Reliability of transcranial magnetic stimulation measurements of maximum activation of the knee extensors in young adult males

Rachel Malcolm[1] PhD, Simon Cooper[1] PhD, Jonathan Folland[2] PhD, Christopher Tyler [3] PhD, Ricci Hannah[4] PhD, Caroline Sunderland[1] PhD

[1]Sport, Health and Performance Enhancement (SHAPE) Research Centre, Department of Sport Sciences, Nottingham Trent University, UK. [2]School of Sport, Exercise and Health Sciences, Loughborough University, UK. [3]Department of Life Sciences, University of Roehampton, UK. [4]Department of Psychology, University of California, San Diego, USA.

Address for correspondence: Dr Rachel Malcolm

Sport, Health and Performance Enhancement Research Centre, Department of Sports Science, Nottingham Trent University, Clifton Lane, Nottingham. NG11 8NS. United Kingdom.

ORCID ID: 0000-0003-3494-3835

Email: rachel.malcolm@ntu.ac.uk

Abstract

Purpose: Transcranial magnetic stimulation (TMS) provides an indication of changes occurring in the corticospinal pathway. This study aimed to determine the between-day (trials 1 week apart) and within-day (trials 1h apart) reliability of TMS and peripheral nerve stimulation. Methods: 22 male participants (age 23±4 years; height 1.80±0.07 m; body mass 75.1±11.7 kg; body mass index 23.1±2.5 $kg.m^{-2}$) completed 2 familiarisation sessions and 3 experimental trials (trial 2 and 3 split by 1h). The interpolated twitch technique was used to determine TMS-assessed voluntary activations ($VA\text{-}_{TMS}$) superimposed on submaximal and maximal leg extension performed on a custom-built dynamometer. Reliability was assessed using equivalence tests, systematic error, 95 % limits of agreement, intraclass correlation coefficient (ICC) and coefficient of variation (CV). Results: $VA\text{-}_{TMS}$ was equivalent between-day (94.1±4.4% versus 93.7±4.9%, *P<0.01*) and within-day (93.7±4.9% versus 93.7±4.8%, *P*<0.01). Systematic error (95% limits of agreement) for $VA\text{-}_{TMS}$ was −0.5% (−5.1%, 4.2%) for between-day and -0.0% (−5.3%, 5.4%) for within-day. ICC and CV values demonstrated high reliability between-day (ICC=0.93, CV=2.5%) and within-day (ICC=0.92, CV=2.9%). Conclusion: Results indicate that TMS can reliably estimate the output of the motor cortex to the knee extensors, both between-day and within-day. The findings have been used to estimate sample sizes for this technique for future research.

Keywords: Cortical, force, neuromuscular, quadriceps, TMS, voluntary activation.

1. Introduction

Voluntary muscle activation, a key factor in neuromuscular function and thus the changes that occur with training, ageing, fatigue and injury, is widely assessed with the interpolated twitch technique (ITT) in order to quantify and express neural drive to the agonist muscles (Merton 1954; Gandevia 2001). The interpolated twitch technique involves comparing the force response to a post-contraction potentiated twitch at rest to the response when a stimulus is superimposed on top of a maximal voluntary contraction (Behm et al. 1996; Gandevia 2001). The stimulus can be delivered at either the peripheral or cortical level in order to measure peripheral or transcranial magnetic stimulation assessed voluntary activation (VA-$_{TMS}$) (Herbert & Gandevia, 1996; Gandevia et al. 1996). VA-$_{TMS}$ is estimated using transcranial magnetic stimulation (TMS) and provides an indication of the lack of drive from the cortical and subcortical structures to the motor cortex, which is known to be involved in volitional movement (Goodall et al. 2015a; Goodall et al. 2015b; Goodall et al. 2017; Goodall et al. 2014; Sidhu et al. 2009; Temesi et al. 2014; Temesi et al. 2017). Specifically, understanding the mechanisms of fatigue for the knee extensors is extremely important and has implications for exercise performance and health. Peripheral voluntary activation can be used to reflect modulation at any level of the central nervous system but is limited to distinguish these specific mechanisms (Gandevia 1992).

TMS coupled with voluntary contractions has been used extensively in the assessment of cortical function and central fatigue (Dekerle et al. 2019; Goodall et al. 2015a; Goodall et al. 2015b; Goodall et al. 2017; Goodall et al. 2014; Klass et al. 2016; Ross et al. 2010; Ross et al. 2007; Sidhu et al. 2009; Temesi et al. 2014; Temesi et al. 2017). This technique utilises the superimposed twitch response (SIT) upon contractions of various intensities to create a linear regression (Todd et al. 2003). The SIT represents the force difference between the peak

63 twitch torque and the force immediately prior to stimulation (e.g. the voluntary contraction).
64 This method is required due to the lack of corticospinal excitability at rest, meaning the
65 resting twitch would not accurately depict a maximum twitch. Extrapolating the regression,
66 using Y=0, provides an estimated resting twitch (ERT), to demonstrate the response if the
67 muscle was at rest.

68 Goodall et al (2014) found a high within day reliability for the estimation of VA-$_{TMS}$
69 (CV=10.2%, ICC=0.82) and between day reliability (CV=2.2%, ICC=0.87) (Goodall et al.
70 2017). However, despite demonstrating a good level of reliability, these studies used 7 and 8
71 participants, respectively. Two further studies have assessed the reliability of VA-$_{TMS}$ in this
72 muscle group in greater detail (Goodall et al. 2009; Sidhu et al. 2009) and similarly, these
73 studies had low participant numbers and did not control for time of day, despite known
74 variations in both isometric knee extensor torque (Callard et al. 2000; Racinais et al. 2005)
75 and corticospinal excitability (Tamm et al. 2009) across the day. Sidhu et al. (2009) only
76 assessed between day reliability, whereas Goodall et al. (2009) was the first study to assess
77 between and within day reliability in this muscle group. A high degree of within day
78 reliability (CV=3.7% (Goodall et al. 2009) and CV=3.1% (Sidhu et al. 2009)) and between
79 day reliability (Goodall et al. 2009) has previously been demonstrated. More recently Dekerle
80 et al. (2019) found between day measures of VA$_{TMS}$ to be highly reliable. However, one issue
81 prevailing across current TMS related reliability studies is the limited participant numbers
82 (n=8-12) (Dekerle et al. 2019; Goodall et al. 2009; Heroux et al. 2015; Ngomo et al. 2012;
83 Sidhu et al. 2009). This has been highlighted as a major flaw in neurophysiology research in
84 the review by Heroux et al. (2015) lending to irreproducible results. Thus, based upon the
85 recommendations of Atkinson and Nevill (Atkinson and Nevill 2001) a reliability study
86 employing an appropriate sample size >20, alongside controlling for time of day is warranted
87 in order to reflect the true variability of a technique.

A further issue for the estimation of VA_{TMS} is the means by which the optimal intensity for stimulation is established. A way in which optimal intensity can be established is through the production of individualized stimulus-response curves, however, this has yet to be thoroughly tested. The need to establish the optimal stimulation intensity is due to risk of biasing your estimated of VA_{TMS} if the stimulus is not supramaximal or if the stimulus recruits too much hamstring activation. A preliminary aim was to set up a study in which optimal stimulus intensity was determined, in order to inform the reliability study.

