</td>
</tr>
<tr>
<td>EDL (n = 24)</td>
<td>25</td>
<td>57</td>
<td>1.1</td>
<td>4.36</td>
</tr>
</tbody>
</table>

Table 1. Muscle model parameters determined from experiments on wildtype and SOD1G93A mice. The numbers of each type of muscle tested are given by $n$. $N_{\text{min}}$, $N_{\text{max}}$, $P_{\text{min}}$, and $P_{\text{max}}$ are the minimum numbers of motor units estimated in healthy muscles and the maximums. $P_{\text{min}}$ are the minimum estimates of peak twitch tension, and $P_{\text{ratio}}$ are the ratios of largest to smallest motor unit forces within a muscle. TA, tibialis anterior; SOL, soleus; GAS, gastrocnemius; EDL, extensor digitorum longus.

accuracy of the ITS-MUNE method, a linear regression analysis was done on the MUNE scores generated by the seven analysts versus the known motor unit numbers ($N$) in the eight sets of simulated test data. Analysts repeated the scoring of each muscle test set five times, and these five scores were averaged into a single data point. Thus the regression was performed on $n = \{(7 \text{ analysts}) \times (8 \text{ data sets})\} = 56$.

Before regression, the data were tested for normality using the Lilliefors test. The distributions were not significantly different from normal within 99% confidence.

An ideal MUNE would give an estimate that was the same as the true number of motor units in the model. This would result in a regression along the line of identity. Therefore a two-tailed hypothesis test based on the $t$-statistic of the regression slope was carried out to determine if the slope was significantly different from unity ($H_0: m = 1; H_1: m \neq 1$). The regression intercept was not tested because it was obviously nonzero (see RESULTS).

To be confident that the accuracy of this MUNE method is not biased by the knowledge base of the analyst, we performed linear regressions of MUNE versus motor unit number after splitting the results into na"ıve and expert categories. Hypothesis tests based on $t$-statistics were completed to determine if the slopes of na"ıve and expert analysts were significantly different from unity. To test whether there were differences between na"ıve and experts, we used $t$-statistics of both slope and intercept.

To determine whether the variability in MUNE was constant for all magnitudes of MUNE, a linear regression was performed on the SD of MUNE scores versus the mean ($\mu$) MUNE score for each simulated muscle data set. A two-tailed hypothesis test based on the $t$-statistic of this regression slope ($m_{\mu SD}$) was carried out to determine whether it is significantly different from zero. That is, is there a relationship between an averaged MUNE score and the variability within those scores? Having determined that the variability within MUNE scores grows as MUNE increases, further statistical tests were applied to reveal whether the two measures were linearly related.

R E S U L T S

Experimental measurement of MUNE and muscle parameters in mice

There is no way to directly assess whether ITS-MUNE, applied to 98 muscles in 31 healthy mice, gave accurate motor unit number estimates. However, it is possible to determine whether the motor unit twitch estimates obtained using the ITS-MUNE method match motor unit twitch sizes obtained by other, more direct methods. We compared the distribution of motor unit twitches to those recorded by stimulation of ventral root nerve filaments to the TA and GAS muscles (Hegedus and Gordon, unpublished data). The distribution of twitches obtained by the two methods was similar, but the ITS-MUNE method produced a small number of oversized twitch estimates. Two percent of all twitch samples collected using ITS-MUNE were $>19$ mN, which was the largest twitch observed by stimulation of nerve filaments. Aside from this positive skew, the frequency distributions for twitch amplitudes overlapped for the two datasets. This comparison provided confidence that the twitches extracted by ITS-MUNE are usually representative of single motor unit force, but a more rigorous system of validation in the form of a mathematical modeling was pursued to evaluate the performance of ITS-MUNE.

To determine a physiological range of parameters for the mathematical models, we first analyzed experimental data collected from four hindlimb muscles of healthy mice. Motor unit numbers were estimated using the ITS-MUNE protocol. Motor unit twitch properties (peak force and contraction time) were determined by examination of the twitch samples extracted by the ITS-MUNE software. These data were used to constrain the parameters in Eqs. 3–5 and were muscle specific (Table 1).

