Examining the Intra-Rater Reliability of Transcranial Magnetic Stimulation (TMS)-Induced Motor Evoked Potentials (MEPs) Within and Between Sessions: A Step Towards Ensuring Accuracy of Observed MEP Changes in Repeated Measures Studies conducted by Newly Trained TMS Operators

Abubakar Tijjani Salihu 1, *, Keith D Hill 2, Maryam Zoghi 3 and Shapour Jaberzadeh 1

1 Monash Neurmodulation Research Unit, Department of Physiotherapy, School of Primary and Allied Health Care, Faculty of Medicine, Nursing and Health Science, Monash University, Melbourne, Australia
2 Rehabilitation, Ageing and Independent Living Research Centre, School of Primary and Allied Health Care, Monash University, Frankston, Australia
3 Institute of Health and Wellbeing, Federation University Australia, Ballarat, Australia

* Corresponding author: Monash Neurmodulation Research Unit, Department of Physiotherapy, School of Primary and Allied Health Care, Faculty of Medicine, Nursing and Health Science, Monash University, Melbourne, Australia. Email: abubakar.salihu@monash.edu

Received 2023 July 01; Revised 2023 August 29; Accepted 2023 September 04.

Abstract

**Background:** An essential factor in the validity of motor evoked potential (MEP)s recorded by transcranial magnetic stimulation (TMS) over multiple times is their test-retest reliability which to a large extent depends on the accuracy and competence of the assessor (intra-rater reliability). However, intra-rater reliability is infrequently reported in TMS studies suggesting that this is rarely done.

**Objectives:** This study was conducted to determine the intra-rater within and between-session reliability of a newly trained TMS assessor prior to a main TMS study and report on the methodology used to encourage similar practice.

**Methods:** Fourteen (10 males, 4 females; mean age: 32 ± 5.8 years) participants took part in the study. Motor evoked potentials were elicited from a relaxed, right first dorsal interosseous (FDI) muscle three times (T1, T2 and T3) across two testing sessions at least 48 hours apart. During the first session, MEPs were recorded twice (T1 and T2) within an interval of 20 minutes to determine the within (intra) session reliability of the assessor. During the second session, a single measurement was carried out (T3) which was compared to T1 to determine the inter-session reliability.

**Results:** Repeated measure analysis of variance (ANOVA) did not reveal significant difference in the amplitude of the MEPs obtained across the three time periods (P = 0.196) demonstrating agreement in the MEPs and hence the reliability of the assessor. Additionally, the intraclass correlation coefficient (ICC) between T1 and T2; and T1 and T3 were 0.952 (P < 0.001) and 0.833 (P = 0.001) respectively further indicating the within and between sessions reliability of the assessor.

**Conclusions:** The agreement between the three measured MEPs amplitude and the significant ICC demonstrates the reliability of the assessor in this study to use TMS for research. We suggest that the intra-rater reliability of new TMS operators should be established using the methodology in this report prior to main TMS studies.

**Keywords:** Motor Evoked Potentials, Transcranial Magnetic Stimulation, Intra-Rater Reliability, Corticospinal Excitability

1. Background

Transcranial magnetic stimulation (TMS) is a non-invasive technique used to deliver magnetic fields across the scalp and skull to activate the neurons within the brain (1). When applied at sufficient intensity over the primary motor cortex (M1), TMS produces a twitch in the muscle corresponding to the stimulated part of the M1 which can be recorded as a motor evoked potential (MEP) using surface electromyography over the target muscle (1, 2). Motor evoked potentials provide an important measure of corticospinal excitability (CSE) or the responsiveness of the corticospinal neurons and pathways (3). Components of the MEP such as the MEP amplitude have been used as...
outcomes for the assessment of CSE changes in different areas of research relating to the central nervous system (4). They are employed in clinical research to comprehend the alterations in the functional integrity of the corticospinal pathway caused by brain diseases and/or recovery, with the aim of developing potential diagnostic, prognostic, and treatment monitoring parameters for these conditions (5, 6). Additionally, research using interventions like non-invasive brain stimulation, exercise training, and other rehabilitation techniques to modulate brain function and improve human performance frequently rely on measures of corticospinal excitability as biomarkers (7-11). These biomarkers are instrumental in comprehending the neuroplastic changes in the brain that underlie the functional improvements linked to these interventions (2, 10-13). Similarly, CSE serves as a valuable tool for understanding the physiology behind several sensory motor processes and how they are influenced by factors such as age, gender, sleep, arousal, motivation, and fatigue (2, 14-18).