Electromyography (EMG) responses also provide vital information regarding the stimulation conditions (Ngomo et al. 2012). The EMG responses to TMS show a motor evoked potential (MEP), which can provide further information regarding the corticospinal excitability of a muscle (Goodall et al. 2012). Therefore, documenting the reliability of the stimulation conditions during contractions which are greater than 50% MVC may be useful in extending the understanding of the mechanisms involved in fatigue.

The aim of the present study was to assess the between-day and within-day reliability of VA_{TMS}, whilst also assessing the influence of familiarisation on reliability. It was hypothesised that these measures would be highly reproducible, and that the inclusion of two familiarisation sessions would improve the reliability of these measures.

A prominent issue across neurophysiology research is inadequate sample sizes (Heroux et al. 2016). Therefore, an additional aim was to provide researchers with sample size and power analysis using the intra-class correlations from our study to inform good practice in future research.

2. Methods

2.1.Participants

Twenty-two recreationally active males (age 23±4 years; height 1.80±0.07 m; body mass 75.1±11.7 kg; body mass index 23.1±2.5 $kg.m^{-2}$) participated in the study. Volunteers completed a health screen questionnaire and provided written informed consent prior to their participation in this study that was approved by the University Human Ethical Review Committee.

2.2.Study Design

2.2.1. Stimulus Response Study

A preliminary study was completed in 10 participants (mean ± SD): age 25 ± 2 years, height 181 ± 1 cm, body mass 77 ± 2 kg and body mass index 24 ± 1 $kg.m^2$ to establish the optimal stimulation intensity. Participants visited the laboratory for 90 min on two separate occasions to complete a familiarisation and one test session.

2.2.2. Reliability Study

Participants were required to attend the laboratory at the same time of day on four separate occasions to complete two familiarisation sessions (where any contraindications to TMS were identified) and two experimental sessions. The familiarisation sessions involved an identical protocol as the experimental sessions. Between day reliability was assessed between the second familiarisation and experimental session one, to establish the need for more extensive familiarisation protocols. Session one and session two were separated by exactly one week to further assess between-day reliability. Experimental session two involved two repeats of the protocol, separated by 1h, allowing for the assessment of within-day reliability.

2.3. Measurements

2.3.1. Knee extension force

A custom-built dynamometer was used to measure knee extensor torque (Johnson et al. 2015). The participant was sat with hip and knee joints at 100˚(full extension=180˚). Participants were strapped to the chair across the chest and pelvis, in order to prevent displacement during contractions. The dominant leg was strapped to a strain gauge (615, Teda-Huntleigh, Herzliya, Israel), positioned perpendicular to the tibia, using an ankle cuff which was 2 cm proximal to the medial malleolus. The force signal was amplified (×1000) and sampled at 2000 Hz using an external A/D converter (1401; CED, Cambridge, UK), interfaced with a personal computer (PC) using Signal 5.08 software.

2.3.2. Electromyography (EMG)

Electromyographic signals were recorded form the vastus lateralis (VL), rectus femoris (RF), vastus medialis (VM) and biceps femoris (BF). Bipolar surface electrodes (2.5cm between-electrode distance; silver/silver chloride, 95mm^2 area, Ambu, Ballerup, Denmark) were attached to each muscle at distances based on percentages of thigh length measured from the knee joint space to greater trochanter: VL, 45%; RF, 55%; VM, 25%; BF, 50%. Electrodes were placed parallel to the assumed orientation of the muscle fibres and medio-laterally over the belly of the muscle. EMG signals were pre-amplified by active EMG leads (input impedance 100M, CMMR > 100 dB. Base gain 500, 1st order high pass filter set to 10 Hz; Noraxan, Scottsdale, U.S.A) connected in series to a custom-built junction box and subsequently to the same A/D converter and PC software that enabled synchronisation with the force data. The signals were sampled at 2000 Hz. EMG data was band-pass filtered in both directions between 20 and 450Hz using a fourth-order zero-lag Butterworth filter prior to analysis.

155 **2.3.3. Electrical stimulation of the femoral nerve**

156 Electrical stimuli were delivered via percutaneous stimulation over the femoral nerve. A
157 cathode stimulation probe (1cm diameter,), complete with an anode (4 × 7cm carbon rubber
158 electrode, Electro Medcal Supplies), was coated in electrode gel and pressed into the femoral
159 triangle to achieve stimulation. Square wave pulses, 200μs in duration were delivered via a
160 constant current variable voltage stimulator (Model DS7AH, Digitimer, Ltd, Welwyn Garden
161 City, UK), using a submaximal electrical current (approximately 40mA). The precise location
162 of the femoral nerve was determined by both the greatest twitch response and the M-wave
163 response to this current and marked with indelible ink to ensure reproducibility of placement.
164 In order to ensure supramaximal stimulation, the current was increased in a step-wise manner
165 until the amplitude of twitch force and M-waves plateaued. The current at this plateau was
166 multiplied by 1.2 and used for the remainder of the testing (Buckthorpe et al. 2012).

167 **2.3.4. Transcranial magnetic stimulation of the motor cortex**

168 TMS of the motor cortex was achieved using a double-cone coil (11cm diameter) and
169 delivered via a magnetic stimulator (Magstim 200^2, Magstim Company Ltd, Dyfed, UK). The
170 coil was held manually over the motor cortex on the side contralateral to the involved leg and
171 oriented in order to induce a posterior-to-anterior current in the brain. The optimal coil
172 position on the scalp (hotspot) to maximally stimulate the knee extensor muscles was located
173 by initiating stimulation during submaximal voluntary contractions (20% MVF), with a
174 stimulation intensity of approximately 80% of maximum output (MO). The optimal position
175 was defined as the one producing the greatest SIT force and MEP amplitude, and was marked
176 using indelible ink.

177 The resting motor threshold was determined by decreasing the stimulation intensity in 1%
178 increments, starting at 100% of maximum output until the MEP amplitude of the knee

179 extensors exceeded 50 μV in approximately 50% of stimuli across 5 stimuli. The active
180 motor threshold was determined whilst the participant performed 20% MVF contraction, and
181 by decreasing the stimulation intensity in 1% increments starting at the resting motor
182 threshold. The active motor threshold was defined as the point where MEP amplitude of the
183 knee extensors was just over 200μV in 50% of stimuli. The stimulus intensity at active motor
184 threshold was multiplied by 1.4 and used throughout the remainder of the testing.

185 **2.4.Stimulus Response Study Protocol**

186 18 sets of contractions (80, 60, 40, 20% MVF) were performed with TMS delivered at the
187 plateau of each contraction and at rest. Sets were organised into 3 blocks of 6. For each set
188 within a block, TMS was delivered at one of six intensities (50 – 100% MSO), such that three
189 sets were performed at each stimulus intensity.