$N_{\text{min}}$, and $N_{\text{max}}$ are estimates of the minimum and maximum numbers of motor units in each muscle, respectively. When simulating force data from a particular muscle, the number of motor units in a model was chosen randomly from within these bounds. $P_{\text{min}}$ is the smallest motor unit twitch force, and $P_{\text{ratio}}$ is the ratio of largest to smallest motor unit twitch forces. Based on $P_{\text{ratio}}$, the four muscles were qualitatively assigned to two groups: low force (SOL and EDL: $P_{\text{ratio}} < 6$) and high force (TA and GAS: $P_{\text{ratio}} > 10$). In the computational model of GAS, for example, the largest motor unit twitch was $P_{\text{min}} \times P_{\text{ratio}} = 1.7 \times 10.2 = 17.3$ mN, whereas in EDL, the largest motor unit twitch was 4.8 mN. The smallest motor unit twitch included in any of the simulated muscles was 1.1 mN, which is larger than the ITS-MUNE minimum twitch criterion of 0.5 mN. In the simulated study, therefore, it was impossible for ITS-MUNE to mistakenly reject true twitches using the cut-off of 0.5 mN for the smallest acceptable twitch size.

Contractile speed of motor units varied within individual muscles. However, the contraction times measured for each muscle group overlapped and were not significantly different. Using the composite experimental data from all muscles, contraction times ranged from 23 ms in the slowest to 17 ms in the fastest units. Thus the coefficient $c$ in Eq. 5 was defined as $c = \log_{\text{Twitch}}(P_{\text{ratio}})$, where $T_{\text{ratio}} = 23/17$ for all four muscles.

To test the parameterized models, the smallest and largest whole muscle twitches generated by simulation were compared with the smallest and largest whole muscle twitches observed experimentally (Table 2). The simulations correctly predicted a range of whole muscle forces that overlapped with the corresponding values collected by experiments on the same muscle.

Testing for sampling bias in experiments

During the incremental stimulation phase of the ITS-MUNE protocol, the stimulus intensity is gradually increased until the recorded twitch is about two thirds of whole muscle force. If electrical stimulation of motor axons preferentially activates the largest axons, as is often supposed based on historical observations and theoretical calculations (Erlanger 1937; McNee 1976), the smallest motor units would not be activated during submaximal stimulation. This would result in overre-
A random sample of simulated muscles produced whole muscle twitches in approximate agreement with those of the healthy mouse experimental data. In comparison to the smallest and largest whole muscle forces extracted from n random samples, the widest possible range whole muscle forces are as follows: TA, 263–620 mN; SOL, 91–174 mN; GAS, 461–952 mN; EDL, 63–144 mN. These wider force ranges correspond to the smallest and largest possible simulated motor unit numbers (N_min and N_max in Table 1). Abbreviations are as shown in Table 1.

Table 2. Whole muscle twitch: simulation vs. experiment

<table>
<thead>
<tr>
<th>Muscle</th>
<th>Simulation</th>
<th>Experiment</th>
</tr>
</thead>
<tbody>
<tr>
<td>TA (n = 31)</td>
<td>292–614</td>
<td>259–510</td>
</tr>
<tr>
<td>SOL (n = 20)</td>
<td>94–171</td>
<td>81–199</td>
</tr>
<tr>
<td>GAS (n = 23)</td>
<td>461–898</td>
<td>424–768</td>
</tr>
<tr>
<td>EDL (n = 24)</td>
<td>63–128</td>
<td>74–155</td>
</tr>
</tbody>
</table>

To assess the relationship between motor unit force and recruitment order by electrical stimulation of the mouse sciatic nerve, we plotted the amplitudes of individual motor unit twitches (as estimated by ITS-MUNE) relative to the peak of the whole muscle twitch force from which it was extracted (Fig. 6). Because the whole muscle twitch force is an indirect measure of stimulus amplitude, if there were a bias for recruiting large diameter axons, i.e., large twitch motor units, at low stimulus amplitudes this plot would show a negative correlation. Instead, motor unit twitch forces seem random with respect to the incremental muscle force. This experimental result suggests that all sizes of motor unit twitch force are sampled within the lower two thirds of whole muscle force output. Furthermore, this finding justifies the random distribution of axon thresholds with respect to twitch force used in the mathematical model.

