Reliability of lower limb motor evoked potentials in stroke and healthy populations:

How many responses are needed?

Gwyn N Lewis*[1], Nada Signal[1], Denise Taylor[1]

[1]Health and Rehabilitation Research Institute, AUT University, Auckland, New Zealand

*Author for correspondence
Health and Rehabilitation Research Institute
AUT University
Private Bag 92006
Auckland 1142
New Zealand

Ph +64 9 921 9999
Fax +64 9 921 9620
Email: gwyn.lewis@aut.ac.nz

Key words: reliability, transcranial magnetic stimulation, stroke, lower limb

Acknowledgements: The authors would like to thank Elizabeth Gerdis for her assistance with data analysis. Funding for the study was provided from STAR Research Grant to NS.

Highlights

- Intrasession reliability of lower limb motor evoked potentials (MEPs) is excellent in patients with stroke when using as few as 6 responses
- Intersession reliability of MEPs is poor in patients with stroke, questioning the ability of this measure to reliably capture changes in corticomotor excitability over time
- Measuring the amplitude of each single MEP or using MEP area provides a more reliable measure of MEP size than averaging non-rectified responses

Abstract

Objective: To determine the intra- and inter-session reliability of motor evoked potential (MEP) size parameters in the lower limb of patients with stroke, focussing on the number of MEPs collected and the method of measuring MEP size. **Methods:** Transcranial magnetic stimulation was used to elicit MEPs in the soleus muscle of patients with stroke (*n*=13) and age-matched healthy participants (*n*=13) during low level muscle activation. Two sets of 10 responses were collected in the first session and a further 10 responses collected in a second session held 7 days later. Four MEP size measurements were made using 4, 6, 8, or all 10 of the MEPs collected. Intra- and inter-session reliability was examined using intraclass correlation coefficients (ICC) and typical percentage error. **Results:** Intrasession ICC statistics using 6 or more MEPs were >0.85 in the stroke group but intersession values were all <0.5. Reliability was best when measuring parameters from individual MEPs rather than averaged responses. **Conclusions:** Reliability of intrasession MEP size is excellent in the lower limb of patients with stroke using as few as 6 MEPs but intersession reliability is poor. **Significance:** Comparing MEP size measures across two or more sessions is questionable in the lower limb of patients with stroke.

Introduction

Studies that have used transcranial magnetic stimulation (TMS) to evaluate the corticomotor system in patients with stroke number in the hundreds. TMS has been used as a tool for predicting recovery of function (Stinear et al., 2007, Trompetto et al., 2000), to determine the type or location of reorganisation occurring within the cortex (Caramia et al., 2000, Hamzei et al., 2006, Shimizu et al., 2002, Turton et al., 1996), or to investigate the neurophysiological effect of interventions designed to enhance recovery (Brouwer and Ambury, 1994, Liepert et al., 2000, Wittenberg et al., 2003). The response to TMS is primarily recorded using motor evoked potentials (MEPs) obtained from musculature in the periphery. While undoubtedly useful for probing the effects of stroke on the motor system, the establishment of measurement reliability and the importance of using reliability data to determine sample and effect sizes in research trials are paramount to good scientific practice. The inherent physiological variability of the corticomotor system is evident in the wide variability of MEP amplitudes that are elicited in a train of stimuli, even in healthy individuals. Acknowledgement of this variability is seen in the accepted practice of collecting 6-10 stimuli per condition so that an average response size can be obtained.

Previous studies examining the reliability of MEPs have mainly focused on hand and forearm muscles (Bastani and Jaberzadeh, 2012, Carroll et al., 2001, Christie et al., 2007, Kamen, 2004, Livingston and Ingersoll, 2008, McDonnell et al., 2004, Mortifee et al., 1994, Ngomo et al., 2012), with a smaller number of studies examining muscles from the lower limb (Cacchio et al., 2009, Cacchio et al., 2011, Tallent et al., 2012, van Hedel et al., 2007, Wheaton et al., 2009). Most authors have reported good or excellent reliability of MEP amplitude measurements within and between sessions but this is not always consistent (Livingston and Ingersoll, 2008, McDonnell, Ridding, 2004). The reliability of MEP characteristics is reduced

between sessions compared to within a session (Bastani and Jaberzadeh, 2012, Doeltgen et al., 2009), in more distal upper limb muscles compared to proximal (Kamen, 2004, Malcolm et al., 2006), and in populations with neurological conditions compared to healthy individuals (Cacchio, Paoloni, 2011, Wheaton, Villagra, 2009). Of note, the two studies examining responses in patients with stroke found poor-average reliability of MEP amplitude in the lower limb muscles on the paretic side. This factor may be compounded by the fact that muscle activation is often required to elicit MEPs in paretic muscles, as MEP reliability is reduced during muscle activity compared to resting conditions (Kamen, 2004, Ngomo, Leonard, 2012, Tallent, Goodall, 2012, van Hedel, Murer, 2007).