190 Two-way repeated measures ANOVA was used to evaluate the effect of voluntary force level
191 [0, 20, 40, 60 and 80% MVF] and stimulus intensity [50, 60, 70, 80, 90, 100%] on absolute
192 TMS-evoked twitch force, and normalised knee extensor and flexor MEP amplitudes.
193 Significant interactions were followed-up by using separate one-way repeated measures
194 ANOVA to evaluate the influence of voluntary force level and stimulus intensity on TMS-
195 evoked twitch force, knee extensor and flexor MEPs, using Bonferroni corrected t-tests to
196 locate specific differences.

197 Linear regressions were performed to evaluate the relationship between voluntary force level,
198 between 60-80% MVF, and TMS-evoked force at the same force levels, for each of the 6
199 TMS stimulus intensities. For each stimulus intensity, the linear regressions involved using
200 the three stimuli at each force level, such that regressions comprised of 6 data points in total.
201 The explained variance (R^2) were calculated for each stimulus intensity and compared using

one-way repeated measures ANOVA with Bonferroni corrected t-tests to locate specific differences between stimulus intensities.

The intensity chosen for the reliability study was due to this preliminary stimulus response data demonstrating a plateau in stimulus response at approximately 130% of AMT, therefore this intensity ensures stimulation is supramaximal.

2.5. Reliability Study Protocol

2.5.1. Preliminary measures

Once seated, the skin was prepared, and EMG electrodes applied. The femoral nerve was located at rest and stimulation thresholds determined. 8 (2 at each intensity) submaximal voluntary warm-up contractions were completed at 55%, 70%, 85% and 90% of estimated MVF.

2.5.2. Maximal voluntary force

Prior to and following the 6 sets of voluntary contractions, 2 MVCs of the knee extensors, each separated by 20s rest, were performed. Participants were instructed to contract as hard as possible for 3s, whilst being provided with biofeedback and encouragement. This process, allowed the assessment of whether the protocol had resulted in significant levels of fatigue, defined as a reduction in MVF.

Hotspot location was then determined prior to establishing motor threshold during active contraction (20% of MVF) and at rest.

2.5.3. Voluntary contractions with electrical stimulation and TMS

Participants completed 6 sets of voluntary contractions, lasting approximately 3 s and separated by 10s (Figure 1). Three sets (sets 1, 3 and 5) consisted of a single contraction at

224 100%, 85%, 70% and 55% MVF with magnetic stimulation superimposed on the peak of the
225 contraction. The remaining 3 sets (sets 2, 4 and 6) consisted of one contraction at 100% MVF
226 with electrical stimulation superimposed on the peak of the contraction and immediately after
227 the contraction. Each set was separated from the preceding set by 120s. Measurements
228 included volitional force at the onset and peak of the superimposed twitch, in order to
229 calculate twitch magnitude, and peak to peak amplitudes of the MEP (TMS) and M-wave
230 (electrical stimulation). An average value amplitude was calculated across the 3 sets for each
231 session. These values were normalised to Mmax, which is defined as the maximum M-wave
232 potential seen within the session (Gandevia 2001).

233 **2.6. Data and statistical analyses**

234 *Voluntary activation*

235 Peripheral voluntary activation was quantified by measurement of the torque responses to
236 electrical stimulation of the femoral nerve. SIT was determined by placing a cursor at the
237 onset of the twitch and a cursor at the peak of the twitch force, the force at the onset of the
238 twitch was then taken from the peak twitch force. RT was determined using the same method.
239 These values were substituted into the following equation (Morrison et al. 2004):

240 $$\text{Peripheral voluntary activation} = 1 - (\text{SIT}/\text{RT}) * 100$$

241 VA-$_{TMS}$ was quantified by measurement of torque responses to magnetic stimulation of the
242 motor cortex. Estimated resting twitch (ERT) was established by extrapolating the linear
243 regression for SIT against torque, ranging between 55% and 100% of MVC. All Only
244 participants who achieved a linear regression where $r^2 \geq 0.9$ were included for analysis. SIT
245 for VA-$_{TMS}$ was determined using the same method as above, utilising the TMS response

during a maximal contraction (100% MVF). The y-intercept was deemed the ERT. $VA\text{-}_{TMS}$ was determined using the equation (Todd et al. 2003):

$$VA\text{-}_{TMS} = 1-(SIT/ERT) *100$$

Quadriceps MEP values were averaged across contraction sets and sessions. Mean and SDs of all MEP and M-wave amplitudes (p-p) were calculated from each contraction set. MEP values at 100% MVF were normalised to Mmax (e.g. MEP amplitude at 100% MVF/Mmax*100). BF and VL MEP values were normalized to the maximal voluntary EMG values during maximal voluntary knee flexion and extension, respectively.

MVF, measured at the beginning and end of each session, was the highest instantaneous torque achieved across two maximal contractions.

SPSS 22.0 software for Windows (SPSS Inc., Chicago, IL, USA) was used for the statistical analysis. Data are presented as mean ± standard deviation (SD) and the level of statistical significance was set at $P \leq 0.05$. To determine any differences in values between one experimental session and the next, equivalence tests were used with bounds based on 10% of the mean. T test were used to assess the change in MVF across each session to establish alterations in fatigue. ICCs were applied to assess how well electrical stimulation and TMS measurements correlated within-day and between-day. 95% confidence intervals were established. In order to determine and compare the variability in measures between the two conditions, the coefficient of variation (CV) and 95% limits of agreement (Bland and Altman 1986) were established. An ICC close to 1 demonstrates excellent reliability, with anything over 0.9 indicating high reliability and an ICC below 0.8 exhibiting questionable reliability (Atkinson and Nevill 1998). A CV value of less than 10% is also required to demonstrate high reliability.

2.6.1 Computer Simulation

A statistical simulation of 10,000 experiments (https://github.com/keithlohse/power_reliability) was run based upon the ICC from the present study to provide information regarding the statistical power for a range of study designs (independent t-test, paired t-test, within & between 2 x 2 ANOVA) and sample sizes. This enables the use of the robust reliability statistics established within this paper to provide practical recommendations for future studies. Hence helping improve the methodologies of future neuroscience research.

3. Results

3.1.Stimulus Response Study

There was a main effect of stimulus intensity on the R^2 value of the relationship between voluntary force and evoked force ($P < 0.001$). The R^2 was lower for 50% TMS stimuli by comparison with 100% MSO and ES ($P < 0.001$).

A main effect of stimulus intensity ($P < 0.001$) indicated that there were significant differences in the y-intercept of the regression between voluntary force and TMS force (60-80% MVF), with lower values for 50% and 60% stimuli compared to 100% MSO (all $P < 0.01$).

Despite a significant effect of stimulation intensity on relative antagonist MEP amplitude for the 80% MVF contraction ($P < 0.05$), follow-up tests revealed no significant differences between stimulation intensities, likely as a result of the very wide inter-individual variability.

The stimulus response curves demonstrate that there is no decrease in twitch force with increasing stimulation intensity (Figure 2). Hence, an increase in co-activation of the

hamstrings does not influence TMS-evoked twitch size. Additionally, when participants contract to 80 % MVF the SIT doesn't change regardless of stimulation intensity (Figure 2). The point at which the stimulus response curve plateaued was around 120-140 % AMT, hence we deemed 130 % - 140% of AMT to be supramaximal.

