Published in final edited form as:

J Electromyogr Kinesiol. 2015 October ; 25(5): 754–764. doi:10.1016/j.jelekin.2015.05.006.

Reproducibility of transcranial magnetic stimulation metrics in the study of proximal upper limb muscles

Vishwanath Sankarasubramanian¹, Sarah Roelle¹, Corin E Bonnett¹, Daniel Janini¹, Nicole Varnerin¹, David A Cunningham¹, Jennifer S Sharma¹, Kelsey A Potter-Baker¹, Xiaofeng Wang⁴, Guang H Yue⁵, and Ela B Plow^{1,2,3}

¹Department of Biomedical Engineering, Cleveland Clinic, Cleveland, Ohio

²Department of Physical Medicine and Rehabilitation, Cleveland Clinic, Cleveland, Ohio

³Center for Neurological Restoration, Cleveland Clinic, Cleveland, Ohio

⁴Department of Quantitative Health Sciences, Cleveland Clinic, Cleveland, Ohio

⁵Human Performance and Engineering Research, Kessler Foundation, West Orange, New Jersey

Abstract

Objective—Reproducibility of transcranial magnetic stimulation (TMS) metrics is essential in accurately tracking recovery and disease. However, majority of evidence pertains to reproducibility of metrics for distal upper limb muscles. We investigate for the first time, reliability of corticospinal physiology for a large proximal muscle—the biceps brachii and relate how varying statistical analyses can influence interpretations.

Methods—14 young right-handed healthy participants completed two sessions assessing resting motor threshold (RMT), motor evoked potentials (MEPs), motor map and intra-cortical inhibition (ICI) from the left biceps brachii. Analyses included paired t-tests, Pearson's, intra-class (ICC) and concordance correlation coefficients (CCC) and Bland-Altman plots.

Results—Unlike paired t-tests, ICC, CCC and Pearson's were >0.6 indicating good reliability for RMTs, MEP intensities and locations of map; however values were <0.3 for MEP responses and ICI.

Conclusions—Corticospinal physiology, defining excitability and output in terms of intensity of the TMS device, and spatial loci are the most reliable metrics for the biceps. MEPs and variables based on MEPs are less reliable since biceps receives fewer cortico-motor-neuronal projections. Statistical tests of agreement and associations are more powerful reliability indices than inferential tests.

Corresponding Author: Ela B Plow PhD PT, Assistant Staff, Department of Biomedical Engineering 9500 Euclid Ave, ND20, Cleveland Clinic, Cleveland, OH 44195, Phone No: 216-445-4589, Fax No: 216-444-9198, plowe2@ccf.org.

Conflicts of Interest: None of the authors have potential conflicts of interest to be disclosed

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Significance—Reliable metrics of proximal muscles when translated to a larger number of participants would serve to sensitively track and prognosticate function in neurological disorders such as stroke where proximal recovery precedes distal.

Keywords

Transcranial magnetic stimulation; Reliability; Primary motor cortex; Proximal muscles; Corticospinal; Motor evoked potential

1. Introduction

Transcranial magnetic stimulation (TMS) is a non-invasive neurophysiologic technique to measure excitability of cortices and their output pathways in the brain. It's most popular use concerns the study of physiology of the primary motor cortex (M1) and emergent corticospinal projections. By measuring the amplitude of evoked motor potentials (MEPs) in muscles in response to TMS, one can estimate corticospinal excitability and output devoted to the muscle (Baker, 1985; Barker et al., 1987; Liepert et al., 2000; Rossini et al., 1994; Bastani and Jaberzadeh, 2012). The entire corticospinal output projecting to the muscle can also be plotted as what is commonly referred to as the motor map. Finally, TMS can be used to determine parameters of cortico-cortical physiology that shape corticospinal excitability and output, such as intra-cortical inhibition (ICI), which represents the influence of cortical interneurons (Kujirai et al., 1993). Because of its ability to define even subtle changes in motor cortical and corticospinal physiology, TMS is fast becoming popular in clinical applications. Hundreds of studies in neurologic populations such as stroke and spinal cord injury attribute time-varying change in TMS metrics to physiologic mechanisms underlying disease and recovery (Stinear et al., 2007; Trompetto et al., 2000; Liepert et al., 2000; Wittenberg et al., 2003). However, to ensure that TMS can indeed serve as a clinical tool to track longitudinal processes, it is first critical to understand test-retest reliability of its metrics in healthy individuals (Malcolm et al., 2006; Pourtney and Watkins, 2000; Christie et al., 2007; Doeltgen et al., 2009; De Vet et al., 2006; Mills and Nithi 1997; Kimiskidis et al. 2004).

At this time, several different studies have reported that TMS metrics are generally reliable. However, the majority defines reliability of metrics for distal muscles of the upper limb (Christie et al., 2007; Kamen, 2004; Livingston and Ingersolle, 2008; Malcolm et al., 2006; McDonnell et al., 2004; Carroll et al., 2001; Bastani and Jaberzadeh, 2012). Distal muscles have been the muscles of choice because they are afforded with prominent cortical representations and substantial corticospinal projections (Malcolm et al., 2006; Kamen, 2004; Brasil-Neto et al., 1992). Thus, they are extremely responsive to TMS such that MEPs are easy to acquire and are sufficiently large to study (van Kuijk et al., 2009; Roick et al., 1993; Cantello et al., 1992; Rossini et al., 1994; Mills and Nithi, 1997). Test-retest reliability of metrics for proximal muscles is, however, lacking. Proximal muscles are often as relevant as distal in tracking neurologic recovery; stronger proximal muscles typically serve to compensate for poor dexterity (Canning et al., 2000). In addition, in neurological conditions, such as stroke and cervical spinal cord injury, the recovery pattern generally shows restoration of proximal function before distal, which supports a majority of the acute, and

sub-acute recovery, and initial functional independence (Colebatch et al., 1990; Rudhe et al., 2009). Therefore, understanding reliability of TMS metrics for proximal muscles becomes a priority for realizing its indication for clinical use.

Van Kuijk and colleagues (van Kuijk et al., 2009) have compared TMS metrics between a proximal muscle (biceps brachii) and a distal intrinsic muscle of the hand (abductor digiti minimi, ADM). They have concluded that corticospinal physiology for biceps brachii shows high inter-individual variability compared to that for ADM. Their finding raises an important question- does the high inter-subject variability of TMS metrics predispose proximal muscles to poor test-retest reliability? To answer this, for the first time, we investigate reproducibility of several key TMS metrics defining parameters as corticospinal excitability and output, ICI, and physiology of motor maps for biceps brachii. Our emphasis is novel because proximal muscles are typically less studied than distal owing to the challenge in assaying with TMS. For instance, fewer cortico-motor-neuronal projections that are spread over a relatively wider area (Brasil-Neto et al., 1992), we have recently reported, render their MEPs small and extremely variable (Plow et al., 2013, 2014). Along similar lines, examining reliability of all major metrics is important, because in the study of distal muscles, a divergence is identified, where certain measures are more reliable than others (Malcolm et al., 2006).

Since, to the best of our knowledge, our study remains the first to explore test-retest reliability of TMS metrics for a large, proximal muscle, it is critical for us to compare interpretations across all- standard and novel methods of reliability analyses. Therefore, our emphasis is also novel because unlike prior studies (van Kuijk et al., 2009) we relate how varying statistical methods of test-retest reliability can influence interpretations of stability of outcomes. By knowing which metrics are most reliable for proximal muscles and which statistical tests aide in establishing reliability, would improve planning and design of longitudinal studies tracking recovery and prognosticating upper extremity function in clinical populations.