In addition to biological variability, experimenter and equipment factors can also contribute to the variability of MEPs, including the coil type and positioning, recording electrode placement, and MEP collection and processing procedures. While the use of neuronavigation systems for coil positioning have not yielded notable improvements in reliability (Fleming et al., 2012, Jung et al., 2010), it has been consistently reported that reliability is enhanced when a greater number of responses to TMS are collected (Bastani and Jaberzadeh, 2012, Christie, Fling, 2007, Doeltgen, Ridding, 2009, Kamen, 2004). The minimum number of MEPs required to obtain reliable results has been recommended at five for intrasession (Bastani and Jaberzadeh, 2012, Christie, Fling, 2007, Doeltgen, Ridding, 2009) and ten for intersession comparisons (Bastani and Jaberzadeh, 2012, Doeltgen, Ridding, 2009). Only one study has investigated the effect of different MEP processing and analysis procedures and showed no difference in reliability between the use of MEP area and MEP amplitude outcome measures to represent the magnitude of the MEP (McDonnell, Ridding, 2004). This study was undertaken in the upper limb and the findings may be

different in lower limb muscles where there is less synchronisation of the descending cortical volley and polyphasic MEPs are often obtained (Marchand-Pauvert et al., 1999, Soto et al., 2006).

Combined, these factors suggest that additional responses should be collected when analysing MEPs from neurological populations or making comparisons across sessions; however, prolonged sessions can be fatiguing and uncomfortable for participants. Hence, it would be useful to know the minimum number of stimuli and the optimal MEP processing and measurement procedures that give rise to the most reliable MEP parameters in patients with neurological impairments. The aim of the current study was to determine the intra- and inter-session reliability of MEP size parameters in the lower limb of patients with stroke during muscle activation. We specifically examined the effect of altering the number of MEPs collected and the method of processing and analysing the evoked responses.

Methods

Table 1 provides a summary of the methodological quality of the study procedure using the checklist recommended by Chipchase and colleagues (2012). Criteria relating to paired-pulse stimulation have been removed.

Participants

The participants were 13 patients with stroke and 13 age-matched healthy controls. The participants with stroke were required to be aged over 18 years, have had a single stroke more than 6 months previously, and have a residual impairment in their ability to walk. The mean age of the stroke group was 57 years (range 22-78; seven male) and the mean age of the control group was 55 years (range 26-73; five male). Seven participants in the stroke group had left hemiplegia, the mean time since onset of stroke was 53 months (range 7-

136), and the mean comfortable treadmill walking speed was 0.72 m/s (range 0.28-1.14 m/s). Participants from both groups were excluded if they were unable to engage in the testing due to cognitive or communication deficits, had another medical condition that could impact the results, had an uncontrolled medical problem that prevented maximum muscle strength testing, had any contraindications to TMS (including medications influencing CNS excitability), or were unable to elicit a response to TMS in the soleus muscle. Nine of the participants with stroke and three of the control participants were taking medication, predominantly for cardiovascular disease. All participants provided informed, written consent prior to inclusion in the study.

Protocol

Participants attended two data collection sessions 7 days apart. Participant testing was conducted at the same approximate time of day (e.g. morning, afternoon) in the two test sessions. In both sessions, responses to TMS were recorded in the soleus muscle during activation at 10% of maximum voluntary contraction (MVC). The most affected leg of the stroke participants and a randomly assigned limb of the control participants were tested. In the first session, responses to TMS were obtained twice (Measurement 1 and 2), while one set of responses was collected in the second session (Measurement 3). During all testing procedures, participants were seated in a custom chair with the test leg extended (hip 90°, knee 120°, ankle 0°) and fully supported, with the foot strapped to a rigid support that allowed isometric ankle plantarflexion. Ankle plantarflexor MVC was collected at the start of each session prior to TMS testing.

Electromyography

Bipolar surface electrodes (Norotrode 20, Myotronics Inc., USA) were applied to the soleus muscle of the selected leg following standard skin preparation (exfoliation, light abrasion, and cleansing with alcohol). Following SENIAM guideline recommendations (Merletti et al., 2001), the electrodes were placed 2/3 of the way between the medial condyle of the femur and the medial malleolus and aligned in the direction of muscle fibres. Skin preparation and electrode placement were repeated if the impedance was greater than 5 kΩ (Groppa et al., 2012). Electromyography (EMG) signals were amplified (AMT-8, Bortec Biomedical, Canada), bandpass filtered (10-1000 Hz) and sampled at 5,000 Hz using a data acquisition board (Micro 1401, CED, UK) and Signal software (CED, UK).

MVC testing

A single axis loadcell (Precision Transducers Ltd, New Zealand) attached to the foot support enabled collection of ankle plantarflexor force. Participants were asked to push against the foot plate as hard as they could for 3-5 s. This was repeated two further times with a 3 min break between each repetition. The maximum force recorded during any of the three trials was established as the MVC. Constant verbal encouragement was provided during all MVC testing. During all subsequent TMS testing, participants were provided with visual feedback of plantarflexor force levels along with a target force level of 10% MVC. Magnetic stimuli were delivered when the force level was within the target range of 10±2% MVC.

TMS procedures

A tightly fitting neoprene cap marked with a 1x1 cm grid relative to the vertex was fitted to the head to ensure maintenance of the coil position within sessions. TMS was delivered to the selected hemisphere using a Magstim 200^2 (Magstim Co, Dyfed, UK) using a double cone

coil and monophasic pulses. The juncture of the coil was initially placed over the mid-sagittal plane approximately 2 cm posterior to the vertex and 1 cm contralateral to the tested leg with a posterior-anterior direction of current flow. The soleus muscle “hot spot” was identified by moving the coil around the scalp until the site eliciting MEPs of the largest amplitude was detected. This spot was marked on the neoprene cap and all further stimuli were delivered over this location. Active motor threshold (AMT) was then determined as lowest stimulus intensity to give rise to at least four discernible MEPs in a train of eight stimuli while generating plantarflexor force at 10% MVC.