3.2.Reliability Study

3.2.1. Familiarisation

The between-day reliability for the second familiarisation session and the first experimental session show lower levels of reliability for VA-$_{TMS}$ (91.3±5.4% vs 94.1±4.4%, CV = 3.9 % and ICC = 0.78) than the between-day reliability for experimental session 1 and 2 (94.1 ± 4.4% versus 93.7±4.9%, CV = 2.5 % and ICC = 0.93) (Table 1 & 2; Figure 3).

3.2.2. Voluntary activation

There were no systematic differences in maximal VA-$_{TMS}$ either between-day (94.1 ± 4.4% versus 93.7±4.9%, CV = 2.5 %) or within-day (93.7±4.9% versus 93.7±4.8%, CV = 2.9 %) (Table 1; Figure 3). There were no systematic differences in maximal peripheral voluntary activation either between-day (94.0±3.8% versus 94.0±3.6%, CV = 2.2 %) or within-day (94.0±3.6% versus 93.0±3.5%, CV = 2.7 %). (*See supplementary material for representative traces*).

3.2.3. Potentiated twitch

Potentiated resting twitch force demonstrated no systematic difference between-day (186±31N versus 190±31N, CV = 7.4 %*;* Table 3), however a difference was seen within-day, with a lower potentiated twitch in the second session (190±31N versus 178±36 N, CV = 15 %).

3.2.4. MEP characteristics

Between-day and within-day CV values for MEP amplitude during a maximal voluntary contraction ranged from 16.7% to 43.3% of Mmax. Between-day and within-day CV values for Mmax amplitude ranged from 10.8% to 47.7% and Between-day and within-day CV values for Mmax area ranged from 10.2% to 52.5%.

Peak-to-peak MEP amplitude in all three quadriceps muscles at 55% of MVC was VM=63.9%, RF=74.3%, VL=64.9%. As contraction intensity continued to increase up to 100%, MEP amplitudes decreased. (*See supplementary material for representative traces).*

Thresholds

No systematic difference existed between-day or within-day for any RMT, AMT and ES threshold (Table 3). Within-day CV was 2.3 % for RMT, for 2.7 % AMT and 5.7 % for ES threshold. Between- day CV was 3.8 % for RMT, 6.6 % for AMT and 10.9 % for ES threshold.

3.2.5. MVF

Measures of reliability for MVFs can be found in Table 4. Between-day and within-day CV values for MVF were 6.8% and 5.4% respectively. Group mean MVF before the stimulation protocol when compared to after the protocol were not significantly different for experimental session 1 ($P = 0.31$), 2 ($P = 0.13$) or 3 ($P = 0.84$).

3.2.6. Computer simulation

Results of the computer simulations of 1,0,000 experiments can be found in Tables 5 and 6. The findings highlight the participant numbers required, dependent on the study design and

335 methodology (e.g. TMS or electrical stimulation), to ensure results are reliable for that
336 technique.

337 4. Discussion

338 4.1.Stimulus Response Study

339 The stimulus-response confirmed that ERT was better estimated, indicated by the greater R^2
340 value, with a greater stimulator output. It was also proven that any co-activation of the
341 hamstrings which may have accompanied the increase in stimulation intensity did not
342 influence TMS-evoked twitch size. This preliminary study provided the rational for the
343 intensities used within the main study to evaluate the reliability of TMS in the determination
344 of voluntary activation.

345 4.2. Reliability Study

346 The main findings of this study demonstrate that when participants are extensively
347 familiarised (two familiarisation sessions) and conditions, such as time of day and prior
348 exercise, are tightly controlled, TMS provides a reliable between-day (CV=2.5%) and within-
349 day (CV=2.9%) estimate of $VA\text{-}_{TMS}$ for the knee extensors. The level of reliability seen for
350 TMS was comparable to that of electrical stimulation (CV=2.2 & 2.7%, between-day and
351 within-day, respectively) providing rationale for its use in conjunction with electrical
352 stimulation to provide a more detailed understanding of muscle function and specifically the
353 contribution of the motor cortex and subcortical structures (Table 1). The incorporation of a
354 between-day and within-day design also supports the use of this technique in acute and
355 repeated-measures research designs. Secondary measures which were assessed showed
356 variable levels of reliability, with motor thresholds demonstrating high within- and between-
357 day reliability (Table 3), whereas M-wave and MEP responses demonstrated a substantial

amount of between- and within-day variability. Therefore, care must be taken when interpreting this data.

The CV values in previous reliability studies (3.7% & 3.1%), with participant numbers of 9 and 8, were higher than that of the present study (Goodall et al. 2009; Sidhu et al. 2009; Dekerle et al. 2019). Whereas the CV values in the present study demonstrate better reliability, highlighting the important of greater statistical power. However, Dekerle et al (2019) disproved this with equally low CV values with a participant number of 10.

The results of the computer simulation suggest that for studies using a between day design, a participant number of 20 is necessary for both cortical voluntary activation and peripheral voluntary activation. Whereas, for a within day design, particularly for peripheral voluntary activation, a participant number of greater than 20 is required in order to provide an adequately powered estimation of activation.

4.2.1. Voluntary activation

The random error component displayed in this study is smaller than that of Goodall et al. (2009) therefore demonstrating that the TMS technique is more reliable at determining VA-$_{TMS}$ than initially indicated in a smaller participant group. Further the difference in systematic bias between our second familiarisation trial and the first main visit is evidence that more extensive familiarisation protocols are necessary in studies related to TMS and peripheral nerve stimulation.

Our reliability coefficients exhibit good levels of reliability with an ICC of 0.93 for between-day and 0.92 for within-day (Table 1). These values are similar to those found by both Goodall et al. (2009) and Sidhu et al. (2009) ICC=0.94 and ICC=0.95, respectively, for between-day VA-$_{TMS}$. Collectively our findings, and those of Sidhu et al. (2009) and Goodall

et al. (2009), indicate that the interpolated twitch technique using TMS can provide a reliable estimation of maximal VA-$_{TMS}$ of the knee extensors, therefore providing important implications for research involving locomotion or full-body exercise, where understanding the activation deficits of this muscle group is particularly important for performance.

In agreement with Sidhu et al. (2009) levels of voluntary activation measured with motor nerve stimulation were similar to those measured with TMS (between-day mean for VA-$_{TMS}$ was 93.9% versus 94.0% for peripheral and within-day mean for VA-$_{TMS}$ was 93.7% versus 93.5% for peripheral). Therefore, the high levels of reliability demonstrated for peripheral and VA-$_{TMS}$ both in the current study and in the literature (Amann et al. 2013; Goodall et al. 2015a; Sidhu et al. 2009) support the use of electrical stimulation and TMS in parallel to provide an extensive understanding of the diverse contributions to fatigue. This will provide information regarding both spinal and supraspinal contributions to fatigue (Dekerle et al. 2019).