2. Methods

2.1 Subjects

Twenty young (23 ± 4.02 years, 10 females), healthy participants were enrolled. All subjects were right-hand dominant, confirmed by the Oldfield handedness test (Oldfield, 1971), and were not involved in any systematic upper limb training for a period of 5 years before enrollment. Exclusion criteria were established based on contraindications to TMS (Rossi et al., 2009) and were intended to remove any confound of neurological or musculoskeletal condition affecting upper limbs. All subjects provided informed consent prior to participation. The Institutional Review Board of the Cleveland Clinic approved the experimental protocol.

2.2 Overview of Procedures

We conducted a pilot study assessing the test-retest reliability of several of the key TMS metrics defining cortico-cortical and corticospinal physiology for the left biceps brachii muscle. Subjects underwent two identical sessions, namely Test 1 and Test 2, separated by

at least eight weeks (Figure 1), wherein they were asked to refrain from any training or intervention in the interim to ensure that differences in measures from the tests were mainly related to TMS methodology. Within each session, single-pulse and paired-pulse TMS were applied to the right hemisphere for studying metrics for left biceps brachii. We chose to investigate metrics for the non-dominant, left biceps brachii, because this study was part of a larger study evaluating the weaker left arm in healthy young versus an aged population (Plow et al., 2014), where the differences in corticospinal excitation between young and old were most accentuated on the left side in right-handed participants.

2.3 TMS Procedures

Participants were seated in a chair with both arms resting in slight shoulder abduction ($\sim 10^\circ$), elbow flexion (90°), and forearm in neutral position, between pronation and supination. TMS was applied using a figure-of-eight coil (70 mm diameter) connected to one or two Magstim devices (200² and Bistim device, Magstim Co., Dyfed, UK). The coil was held tangential to the scalp with the handle oriented backwards and laterally at 45° from the midsagittal axis, which is approximately perpendicular to the central sulcus and M1, and is believed to optimally stimulate the corticospinal tracts (Di Lazzaro et al., 2004). The position of the coil was guided by frameless stereotaxy (Brainsight, Rogue Research Inc., Montreal, Canada) using a magnetic resonance imaging (MRI) template as a reference. Each subject's head was registered to the MRI template using defined cranial landmarks, to ensure the coil position and orientation was realized with respect to that of the MRI template (Ruohonen and Karhu, 2010).

Surface electromyography (EMG) electrodes (Ag/AgCl, 45 mm diameter) were positioned over the middle of the muscle belly of the left biceps brachii. A bipolar montage was used while a reference electrode was placed over the acromion. MEP signals were amplified, band-pass filtered (10 Hz-2 kHz), digitized (4 kHz; PowerLab 4/25T, AD Instruments Inc., Colorado Springs, CO), and stored on a computer for offline analysis (Scope, version 4.0.8).

Single-Pulse TMS—We identified the optimal site for targeting biceps brachii in the right hemisphere. The optimal site, termed the “hotspot”, was the site that elicited MEPs with peak-to-peak amplitude of at least 50 μ V, in 3 out of 5 trials at the lowest intensity. The lowest intensity was defined as the resting motor threshold (RMT) and was determined using the classical approach as suggested by Rossini et al (Rossini et al., 1994). TMS was started at sub-threshold intensity, as percentages of maximum stimulator output (MSO) and gradually increased in steps of 5% MSO until TMS consistently evoked MEPs of more than 50 μ V in each trial. Thereafter, stimulus intensity was gradually lowered in steps of 1% MSO until less than 3 out of 5 trials met the criterion peak-to-peak amplitude. Once this level was identified, the stimulus intensity was raised in steps of 1% MSO until at least 3 out of 5 trials evoked the criterion response. This intensity was defined as the RMT. Subsequently, supra-maximal responses were elicited using higher (supra-threshold) intensities. Supra-threshold intensities also referred to in this paper as supra-maximal MEP intensities (Table 1), signified intensities ranging up to 95% of MSO that generated supra-maximal MEPs in the range of 0.1-0.7 mV in 3 out of 5 trials. MEPs within the range of 0.1-0.4 mV were categorized as ‘MEPs with lower amplitude’ whereas those within 0.4-0.7

mV were categorized as ‘MEPs with higher amplitude’. The criterion ranges of supra-maximal MEPs were set lower than that which is generally adopted for study of distal hand muscles (~1 mV) because it is difficult to evoke larger MEPs from biceps brachii (Chen et al., 1998; Harris-Love et al., 2007; Plow et al., 2014). A wider range was chosen because MEPs elicited in biceps brachii are more variable (Brasil-Neto et al., 1992; Plow et al., 2013,) than those in distal muscles (van Kuijk et al., 2009).

Motor Map—Using single-pulse TMS, a motor map was created over the right hemisphere. Stimulation was applied at various scalp sites at an intensity of 110% RMT (Sawaki et al., 2008), while the left biceps brachii remained at rest. Starting from the hotspot, sites at incremental distances of 3 mm were targeted in eight radial directions. Our mapping procedure involves a higher-than-typical spatial resolution (3 vs 10 mm) because Brasil-Neto and colleagues have recommended that weaker cortical representations of proximal muscles can be accurately studied only at high spatial resolution than that used for study of distal (Brasil-Neto et al., 1992). Sites were deemed responsive if they produced peak-to-peak MEPs of at least 10 μ V, over two trials as discussed in our recent work (Plow et al., 2014) and as used by others (Streletz et al., 1995; Marconi et al., 2011). The reason the criterion for motor mapping (MEPs of at least 10 μ V) was set at a lower amplitude than that which is typically used for RMT was that in a proximal muscle, such as biceps brachii, it is difficult to obtain MEPs of higher amplitude (Chen et al. 1998 and Harris-Love et al. 2007) unlike in a distal muscle (Rossini et al. 1994 and Mills et al. 1997). By choosing a criterion of 10 μ V thus, we intended to capture the full extent of the representation of the corticospinal output devoted to the biceps brachii muscle. Mapping was continued in each radial direction until two consecutive sites generated no discernible response.

Cortico-cortical Interactions: intra-cortical physiology—Using paired-pulse TMS, we investigated ICI within the right hemisphere. When paired pulses are delivered at short inter-pulse intervals (1–5 ms), MEP amplitude generated by the 2nd pulse can be inhibited (Ziemann et al., 1996). The 1st pulse, or the conditioning pulse, was sub-threshold and was set at an intensity of 90% RMT (Chen et al., 1998). The subsequent pulse (test stimulus) was delivered at the supra-threshold intensity that was used to evoke supra-maximal MEP (in this case 0.1–0.7 mV). We delivered paired pulses at intervals of 1 through 5 ms (Perez et al., 2004). However, we only report ICI at 2 ms because most of our subjects exhibited peak inhibition at this interval as also noted in our previous report (Plow et al., 2013).

2.4 TMS Metrics and TMS Parameters

We studied TMS metrics shown in Figure 1, definitions and analyses for which are presented in Table 1. TMS parameters described by each metric are shown in Figure 2.

2.5 Statistical analysis

All data was tested for normality using the Shapiro-Wilk test. A paired t-test was used to ensure that the range of supra-maximal MEPs used for ICIs were comparable between the two sessions.