Following the establishment of AMT, participants received ten stimuli at an intensity of 120% AMT while at 10% MVC. At the first session, a further set of ten stimuli were delivered approximately 15 minutes later.

Data processing

MEPs from the three measurement periods were analysed in four ways. MEP_{mean} amplitude was determined by averaging the ten MEPs and then measuring the maximum peak-to-peak amplitude of the averaged response. MEP_{single} amplitude was determined by measuring the maximum peak-to-peak amplitude of each individual MEP and then averaging these amplitudes to provide a single value. MEP_{mean} area was determined by rectifying the EMG signals, averaging the ten responses, and then measuring the root mean square amplitude (rms) of EMG activity in a 30 ms window from MEP onset of the averaged response. MEP onset was defined as the first point at which EMG activity exceeded 3 standard deviations (SD) of a 30 ms window of pre-stimulus EMG activity. MEP_{single} area was determined by measuring the rms amplitude in a 30 ms window of each individual MEP and then averaging these to provide a single value. To investigate the influence of the number of MEPs

recorded, these procedures were also conducted using the first 8, 6, and 4 MEPs in each block of 10.

Statistical analysis

The intra- and inter-session intraclass correlation coefficient (ICC) and typical percentage error (TPE) were calculated for the four MEP measurements using 4, 6, 8, and 10 responses. The ICC is a measure of consistency of the data that is based on both the within- and between-subject variation. ICC values are bound from 0-1 with those closer to 1 reflecting greater reliability. TPE is a relative error score that provides a measure of reproducibility within an individual. TPE was calculated using (Hopkins, 2000):

$100 \times (s_{diff}/\sqrt{2}) / \bar{x}$

where s_{diff} is the standard deviation of the individual difference scores and $\bar{x}$ is the Measurement 1 mean. Intersession ICC and TPE also were determined for the ankle plantarflexor MVC and soleus AMT. To examine the relationship between the amplitude and area MEP measurements, Pearson correlations were performed between MEP_{mean} amplitude and MEP_{mean} area values from Measurement 1. To examine the relationship between measurements from individual and averaged responses, Pearson correlations were performed between MEP_{single} and MEP_{mean} amplitude data from Measurement 1. Additionally, separate one-way repeated measures ANOVAs were conducted for the stroke and control groups comparing the soleus pre-stimulus EMG rms values among the three measurements to verify that a consistent level of soleus activation was achieved during testing.

Results

Measurement 3 data from one stroke participant were unable to be analysed due to signal noise introduced by a faulty cable. In addition, only 8 MEPs were able to be analysed from two stroke and one control participant at Measurement 1 due to problems with data collection. The background EMG rms was not different among the three measurement periods in the stroke ($F_{2,22}$ =1.5; *P*=0.25) or control ($F_{2,24}$ =9.8; *P*=0.36) groups. While MEP size measurements were smaller in the stroke group for all measurement methods, the difference from the control group was only significant for MEP_{mean} amplitude (*P* = 0.02).

Reliability of MEP amplitude and area measurements

Intrasession ICC and TPE data from the stroke and control groups are shown in Table 2. According to Fleiss (Fleiss, 1986), all of the ICC values are excellent except for MEP_{mean} amplitude in the control group using 4 responses. Overall, the ICC values were higher and TPE values lower when more MEPs were collected. ICC and TPE values were consistently less reliable for MEP_{mean} amplitude and area compared to MEP_{single} amplitude and area, but the difference between amplitude and area measures was negligible. ICCs were also higher in the control group compared to the stroke group. In contrast, TPE was lower in the stroke group across all measurements.

Intersession ICC and TPE data from the stroke and control groups are shown in Table 3. All of the ICC values for the control group are excellent except for MEP_{mean} amplitude using 4 or 6 responses. The ICC and TPE values for this group are slightly reduced in comparison to the intrasession results but still show a pattern of reduced values when collecting fewer responses and when measuring the averaged response compared to individual responses. In

contrast to the control data, all of the ICC values for the stroke group are classified as poor and the TPE are substantial (all ≥ 70%) across all four measurement methods.

Relationships among MEP processing procedures

Figure 1 shows scatter graphs of the relationships among the MEP processing and measurement procedures. The Pearson correlation coefficients are shown on each graph. There is a strong relationship between the amplitude and area measures in both groups, but this is particularly marked in the controls. The lower two graphs show that MEP_{single} amplitude resulted in larger amplitude values compared to the ensemble MEP_{mean} amplitude. This is more marked in the stroke group.

Reliability of MVC and AMT

Group data and intersession reliability statistics for MVC and motor threshold are shown in Table 4. The motor threshold ICC values are excellent for both groups and were associated with very low TPE. For the control group, the MVC ICC is also excellent but the TPE is substantial. Reliability of MVC in the stroke group is only fair to good and the TPE is also substantial. There was no significant difference in AMT between the two groups ($P = 0.2$). Although the MVC was lower in the stroke group, this was not statistically significant ($P = 0.06$).