4.2.2. Motor evoked potentials

The largest MEP amplitude was seen at a contraction intensity of 55% of MVF. In agreement with previous literature in the knee extensors (Goodall et al. 2009; Sidhu et al. 2009), elbow flexors (Todd et al. 2003; Todd et al. 2004) and wrist extensors (Lee et al. 2008) increasing the contraction intensity caused minimal changes in MEP amplitude beyond the value achieved at 55% MVF. This suggests that at 55% MVF the greatest activation of motoneurons occurs. Motoneuronal output may be decreasing in response to the stimulus as contraction strength increases, causing minimal change in MEP characteristics. The findings of this study suggest a large amount of variability exists for MEP characteristics both within-day and between-day. Between-day mean amplitude value for ICC was 0.59. Within-day mean amplitude ICC was 0.84. The values for systematic bias and random error in the current

study suggest EMG both between-day and within-day may not provide a reliable indication of corticospinal excitability. The variation in MEP data is greater than that seen in Goodall et al. (2009), but similar to that of Sidhu et al. (2009). These results demonstrate questionable levels of reliability (Atkinson and Nevill 1998), as all values fall below the 0.9 threshold associated with high reliability. Mathur et al. (2005) assessed the reliability of EMG measurements in the knee extensors and reported ICC values ranging from 0.58 to 0.99 for amplitude. The findings highlight variability exists and care must be taken when utilising EMG data to assess corticospinal excitability. This variability could be linked to the variability in Mmax shown in the current study, which may have influenced results during normalization of MEPs.

The size of an MEP can vary greatly between one stimulation and the next (Magistris et al. 1998). The variability may be due to the large number of contributing factors influencing MEP characteristics, including the excitability of the motor cortex and nerve roots and the conduction along the peripheral motor pathway of the muscles (Kobayashi and Pascual-Leone 2003). Therefore, the intrinsic fluctuations in neural excitability at both the cortical and spinal levels may have resulted in highly variable MEP amplitudes (Rossini et al. 1994). Variation in EMG recordings has also been commonly attributed to changes in the orientation of the recording electrodes and minor difference in skin preparation (Mathur et al. 2005) and a number of other non-physilogical factors such as motor unit synchronisation and signal cancellation (Buckthorpe et al. 2012). Despite effort being made to eradicate these issues in the present study, it is still possible that these factors may have influenced the reliability of the technique. Therefore, in agreement with Buckthorpe et al. (2012), individual EMG data is highly variable between measurement sessions, whether it has been normalised or not.

4.2.3. Limitations and future considerations

429 The TMS intensity in the present study is higher than that used in a number of other studies
430 (Goodall et al. 2009; Sidhu et al. 2009). Participants in this study were not screened to ensure
431 a low stimulation threshold, which may account for the higher stimulation intensity observed.
432 Some research suggests that such a high stimulation intensity may cause antagonist muscle
433 co-activation. However, co-activation should not influence results as the thresholds are likely
434 to be similarly high for both the quadriceps and hamstrings, hence not evoking a
435 disproportionate amount of hamstring activity, as previously suggested with high stimulation
436 intensities. Additionally, the intensity of thresholds may inevitably vary to some extent
437 between stimulators. Finally, further larger scale reliability studies are still warranted in this
438 area, which also extend to both females and the elderly.

439 **5. Conclusions**

440 The findings of the present study suggest that VA-$_{TMS}$ of the knee extensors can be reliably
441 estimated between-day (CV=2.5%) and within-day (CV=2.9%) using TMS following an
442 extensive familiarisation process. Further, the incorporation of the computer simulation
443 techniques enables the findings of this study to be applied to future research to enable
444 justification of adequate participant numbers. The use of both a between-day and within-day
445 design also supports the use of this technique in research which may require repeated
446 measures during acute and chronic research protocols and designs. Therefore, TMS can be
447 used reliably to estimate the extent to which output from the motor cortex and subcortical
448 structures influences muscle fatigue during lower body exercise.

449

Conflicts of Interest

The authors declare that they have no competing interests.

Author Contributions

All data was collected by RM. RM, SC, JF, CT, & RH contributed to study design and revising the manuscript. All authors approved the final manuscript. All authors listed qualify for authorship and agree with the order of authorship.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Acknowledgements

The authors would like to thank all the participants for their contributions and commitment to the study.

References

Amann M, Goodall S, Twomey R, Subudhi AW, Lovering AT, Roach RC (2013) AltitudeOmics: on the consequences of high-altitude acclimatization for the development of fatigue during locomotor exercise in humans. J Appl Physiol (1985) 115:634-642 doi:10.1152/japplphysiol.00606.2013

Atkinson G, Nevill AM (1998) Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. Sports Med 26:217-238

Atkinson G, Nevill AM (2001) Selected issues in the design and analysis of sport performance research. J Sports Sci 19:811-827

Behm DG, St-Pierre DM, Perez D (1996) Muscle inactivation: assessment of interpolated twitch technique. J Appl Physiol 81:2267-2273

Bland JM, Altman DG (1986) Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1:307-310

Buckthorpe MW, Hannah R, Pain TG, Folland JP (2012) Reliability of neuromuscular measurements during explosive isometric contractions, with special reference to electromyography normalization techniques. Muscle Nerve 46:566-576 doi:10.1002/mus.23322

Callard D, Davenne D, Gauthier A, Lagarde D, Van Hoecke J (2000) Circadian rhythms in human muscular efficiency: continuous physical exercise versus continuous rest. A crossover study. Chronobiology international 17:693-704

Dekerle J, Ansdell P, Schäfer L, Greenhouse-Tucknott A, & Wrightson J (2019) Methodological issues with the assessment of voluntary activation using transcranial magnetic stimulation in the knee extensors. European journal of applied physiology, 119(4): 991-1005.

Gandevia SC (1992) Some central and peripheral factors affecting human motoneuronal output in neuromuscular fatigue. Sports Med 13:93-98

Gandevia SC (2001) Spinal and supraspinal factors in human muscle fatigue. Physiol Rev 81:1725-1789

Gandevia SC, Allen GM, Butler JE, & Taylor JL (1996) Supraspinal factors in human muscle fatigue: evidence for suboptimal output from the motor cortex. The Journal of physiology 490(2): 529-536.

Goodall S, Charlton K, Hignett C, Prichard J, Barwood M, Howatson G, Thomas K (2015a) Augmented supraspinal fatigue following constant-load cycling in the heat. Scand J Med Sci Sports 25 Suppl 1:164-172 doi:10.1111/sms.12370

Goodall S, Charlton K, Howatson G, Thomas K (2015b) Neuromuscular fatigability during repeated-sprint exercise in male athletes. Med Sci Sports Exerc 47:528-536 doi:10.1249/MSS.0000000000000443

Goodall S, Howatson G, Romer L, Ross E (2012) Transcranial magnetic stimulation in sport science: A commentary. European Journal of Sport Science 14:S332-S340 doi:10.1080/17461391.2012.704079

504 Goodall S, Romer LM, Ross EZ (2009) Voluntary activation of human knee extensors
505 measured using transcranial magnetic stimulation. Exp Physiol 94:995-1004
506 doi:10.1113/expphysiol.2009.047902

507 Goodall S et al. (2017) The assessment of neuromuscular fatigue during 120 min of simulated
508 soccer exercise. Eur J Appl Physiol 117:687-697 doi:10.1007/s00421-017-3561-9

509 Goodall S et al. (2014) AltitudeOmics: exercise-induced supraspinal fatigue is attenuated in
510 healthy humans after acclimatization to high altitude. Acta Physiol (Oxf) 210:875-888
511 doi:10.1111/apha.12241

512 Herbert RD & Gandevia SC (1996) Muscle activation in unilateral and bilateral efforts
513 assessed by motor nerve and cortical stimulation. Journal of Applied Physiology,
514 80(4): 1351-1356.