A common feature of the studies using TMS to assess CSE for different research purposes is the need to carry out repeated measurements of CSE parameters within and across different sessions of the studies (4). For instance, the use of TMS to understand the changes in the functional integrity of the corticospinal neurons after certain rehabilitation techniques aimed at improving functional abilities of patients with brain diseases would require multiple assessments of the CSE (10-12). These assessments encompass pre-to-post intervention measurements of the CSE, as well measurements at different time periods after the experimental and control interventions over several weeks or months to enable appropriate comparisons that would reveal the positive effects or otherwise of the interventions (9-11). This need for multiple assessments brings the question of the test-retest reliability of the CSE measures especially in relation to the intra-rater reliability or the reliability of the person using TMS to elicit MEPs (4). In many ways, this concern becomes particularly pertinent when the individual is new to the TMS device and technique. Several factors, including the TMS device, TMS device operator, participants, and recording or analysis set-up, can influence the elicited MEPs in TMS research (4, 19, 20). It is therefore crucial to tightly control these factors to ensure that observed changes in MEPs over time accurately reflect the effects of the research intervention rather than errors resulting from these factors (4). The validity and reliability of the TMS device is usually already assessed and confirmed by the manufacturer before making it commercially available (4). Therefore, the test-retest reliability of TMS induced MEPs for assessment of CSE is largely dependent on the remaining three factors (4, 19, 20). In addition to ensuring that the TMS rater has avoided any errors that may affect the elicited MEPs during the measurement, he/she also has a duty to ensure adequate control of the participant’s related factors as well as the set-up related factors during both the recording and analysis phases (4). Thus, the TMS operator has a central role to play in controlling the remaining factors that may affect the accuracy of the elicited MEPs in studies involving repeated TMS. However, in the majority of TMS studies, the intra-rater reliability is rarely reported suggesting that this may not be done in most studies (4). Given that researchers, especially graduate students, who are using TMS to induce MEPs for various research objectives, might have limited familiarity with both the device and the technique, it becomes necessary to adequately train them and also assess the reliability of the MEPs they evoke during preliminary sessions (both within session and across different sessions), before embarking on their primary research studies. In view of this, the main aim in this study was to investigate the intra-rater within- and between-session reliability of TMS induced MEPs for assessment of CSE by a newly trained operator. Furthermore, we present the methodology used and discussed the limitations of not ensuring the reliability of newly trained TMS operators. This could significantly contribute to promoting analogous practices, thereby improving the precision of changes in corticospinal excitability when conducting repeated measurements in various research types that involve transcranial magnetic stimulation.

2. Methods

2.1. Training of the New Operator (Rater)

The new rater in this study is a PhD student who will use TMS to study the effects of a particular intervention on corticospinal excitability as part of his PhD research. Considering that the new rater has no prior experience in using TMS, there is a need to adequately train him and establish the reliability of MEPs he evoked both within and across sessions prior to the main study. As a first step, the rater was equipped with adequate theoretical knowledge of the TMS through several teaching and demonstration sessions over a two-week period (3 × 2 hours sessions per week) (19, 20). During these sessions, several topics including the basic principles of transcranial magnetic stimulation, the procedure for assessment of corticospinal and cortico-cortical excitability using single and paired pulse TMS, safety considerations when using TMS, as well as the various TMS parameters and CSE outcome measures were described to the rater.
2.2. Participants

Subsequently, the rater was taken through practical sessions where he conducted the various steps that he learnt during the theoretical sessions. These sessions were repeated six times over another two weeks period under close supervision of the laboratory head up to the point where the rater was deemed confident in TMS assessment of CSE (19, 20). Thereafter, the rater was allowed to independently perform other practical sessions on as many people as possible over another four weeks to further build his competence and confidence in the assessment method (19, 20). From this point, the reliability study was started to assess the intra and inter-session reliability of the rater.

2.2. Participants

Fourteen (10 males, 4 females; age: 32 ± 5.8 years) participants attended two TMS sessions at least 48 hours apart (19, 20). All participants were right-handed and had no known neurological or psychological symptoms. Participants were also not taking any medications (including caffeine) that could influence corticospinal excitability. The participants completed a TMS safety questionnaire before the first session to ensure that they were safe for the study procedures, including questions relating to previous history of seizure or epileptic fit, presence of metal in the head such as surgical clip and other implanted devices such as cardiac pacemaker. Written informed consent was provided by each participant prior to inclusion in the study and all study procedures were conducted in accordance with the Helsinki declaration. The study was approved by the Monash University Human Research Ethics Committee (project ID: 27394).

