Brain Stimulation 10 (2017) 196–213

Contents lists available at ScienceDirect

Brain Stimulation

journal homepage: http://www.journals.elsevier.com/brain-stimulation

Reliability and minimal detectable change of transcranial magnetic stimulation outcomes in healthy adults: A systematic review

Louis-David Beaulieu [a, b, *], Véronique H. Flamand [b, c], Hugo Massé-Alarie [a, b], Cyril Schneider [a, b]

[a] Clinical Neuroscience and Neurostimulation Laboratory, CHU de Québec Research Center - Neuroscience Division, Quebec City, Qc, Canada
[b] Department of Rehabilitation, Faculty of Medicine, Université Laval, Quebec City, Qc, Canada
[c] Center for Interdisciplinary Research in Rehabilitation and Social Integration, Quebec City, Qc, Canada

ARTICLE INFO

Article history:
Received 20 April 2016
Received in revised form
12 December 2016
Accepted 16 December 2016
Available online 19 December 2016

Keywords:
Transcranial magnetic stimulation
Reliability
Minimal detectable change
Primary motor cortex
Healthy humans
Systematic review

ABSTRACT

Background: Transcranial magnetic stimulation (TMS) is used worldwide for noninvasively testing human motor systems but its psychometric properties remain unclear.
Objective/Hypothesis: This work systematically reviewed studies on the reliability of TMS outcome measures of primary motor cortex (M1) excitability in healthy humans, with an emphasis on retrieving minimal detectable changes (MDC).
Methods: The literature search was performed in three databases (Pubmed, CINAHL, Embase) up to June 2016 and additional studies were identified through hand-searching. French and English-written studies had to report the reliability of at least one TMS outcome of M1 in healthy humans. Two independent raters assessed the eligibility of potential studies, and eligible articles were reviewed using a structured data extraction form and two critical appraisal scales.
Results: A total of 34 articles met the selection criteria, which tested the intra- and inter-rater reliability (relative and absolute subtypes) of several TMS outcomes. However, our critical appraisal of studies raised concerns on the applicability and generalization of results because of methodological and statistical pitfalls. Importantly, MDC were generally large and likely affected by various factors, especially time elapsed between sessions and number of stimuli delivered.
Conclusions: This systematic review underlined that the evidence about the reliability of TMS outcomes is scarce and affected by several methodological and statistical problems. Data and knowledge of the review provided however relevant insights on the ability of TMS outcomes to track plastic changes within an individual or within a group, and recommendations were made to level up the quality of future work in the field.

1. Introduction

The technology of transcranial magnetic stimulation (TMS) was first introduced in 1985 [1] and since, it has been extensively used as a non-invasive, painless and safe brain stimulation technique [2]. When applied over the primary motor cortex (M1), TMS allows the investigation of motor cortical physiology and integrity of the motor systems, including basic mechanisms involved in motor control and neural plasticity in both normal and pathological conditions [3–8]. Hence, TMS is used for diagnosis-prognosis [8–10] or for evaluating change [11–14]. However, despite a widespread use, evidence regarding the psychometric properties of TMS-related measures (i.e. validity, reliability, responsiveness [15]) is still lacking [3]. One particular issue is the reliability of the different TMS outcome measures, namely the degree to which measures are free from error and consistent in stable individuals [15,16]. Indeed, a large variability of TMS outcomes has been frequently reported [11,17–29], resulting from several factors. It is acknowledged that the intrinsic fluctuations of neural excitability at the cortical and spinal levels (see Ref. [29]), in addition to a considerable amount of methodological and subject-related factors (e.g. age and medication) [8,11,17,29] can influence the trial-to-trial

* Corresponding author. Clinical Neuroscience and Neurostimulation Laboratory, CHU de Québec Research Center, Neuroscience Division, 2705 boul. Laurier, RC-9800, Quebec City, QC, G1V 4G2, Canada.
E-mail address: louis-david.beaulieu.1@ulaval.ca (L.-D. Beaulieu).

http://dx.doi.org/10.1016/j.brs.2016.12.008
1935-861X/

variability, thus challenging the reliability of TMS outcomes. Testing reliability is a crucial step in the thorough process of validating an instrument for diagnosis-prognosis and evaluative purposes because unreliable measures prone to large systematic and random errors would logically never give valid measurements [30].