515 Heroux ME, Taylor JL, Gandevia SC (2015) The Use and Abuse of Transcranial Magnetic
516 Stimulation to Modulate Corticospinal Excitability in Humans. PLoS One
517 10:e0144151 doi:10.1371/journal.pone.0144151

518 Johnson MA, Sharpe GR, Williams NC, Hannah R (2015) Locomotor muscle fatigue is not
519 critically regulated after prior upper body exercise. J Appl Physiol (1985) 119:840-
520 850 doi:10.1152/japplphysiol.00072.2015

521 Klass M, Duchateau J, Rabec S, Meeusen R, Roelands B (2016) Noradrenaline Reuptake
522 Inhibition Impairs Cortical Output and Limits Endurance Time. Med Sci Sports Exerc
523 48:1014-1023 doi:10.1249/MSS.0000000000000879

524 Kobayashi M, Pascual-Leone A (2003) Transcranial magnetic stimulation in neurology.
525 Lancet Neurol 2:145-156

526 Lee M, Gandevia SC, Carroll TJ (2008) Cortical voluntary activation can be reliably
527 measured in human wrist extensors using transcranial magnetic stimulation. Clin
528 Neurophysiol 119:1130-1138 doi:10.1016/j.clinph.2007.12.018

529 Magistris MR, Rosler KM, Truffert A, Myers JP (1998) Transcranial stimulation excites
530 virtually all motor neurons supplying the target muscle. A demonstration and a
531 method improving the study of motor evoked potentials. Brain 121 (Pt 3):437-450

532 Mathur S, Eng JJ, MacIntyre DL (2005) Reliability of surface EMG during sustained
533 contractions of the quadriceps. Journal of electromyography and kinesiology : official
534 journal of the International Society of Electrophysiological Kinesiology 15:102-110
535 doi:10.1016/j.jelekin.2004.06.003

536 Morrison S, Sleivert GG, Cheung SS (2004) Passive hyperthermia reduces voluntary
537 activation and isometric force production. European Journal of Applied Physiology
538 91:729-736 doi:10.1007/s00421-004-1063-z

539 Ngomo S, Leonard G, Moffet H, Mercier C (2012) Comparison of transcranial magnetic
540 stimulation measures obtained at rest and under active conditions and their reliability.
541 J Neurosci Methods 205:65-71 doi:10.1016/j.jneumeth.2011.12.012

542 Racinais S, Blonc S, Jonville S, Hue O (2005) Time of day influences the environmental
543 effects on muscle force and contractility. Med Sci Sports Exerc 37:256-261

544 Ross EZ, Gregson W, Williams K, Robertson C, George K (2010) Muscle contractile
545 function and neural control after repetitive endurance cycling. Med Sci Sports Exerc
546 42:206-212 doi:10.1249/MSS.0b013e3181b07a18

547 Ross EZ, Middleton N, Shave R, George K, Nowicky A (2007) Corticomotor excitability
548 contributes to neuromuscular fatigue following marathon running in man. Exp Physiol
549 92:417-426 doi:10.1113/expphysiol.2006.035972

550 Rossini PM et al. (1994) Non-invasive electrical and magnetic stimulation of the brain, spinal
551 cord and roots: basic principles and procedures for routine clinical application. Report
552 of an IFCN committee. Electroencephalogr Clin Neurophysiol 91:79-92

553 Sidhu SK, Bentley DJ, Carroll TJ (2009) Cortical voluntary activation of the human knee
554 extensors can be reliably estimated using transcranial magnetic stimulation. Muscle
555 Nerve 39:186-196 doi:10.1002/mus.21064

556 Tamm AS, Lagerquist O, Ley AL, & Collins DF (2009) Chronotype influences diurnal
557 variations in the excitability of the human motor cortex and the ability to generate
558 torque during a maximum voluntary contraction. Journal of Biological Rhythms,
559 24(3), 211-224.

560 Temesi J, Gruet M, Rupp T, Verges S, Millet GY (2014) Resting and active motor thresholds
561 versus stimulus-response curves to determine transcranial magnetic stimulation
562 intensity in quadriceps femoris. J Neuroeng Rehabil 11:40 doi:10.1186/1743-0003-
563 11-40

564 Temesi J, Ly SN, Millet GY (2017) Reliability of single- and paired-pulse transcranial
565 magnetic stimulation for the assessment of knee extensor muscle function. J Neurol
566 Sci 375:442-449 doi:10.1016/j.jns.2017.02.037

567 Todd G, Taylor JL, Gandevia SC (2003) Measurement of voluntary activation of fresh and
568 fatigued human muscles using transcranial magnetic stimulation. J Physiol 551:661-
569 671 doi:10.1113/jphysiol.2003.044099

570 Todd G, Taylor JL, Gandevia SC (2004) Reproducible measurement of voluntary activation
571 of human elbow flexors with motor cortical stimulation. J Appl Physiol (1985)
572 97:236-242 doi:10.1152/japplphysiol.01336.2003

573

574

Abbreviations

AMT – Active motor threshold

BF – Biceps Femoris

CV – Coefficient of Variation

EMG - Electromyography

ERT – Estimated resting twitch

ES – Electrical stimulation

ICC – Intraclass correlation coefficient

ITT – Interpolated twitch technique

MEP – Motor evoked potential

MO – Maximum output

MVC – Maximal voluntary contraction

MVF – Maximal voluntary force

RF – Rectus femoris

RMT – Resting motor threshold

RT – Resting twitch

SIT – Superimposed twitch

TMS- Transcranial magnetic stimulation

VA-$_{TMS}$ – TMS assessed voluntary activation

VM – Vastus medialis

VL – Vastus lateralis

Table 1. Mean values for TMS-assessed voluntary activation (VA-$_{TMS}$) and peripheral voluntary activation (VA) and potentiated twitch and measures of between-day and within-day reliability. Mean values are displayed ± standard deviation. LoA= limits of agreement; ICC=intraclass correlation coefficient; CI=confidence intervals; CV=coefficient of variation.