2.3. Procedure

In the first session, two TMS assessments were conducted with an interval of 20 minutes between the first (T1) and the second (T2) assessments. Throughout the 20-minute resting period, participants were instructed to refrain from any contraction of the target muscle (right first dorsal interosseous (FDI) muscle) as well as any hand or finger movements. Participants subsequently attended the second session for the third TMS assessment (T3), after a minimum gap of 48 hours (2 days) from the initial session (Figure 1).

The first and second assessments (T1 and T2) were used to investigate the intra-session reliability of the rater, while the first and third assessments (T1 and T3) were used to assess the inter-session reliability of the rater (19, 20). In this way, the intra-rater within and between sessions reliability was established. Participants were asked to ensure adequate sleep the night before, avoid drugs that may affect their CNS function as well as participation in rigorous physical activity before and in between the two sessions. Before each session, participants were asked questions relating to these factors to ensure adequate compliance (4). Additionally, all assessments were done at approximately the same time of the day to control circadian effects.

2.4. Electromyography

During each session, participants were positioned comfortably on a chair to ensure that they were relaxed, with their head well positioned and the right forearm supported in a pronated position on the arm rest of the chair (20, 21). To ensure good surface contact and minimize skin resistance, the participant’s skin over the FDI muscle and the dorsum of the hand was abraded and cleaned (22). First dorsal interosseous location was determined through palpation during a manually resisted abduction of the second digit (20). Two pre-gelled self-adhesive bipolar Ag/AgCl disposable surface electrodes were then placed over the FDI with an inter-electrode distance of 2 cm and a ground electrode was placed over the dorsum of the hand. These electrodes were used to record the electromyographic (EMG) activity from the FDI during induction of MEPs. First dorsal interosseous was chosen because the newly trained rater is going to elicit MEPs from the same muscle in his subsequent study using TMS. This muscle is commonly used to evaluate CSE due to the strong corticospinal projection to the spinal motor neuron pool innervating it and therefore ease of stimulation (23). Electromyographic signals were filtered (10 – 500 Hz), amplified (×1000) and sampled at 1000 Hz. All data were recorded on PC using LabChart™ software (ADInstruments, Australia) via a laboratory analogue-digital interface (PowerLab, ADInstruments, Australia) for later off-line analysis.

2.5. Induction of MEPs by TMS

During each session, MEPs were elicited from the right FDI at rest through single pulse magnetic stimulation of the right FDI M1 administered via 70 mm figure-of-eight magnetic coil (Magstim Company Limited, UK). The vertex was first located by measuring the mid-point between the nasion and the inion and the two preauricular areas. 5 cm was then measured medio-laterally from the vertex to locate part of the left (dominant) M1 corresponding to the right FDI. The coil was then placed over this area and oriented at 45° to the midline with the handle of the coil pointing backward to induce current flow in the posterior-anterior direction (20, 24). The coil was...
then carefully moved by about 1 cm in each one of these four directions (i.e., anteriorly, posteriorly, medially, and laterally) eliciting 3 MEPs in each site and taking note of the size of the MEP (226). The site that produced MEPs with largest peak-to-peak amplitude was used as the hotspot for the stimulation of the target muscle (FDI) (227). This area was marked with a non-permanent marker to ensure consistent placement of the coil throughout the entire session (20). Subsequently, the marked was cleaned with wipes after each session.

The resting motor threshold (RMT) was then determined through Adaptive threshold-hunting based on parameter estimation by sequential testing (PEST) method using a freely available computer program called motor threshold assessment tool (MTAT). The stimulator output (test intensity) was then set at 120% of the RMT and 25 MEPs with inter-pulse interval of 6 seconds were recorded during each of the three TMS assessment time points (T1, T2 and T3).