To determine whether the sample of motor unit twitches collected in the experiments gave a good representation of the population, we plotted histograms of the sample of twitch forces for the high- and low-force muscles (Fig. 6, insets). The histograms reveal an exponential distribution of twitch sizes, skewed toward small motor units as expected (Burke 1981). These twitch distributions are also well fit by Eq. 4, which was used to generate representative twitch amplitudes throughout the motor unit pools for use in the ITS-MUNE validation model (Fig. 6, inset curves).

Although it is not apparent by visual inspection of Fig. 6, statistical analysis on a log-transformed version of the twitch samples revealed that there was a statistically significant trend toward estimation of larger twitches at higher stimulus intensities. One-way ANOVA on twitches versus the force level at which they were activated reveals that, in high-force muscles, twitch estimates grew exponentially by 0.07–0.12% (within SE) for each millinewton increase in incremental force (P = 0.0004, n = 898). In low-force muscles, twitch estimates grew by 0.31% (0.22–0.40%) for each millinewton (P = 0.0008, n = 489).

Rigorous examination of this effect is deferred to the Discussion. It is important to note that the correlation between twitch estimate and incremental force is very weak (9.5% for high-force muscles, 1.2% for low-force muscles). Most of the variability in twitch estimates is not related to the force level at which they were elicited (Fig. 6), but presumably because of natural variation in motor unit sizes.

Validation of the ITS-MUNE algorithm

To determine whether the proposed ITS-MUNE paradigm and analysis algorithm accurately measured the number of motor units in a muscle, we generated a test database using the computational models. The test database consisted of eight sets of data, two for each of the four muscles, with a random number of motor units chosen from the ranges in Table 1. The identity of the muscle and the number of motor units in each test set was not known by any of the analysts.

The validation procedure showed that the ITS-MUNE method provided an accurate and sensitive measure of the number of motor units in the test database. This is shown by the linear relationship between the estimated and true number (N) of motor units in the test database. 

Hypothesis testing revealed that the slope (m = 1.19) was significantly different from unity (the null hypothesis of m = 1 is rejected with P < 0.0001). There was a bias toward underestimation of the true number of units, as shown by the negative y-intercept. Because the slope of the regression line is greater than unity, the negative bias is most evident in muscles with fewer motor units.

Separate regression analysis of expert and naive analysts also showed slopes near, but significantly different from unity. The null hypotheses of m_ex = 1 and m_na = 1 were rejected with P < 0.0001 and P = 0.011, respectively. Although the expert regression line (m_ex = 1.26) was steeper than naive (m_na = 1.13), t-test did not reveal significant differences in the slope or y-intercept (b) of the two analyst groups. The null hypotheses...
The first issue to critically examine is whether the models capture the essential neuromuscular physiology, we conclude that the ITS-MUNE method gives a good estimate of the number of motor units in our models. To the extent that our models are confident that the force output of the mouse hindlimb muscles has been well described by the model.

The stochastic axon threshold model includes assumptions that would benefit from more extensive experimental support. Experiments have established that the threshold of a single motor axon follows a curve that is well fit by a cumulative Gaussian function (Bergmans 1970; Brown and Milner-Brown 1976; Hales et al. 2004; Milner-Brown and Brown 1976; Stein and Yang 1990). The cumulative Gaussian function has two parameters, mean and SD, and appropriate values for these parameters across an entire pool of motor axons have not been measured. We generated random mean thresholds from a normal distribution that were constrained to the range of stimulus values in experimental stimulus-response curves. The choice of a normal distribution of mean thresholds, rather than a uniform distribution for example, was made because it resulted in a sigmoid relationship between stimulus amplitude and muscle force, whereas a uniform distribution of mean thresholds resulted in a linear relationship. On average, the experimental data showed a sigmoid relationship, therefore the choice of a normal distribution.