Discussion

This study is the first to compare the reliability of different MEP collection and measurement procedures in the lower limb of patients with stroke and healthy participants. The main reliability findings were that MEP_{mean} amplitude and area measures were less reliable than MEP_{single} amplitude and area measures, reliability was reduced when fewer responses were collected, intrasession MEP reliability was excellent both for the control and

stroke groups, and intersession reliability was excellent for the control group but poor for the stroke group. There were strong relationships between the area and amplitude estimates of MEP size; however, MEP_{mean} amplitude gave rise to comparatively smaller responses than MEP_{single} amplitude. These findings are discussed in more detail below.

Reliability in stroke and control groups

The ICC values were higher overall for the control group but both groups showed excellent intrasession values, suggesting that MEP size measures in the lower limb are reliable in both stroke and control participants within a testing session. The control group ICCs were comparable to those reported by other studies investigating intrasession reliability of MEPs (McDonnell, Ridding, 2004, van Hedel, Murer, 2007). The intersession ICC values were again excellent for the control group, indicating that MEPs are a reliable outcome measure to use across multiple testing sessions in a healthy population. In contrast, the intersession ICC values showed poor reliability in the stroke group.

Our finding of a poor intersession reliability is in line with two other studies that have examined the reliability of lower limb MEP amplitude in patients with stroke. Cacchio et al. (2011) reported an intersession ICC value in the paretic leg that had a wide confidence interval that spanned 0, while Wheaton et al. (2009) reported intersession ICCs in two quadriceps muscles ranging from 0.21-0.54. Both of these studies also required participants to generate a low level of muscle activity during collection of MEPs. These consistently low ICC values across studies question the ability to reliably monitor responses to TMS over time in patients with stroke, even though this is routinely undertaken. Muscle fatigue potentially influences the reliability of MEPs collected during muscle activation as there are known changes in corticomotor excitability during fatiguing contractions (Taylor et al., 1996).

However, neuromuscular fatigue is likely to be minimal at the low level of activation used in our study and evidence suggests that any difference between stroke and control participants in terms of neuromuscular fatigue is likely to be negligible (Lindström et al., 1998, Sunnerhagen et al., 1999). Therefore, fatigue is unlikely to account for differences in reliability measures between the two populations.

The intersession TPE of MEP data ranged from 27-150% across both groups; this indicates that much of the poor intersession reliability may be explained by biological variability. TPE can be interpreted as the amount of error in the measure or the amount the measure would have to change by in order to exceed the inherent biological variability. It is generally accepted that a TPE of less than 5% reflects a reliable measure (Hopkins, 2011). Our TPE values were substantially higher than this, even in the control participants. Our data suggest that intersession changes in MEP amplitude need to be greater than 25-35%, depending on the MEP processing procedure used, in healthy participants to be considered real changes; the equivalent change in patients with stroke is approximately 75%. While modulations of MEP size of such magnitude are not unrealistic in the stroke population, it does further question the ability of MEPs to reliably capture changes in corticomotoneuronal excitability over multiple sessions.

Similar to other studies (Cacchio, Paoloni, 2011, Carroll, Riek, 2001, Ngomo, Leonard, 2012, Wheaton, Villagra, 2009), we found that active motor threshold demonstrated good reliability and was superior to the MEP size measures overall. This indicates that differences in threshold between sessions did not contribute substantially to the poor intersession reliability of MEP size in the stroke group. Motor threshold reflects membrane excitability of intracortical neurons as well as the excitability of corticospinal and lower motoneurons,

whereas measures of MEP size reflect the excitability and efficacy of the corticomotoneuronal tract. A large cross-sectional study (Wassermann, 2002) demonstrated a substantial range in interindividual motor threshold in a group of healthy participants but repeated assessments also show a low variability across days. Given the stability of our threshold measures, we suggest that variability in corticospinal tract efficacy between days contributes to the poor reliability of this measure in the stroke group. Of note, the intersession reliability of MVC was only fair to good in the stroke group, a common finding for maximum effort testing of ankle plantarflexor muscles in this population (Hsu et al., 2002, Pohl et al., 2000). This suggests that the stroke participants were likely to be recruiting a differing extent of the motoneuron pool between the two sessions, as the target level of soleus activation was based on a set percentage of MVC, reflecting differing levels of corticospinal tract activation. Indirect descending pathways from the brainstem or spinal level influences such as the lumbar propriospinal interneurons may also have contributed to muscle activation to varying extents. The fact that active motor threshold remained reliable despite these factors indicates that such variation in the level of background muscle activation may not be important in terms of overall reliability.

MEP processing procedures

The reliability statistics for the four processing procedures were similar but amplitude and area measures obtained from single MEPs were consistently superior to those from the averaged responses. This difference between the single and averaged measures was more marked in the amplitude measures and in the stroke group, suggesting that phase cancellation that occurs when averaging responses may make averaged MEP amplitude measurements less reliable. Phase cancellation does not occur when the data are rectified

prior to averaging and the response area or rms measured. Our finding that MEP measures from averaged responses were less reliable outcome measures contrasts with McDonnell and colleagues (McDonnell, Ridding, 2004), who did not find any consistent differences in reliability between ensemble average or individual MEP measures in healthy participants. However, their ICC values were very low overall compared to our intersession values, which may have masked any subtle differences between processing techniques.