2.6 Reliability Analysis

We studied the test-retest reliability of all of the mentioned metrics of TMS using the statistical software, R version 3.1.1 (R Foundation for Statistical Computing, Vienna, Austria). We first used paired t-test to assess whether the systematic changes in means of the two sessions were statistically different from each other (Portney and Watkins, 2009). Secondly, the correlation between their measurements was assessed using Pearson's correlation coefficient, abbreviated as 'Pearson's'. To assess test-retest reliability, we used three well-established statistical correlation approaches: intra-class coefficient (ICC), concordance correlation coefficient (CCC), as well as Bland-Altman graphical analysis. ICC is defined as the ratio of between-subject variance to the total variance and can be interpreted as a correlation coefficient. On the other hand, CCC, considered as a statistical relative of ICC, is a non-parametric moment method that evaluates the degree to which pairs of observations fall on the 45 degree line through the origin. Although the CCC is a non-parametric method, it is compared against a parametric method as in ICC. This is because the CCC takes into account not only mean differences between the first and the second measurement such as ICCs, but also takes into account variance differences between the first and the second measurement by reducing the magnitude of the resulting test-retest reliability estimate. Both ICC and CCC are scaled statistical indices attaining values between -1 and 1. The Bland-Altman plot (Bland and Altman, 1999) is a graphical method to evaluate the agreement between test-retest data, where the differences between two measurements are plotted against the averages of the two measurements. It allows us to investigate the existence of any systematic difference between test-retest (i.e., fixed bias) and to identify possible outliers. We also presented 95% limits of agreement (average difference $\pm$ 2 standard deviation of the difference), which tells us how far apart measurements by two methods were more likely to be for most individuals. Finally, to further quantify the reliability of the method, the measurement error (ME) was used. Because the ICC does not allow us to fully appreciate the magnitude of the within-subject variance, we also calculated the ME. It is the value below which the absolute differences between two measurements would lie with 0.95 probabilities. The ME was estimated by dividing the standard deviation of the mean differences by the square root of 2, and was reported when good to excellent reproducibility was found (ICC>0.6).

3. Results

Of the 20 young participants enrolled in the study, 14 completed test 1 and test 2. Motor maps could be recorded only from 12 subjects because 2 subjects had no other scalp sites except the TMS hotspot that elicited any level of criterion MEP response in left biceps brachii. Reliability results for TMS metrics based on each statistical test are presented in Table 2.

3.1 Corticospinal excitability and output

As shown in Figure 3(a) corticospinal excitability, defined as RMT intensity (%) and (b) corticospinal output, defined as supra-maximal MEP intensity (%) were more reliable than (c) corticospinal output, quantified as supra-maximal MEP (mV). ICC, CCC and Pearson's showed values >0.6 (Table 2), which supported the illustrations demonstrating a good

reliability for RMT intensity and supra-maximal MEP intensity. However, paired t-test showed that RMT intensity varied significantly from one test to the next because the p-value was less than 0.05 (0.011) (Table 2).

Contrarily, while the t-test did not reveal test-retest difference for supra-maximal MEPs (p-value >0.05), ICC, CCC and Pearson's showed values <0.3 (Table 2) that indicated poor reliability of supra-maximal MEPs (mV). Moreover, from the sample MEPs obtained from points 1 and 2 demonstrated in figure 3(c)), it is evident that supra-maximal MEPs with higher amplitudes were less reliable than supra-maximal MEPs with lower amplitudes (Figure 3, bottom).

3.2 Motor map output and spatial distribution

Depicted are plots of motor map area (Figure 4(a)) and motor map volume (Figure 4(b)). While t-test values indicated good reliability of motor map area and normalized map volume (Table 2), ICC, CCC and Pearson's showed weak to moderate test-retest reliability that coincided with the illustration suggesting only moderate reliability (Figure 4(a), 4(b)). Also shown are the contour plots (Figure 4(c)) that depict the distribution of map volume for a subject with less reliable map volume from test 1 to test 2. It was observed that the region of highest MEP amplitudes (the area in red on contour plots) on the map moved approximately 10 mm lateral and 10 mm posterior from test 1 to test 2. Also, on the map of test 2, the region of highest MEP amplitude was larger and focused to one location as compared to the map of test 1. These changes show that although the overall map volume seem to have changed from test 1 to test 2 indicating poor reliability, the region of map volume with highest MEP amplitude remained generally around the same location.

COG was found to be more reliable (Figure 5(c) and 5(d)) than the location of the TMS hotspot (Figure 5(a) and 5(b)). Also, the antero-posterior (y) coordinates of COG (Figure 5(d)) and hotspot (Figure 5(b)) were more reliable than medio-lateral (x) coordinates (Figure 5(c)) and (Figure 5(a)). ICC, CCC and Pearson's showed values >0.6 (Table 2) that suggest good reliability of COG-y and Hotspot-y. This difference in reliability of x vs y coordinate was also ascertained from Figure 4(c), where the contour plot from Test 2 showed more sample points converging in a particular y area but more spread out in the x direction.

3.3 Paired-pulse interactions

The Percent ICI plot (Figure 6) showed that there are two data points outside the acceptable range, which could be the cause of the low reliability results, as also seen from ICC, CCC and Pearson's value in table 2. However, when these two outliers were removed, ICI agreement did not improve between test 1 and test 2 and hence are not shown.

4. Discussion

The purpose of the study was to assess test-retest reliability of several of key TMS-based neurophysiologic metrics defining cortico-cortical and corticospinal physiology for the biceps brachii muscle. We have found that corticospinal excitability and output when expressed as RMTs and supra-maximal MEP intensities are the most reliable metrics. Reliability of corticospinal output expressed as amplitude of MEPs is however compromised

due to the variable nature of MEPs elicited in biceps brachii - only smaller MEP responses are reliable suggesting that setting a lower MEP criterion within a set range (0.1-0.4 mV) would be more sensitive for proximal muscles. Map volume and intra-cortical metrics that are inherently related to amplitude of MEPs are thereby poorly reproduced as well. COG was more reliable than using a single hotspot locus to define the spatial distribution of corticospinal excitability. Comparison across statistical approaches reveals that while paired-t tests report good reliability for most metrics, tests that are not influenced by means would be accurate and more suitable indices for reliability. These include tests of correlation that assess either association (Pearson's) or association and systematic differences (ICC, CCC) between the metrics. Biceps brachii, being a proximal large muscle, is especially important in founding subsequent recovery of the hand in neurologic populations (Kumar et al., 1989). However, based on present findings and previous evidence from our work and of others (Plow et al., 2013; Plow et al., 2014; van Kuijk et al., 2009; Brasil-Neto et al., 1992), it becomes clear that the variable nature of its MEPs renders test-retest measurements of its TMS metrics weakly reproducible. Nevertheless, choosing metrics that define its excitability and output in terms of intensity of TMS device (RMTs or supra-maximal MEP intensities) and/or spatial spread of excitability (COGs and hotspots) would serve as robust markers to longitudinally track recovery and prognosticate upper extremity function across populations, such as stroke, and spinal cord injury.