Hales et al. (2004) measured the relative spread (SD/mean) of a sample of human motor axons. We used their value of 1.65% to calculate the appropriate SD for the cumulative Gaussian threshold function of each axon in the population (Eqs. 1 and 2). While this is currently the best empirical data available for modeling this parameter, it is limited. The sample was small (n = 9) and obtained from multiple subjects, and all the axons tested were the lowest threshold in the nerve. Ideally relative spread data from a larger number of motor axons, in the same individual, with a range of mean thresholds should be acquired to test the validity of extrapolating the value of 1.65% to the entire population of healthy motor axons. To extend the model to neurodegenerative disease, peripheral neuropathy, or

Advantages and disadvantages of the models

The first issue to critically examine is whether the models are a good representation of the neuromuscular physiology and methodological paradigm used in ITS-MUNE. The muscle twitch model (Eq. 3) is the impulse response function of a critically damped, second-order visco-elastic system that has been previously tested under isometric conditions (Mannard and Stein 1973; Stein et al. 1972). This model provides a good description of the linear behavior of mammalian skeletal muscle in response to a single stimulus in isometric conditions, as in this study. The free parameters of the muscle model were fit to the experimental data gathered from the mouse, and we tested the validity of the model by comparing the range of whole muscle forces (Table 2). Anecdotally, after addition of the simulated “noise” to the twitch forces (Fig. 1B), the compound muscle forces examined in the ITS-MUNE software were indistinguishable from the experimental data. Thus we are confident that the force output of the mouse hindlimb muscles has been well described by the model.

The results show that the proposed “incremental twitch subtraction” motor unit number estimation method (ITS-MUNE) is an efficient method to determine the number of motor units in our models. To the extent that our models are available for modeling this parameter, it is limited. The sample was small (n = 9) and obtained from multiple subjects, and all the axons tested were the lowest threshold in the nerve. Ideally relative spread data from a larger number of motor axons, in the same individual, with a range of mean thresholds should be acquired to test the validity of extrapolating the value of 1.65% to the entire population of healthy motor axons. To extend the model to neurodegenerative disease, peripheral neuropathy, or
injury, the values of mean threshold and relative spread would have to be measured to determine if there are changes relative to healthy axons.

Having pointed out these limitations in modeling the axon threshold, it should be noted that the stimulus-response characteristics of the model produce data with qualitative characteristics that match experimental results. Therefore, we suggest that the models capture the essential neuromuscular physiology and are a good test of the validity of the ITS-MUNE method applied to the experimental mouse data.

**MUNE in animal studies**

This work extends previous rodent studies that have estimated the number of motor units using anatomical and physiological methods. Anatomical methods have counted axons (Eisen et al. 1974; Pun et al. 2006) or immunohistologically stained ventral horn neurons (Arasaki et al. 1997; McHanwell and Biscoe 1981). In addition, there have been qualitative counts of motor neurons used to test possible treatments of motor neuron degeneration (Gao et al. 2005; Storkebaum et al. 2005). The primary limitation of the anatomical methods is, that while they accurately measure the number of intact axons, they do not reveal if the neuromuscular junctions are functional. Therefore, if a novel treatment exhibits potential based on anatomical studies, a logical next step would be to show the number of axons that successfully form functional neuromuscular junctions.

Physiological measurements of functional connectivity have used force output (Fu and Gordon 1995a, b; Kieran et al. 2004; Sharp et al. 2006); electromyography (Azzouz et al. 1997; Eisen et al. 1974; Kennel et al. 1996; Shefner 2001; Shefner et al. 1999), or both (Arasaki et al. 1997) as the outcome measure. The majority of these studies used a variant of McComas’ original incremental MUNE method that is confounded by the problem of stimulus-to-stimulus alternation in axon responses (McComas 1998; McComas et al. 1971). In low-force muscles, such as the mouse EDL, the motor unit number estimates ranged from 24 to 28, (Kieran et al. 2004; Sharp et al. 2006), which is consistent with our lowest estimates (Table 1). In those studies, the number of force increments was visually assessed between threshold and maximum using manual increments in stimulus amplitude. It is possible that small motor unit twitches, not easy to recognize visually, are missed in the assessment leading to an underestimate of the final MUNE and the difference with our results. The advantage of the ITS-MUNE over the visual counting methodology becomes apparent with larger muscles, where it is not possible to visually resolve and count stepwise increments in force over the full range of muscle force output (e.g., Fig. 4, left).