Consistent with other studies (Kiers et al., 1995), our overall correlations between MEP amplitude and MEP area measurements were very high, suggesting that these two outcomes provide a similar reflection of MEP size and corticomotor excitability. The lower correlation in the stroke group is suggestive of a greater dispersion of the descending volley or a lack of synchrony in the MEPs in the stroke population, creating a distinction between MEP amplitude (reduced by lack of synchrony) and MEP area (less affected by synchrony) measurements (Groppa, Oliviero, 2012). This variability in responses also explains the smaller MEP amplitude when using the ensemble average compared to individual responses (Rábago et al., 2009). Thus, from our results, we recommend using either MEP_{single} area or MEP_{single} amplitude as measures of corticomotor excitability.

Number of responses

As predicted, the ICC values were higher and the TPE values lower when more MEPs were obtained. Despite this, the reliability measures were not markedly impaired when as few as six MEPs were analysed. With the exception of the stroke group intersession values, in all cases MEP_{single} area and amplitude and MEP_{mean} area ICC values were greater than 0.80 when using six or more MEPs. From these data, and in agreement with previous studies in healthy individuals (Bastani and Jaberzadeh, 2012, Christie, Fling, 2007, Doeltgen, Ridding,

2009) and guidelines for clinical use of TMS (Groppa, Oliviero, 2012), we recommend that as few as six MEPs can be used to achieve reliable within session measures for patients with stroke. Six MEPs should also provide reliable intersession measures for people with an intact nervous system. This is fewer than others have recommended (Bastani and Jaberzadeh, 2012, Doeltgen, Ridding, 2009) but may be influenced by the specific muscle tested.

Study limitations

We evaluated the methodological quality of our study using the consensus checklist for TMS studies (Table 1). The study scored highly with 21 out of a possible 25 criteria controlled. We did not control the hand dominance of participants but instead included people with both dominant and non-dominant side lesions. While we restricted people taking medication known to influence CNS excitability from participating, other forms of medication were allowed but were documented. The participants' history of lower limb activity was not recorded or controlled. While altered cortical representation has been documented in people with a history of skilled motor training, e.g. braille readers (Pascual-Leone et al., 1993), this is unlikely to be a marked factor in the soleus muscle. A limitation of the study methodology was that we did not control the level of activation of other lower limb muscles during the motor task, and therefore the recruitment of synergist or antagonist muscles could have altered the excitability of the target soleus muscle. One further limitation of our study is that, due to our method of participant recruitment, we were unable to determine the mechanism or location of stroke in our participants. Thus, we are unable to evaluate if the reliability of responses is different between patients with cortical and subcortical lesions.

Conclusion

Overall, we found excellent intrasession reliability for MEP measures in stroke patients when recorded within a session. We recommend using a minimum of six stimuli and using MEP_{single} amplitude or area to analyse the data. Given the poor reliability of intersession data, we caution the use of MEPs to monitor changes in lower limb corticomotor excitability over time in this population.

References

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.

Brouwer B, Ambury P. Upper extremity weight-bearing effect on corticospinal excitability following stroke. Arch Phys Med Rehabil. 1994;75:861-6.

Cacchio A, Cimini N, Alosi P, Santilli V, Marrelli A. Reliability of transcranial magnetic stimulation-related measurements of tibialis anterior muscle in healthy subjects. Clin Neurophysiol. 2009;120:414-9.

Cacchio A, Paoloni M, Cimini N, Mangone M, Liris G, Aloisi P, et al. Reliability of TMS-related measures of tibialis anterior muscle in patients with chronic stroke and healthy subjects. J Neurol Sci. 2011;303:90-4.

Caramia MD, Palmieri MG, Giacomini P, Iani C, Dally L, Silvestrini M. Ipsilateral activation of the unaffected motor cortex in patients with hemiparetic stroke. Clin Neurophysiol. 2000;111:1990-6.

Carroll TJ, Riek S, Carson RG. Reliability of the input–output properties of the cortico-spinal pathway obtained from transcranial magnetic and electrical stimulation. J Neurosci Methods. 2001;112:193-202.

Chipchase L, Schabrun S, Cohen L, Hodges P, Ridding M, Rothwell J, et al. A checklist for assessing the methodological quality of studies using transcranial magnetic stimulation to study the motor system: An international consensus study. Clin Neurophysiol. 2012;123:1698-704.

Christie A, Fling B, Crews RT, Mulwitz LA, Kamen G. Reliability of motor-evoked potentials in the ADM muscle of older adults. J Neurosci Methods. 2007;164:320-4.

Doeltgen SH, Ridding MC, O'Beirne GA, Dalrymple-Alford J, Huckabee M-L. Test–retest reliability of motor evoked potentials (MEPs) at the submental muscle group during volitional swallowing. J Neurosci Methods. 2009;178:134-7.

Fleiss J. The design and analysis of clinical experiments. New York: John Wiley; 1986.

Fleming MK, Sorinola IO, Newham DJ, Roberts-Lewis SF, Bergmann JH. The effect of coil type and navigation on the reliability of transcranial magnetic stimulation. IEEE Trans Neural Syst Rehabil Eng. 2012;20:617-25.

Groppa S, Oliviero A, Eisen A, Quartarone A, Cohen LG, Mall V, et al. A practical guide to diagnostic transcranial magnetic stimulation: Report of an IFCN committee. Clin Neurophysiol. 2012;123:858-82.

Hamzei F, Liepert J, Dettmers C, Weiller C, Rijntjes M. Two different reorganization patterns after rehabilitative therapy: An exploratory study with fMRI and TMS. Neuroimage. 2006;31:710-20.

Hopkins WG. Measures of reliability in sports medicine and science. Sports Med. 2000;30:1-15.