Measure	**Mean 1**	**Mean 2**	**Equivalence (P value)**	**Mean systematic bias**	**95% LoA**	**CI**	**ICC**	**CV (%)**
Between-day reliability								
VA-$_{TMS}$ (%)	94.1 ± 4.4	93.7 ± 4.9	<0.01	−0.5	−5.1, 4.2	92.9, 95.0	0.93	2.5
Peripheral VA (%)	94.0 ± 3.8	94.0 ± 3.6	<0.01	0.0	−4.1, 4.1	93.1, 94.9	0.92	2.2
Potentiated twitch (N)	186 ± 31	190 ± 31	<0.01	4.1	−23.4, 31.6	182.6, 195.1	0.95	7.4
Within-day reliability								
VA-$_{TMS}$ (%)	93.7 ±4.9	93.7 ±4.8	<0.01	0.0	−5.3, 5.4	92.5, 94.9	0.92	2.9
Peripheral VA (%)	94.0 ± 3.6	93.0 ± 3.5	<0.01	−1.0	-5.9 , 3.8	92.4, 94.6	0.84	2.7
Potentiated twitch (N)	190 ± 31	178 ± 36	0.10	−12.5	−66.7, 41.7	172.4, 196.9	0.78	15.0

Table 2. Mean values and between-day measures of reliability between the familiarisation session and the first experimental session for VA-$_{TMS}$ and peripheral voluntary activation. Mean values are displayed ± standard deviation. LoA= limits of agreement; CI=confidence intervals; ICC=intraclass correlation coefficient; CV=coefficient of variation.

Measure	Familiarisation mean (%)	Session 1 mean (%)	Equivalence (P value)	Mean systematic bias (%)	95% LoA (%)	CI	ICC	CV (%)
VA-$_{TMS}$ (%)	91.3 ± 5.4	94.1 ± 4.4	<0.01	2.8	−4.2, 9.8	91.1, 94.3	0.78	3.9
Peripheral voluntary activation (%)	93.2 ± 4.4	94.0 ± 3.8	<0.01	0.8	−4.9, 6.6	92.3, 94.9	0.85	3.1

Table 3. Mean values for electrical stimulation threshold, resting motor threshold and active motor threshold and between-day and within-day measures of reliability. Mean values are displayed ± standard deviation. LoA= limits of agreement; CI=confidence intervals; ICC=intraclass correlation coefficient; CV=coefficient of variation.

Measure	Mean 1	Mean 2	Equivalence (P value)	Mean systematic bias	95% LoA	CI	ICC	CV (%)
Between-day reliability								
Electrical stimulation threshold (mA)	79 ± 19	81 ± 18	<0.01	2.7	−14.4, 19.7	76, 84	0.94	10.9
Resting motor threshold (mA)	85 ± 11	86 ± 12	<0.01	0.8	−5.6, 7.2	84, 87	0.98	3.8
Active motor threshold (mA)	67 ± 15	66 ± 15	<0.01	−0.2	−8.8, 8.3	65, 68	0.98	6.6
Within-day reliability								
Electrical stimulation threshold (mA)	81 ± 18	82 ± 17	<0.01	0.3	-8.8, 9.5	79, 83	0.98	5.7
Resting motor threshold (mA)	86 ± 12	86 ± 12	<0.01	0.4	−3.6, 4.3	85, 87	0.99	2.3
Active motor threshold (mA)	66 ± 15	66 ± 15	<0.01	0.1	−3.5, 3.6	66, 67	0.99	2.7

Table 4. Mean values and between-day and within-day measures of reliability for MVF. Mean values are displayed ± standard deviation. LoA= limits of agreement; CI=confidence intervals; ICC=intraclass correlation coefficient; CV=coefficient of variation.

Measure	Mean 1 (N)	Mean 2 (N)	Equivalence (P value)	Mean systematic bias (N)	95% LoA (N)	CI	ICC	CV (%)
Between-day reliability								
MVF	661 ± 133	688 ± 134	<0.01	27.8	−62.4, 118.0	654, 695	0.96	6.82
Within-day reliability								
MVF	688 ± 134	650 ± 142	<0.01	- 38.4	−108.7, 31.9	653, 685	0.97	5.36

Table 5. Percentage of significant results out of 10000 simulations for between day and within day $VA\text{-}_{TMS}$ for n=5-50 with a Cohen's d effect size set at either 0.5 or 0.8.

Cohen's d	n= 5 (%)	n= 8 (%)	n= 10 (%)	n= 12 (%)	n= 20 (%)	n= 50 (%)
Between day						
Paired Samples t test						
0.5	33.0	58.1	70.1	79.7	96.1	100.0
0.8	64.9	92.4	97.7	99.3	100.0	100.0
Independent t test						
0.5	24.0	39.3	47.7	55.4	78.5	99.4
0.8	51.0	76.2	86.0	91.6	99.3	100.0
ANOVA						
0.5	24.0	39.3	47.7	55.4	78.5	99.4
0.8	50.8	76.2	86.0	91.6	99.3	100.0
Within day						
Paired Samples t test						
0.5	29.1	49.9	64.1	73.0	94.1	100.0
0.8	60.2	89.7	95.5	98.5	100.0	100.0
Independent t test						
0.5	32.6	53.1	62.9	71.4	91.6	100.0
0.8	66.3	89.5	95.4	98.3	100.0	100.0
ANOVA						
0.5	21.6	35.0	42.4	50.0	72.8	98.7
0.8	45.6	70.6	80.6	87.7	98.7	100.0

Table 6. Percentage of significant results out of 10000 simulations for between day and within day peripheral VA for n=5-50 with a Cohen's d effect size set at either 0.5 or 0.8.

Cohen's d	n= 5 (%)	n= 8 (%)	n= 10 (%)	n= 12 (%)	n= 20 (%)	n= 50 (%)
Between day						
Paired Samples t test						
0.5	29.1	52.7	65.6	39.5	94.0	100.0
0.8	61.3	89.3	96.1	98.7	100.0	100.0
Independent t test						
0.5	24.5	39.3	46.9	55.3	100.0	99.4
0.8	51.2	75.8	85.4	91.7	99.3	100.0
ANOVA						
0.5	21.7	34.7	41.8	49.0	72.1	98.5
0.8	45.3	69.7	79.8	87.3	98.3	100.0
Within day						
Paired Samples t test						
0.5	16.2	25.3	33.3	39.5	62.0	95.9
0.8	33.2	55.6	68.1	77.6	95.0	100.0
Independent t test						
0.5	15.0	24.3	29.2	34.6	55.4	91.2
0.8	31.8	52.2	62.6	71.9	91.7	100.0
ANOVA						
0.5	11.3	17.0	20.0	23.1	16.1	74.2
0.8	21.9	35.9	43.8	51.9	76.0	99.0