4.1 Corticospinal excitability and output

Our results demonstrate that RMTs remained relatively reliable when expressed as percentage of the maximum stimulator output (Figure 3 and Table 2). Although the paired t-test showed a p value < 0.5, and a high ME (>5% of the MSO), the ICC, Pearson's, and CCC demonstrated excellent reliabilities for RMT. Our studies are generally in line with studies showing moderate-to-excellent RMTs of distal muscles (Mills and Nithi, 1997; Stewart et al., 2001; Cicinelli et al., 1997; Maeda et al., 2002; Malcolm et al., 2006; Wolf et al., 2004; Corneal et al., 2005; Kimiskidis et al., 2004). The discrepancies in the absolute values of the RMTs and the MEs between studies could be due to the following reasons- differences in the type of musculature (Brasil-Neto et al., 1992; Chen et al., 1998), time interval between testing sessions, hemisphere (right/left) stimulated, and type of statistical indices used. Despite this, good test-retest reliability of RMTs suggests that they have the potential to be developed as a reliable outcome measure to illustrate corticospinal excitability for biceps brachii and perhaps even other proximal extremity muscles. It is also worth acknowledging that different validated scientific methods, especially the adaptive techniques utilizing threshold-tracking algorithms could provide greater advantage over the classical method (Groppa et al. 2012) in (a) alleviating RMT differences with proximal muscles, and (b) faster RMT estimation with fewer stimuli. However, because the classical method is employed in a majority of TMS studies to date and has become a standard technique of RMT estimation, we adopted to study TMS reliability metrics based on this standard.

Unlike RMT, we found poor reliability of corticospinal output expressed as amplitude of supra-maximal MEPs (Figure 3(c)). Variability of MEPs from biceps brachii is higher than that of distal muscles (van Kuijk et al., 2009). The reasons may be that for proximal

muscles, MEP responses increase more gradually with supra-maximal intensities, i.e. the gain of the corticospinal system is smaller and more variable than in distal. This lower gain, which reflects the density of cortico-motor-neuronal projections onto the spinal motor neurons, is potentially why we observe poor test-retest reliability of supra-maximal MEPs.

Interestingly, lower ranges of MEP amplitudes were more reliable than larger ones (Figure 3 bottom). Setting a lower criterion range of MEP amplitude, in this case, 0.1-0.4 mV for a proximal muscle as biceps brachii would be more feasible, yet robust. Our suggested range aligns with Chen et al. who set a lower limit for their MEPs for biceps because in many subjects, it is difficult to obtain MEPs of higher amplitude (Chen et al., 1998). Their reported values were 0.18 ± 0.04 mV. Similarly, Harris-Love et al. discuss that they set a lower range up to 0.3 mV to record from biceps brachii (Harris-Love et al. 2007). Such a range is important for recording from biceps because its representation has such scarce corticospinal neurons distributed over a relatively wider area. Consequently, compared to the more excitable distal muscles, MEPs show greater variability from trial to trial (Brasil-Neto et al., 1992).

4.2 Motor map output and spatial distribution

Motor map areas and normalized map volumes for biceps brachii showed poor reliability (Figure 4). Our results differ from studies with distal muscles that have demonstrated moderate reproducibility (Malcolm et al., 2006; Wolf et al., 2004). By definition, map volume is a function of MEP response and remains inherently less reliable (Malcolm et al., 2006). Therefore poor MEP reliability of biceps brachii naturally engenders weak reliability for map volumes. Alternatively, alterations in motor map area may occur due to spontaneous fluctuations in the M1 cortical outflow or to variations in general activity patterns over time (Malcolm et al., 2006).

Still, there are several potential reasons, both neurophysiological and methodological, why maps of biceps brachii are weakly reliable. Motor maps are reproducible if electrodes are left in place between testing sessions, while values are significantly different when they are removed (McMillan et al., 1998). Therefore, to ensure good reliability of maps from proximal muscles, precise marking and recording of electrode placement is necessary. A common concern in TMS mapping also pertains to selection of TMS site. While mapping in a rectangular grid is common, radial mapping used here may enable a greater number of data points to be collected at a better resolution that carries significance for a proximal muscle, such as biceps, with fewer cortical representation areas (Plow et al., 2014). However, testing the scalp sites between the radial directions, as well, may improve reliability further. Regardless, maps of distal muscles show better reliability than maps of biceps since their representations occupy a larger proportion of motor cortices (Malcolm et al., 2004; Mortifee et al., 1994; Uy et al., 2002).

COGs demonstrated greater reliability of motor map spatial distribution than hotspot location (Figure 5). Hotspot is considered the most excitable point of M1 and it provides a reference for setting the stimulation intensity for recording other parameters. Moreover, since hotspot location is determined during the course of the TMS session, it appeals as a diagnostically suitable measure. However, localization of hotspot is challenging and

methodology-driven. This can affect reliability of RMTs and MEPs subsequently elicited from it. In contrast, COG represents the weighted-average location of corticospinal output devoted to a muscle, and hence is likely to be more accurate than a single hotspot locus. To reinforce this finding, our results and illustrations (Figure 4(c)) demonstrate how the most weighted regions of a map tend to stay reliable over time.

4.3 Paired pulse interactions

ICI of biceps brachii showed weak agreement between the tests (Figure 6). While no studies have tested reliability of these parameters for biceps brachii, the few in distal muscles (Boroojerdi et al., 2000; Maeda et al., 2002; Orth et al., 2003) have shown mixed results. Boroojerdi et al. showed poor inter-session reliability for ICI, while Maeda et al. found ICI to be reproducible. The difference in interpretations of reliability across studies could be due to the difference in intervals used between the paired pulses. Inhibition, which is calculated as the ratio of conditioned to supra-maximal MEPs, is also dependent mostly on the MEPs. Therefore, in our study, the poor reliability of MEP response of biceps brachii could have had a major effect on the reliability of ICI measurement.

4.4 Significance of statistical measures

Paired t-test used to measure the influence of means, and the Pearson's correlation measure both suffer from many drawbacks in that neither of them alone can fully assess the desired reproducibility characteristics (Lin, 1989; Lin et al., 2002; Barnhart et al., 2007; Obuchowski et al., 2014). While the paired t-test can reject a highly reproducible data due to very small variance, the Pearson's correlation coefficient captures only the correlation or association between the measures, and not whether the measures are systematically different. In this regard, the ICC and the CCC were preferred measures for quantifying test-retest reliability. Excellent reliability was demonstrated for the following TMS metrics, namely RMTs, supra-maximal MEP intensities and COGs. Despite showing excellent reliability indicated by the respective statistical approaches, their absolute values differed. The reasons are as follows: perfect correlation does not necessarily mean complete agreement between the test-retest scores. For example with RMT, Pearson's showed a very high value of 0.843, indicating excellent correlation (Table 2). This correlation is a reflection of how closely the set of paired observations followed a straight line regardless of the slope of the line. This is not a suitable property for a reliability index since Pearson's does not provide any insight into the systematic errors that may be inherent in the measurement. Therefore, this excellent correlation can be often mistaken for excellent one-to-one agreement between the scores, which is clearly not the case, as CCC shows a value of only 0.636 (Table 2). This is because CCC indicates the systematic variation between the observed data and the 45 degree line, and not between the data and the best-fitting line, as indicated by the Pearson's. ICC, although not an ideal measure of absolute correlation, can sometimes be reported in place of the Pearson's. This is because, unlike Pearson's, the ICC accounts for both consistency of performances from test to retest (within-subject change), as well as change in average performance of participants over time (systematic change in mean). As a result, the value of ICC is decreased in situations where there is a large mean difference ($p=0.011$) between the measurements – here, ICC shows a value of 0.745 for RMT (Table 2), which is still lower than the Pearson's. Finally, a high ICC as in 0.745 for

RMT can be indicative of the heterogeneity of the participant groups (since the ICC takes into account the ratio of between subject variance to total variance). In this regard, by comparing ICC and CCC, we believe to have accounted for heterogeneity. There can be large differences in ICC and CCC scores, especially in studies with heterogeneous groups. The similar scores found in our study reflect that both coefficients worked well in this population. Therefore, in reliability studies, before selecting a statistical tool for analysis, it is important that the characteristics (homogeneity or inhomogeneity) of the participants or group and the systematic errors in measurements be reviewed. By reporting several clinically meaningful metrics (paired t-tests, Pearson's, ICC, ME) and adopting new ones as well (CCC, Bland-Altman) in our study, we believed to have discussed their differential strengths and weakness.