Hopkins WG. A new view of statistics. 2011.

Hsu A-L, Tang P-F, Jan M-H. Test-retest reliability of isokinetic muscle strength of the lower extremities in patients with stroke. Arch Phys Med Rehabil. 2002;83:1130-7.

Jung NH, Delvendahl I, Kuhnke NG, Hauschke D, Stolle S, Mall V. Navigated transcranial magnetic stimulation does not decrease the variability of motor-evoked potentials. Brain Stimulat. 2010;3:87-94.

Kamen G. Reliability of motor-evoked potentials during resting and active contraction conditions. Med Sci Sports Exerc. 2004;36:1574-9.

Kiers L, Clouston P, Chiappa KH, Cros D. Assessment of cortical motor output: compound muscle action potential versus twitch force recording. Electroencephalogr Clin Neurophysiol. 1995;97:131-9.

Liepert J, Bauder H, Miltner WHR, Taub E, Weiller C. Treatment-induced cortical reorganisation after stroke in humans. Stroke. 2000;31:1210-6.

Lindström B, Gerdle B, Forsgren L. Repeated maximum reciprocal knee movements in patients with minimal overt symptoms after ischaemic stroke: An evaluation of mechanical performance and EMG. Scand J Rehabil Med. 1998;30:47-54.

Livingston SC, Ingersoll CD. Intra-rater reliability of a transcranial magnetic stimulation technique to obtain motor evoked potentials. Int J Neurosci. 2008;118:239-56.

Malcolm MP, Triggs WJ, Light KE, Shechtman O, Khandekar G, Gonzalez Rothi LJ. Reliability of motor cortex transcranial magnetic stimulation in four muscle representations. Clin Neurophysiol. 2006;117:1037-46.

Marchand-Pauvert V, Simonetta-Moreau M, Pierrot-Deseilligny E. Cortical control of spinal pathways mediating group II excitation to human thigh motoneurones. J Physiol. 1999;517:301-13.

McDonnell MN, Ridding MC, Miles TS. Do alternate methods of analysing motor evoked potentials give comparable results? J Neurosci Methods. 2004;136:63-7.

Merletti R, Farina D, Gazzoni M, Merlo A, Ossola P, Rainoldi A. Surface electromyography: a window on the muscle, a glimpse on the central nervous system. Europa Medicophysica. 2001;37:57-68.

Mortifee P, Stewart H, Schulzer M, Eisen A. Reliability of transcranial magnetic stimulation for mapping the human motor cortex. Electroencephalogr Clin Neurophysiol. 1994;93:131-7.

Ngomo S, Leonard G, Moffet H, Mercier C. Comparison of transcranial magnetic stimulation measures obtained at rest and under active conditions and their reliability. J Neurosci Methods. 2012;205:65-71.

Pascual-Leone A, Cammarota A, Wassermann EM, Brasil-Neto JP, Cohen LG, Hallett M. Modulation of motor cortical outputs to the reading hand of braille readers. Ann Neurol. 1993;34:33-7.

Pohl PS, Startzell JK, Duncan PW, Wallace D. Reliability of lower extremity isokinetic strength testing in adults with stroke. Clin Rehabil. 2000;14:601-7.

Rábago CA, Lancaster JL, Narayana S, Zhang W, Fox PT. Automated-parameterization of the motor evoked potential and cortical silent period induced by transcranial magnetic stimulation. Clin Neurophysiol. 2009;120:1577-87.

Shimizu T, Hosaki A, Hino T, Sato M, Komori T, Hirai S, et al. Motor cortical disinhibition in the unaffected hemisphere after unilateral cortical stroke. Brain. 2002;125:1896-907.

Soto O, Valls-Solé J, Shanahan P, Rothwell J. Reduction of Intracortical Inhibition in Soleus Muscle During Postural Activity. J Neurophysiol. 2006;96:1711-7.

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.

Sunnerhagen KS, Svantesson U, Lönn L, Krotkiewski M, Grimby G. Upper motor neuron lesions: Their effect on muscle performance and appearance in stroke patients with minor motor impairment. Arch Phys Med Rehabil. 1999;80:155-61.

Tallent J, Goodall S, Hortobágyi T, St Clair Gibson A, French DN, Howatson G. Repeatability of corticospinal and spinal measures during lengthening and shortening contractions in the human tibialis anterior muscle. PLoS ONE. 2012;7:e35930.

Taylor JL, Butler JE, M. AG, Gandevia SC. Changes in motor cortical excitability during human muscle fatigue. J Physiol. 1996;490:519-28.

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.

Turton A, Wroe S, Trepte N, Fraser C, Lemon RN. Contralateral and ipsilateral EMG responses to transcranial magnetic stimulation during recovery of arm and hand function after stroke. Electroencephalogr Clin Neurophysiol. 1996;101:316-28.

van Hedel HJA, Murer C, Dietz V, Curt A. The amplitude of lower leg motor evoked potentials is a reliable measure when controlled for torque and motor task. J Neurol. 2007;254:1089-98.

Wassermann EM. Variation in the response to transcranial magnetic brain stimulation in the general population. Clin Neurophysiol. 2002;113:1165-71.

Wheaton LA, Villagra F, Hanley DF, Macko RF, Forrester LW. Reliability of TMS motor evoked potentials in quadriceps of subjects with chronic hemiparesis after stroke. J Neurol Sci. 2009;276:115-7.