In the same way that we have used mathematical models as a means of providing validation for the ITS-MUNE methodology, two previous rodent studies included validation of alternative MUNE methodologies using anatomical data (Arasaki et al. 1997; Eisen et al. 1974). Using an incremental MUNE method, the average motor unit response in the rat soleus muscle was estimated from the first 10 incremental responses to increasing stimulus amplitudes and compared with anatomical counts of axons (Eisen et al. 1974). The ITS-MUNE improves on this methodology by estimating the average response from a random sample of a broad range of data. In addition, the ITS-MUNE method uses the presence of alternation to rank order and subtract responses to estimate twitch force of a single motor unit.

The intraneural microstimulation method is immune to the problem of alternation (Arasaki et al. 1997). Using this method, the number of motor units in the rat medial gastrocnemius muscle was estimated to be 93 ± 22 using EMG and 117 ± 21 using force output. These values were in good agreement with the value of 103 ± 16 motor neuron somata that were counted in the lumbar spinal cord after retrograde labeling in a separate population of rats (Arasaki et al. 1997). The intraneural microstimulation method, however, requires isolating single fibers from the ventral root projecting to the muscle of interest. Thus the surgical dexterity required is much greater than that needed for ITS-MUNE, limiting widespread use.

**Limitations of ITS-MUNE**

The ITS-MUNE method is not without some shortcomings. First, the MUNE does not match the absolute number of motor units because there is a clear underestimation bias (Fig. 7). The presence of underestimation indicates that our estimate of the average twitch force is too large. One reason for the overestimation could be poor signal-to-noise ratio for twitches <2 mN. Sampling of small twitches could be enhanced by improved isolation of the experimental setup from vibrations and electromagnetic interference or introducing additional signal processing before calculating the subtracted twitch force.

Motor unit twitch force would also be overestimated whenever the difference between two subtracted compound muscle force waveforms is greater than a single unit. For example, if the threshold functions for three units overlap at a given stimulus intensity, there are seven possible force responses that could result (Fig. 6 in Major and Jones 2005). The largest peak force response would result from synchronous stimulation of all three units: A + B + C. If the next smallest waveform happens to be the twitch response of unit A, the subtracted response is the net effect of two units rather than one.

On investigating the nature of twitch overestimation by ITS-MUNE, it was discovered that motor unit twitch estimates are biased by the incremental force level at which the units are recruited (see Testing for sampling bias). There are three factors that change with stimulus intensity that could potentially contribute to the observed trend: 1) decreased resolution of the A/D conversion (ADC) as incremental force increased, 2) systematic instead of random recruitment order, and/or 3) the number of alternating motor units. At the lowest amplifier sensitivity of 20 mN/V, the resolution was 0.25 mN. This could interfere with the identification of very small twitches but did not preclude finding small twitches at high intensities. In high-force muscles, the smallest twitches extracted were 0.68, 0.62, 0.88, and 0.54 mN at gain settings of 20, 40, 100, and 200 mN/V, respectively. Furthermore, when twitches <2 mN were excluded from the data set, the trend remained strong—twitch estimates grew by 0.04 to 0.09% for each millinewton (P = 0.0036, n = 728). We conclude that ADC resolution may have contributed to the sampling bias, but does not account for it completely.

Based on the theoretical prediction of reverse recruitment and experimental evidence of random recruitment, it is doub-
ful that orderly recruitment caused the positive correlation between motor unit twitches and force level (Erlanger 1937; McNeal 1976). Moreover, twitches extracted using ITS-MUNE from simulated high-force muscle data showed a similar, albeit weaker, trend where twitch estimates grew by 0.03 to 0.07%/mN ($P = 0.0087$, $n = 325$). The mathematical model used to generate simulated force data did not feature orderly recruitment—recruitment was random with respect to stimulus intensity—so in this case, the correlation between twitch estimates and incremental force level was caused by some other factor. Nor did the mathematical model include the effect of ADC resolution—there was no ADC quantization involved, and equal amounts artificial noise were added to every force trace (Fig. 1B). Therefore the trend observed within ITS-MUNE twitch estimates from simulated force data were certainly caused by the only other possibility, motor unit alternation.

All considered, it is reasonable to conclude that alternation was the most significant factor that led ITS-MUNE to detect larger twitches at higher intensities. The underlying mechanism of this effect is not mysterious: Motor unit thresholds are usually well separated at low stimulus intensities and more clustered about the intermediate portion of the submaximal stimulus range (the frequency distribution of mean thresholds is considered to be bell-shaped). The greater the number of motor unit thresholds near a particular stimulus strength, the more instances of response alternation will ensue when that stimulus is applied repeatedly. Given a fixed number of stimulus-response observations, a growing number of alternating motor units drastically reduces the probability of capturing a true motor unit force increment while increasing the occurrence of increments that are caused by concurrent activation of more than one unit. Therefore the chance of twitch overestimation increases as stimulus is increased from minimum to mid-range.