4.5 Limitations

Limitations of the study include limited data of test 1 and test 2 measurements from a relatively small number of participants (14), the age group (young) and the time interval (8 weeks) between the test sessions. We recruited an initial sample size ($n=20$) that is comparable to the sample sizes of several of the TMS reliability studies (van Kuijk et al. 2009, Corneal et al. 2005, Wolf et al. 2004, Maeda et al. 2002, Ngomo et al. 2012). Moreover, we adopted a larger time window between the test-retest sessions (8 weeks) as compared to most studies that employ only a smaller window (Bastani and Jaberzadeh, 2012). A large time window was chosen to mimic therapy duration generally adopted for longitudinal studies. This however reduced the number of participants from 20 who completed test 1 to 14 who completed both tests, leaving only a small sample size to conduct analysis. Although an absolute therapeutic window is unstated and unlikely, we believe to have used a relevant and consistent time frame to offer a more reasonable estimate of long-term reliability of tested parameters.

4.6 Recommendations and future work

As the first group to test reliability of a comprehensive set of TMS metrics defining physiology of a proximal muscle, namely the biceps brachii, we wanted to set forth some suggestions for future TMS studies seeking high reliability for the corticospinal physiology measurements of proximal upper extremity muscles, such as for evaluating intervention effects.

Intensity vs amplitude of MEPs—Based on the observation of poor reliability of supra-maximal MEP amplitudes and strong reliability of supra-maximal MEP intensities for a proximal muscle, we suggest that, focusing on intensity used to generate such MEPs may be more meaningful. Also, we suggest that by setting a lower range of MEP size, in this case a (0.1-0.4 mV) criterion for a proximal muscle such as biceps brachii, and measuring the corresponding changes in intensity, we can arrive at a more sensitive and reproducible marker for longitudinal recovery.

Active motor thresholds and MEPs—Recording MEPs when subjects maintain a slight voluntary contraction of the muscle can significantly decrease excitation thresholds and reduce variability in MEP amplitudes (Rossini et al., 1994; Chen et al., 1998; Thompson

et al., 1991). For distal hand muscles, MEP amplitude reaches the maximal value by approximately 10-20% of maximal voluntary contraction, whereas this is more gradual (50%) for proximal muscles (Kamen, 2004). Therefore, we suggest using MEP amplitudes of proximal muscles for clinical evaluation if the required level of voluntary contraction can be achieved and accurately monitored.

Mapping criteria—Based on the observation of poor reliability of motor map areas and map volumes, and the improved reliability of COGs to hotspots, we recommend the following: (a) rather than mapping at a pre-set intensity, mapping to generate a criterion MEP (such as 0.1 mV) instead and capturing the changes in intensity at different scalp sites may afford greater reliability for proximal muscles and even improve comparison of its maps with any other muscle maps. (b) COGs determined across different mapping methods may be more reliable than the respective hotspots. In other words, mapping from 2 different hotspots may end up resulting in the same COG, but maps with similar COGs may not necessarily have the same hotspot. Hence, weighted average location of excitability is suggested as a better reliable measure when compared to a single hotspot location.

Selection of intensities for ICI—The selection of intensities of the conditioning and the test pulse may be crucial for improving the reliability of ICI. Also, as discussed earlier, by setting a lower range of MEP size criterion for a proximal muscle, and delivering the corresponding intensity of the test pulse to generate such response, may reduce the variability in inhibition measured between the sessions.

4.7 Conclusions

This study demonstrated that in young healthy subjects, resting motor thresholds and supra-maximal MEP intensities, locations of center of gravity and hotspot are the key reliable neurophysiologic metrics that define corticocortical and corticospinal physiology of the biceps brachii muscle. Given the poor reproducibility of MEP amplitudes, we suggest that setting a lower MEP criterion within the set range of (0.1–0.4 mV) would be more sensitive for proximal muscles. Comparison across statistical approaches showed that correlation tests that assess either association (Pearson's) or association and systematic differences between the metrics (ICC, CCC) would be more accurate and powerful indices for reliability than tests measuring means (paired t-tests). Hence, for a large proximal muscle such as biceps brachii, we recommend choosing metrics that define corticospinal excitability and output in terms of intensity (resting motor thresholds or supra-maximal MEPs) and/or spatial spread of excitability (center of gravity and hotspot) would serve as robust and reproducible markers to longitudinally track recovery and prognosticate upper extremity function across neurological populations such as stroke, and spinal cord injury.

Acknowledgments

This work was supported by the National Institutes of Health (NIH) grant R01-NS-35130 to GHY, NIH grant 1K01HD069504 and American Heart Association grant 13BGIA17120055 to EP and Clinical & Translational Science Collaborative grant RPC 2014-1067 to DC.

References

- Baker A. Non-invasive magnetic stimulation of the human motor cortex. *Lancet*. 1985;1106–7.
- Barker A, Freeston I, Jalinous R, Jarratt J. Magnetic stimulation of the human brain and peripheral nervous system: an introduction and the results of an initial clinical evaluation. *Neurosurgery*. 1987; 20:100–9. [PubMed: 3808249]
- Barnhart HX, Haber MJ, Lin LI. An Overview on Assessing Agreement with Continuous Measurements. *J Biopharm Stat*. 2007; 17:529–69. [PubMed: 17613641]
- Bastani A, Jaberzadeh S. A Higher Number of TMS-Elicited MEP from a Combined Hotspot Improves Intra- and Inter-Session Reliability of the Upper Limb Muscles in Healthy Individuals. *PLoS One*. 2012; 7:e47582. [PubMed: 23077645]
- Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res*. 1999; 8:135–60. [PubMed: 10501650]
- Borojerd B, Kopylev L, Battaglia F, Facchini S, Ziemann U, Muellbacher W, Cohen LG. Reproducibility of intracortical inhibition and facilitation using the paired-pulse paradigm. *Muscle Nerve*. 2000; 23:1594–7. [PubMed: 11003798]
- Brasil-Neto JP, McShane LM, Fuhr P, Hallett M, Cohen LG. Topographic mapping of the human motor cortex with magnetic stimulation: factors affecting accuracy and reproducibility. *Electroenceph Clin Neurophysiol*. 1992; 85:9–16. [PubMed: 1371748]
- Canning CG, Ada L, O'Dwyer NJ. Abnormal muscle activation characteristics associated with loss of dexterity after stroke. *J Neurol Sci*. 2000; 176:45–56. [PubMed: 10865092]
- Cantello R, Gianelli M, Civardi C, Mutani R. Magnetic brain stimulation: The silent period after the motor evoked potential. *Neurology*. 1992; 42:1951–9. [PubMed: 1407578]
- Carroll TJ, Riek S, Carson RG. Reliability of the input-output properties of the corticospinal pathway obtained from transcranial magnetic stimulation and electrical stimulation. *J Neurosci Methods*. 2001; 112:193–202. [PubMed: 11716954]
- Chen R, Tam A, Butefisch C, Corwell B, Ziemann U, Rothwell JC, et al. Intracortical inhibition and facilitation in different representations of the human motor cortex. *J Neurophysiol*. 1998; 80:2870–81. [PubMed: 9862891]
- Colebatch JG, Rothwell JC, Day BL, Thompson PD, Marsden CD. Cortical outflow to proximal arm muscles in man. *Brain*. 1990; 113:1843–56. [PubMed: 2276047]
- Corneal SF, Butler AJ, Wolf SL. Intra- and intersubject reliability of abductor pollicis brevis muscle motor map characteristics with transcranial magnetic stimulation. *Arch Phys Med Rehabil*. 2005; 86:1670–5. [PubMed: 16084825]
- Christie A, Fling B, Crews RT, Mulwitz LA, Kamen. Reliability of motor-evoked potentials in the ADM muscle of older adults. *J Neurosci Methods*. 2007; 164:320–4. [PubMed: 17588673]
- Cicinelli P, Traversa R, Rossini PM. Post-stroke reorganization of brain motor output to the hand: a 2–4 month follow-up with focal magnetic transcranial stimulation. *Electroenceph Clin Neurophysiol*. 1997; 105:438–50. [PubMed: 9448645]
- De Vet HCW, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. *J Clin Epidemiol*. 2006; 59:1033–9. [PubMed: 16980142]
- Di Lazzaro V, Oliviero A, Pilato F, et al. Motor cortex hyperexcitability in transcranial magnetic stimulation in Alzheimer's disease. *J Neurol Neurosurg Psychiatry*. 2004; 75:555–9. [PubMed: 15026495]
- Doeltgen SH, Ridding MC, O'Beirne GA, Dalrymple-Alford J, Huckabee ML. Test-retest reliability of motor evoked potentials (MEPs) at the submental muscle group during volitional swallowing. *J Neurosci Methods*. 2009; 178:134–7. [PubMed: 19118575]
- Groppa S, Oliviero A, Eisen A, Quartarone A, Cohen LG, Mall V, Kaelin-Lang A, et al. A practical guide to diagnostic transcranial magnetic stimulation. Report of an ICFN committee. 2012; 123:858–882.
- Harris-Love ML, Perez MA, Chen R, Cohen LG. Interhemispheric inhibition in distal and proximal arm representations in the primary motor cortex. *J Neurophysiol*. 2007; 97:2511–5. [PubMed: 17215494]