Wittenberg G, Chen R, Ishii K, Bushara O, Taub E, Gerber LH, et al. Constraint-induced therapy in stroke: magnetic-stimulation motor maps and cerebral activation. Neurorehabil Neural Repair. 2003;17:48-57.

Figure Legend

Figure 1. Graphs showing the relationship between the motor evoked potential (MEP) measurement procedures in the healthy control (left) and stroke (right) groups. The Pearson correlation coefficient is shown for each relationship. The top figures compare amplitude and area measurements, while the lower figures compare amplitude measures from single and averaged responses.

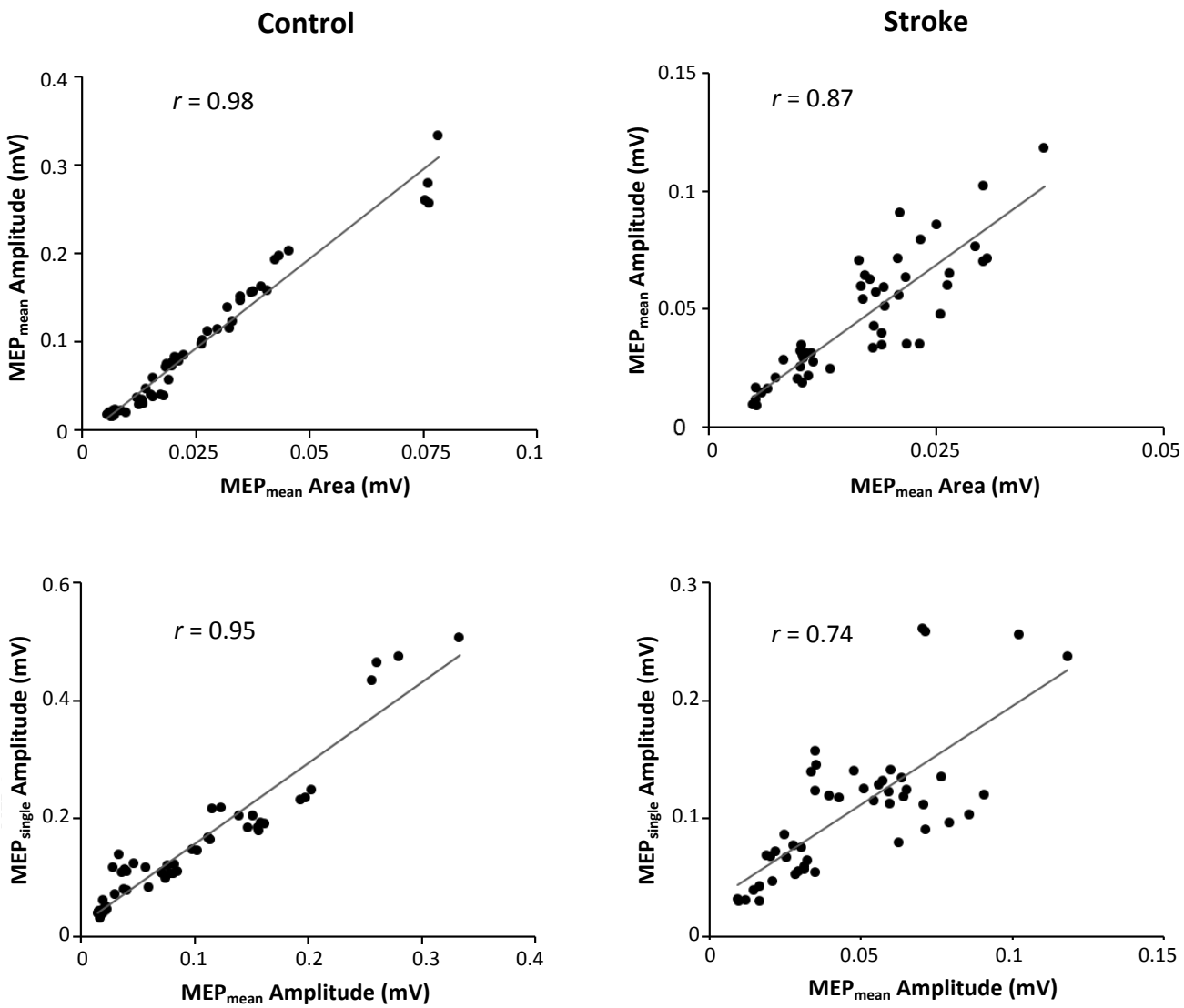
Control
Stroke
r = 0.98
r = 0.87
r = 0.95
r = 0.74
MEP mean Amplitude (mV)
MEP mean Area (mV)
MEP single Amplitude (mV)
MEP mean Amplitude (mV)
0
0.025
0.05
0.075
0.1
0.15
0.2
0.3
0.4
0.6

Table 1. Checklist of transcranial magnetic stimulation (TMS) methodology.