Figure Legends

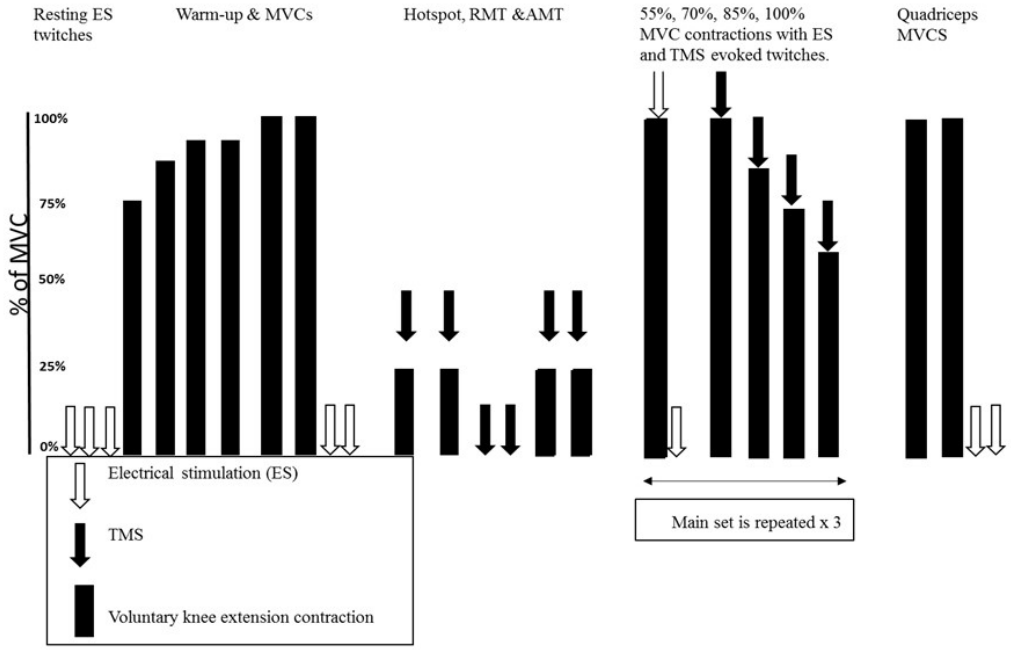


Figure 1. Schematic of the protocol

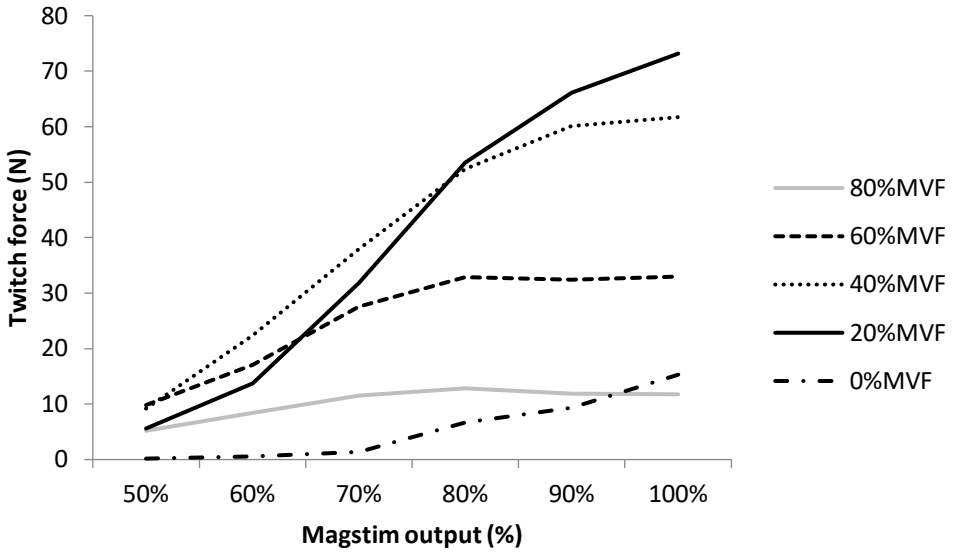
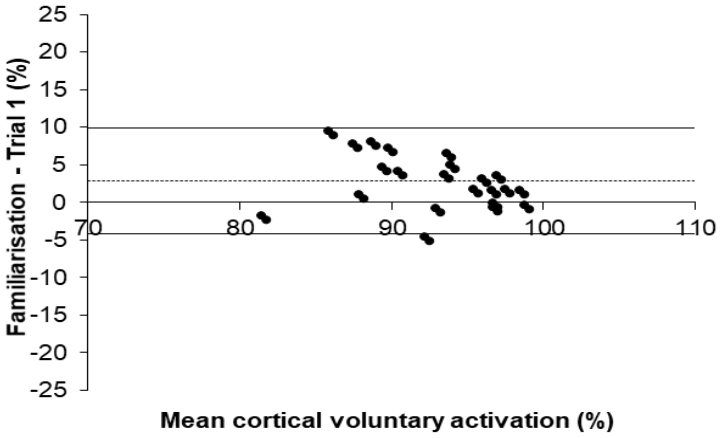
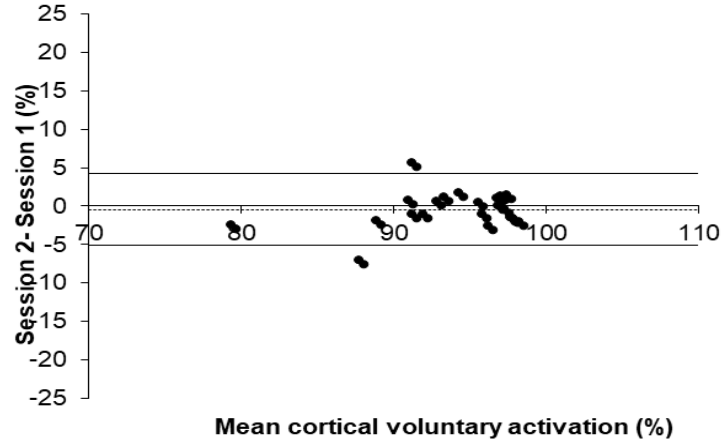


Figure 2. Stimulus response curves for stimulator output against twitch force at varying levels of maximal voluntary force.

a)


Familiarisation - Trial 1 (%)
25
20
15
10
5
0
-5
-10
-15
-20
-25
70
80
90
100
110
Mean cortical voluntary activation (%)

b)


Session 2- Session 1 (%)
25
20
15
10
5
0
-5
-10
-15
-20
-25
70
80
90
100
110
Mean cortical voluntary activation (%)

c)

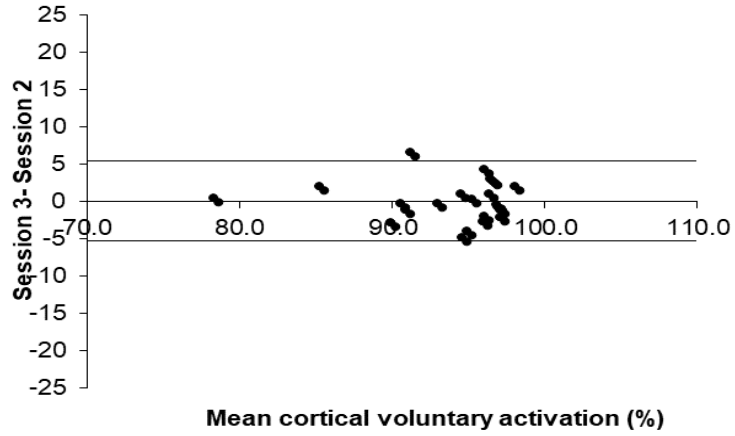

Session 3- Session 2
25
20
15
10
5
0
-5
-10
-15
-20
-25
70.0
80.0
90.0
100.0
110.0
Mean cortical voluntary activation (%)

Figure 3. Bland-Altman plots for VA-$_{TMS}$, demonstrating (A) between-day reliability between the familiarisation session and the first experimental session (B) Between-day reliability between experimental session 1 & 2 and (C) within-day reliability between experimental session 2 & 3. The dotted line represents systematic bias and the bold line represents the limits of agreement.