- Kamen G. Reliability of motor-evoked potentials during resting and active contraction conditions. *Med Sci Sports Exerc.* 2004; 36:1574–9. [PubMed: 15354040]
- Kimiskidis VK, Papagiannopoulos S, Sotirakoglou K, Kazis DA, Dimopoulos G, Kazis A, et al. The repeatability of corticomotor threshold measurements. *Neurophysiologie Clinique/Clinical Neurophysiology.* 2004; 34:259–266. [PubMed: 15890159]
- Kujirai T, Caramia MD, Rothwell JC, Day BL, Thompson PD, Ferbert A, Wroe S, Asselman P, Marsden CD. Corticocortical inhibition in human motor cortex. *J Physiol.* 1993; 471:501–19. [PubMed: 8120818]
- Kumar VP, Satku K, Balasubramaniam P. The role of the long head of biceps brachii in the stabilization of the head of the humerus. *Clin Orthop Relat Res.* 1989; 244:172–175. [PubMed: 2743659]
- Liepert J, Hallett M, Samii A, Oddo D, Celnik P, Cohen LG, et al. Motor cortex excitability in patients with cerebellar degeneration. *Clin Neurophysiol.* 2000; 111:1157–64. [PubMed: 10880788]
- Lin L. A Concordance Correlation Coefficient to Evaluate Reproducibility. *Biometrics.* 1989; 45:255–68. [PubMed: 2720055]
- Lin L, Hedayat AS, Sinha B, Yang M. Statistical Methods in Assessing Agreement. *J Am Stat Assoc.* 2002; 97:257–70.
- Livingston SC, Ingersoll CD. Intra-rater reliability of a transcranial magnetic stimulation technique to obtain motor evoked potentials. *Int J Neurosci.* 2008; 118:239–256. [PubMed: 18205080]
- Maeda F, Gangitano M, Thall M, Pascual-Leone A. Inter-and intra-individual variability of paired-pulse curves with transcranial magnetic stimulation (TMS). *Clin Neurophysiol.* 2002; 113:376–82. [PubMed: 11897538]
- Malcolm M, Triggs W, Light K, Shechtman O, Khandekar G, Rothi L, et al. Reliability of motor cortex transcranial magnetic stimulation in four muscle representations. *Clin Neurophysiol.* 2006; 117:1037–46. [PubMed: 16564206]
- Marconi B, Filippi GM, Koch G, Giacobbe V, Pecchioli C, Versace V, et al. Long-term effects on cortical excitability and motor recovery induced by repeated muscle vibration in chronic stroke patients. *Neurorehabil Neural Repair.* 2011; 25:48–60. [PubMed: 20834043]
- McDonnell MN, Ridding MC, Miles TS. Do alternate methods of analyzing motor evoked potentials give comparable results? *J Neurosci Methods.* 2004; 136:63–7. [PubMed: 15126046]
- McMillan AS, Watson C, Walshaw D. Transcranial magnetic-stimulation mapping of the cortical topography of the human masseter muscle. *Arch Oral Biol.* 1998; 43:925–931. [PubMed: 9877323]
- Mortifee P, Stewart H, Schulzer M, Eisen A. Reliability of transcranial magnetic stimulation for mapping the human motor cortex. *Electroenceph Clin Neurophysiol.* 1994; 93:131–7. [PubMed: 7512919]
- Mills KR, Nithi KA. Corticomotor threshold to magnetic stimulation: Normal values and repeatability. *Muscle Nerve.* 1997; 20:570–576. [PubMed: 9140363]
- Obuchowski NA, Reeves AP, Huang EP, Wang XF, Buckler AJ, Kim HJ, et al. Quantitative imaging biomarkers: A review of statistical methods for computer algorithm comparisons. *Stat Methods Med Res.* 2014 Epub ahead of print.
- Oldfield RC. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychol.* 1971; 9:97–113.
- Orth M, Snijders AH, Rothwell JC. The variability of intracortical inhibition and facilitation. *Clin Neurophysiol.* 2003; 114:2362–9. [PubMed: 14652096]
- Perez MA, Lungholt BK, Nyborg K, Nielson JB. Motor skill training induces changes in the excitability of the leg cortical area in healthy humans. *Exp Brain Res.* 2004; 159:197–205. [PubMed: 15549279]
- Pourtney, LG.; Watkins, MP. *Foundations of clinical research: application to practice.* Connecticut: Appleton and Lange; 2000. p. 892
- Pourtney, LG.; Watkins, MP. *Foundations of clinical research: applications to practice.* Pearson/Prentice Hall; 2009. p. 892