Many research groups previously tested the reliability of different TMS measures of corticomotor excitability and function in both healthy subjects [19,31–33] and in pathological conditions [34–38]. A particularly relevant paper was recently published by Schambra and coworkers [3]. Beyond their original results on TMS reliability in healthy and stroke populations, the authors provided up-to-date knowledge on how reliability of TMS outcomes should be rigorously tested and they pointed out many concerns related to previous studies on TMS reliability [3]. They especially underscored the widespread misunderstanding of reliability subtypes and appropriate statistical assessment. In fact, reliability can be classified in two main subtypes, referred to as reliability$_{MP}$ (or relative reliability; 'MP' standing for measurement property) and measurement error (or absolute reliability) [16]. These subtypes refer to strictly different concepts, hence, different potential applications of the measurement tool. Reliability$_{MP}$ refers to the degree to which stable individuals in a sample maintain their position relative to each other with repeated measurements [16]. In other terms, reliability$_{MP}$ informs on how well a TMS measure can distinguish individuals from one another, which is useful for diagnostic/prognostic and staging purposes [3]. This is peculiarly relevant in pathological conditions, such as stroke where TMS outcome measures are affected by the cerebral lesion [9] and can be used as predictive measures of the level of sensorimotor recovery [39,40]. On the other hand, the measurement error determines the degree to which repeated measurements vary within a sample of stable individuals [30], i.e. “the systematic and random error of a patient's score that is not attributed to true changes in the construct to be measured” [16]. The smaller the measurement error, the more accurate the measure, and the more likely it will be sensitive to change, which serves precisely the evaluative purpose of TMS outcome measures [3]. The intraclass correlation coefficient (ICC) and the standard error of the measurement (SEM$_{eas}$) are respectively recommended for an appropriate estimation of reliability$_{MP}$ and measurement error [30,41,42]. Furthermore, the SEM$_{eas}$ enables the calculation of the minimal detectable change (MDC, a.k.a. smallest detectable change or smallest real difference). MDC represents the minimum difference required to deduce with a relative degree of certainty (most often 95%) that a significant change has occurred, in an individual [3,43]. In other words, a change below the MDC threshold is more likely to result from random variations within the measurement error and may not be considered as a real change. It is therefore crucial to know the MDC of a TMS outcome utilised to monitor plastic changes within the motor system. However, Schambra et al.'s paper underscored that previous studies on the topic did not always use these appropriate statistical methods and that results were frequently misinterpreted [3]. Schambra et al.'s recommendations were crucial to level up the rigor and quality of future studies on TMS reliability but they did not use a standardized and systematic appraisal to score the studies they discussed.

Therefore, the present work aimed at providing an up-to-date analysis of both the quality and content of studies having evaluated the reliability of TMS outcomes in healthy individuals, with a particular focus on retrieving minimal detectable changes. This will be useful in research where the majority of studies use TMS outcome measures for evaluative purposes rather than diagnosis [3]. To this end, two standardized critical appraisal tools were used [11,44], one dedicated to the field of psychometric properties [44] and the other specific to TMS methodology [11]. The quality of the studies included were further characterized by up-to-date concepts related to statistical methods of reliability testing [15,16,30,41–43,45] and MDC results were specifically retrieved and grouped together for studies having similar methodological approaches. In order to get the whole picture of the reliability of TMS outcome measures, we were interested in retrieving data for (i) single pulse TMS (e.g. motor threshold; motor evoked potential amplitude, area & latency; silent period); (ii) recruitment curves (e.g. plateau; mean & peak slopes); (iii) paired-pulse TMS (e.g. intracortical and interhemispheric inhibition or facilitation); and (iv) mapping (e.g. map area & volume; hotspot location & center of gravity).

2. Materials and methods

A systematic review (not related to a registered protocol) was conducted in alignment with Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [46,47].

2.1. Literature search

The literature search was performed in three databases (Pubmed, CINAHL, Embase) using the following keywords and headings (Pubmed MeSH terms, CINAHL Major Headings (MH), or Embase Emtree): ((Transcranial Magnetic Stimulation) OR Transcranial Magnetic Stimulation[MeSH] OR 'transcranial magnetic stimulation'/exp) AND (Reliability OR (Test retest reliability) OR (Intrarater reliability) OR (Interrater reliability) OR Reproducibility OR Variability OR Repeatability OR Reproducibility of Results [MeSH] OR MH “Reliability and Validity+” OR 'statistical parameters'/exp) AND (Healthy OR Healthy volunteers[MeSH]). Additional relevant studies were also hand-searched in the references list of the papers selected for the review. Articles published up to June 2016 were included.

2.2. Selection of the studies

The lead author (LDB) reviewed the title and abstract of each article retrieved by the literature search to determine their eligibility. Articles that met the following inclusion criteria were retained for full-length examination: (i) evaluation of at least one TMS outcome measure of M1 representation of an upper or lower extremity muscle; (ii) tested at least one type of reliability (i.e. intrarater (or test-retest) and interrater) with statistical tests of reliability$_{MP}$ (i.e. correlation coefficients such as the ICC and Pearson's R) or measurement error (for example, SEM$_{eas}$ or coefficient of variation - CV) and reported their results in precise numbers (i.e. not only showed in figures); (iii) full-text paper written in French or English; (iv) inclusion of healthy individuals (but not limited to). The potential articles that met the inclusion criteria, based on the title and abstract, were then further reviewed (full-text) independently by two authors (LDB & HMA) against the inclusion criteria and the following exclusion criteria: (i) systematic reviews or literature reviews; (ii) articles that only tested the reliability of *post-hoc* analysis methods of TMS outcomes (potential overestimation of TMS reliability). The eligibility of each article was discussed under a structured approach to reach consensus (authors LDB, HMA: read the facts in the article, discussed about standards). If no consensus was reached, the opinion of a third author (VHF) was requested using the same structured approach.

2.3. Critical appraisal of studies

Two standardized critical appraisal tools were used for assessing the quality of included studies: the validated scale of Law and