Participant factors	**Controlled**
Age of subjects	Y
Gender of subjects	N/A
Handedness of subjects	N
Subjects prescribed medication	N
Use of CNS active drugs (e.g. anti-convulsants)	Y
Presence of neurological/ psychiatric disorders in healthy subjects	Y
Any medical conditions	Y
History of specific repetitive motor activity	N
Methodological factors	
Position and contact of EMG electrodes	Y
Amount of relaxation/contraction of target muscles	Y
Prior motor activity of the muscle to be tested	Y
Level of relaxation of muscles other than those being tested	N
Coil type (size and geometry)	Y
Coil orientation	Y
Direction of induced current in the brain	Y
Coil location and stability (with or without a neuronavigation system)	Y
Type of stimulator used (e.g. brand)	Y
Stimulation intensity	Y
Pulse shape (monophasic or biphasic)	Y
Determination of optimal hotspot	Y
The time between MEP trials	Y
Time between days of testing	Y
Subject attention (level of arousal) during testing	Y
Method for determining threshold (active/resting)	Y
Number of MEP measures made	Y
Analytical factors	
Method for determining MEP size during analysis	Y

CNS = central nervous system; EMG = electromyography; MEP = motor evoked potential.

Table 2. Intrasession reliability measures for the stroke and control groups.

	10 MEPs		**8 MEPs**		**6 MEPs**		**4 MEPs**	
	ICC	**TPE**	**ICC**	**TPE**	**ICC**	**TPE**	**ICC**	**TPE**
Stroke	*n=11*		*n=13*		*n=13*		*n=13*	
MEP_{mean} amplitude	0.81 (0.26-0.95)	35	0.76 (0.24-0.93)	40	0.86 (0.64-0.96)	34	0.76 (0.27-0.93)	43
MEP_{single} amplitude	0.97 (0.88-0.99)	36	0.95 (0.84-0.98)	38	0.94 (0.80-0.98)	43	0.88 (0.58-0.97)	55
MEP_{mean} area	0.93 (0.74-0.98)	18	0.90 (0.69-0.97)	23	0.88 (0.61-0.96)	26	0.87 (0.56-0.96)	26
MEP_{single} area	0.96 (0.84-0.99)	35	0.94 (0.79-0.98)	38	0.92 (0.73-0.97)	41	0.86 (0.52-0.96)	51
Control	*n=12*		*n=13*		*n=13*		*n=13*	
MEP_{mean} amplitude	0.90 (0.65-0.97)	44	0.88 (0.62-0.96)	44	0.85 (0.51-0.95)	47	0.74 (0.12-0.92)	55
MEP_{single} amplitude	0.94 (0.80-0.98)	30	0.93 (0.78-0.98)	32	0.92 (0.75-0.98)	31	0.84 (0.46-0.95)	41
MEP_{mean} area	0.93 (0.77-0.98)	32	0.92 (0.75-0.98)	33	0.92 (0.75-0.98)	32	0.86 (0.52-0.96)	39
MEP_{single} area	0.94 (0.80-0.98)	30	0.94 (0.80-0.98)	30	0.93 (0.79-0.98)	29	0.84 (0.46-0.93)	39

MEP = motor evoked potential; ICC = intraclass correlation coefficient with 95% confidence interval; TPE = typical percentage error.

Table 3. Intersession reliability measures for the stroke and control groups.

	10 MEPs		8 MEPs		6 MEPs		4 MEPs	
	ICC	**TPE**	**ICC**	**TPE**	**ICC**	**TPE**	**ICC**	**TPE**
Stroke	*n=12*		*n=12*		*n=12*		*n=12*	
MEP_{mean} amplitude	0.30 (-1.2-0.79)	150	0.36 (-1.0-0.81)	127	0.36 (-1.0-0.81)	111	0.28 (-1.2-0.78)	95
MEP_{single} amplitude	-0.1 (-3.4-0.71)	76	-0.1 (-3.3-0.71)	77	-0.1 (-3.5-0.70)	75	0.0 (-2.8-0.72)	72
MEP_{mean} area	0.25 (-1.5-0.78)	79	0.27 (-1.5-0.79)	80	0.33 (-1.3-0.81)	76	0.30 (-1.4-0.80)	74
MEP_{single} area	0.18 (-2.0-0.77)	74	0.14 (-2.1-0.76)	77	0.20 (-1.9-0.77)	74	0.29 (-1.4-0.80)	70
Control	*n=12*		*n=13*		*n=13*		*n=13*	
MEP_{mean} amplitude	0.82 (0.36-0.95)	49	0.78 (0.25-0.93)	47	0.71 (0.01-0.91)	50	0.65 (-0.2-0.90)	49
MEP_{single} amplitude	0.94 (0.79-0.98)	27	0.93 (0.76-0.98)	28	0.92 (0.75-0.98)	29	0.84 (0.48-0.95)	35
MEP_{mean} area	0.89 (0.60-0.97)	35	0.88 (0.60-0.96)	34	0.85 (0.50-0.96)	37	0.75 (0.15-0.92)	44
MEP_{single} area	0.91 (0.68-0.97)	31	0.90 (0.65-0.97)	33	0.88 (0.60-0.96)	35	0.80 (0.31-0.94)	41

MEP = motor evoked potential; ICC = intraclass correlation coefficient with 95% confidence interval; TPE = typical percentage error.

Table 4. Group data (Session 1) and intersession reliability measures for maximum voluntary contraction (MVC) and active motor threshold (AMT).

	Mean±SD	**ICC**	**TPE**
Stroke			
MVC	196±95 N	0.62 (-0.23-0.89)	37
AMT	58**±7%**	0.82 (0.34-0.95)	7.9
Control			
MVC	277±137 N	0.81 (0.37-0.94)	30
AMT	55**±**7%	0.92 (0.75-0.98)	4.8

SD = standard deviation; ICC = intraclass correlation coefficient with 95% confidence interval; TPE = typical percentage error; N = newtons.