- Plow EB, Cunningham DA, Bonnett C, Gohar D, Bayram M, Wyant A, et al. Neurophysiologic Correlates of Aging-Related Muscle Weakness. *J Neurophysiol.* 2013;2563–73. [PubMed: 24027104]
- Plow EB, Varnerin N, Cunningham DA, Janini D, Bonnett C, Wyant A, et al. Age-Related Weakness of Proximal Muscle Studied with Motor Cortical Mapping: A TMS Study. *PLoS One.* 2014; 9(2):e89371. [PubMed: 24586726]
- Roick H, von Giesen HJ, Benecke R. On the origin of the postexcitatory inhibition seen after transcranial magnetic brain stimulation in awake human subjects. *Exp Brain Res.* 1993; 94:489–98. [PubMed: 8359263]
- Rossi S, Hallett M, Rossini PM, Pascual-Leone A. Safety, ethical considerations, and application guidelines for the use of transcranial magnetic stimulation in clinical practice and research. *Clin Neurophysiol.* 2009; 120:2008–39. [PubMed: 19833552]
- Rossini P, Barker A, Berardelli A, Caramia M, Caruso G, Cracco, et al. Non-invasive electrical and magnetic stimulation of the brain, spinal cord and roots: basic principles and procedures for routine clinical application. Report of an IFCN committee. *Electroenceph Clin Neurophysiol.* 1994; 91:79–92. [PubMed: 7519144]
- Rudhe C, van Hedel HJ. Upper extremity function in persons with tetraplegia: relationships between strength, capacity, and the spinal cord independence measure. *Neurorehabil Neural Repair.* 2009; 23:413–21. [PubMed: 19261766]
- Ruohonen J, Karhu J. Navigated transcranial magnetic stimulation. *Clin Neurophysiol.* 2010; 40:7–17.
- Sawaki L, Butler AJ, Leng X, Wassenaar PA, Mohammad YM, Blanton S, et al. Constraint-induced movement therapy results in increased motor map area in subjects 3 to 9 months after stroke. *Neurorehabil Neural Repair.* 2008; 22:505–513. [PubMed: 18780885]
- Stewart LM, Walsh V, Rothwell JC. Motor and phosphene thresholds: a transcranial magnetic stimulation correlation study. *Neuropsychologia.* 2001; 39:415–9. [PubMed: 11164880]
- Stinear CM, Barber PA, Smale PR, Coxon JP, Fleming MK, Byblow WD. Functional potential in chronic stroke patients depends on corticospinal tract integrity. *Brain.* 2007; 130:170–80. [PubMed: 17148468]
- Streletz LJ, Belevich JK, Jones SM, Bhushan A, Shah SH, Herbison GJ. Transcranial magnetic stimulation: cortical motor maps in acute spinal cord injury. *Brain Topogr.* 1995; 7:245–250. [PubMed: 7599023]
- Thompson PD, Day BL, Rothwell JC, Dressler D, Maertens De Noordhout A, Marsden CD. Further observations on the facilitation of muscle responses to cortical stimulation by voluntary contraction. *Electroencephalogr Clin Neurophysiol.* 1991; 81:397–402. [PubMed: 1718726]
- Trompetto C, Assini A, Buccolieri A, Marchese R, Abbruzzese G. Motor recovery following stroke: a transcranial magnetic stimulation study. *Clin Neurophysiol.* 2000; 111:1860–7. [PubMed: 11018503]
- Uy J, Ridding MC, Miles TS. Stability of maps of human motor cortex made with transcranial magnetic stimulation. *Brain Topogr.* 2002; 14:293–7. [PubMed: 12137362]
- van Kuijk AA, Anker LC, Pasman JW, Hendriks JC, van Elswijk G, Geurts AC. Stimulus-response characteristics of motor evoked potentials and silent periods in proximal and distal upper-extremity muscles. *J Electromyogr Kinesiol.* 2009; 19:574–83. [PubMed: 18396413]
- Wittenberg G, Chen R, Ishii K, Bushara O, Taub E, Gerber LH, et al. Constraint-induced therapy instroke: magnetic-stimulation motor maps and cerebral activation. *Neurorehabil Neural Repair.* 2003; 17:48–57. [PubMed: 12645445]
- Wolf SL, Butler AJ, Campana GI, Parris TA, Struys DM, Weinstein SR. Intra-subject reliability of parameters contributing to maps generated by transcranial magnetic stimulation in able-bodied adults. *Clin Neurophysiol.* 2004; 115:1740–7. [PubMed: 15261852]
- Ziemann U, Rothwell JC, Ridding MC. Interaction between intracortical inhibition and facilitation in human motor cortex. *J Physiol.* 1996; 496:873–881. [PubMed: 8930851]

Biographies

Vishwanath Sankarasubramanian graduated in 1998 from Chennai, India and went on to complete his Master of Science in Biomedical Engineering from Aachen University of Technology, Germany in 2006. In 2013 he received his PhD degree in Neuromodulation (specialization-spinal cord stimulation for chronic pain) from the University of Twente, Netherlands. Since March 2014, he is a postdoctoral research fellow at the Biomedical Engineering department of the Cleveland Clinic. His research interests broadly include neurostimulation therapies (both invasive and non-invasive) for stroke and pain, experimental design, resting-state functional MRI and functional connectivity data analysis and interpretation.

Sarah Roelle was born in Cleveland, OH in 1990. She graduated from the University of Toledo in 2013 with a B.S. in Bioengineering. She joined the neural control lab at the Cleveland Clinic as an undergraduate student in 2011 and returned after receiving her degree, where she focused on TMS in healthy and disease states as well as neuroimaging. She is currently a staff member at Case Western Reserve University where her research interests include targeted drug delivery and molecular imaging in cancer.

Corin Bonnett graduated from Marietta College with a Bachelor of Science degree in Biology in 2011. Between 2011 and 2014, she assisted with data collection and study coordination in the laboratory of Dr. Ela Plow at the Cleveland Clinic.

Daniel Janini graduated from Case Western Reserve University in 2011 with a B.S degree in Biology. Since February 2012, he has been working as a research intern at Dr. Ela Plow's lab in the Biomedical Engineering department at the Cleveland Clinic.

Nicole Varnerin completed a degree in Electrical Engineering from Case Western Reserve University in Cleveland, OH in 2013. She has been a research engineer with the Neuromodulatory Laboratory at the Cleveland Clinic since 2010 and continues to develop analyses in the area of TMS and neuroimaging.

David Cunningham received his Bachelor's degree from Ohio University in 2008 and his Master's degree from Texas A&M University in 2011, both in kinesiology. He is currently a doctoral student at Kent State University in the Neuroscience collaborative program with the Cleveland Clinic. His doctoral education employs neurophysiologic methods of brain stimulation in conjunction with neuroimaging techniques in order to study neurological predictors of recovery in patients with chronic stroke. Specifically, he is interested in the variability of neural plasticity in chronic stroke patients that range from mild motor impairments to the most severely impaired.

Originally from Toronto, Ontario, **Jennifer Sharma** moved to Cleveland in 2006 to join Case Western Reserve University where she graduated from their undergraduate program in 2010 with a Bachelor's Degree in Psychology and Chemistry. In 2014, she graduated from the Case Western Reserve University, School of Medicine with an M.D and a Masters of Public Health degree and is currently a first year Neurology resident in Kingston, Ontario. She hoping to further specialize in either Neuro- Oncology or Stroke in the future.

Kelsey Ann Potter-Baker was born in Salt Lake City, Utah, USA in 1986. She graduated with her bachelor's in 2008 and then obtained a PhD in Biomedical Engineering at Case Western Reserve University located in Cleveland, Ohio, USA in 2014. In late 2014, she joined the Cleveland Clinic Foundation Lerner Research Institute. She is currently a post-doctoral research fellow at this institution. Her areas of research interest include neuromodulation, neurodegenerative diseases, spinal cord injury and noninvasive stimulation modalities.

Xiao-Feng Wang received Ph.D. degree in Statistics from Case Western Reserve University, Cleveland, OH, USA, in 2005. He joined the Department of Quantitative Health Sciences at Cleveland Clinic since he graduated. He is currently an Associate Staff Member of Cleveland Clinic Lerner Research Institute and Associate Professor of Medicine, Epidemiology and Biostatistics at Cleveland Clinic Lerner College of Medicine. His research interests include statistical data mining, functional data analysis, quantitative image analysis, bioinformatics and neuroinformatics.