Motor unit alternation is not only more prevalent at mid-range stimulus intensities, it is also related to the total number of motor units in the muscle. Statistical analysis by ANOVA showed that average twitch estimates are affected by muscle size. For high-force muscles, twitch estimates grew by 0.16 to 0.20% for each millinewton increase in whole muscle force ($P = 0$). The growth factor was $0.65–0.80\%$ for each millinewton for low-force muscles ($P = 0$). These observations support the conclusion that ITS-MUNE twitch overestimation is, for the most part, caused by too few stimulus-response observations in the context of severe motor unit alternation.

Because the slope of the regression line in Fig. 7 was 1.19 and significantly greater than 1, the overestimation of the average twitch force is greater for smaller muscles with fewer motor units. This result is apparently contradictory to the conclusion that ITS-MUNE twitch overestimation is worse in muscles with more motor units. In fact, the validation results presented in Fig. 7 are strongly influenced by the number of stimulus-response observations acquired for each muscle. The smallest muscle received only 90 stimuli (3 $\times$ 30 intensities), whereas 240 stimuli were delivered to the largest muscle (8 $\times$ 30 intensities). The simulated protocol was designed this way for two reasons: 1) to mimic the experimental method where smaller muscles are more susceptible to deterioration and cannot tolerate as many stimuli as large muscles and 2) to compensate for more motor unit alternation in large muscles. It is understood in retrospect that three observations per stimulus intensity are not enough to overcome any amount of motor unit alternation.

Two recommendations follow from the analysis presented above: When collecting force data by incremental stimulation for analysis by ITS-MUNE, 1) observe $\geq 10$ responses at each stimulus intensity and 2) restrict the range of stimulus strengths to less than one third of supramaximal. It is also important, as with any MUNE technique, that the experimentalist perform the stimulation protocol consistently across all muscles from which MUNES are to be compared. These recommendations could be implemented by automating the stimulation using a computer controlled stimulator rather than manually changing the stimulus amplitudes. Otherwise, normative data should be assembled on a laboratory-by-laboratory basis.

Perhaps another limitation of ITS-MUNE is that it does not allow for longitudinal studies from the same animal. Longitudinal studies of MUNE in mice have been carried out using EMG as the outcome measure (Shefner 2001; Shefner et al. 1999, 2002). The limitation of these studies was that the EMG measurements come from a large number of muscles and not a single muscle group. Therefore differences in rates of muscle denervation of individual muscles cannot be resolved. One possibility for extending the ITS-MUNE method to longitudinal studies would be to measure intramuscular pressure rather than isometric forces (Davis et al. 2003; Kaufman et al. 2003). The miniaturization of intramuscular pressure sensors makes them an ideal candidate for repeat in vivo measurement of muscle force. Whether the delicate muscles of the mouse hindlimb would withstand repeat insertion of these microsensors remains to be evaluated.

We introduced a new analysis method for assessing the health of the neuromuscular system by estimating the number of functional motor units within muscles of the mouse hindlimb using incremental stimulation. While the method underestimates the absolute number of motor units in our simulated data, it provides a linear estimate from small to large muscles, i.e., 40–120 motor units. The coefficient of variation of the MUNE is constant over this range with a value of 10.6%, indicating good reliability across naïve and expert analysts. The ITS-MUNE method has the potential to provide insights into the pathophysiology of muscle denervation and implications for testing new therapies in animal models.

Acknowledgments

Present address for D. Weber: University of Pittsburgh, Department of Physical Medicine and Rehabilitation, Pittsburgh, PA.

Grants

This study was supported by grants from The Whitaker Foundation, The Canadian Institutes of Health Research, and the National Response Plan. Studentships were supplied by the Alberta Heritage Foundation for Medical Research (AHFMR) and the Natural Sciences and Engineering Research Council of Canada. Kelvin Jones and Tessa Gordon are AHFMR investigators.

References