2.6. Statistical Analysis

Statistical analyses were conducted using the statistical package for the social sciences (SPSS) version 28 (IBM Corp., Armonk, NY, United States). Motor evoked potential values were calculated by measuring the peak-to-peak amplitude and the average of 25 MEPs for each data collection period (T1, T2 and T3) were calculated. The data were screened for normal distribution using the Shapiro-Wilk test and visual inspection of histograms. Because our data was found to be positively skewed, log transformation was performed to correct the skewness before conducting the analysis (20). To assess the agreement between the repeated measurements (T1, T2 and T3), one-way repeated measure analysis of variance (RM-ANOVA) was carried out. Before interpreting RM-ANOVA statistical outcomes, sphericity was verified by Mauchly’s test. Additionally, intra-class correlation coefficient (ICC) were calculated to investigate the correlation between the MEP amplitude values obtained from the three-testing periods (intra-session reliability: T1 and T2; inter-session reliability: T1 and T3). This was to determine the within- and between-session reliability of the TMS-induced MEPs by the operator. Intra-class correlation coefficient values range between 0 and 1, with larger value indicating stronger correlation.

3. Results

All participants completed both session one and two. The mean interval between the first and the second session was 6.71 ± 3.02 days. The mean RMT in the first and second sessions were 37.50 ± 7.94 and 36.79 ± 6.86 percent of the stimulator output respectively.

3.1. Reliability of TMS-evoked MEPs

3.1.1. Intra-Session Reliability

The ICC between the MEP sizes at T1 and T2 was 0.952 (P < 0.001) indicating excellent correlation between the values and hence the intra-session reliability of the rater (Table 1).

3.1.2. Inter-Session Reliability

The ICC between the MEP sizes at T1 and T3 was 0.833 (P = 0.001) similarly indicating strong and significant between sessions reliability of the rater (Table 2).
Table 1. Results of the Intraclass Correlation Coefficient Calculation Between the Motor Evoked Potential Amplitudes Measured at T1 and T2

<table>
<thead>
<tr>
<th>Intraclass Correlation</th>
<th>95% Confidence Interval</th>
<th>F Test with True Value 0</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lower Bound</td>
<td>Upper Bound</td>
</tr>
<tr>
<td>Values</td>
<td>0.952</td>
<td>0.852</td>
</tr>
</tbody>
</table>

Abbreviation: df, degree of freedom.

Table 2. Results of the Intraclass Correlation Coefficient Calculation Between the Motor Evoked Potential Amplitudes Measured at T1 and T3

<table>
<thead>
<tr>
<th>Intraclass Correlation</th>
<th>95% Confidence Interval</th>
<th>F Test with True Value 0</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lower Bound</td>
<td>Upper Bound</td>
</tr>
<tr>
<td>Values</td>
<td>0.833</td>
<td>0.478</td>
</tr>
</tbody>
</table>

Abbreviation: df, degree of freedom.

3.2. Level of Agreement

One-way RM-ANOVA did not reveal significant effect of time on the measured MEP amplitudes (F (2,26) = 1.736, P = 0.196, ηp² = 0.118). Therefore, there was agreement in the MEP amplitude obtained during the three testing periods, further indicating the reliability of the rater (Figure 2).

4. Discussion

The purpose of this study was to assess the intra-rater within and between session reliability of TMS-induced MEPs amplitude. Additionally, we aimed to outline the procedures followed during the assessment in order to promote similar practices among other new TMS operators.