Guang H. Yue received his Ph.D. from the University of Iowa. From 1991 to 1994, he was a Postdoctoral Fellow at the University of Arizona and Cleveland Clinic. From 1994 to 2011 he was a faculty member in the Department of Biomedical Engineering at the Cleveland Clinic and Department of Molecular Medicine at Cleveland Clinic Lerner College of Medicine of Case Western Reserve University. He is now the Director of Human Performance and Engineering Research at Kessler Foundation/Kessler Institute for Rehabilitation, and Professor at Rutgers New Jersey Medical School, Rutgers University. Dr. Yue's research program is focused primarily on human motor control in health and disease and on finding neural mechanisms underlying movement disorders and recovery of motor function as a result of medical intervention.

Ela B Plow is an Assistant Professor at the Cleveland Clinic Lerner College of Medicine and runs her own lab in the Department of Biomedical Engineering, Cleveland Clinic. She specializes in rehabilitation of stroke patients and her research interests broadly include noninvasive brain stimulation, motor control, neuroimaging and clinical neuroscience. Dr. Plow earned her Bachelor of Science degree in Physical therapy from PGIMER, India and a PhD in Rehabilitation Science at the University of Minnesota. She completed a post-doctoral fellowship in Neurology, Harvard Medical School and since then is part of the Cleveland Clinic.

Figure 1.

The schematic representation shows the flow of procedures for the test-retest reliability paradigm. TMS was performed in each session (test 1 and test 2) to obtain the following test metrics - RMT, supra-maximal MEP, motor map and ICI.

Figure 2.

The schematic representation shows the TMS parameters that are described by each TMS metric. Abbreviations: COG = Center of gravity

Figure 3.

Top: Illustration of measures of corticospinal excitability - (a) RMT intensity, and corticospinal output (b) Supra-maximal MEP intensity, and (c) Supra-maximal MEP. Each plot shows the values from test 1 plotted against those from test 2. The line in each plot is a model line, where $y = x$, showing the ideal case of reliability in which the value at test 1 equals that at test 2. Therefore, the distance of the data point from the model line shows its qualitative degree of reliability, or how close it is to perfect reliability. The (+) sign represents an outlier that was excluded from the plot, but was included in the analysis.

Bottom: The sample MEPs refer to specific data points, marked by 1 and 2 in the supra-maximal MEP plot (c). These samples are meant to show how a subject's (subject 1) MEPs with higher amplitude (values are 1.144 mV at test 1 and 0.495 mV at test 2) compare with another subject's (subject 2) MEPs with smaller amplitudes (values are 0.226 mV at test 1 and 0.186 mV at test 2).

Figure 4.

Top: The panel shows plots of (a) Motor map area and (b) Normalized map volume. The values at test 1 are plotted against those at test 2 and compared to the model line. **Bottom:** The panel shows contour plots representing the distribution of map volume from test 1 (c) and test 2 (d) of the subject marked by a star in the top map volume plot. The ordinate shows the y-coordinate location and the abscissa is the x-coordinate location. Each black dot represents a TMS scalp site at that x- and y-coordinate location that elicited an MEP response meeting our criterion (at least 10 μ V for the motor map). The colors represent the responses' relative amplitudes normalized to MEP_{Maxima} according to the scale shown in the color bar. The points surrounding the red area have the highest amplitudes (100% MEP_{Maxima}).

Figure 5.

Representation of (a) x- and (b) y-coordinate locations of the hotspot and (c) x and (d) y-coordinate locations of COG. The values at test 1 are plotted against those at test 2 and are compared to the model line to show their qualitative degree of reliability. The y-coordinates for both variables ((b) and (d)) are more reliable than the x-coordinates ((a) and (c)).

Figure 6.

The Bland-Altman plots for ICI shows the difference between data points at test 1 and at test 2 for each subject in relation to the mean of the two values. This creates a plot showing the agreement between values at test 1 and test 2. If the values fall along the line at $y = 0$, then there is strong agreement. Any data point within the ± 2 standard deviation, marked by the dashed lines, is acceptable. As can be noted, ICI shows weak agreement.

Table 1

TMS metrics with corresponding TMS parameters and their definitions resulting from the procedures of single-pulse and paired-pulse stimulation. Equations are added for reference where necessary. MEP_i = the average of two trials at each responsive scalp site, x_i , y_i = the x- and y-coordinates of the site normalized to the nasion. MEP_{Maxima} = the amplitude of the MEP that elicited the largest response of the entire map.

TMS Metric	TMS Parameter	Definition	Equation
RMT Intensity	Corticospinal excitability	Lowest intensity of TMS to describe corticospinal excitability	
Supra- maximal MEP Intensity	Corticospinal output	Supra-threshold intensity of TMS to describe corticospinal output	
Supra- maximal MEP		Muscle response (0.1–0.7 mV) from supra-threshold intensity to describe corticospinal output	
Motor Map	Hotspot	x- and y-coordinate locations, normalized to the coordinates of the nasion	
	COG	Weighted-average location of spread of corticospinal activation	$COGX = \sum (MEP_i * x_i) / \sum MEP_i,$ $COGY = \sum (MEP_i * y_i) / \sum MEP_i$
	Motor Map Area	Number of scalp sites eliciting MEPs of at least 10 μ V	
	Normalized Map Volume	Sum of normalized MEPs across all responsive scalp sites	$\text{Map Volume} = \sum MEP_i / MEP_{Maxima}$
ICI		Comparison of the conditioned MEP and the supra-maximal MEP expressed as a percentage	$ICI = (\text{Conditioned MEP} / \text{Supra-maximal MEP}) * 100$

Table 2

Results from each statistical test for each TMS metric/TMS parameter. Mean represents the mean difference between the test 1 and test 2 and standard deviation (SD) represents the deviation of the differences as provided by the paired t-test. Also provided from the t-test is the p-value ($p < 0.05$ is significant). The values provided by ICC, CCC, and Pearson's correlation are the r^2 values representing the strength of the correlation (a strong correlation is > 0.6). Values with bold font and asterisk are those representing strong reliability, meaning they have a strong correlation. In addition, the one p-value that is < 0.05 is highlighted to emphasize the only variable (RMT) that is significant according to the t-test.

TMS Metric/Parameter	t-test			ME	ICC	CCC	Pearson's
	Mean	SD	p-value				
RMT Intensity (%)	8.643	10.938	0.011*	7.7	0.745*	0.636*	0.843*
Supra-maximal MEP Intensity (%)	4.857	11.340	0.133	8.02	0.670*	0.629*	0.728*
Supra-maximal MEP (mV)	-0.248	0.993	0.368		0.163	0.155	0.263
Hotspot x-coordinate (mm)	1.956	6.857	0.344		0.556	0.535	0.573
Hotspot y-coordinate (mm)	3.461	6.674	0.100	4.72	0.711*	0.655*	0.712*
COG x-coordinate (mm)	1.170	6.126	0.522		0.564	0.554	0.569
COG y-coordinate (mm)	2.491	5.542	0.148	3.92	0.721*	0.679*	0.732*
Motor Map Area	5.167	21.191	0.416		0.363	0.349	0.364
Normalized Map Volume	139.609	586.63	0.427		0.432	0.418	0.436
Percent ICI (%)	6.472	68.116	0.738		0.357	0.355	0.367