The new rater in this study is a PhD student who will use TMS to study the effects of a particular intervention on corticospinal excitability as part of his PhD research. This study was undertaken with the objective of ensuring that any observed changes in the size of MEPs during subsequent TMS studies were genuinely attributable to the intervention rather than potential errors arising from inconsistent or suboptimal performance of the TMS technique by the rater. Results of the study revealed no significant difference between the MEP amplitudes measured at three time points indicating agreement between the three measurements. Similarly, a strong correlation was observed between the two measurements in the first session (T1 and T2) and between the first measurement in session 1 (T1) and the measurement during the second session (T3) indicating the within and between sessions reliability of the new user respectively. This is in line with the result of previous intra-rater reliability studies conducted using the same methodology described in this study (19, 20). Therefore, the methodology followed in the current and previously mentioned studies is an effective way of training and examining the intra-rater within and between session reliability of new TMS operators (19, 20). In view of this, the rater in this study can be said to be reliable in assessing CSE using TMS and any changes in CSE observed in subsequent studies following intervention and assessment by the same rater can be confidently attributed to the effect of the intervention (4).
Transcranial magnetic stimulation is increasingly being used to evaluate changes in corticospinal excitability in different areas of research relating to the central nervous system in both healthy and clinical populations (25-28). For instance, single and paired pulse TMS is commonly used before and after different types of interventions (e.g., other non-invasive brain stimulation techniques such as transcranial direct current stimulation and exercise training) to understand the brain mechanisms underlying the positive effects of these interventions on cognitive, motor, or affective symptoms in clinical populations in order to develop potential therapies or optimize existing treatment parameters (29-32). However, a significant aspect of the TMS-evoked MEP amplitude which is used as a measure of corticospinal excitability and brain function in these studies is its test-retest reliability (4, 20, 33). Several methodological factors can bring about significant variability in the MEP amplitudes measured at different time points using TMS thereby confounding the true effect of the intervention being investigated (4). Therefore, to ensure that any changes observed following a particular intervention truly reflect the effects of that intervention not other measurement errors, these factors need to be tightly controlled (4). Interestingly, the majority of these factors can be controlled by ensuring the reliability, competence and confidence of the TMS operator (4). Hence, it is critically important to adequately train and confirm the reliability of new TMS operators using the methodology described in this study prior to their use of TMS for assessment of CSE for research purposes. A properly trained and reliable TMS operator would be able to accurately locate the hot spot during each assessment session and ensure that the TMS coil is properly positioned over the hot spot and kept on the same spot during multiple elicitations of the MEPs (4). Indeed, any misjudgement in precise identification of the hot spot or deviations from this spot during the assessment of the CSE will greatly affect the amplitude of the elicited MEPs (34-37). This underscores the importance of ensuring the intra-rater within and between sessions reliability before repeated TMS assessment of CSE by a new user (4). Additionally, several participant’s related factors such as head movement during testing, taking certain drugs or drinks and engaging in rigorous physical exercise/training programs in between sessions may bring about significant variability in the measured MEPs (4, 38-40). While the participant may not be aware of the influence of these factors on the measured CSE, a reliable TMS operator would already be aware of these factors and control them by instructing the participant against head movement during testing and monitoring for it (4). Similarly, he/she will instruct the participant to avoid, as well as report about the use of drugs, drinks or physical training between TMS sessions that may potentially affect the elicited MEPs (4). Other important factors that may significantly affect the reliability of the measured MEPs include the number of recorded MEPs, the inter-pulse interval between the TMS pulses, Intensity of the TMS stimuli, any background contraction, and normalization of MEPs to the size of the M-wave (when the motor neuron excitability at the spinal cord level could be changed by the intervention) (4, 19, 20, 41-44). A TMS operator that passed through the rigour of a reliability study as described in this study would be familiar with these recording and analysis factors that may affect MEP amplitudes and control them before hand in the research he/she is undertaking (4). Therefore, the TMS operator is responsible directly or indirectly for controlling the various factors that may affect CSE measurement using TMS to ensure that any observed changes in research projects are truly due to the effect of the interventions used (4). Hence, with the widespread and increasing use of TMS to assess corticospinal excitability in different research settings, it is critical to establish the reliability of the TMS operator prior to main TMS studies to ensure that any changes in MEPs observed across multiple measurement are truly because of the intervention used in the studies not measurement errors arising from the assessor’s handling of the equipment or participants.

4.1. Conclusions
The agreement between the three measured MEPs and the significant ICC demonstrates the reliability of the assessor in this study to use TMS for research. We suggest that the intra-rater reliability of new TMS operators should be established using the methodology reported here prior to main TMS studies to ensure that any changes in MEPs observed across multiple measurement are truly because of the intervention used in the studies not measurement errors arising from the assessor’s handling of the equipment or participants.

Footnotes

Authors’ Contribution: Abubakar Tijjani Salihu: Conceptualization, methodology, data collection, formal analysis, data curation, writing-original draft preparation; Shapour Jaberzadeh: Conceptualization, methodology, writing-review and editing; Keith D Hill: Conceptualization, methodology, writing-review and editing; Maryam Zoghi: Conceptualization, methodology, writing-review and editing.

Conflict of Interests: The authors declare no conflict of interests.
Data Reproducibility: The dataset presented in the study is available on request from the corresponding author during submission or after publication.

Ethical Approval: Monash University Human Research Ethics Committee (project ID: 27394).

Funding/Support: This study was not funded by any organization or individual.

Informed Consent: Written informed consent was provided by each participant prior to inclusion in the study and all study procedures were conducted in accordance with the Helsinki Declaration.

References


