Smartphone-Based Tapping Frequency as a Surrogate for Perceived Fatigue. An in-the-Wild Feasibility Study in Multiple Sclerosis Patients

LILIANA BARRIOS, Department of Computer Science, ETH Zurich, Switzerland
PIETRO OLDRATI, Faculty of Medicine, University of Zurich, Switzerland
MARC HILTY, Department of Neurology, University Hospital Zurich, Switzerland
DAVID LINDLBAUER, Human-Computer Interaction Institute, Carnegie Mellon University, United States
CHRISTIAN HOLZ, Department of Computer Science, ETH Zurich, Switzerland
ANDREAS LUTTEROTTI, Department of Neurology, University Hospital Zurich, Switzerland

Fatigue is a common symptom in various diseases, including multiple sclerosis (MS). The current standard method to assess fatigue is through questionnaires, which has several shortcomings; questionnaires are subjective, prone to recall bias, and potentially confounded by other symptoms like stress and depression. Thus, there is an unmet medical need to develop objective and reliable methods to evaluate fatigue. Our study seeks to develop an objective and ubiquitous monitoring tool for assessing fatigue. Leveraging a smartphone-based rapid tapping task, we conducted a two-week in-the-wild study with 35 MS patients. We explore the association between tapping derived metrics and perceived fatigue assessed with two standard clinical scales: fatigue severity scale (FSS) and fatigue scale for motor and cognitive function (FSMC). Our novel smartphone-based fatigue metric, mean tapping frequency, objectively ranks perceived fatigue with a mean AUC_{ROC} = .76, CI = [.71, .81] according to the FSMC, and a mean AUC_{ROC} = .81, CI = [.76, .86] according to the FSS. These results demonstrate that our approach is feasible and valid in uncontrolled environments. In this work, we provide a promising tool for objective fatigue monitoring to be used in clinical trials and routine medical care.

CCS Concepts: • **Applied computing** → **Health informatics**; • **Human-centered computing** → **Empirical studies in ubiquitous and mobile computing**; • **Computing methodologies** → Cross-validation.

Additional Key Words and Phrases: fatigue, smartphones, objective-measurement, fatigability, mobile health, remote-monitoring, impairment, multiple sclerosis

ACM Reference Format:
Liliana Barrios, Pietro Oldrati, Marc Hilty, David Lindlbauer, Christian Holz, and Andreas Lutterotti. 2021. Smartphone-Based Tapping Frequency as a Surrogate for Perceived Fatigue. An in-the-Wild Feasibility Study in Multiple Sclerosis Patients. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 3, Article 89 (September 2021), 30 pages. https://doi.org/10.1145/3478098

Authors' addresses: Liliana Barrios, liliana.barrios@inf.ethz.ch, Department of Computer Science, ETH Zurich, Switzerland; Pietro Oldrati, pietro.oldrati@usz.ch, Faculty of Medicine, University of Zurich, Switzerland; Marc Hilty, marcalexandermaconnen.hilty@usz.ch, Department of Neurology, University Hospital Zurich, Switzerland; David Lindlbauer, davidlindlbauer@cmu.edu, Human-Computer Interaction Institute, Carnegie Mellon University, United States; Christian Holz, christian.holz@inf.ethz.ch, Department of Computer Science, ETH Zurich, Switzerland; Andreas Lutterotti, andreas.lutterotti@usz.ch, Department of Neurology, University Hospital Zurich, Switzerland.

2474-9567/2021/9-ART89 $15.00
https://doi.org/10.1145/3478098

1 INTRODUCTION

Fatigue is a common symptom in many diseases, caused by viral infections [17, 57, 79], autoimmunity [6, 33], cancer [52, 72], neurodegenerative [26] and cardiovascular disease [11, 21]. Fatigue has been defined as "a subjective lack of physical and/or mental energy that is perceived by individuals or caregivers to interfere with the usual and desired activities" [30]. Up to 30% of individuals with COVID-19, for example, suffer from this debilitating symptom even weeks following the acute disease [17, 57, 79]. In multiple sclerosis (MS), fatigue affects up to 95% of patients [4, 25, 40, 42], many of them rate it as their most troubling symptom [38] and as the principal cause of patients' reduced work-productivity [38]. Still, the pathogenesis of fatigue remains uncertain, and there is no approved therapy available yet [37, 41, 63].

A critical limitation for understanding fatigue is the lack of objective methods to rate the symptom. The current standard measures to assess fatigue are self-reported questionnaires [56], which have several shortcomings: (1) Questionnaires rely entirely on subjective rating; (2) they evaluate fatigue retrospectively based on the past week or even month and are thus prone to recall bias [12, 32]; (3) they fail to define the symptom, and different conditions such as tiredness, sleepiness, and lack of motivation are reported as fatigue by patients [32]; (4) many questionnaires do not differentiate between specific domains of fatigue, such as cognitive and physical fatigue [12]; (5) they are prone to confounding by other symptoms such as depression [65]. Hence, there is an unmet medical need to develop objective and reliable measures to evaluate and ideally quantify the severity of fatigue. New outcome measures could be used for assessing the efficacy of therapeutic interventions and thus facilitate the development of novel treatments.

With the aim to provide clarity and consistency in the definition of fatigue, a unified taxonomy has been proposed that discriminates between subjective or perceived fatigue and performance fatigability [37]. Both domains add to the overall symptom fatigue, as depicted in Figure 1 (dimensions). A promising way to study fatigue in patients is by assessing performance fatigability, which can be measured objectively. Kluger et al. defined fatigability as "the magnitude or rate of change in a performance criterion relative to a reference value over a given time of task performance" [37]. The relation between perceived fatigue and objective fatigability is still unclear, but a recent meta-analysis of several studies in MS patients supports a correlation between both entities [37, 48]. Establishing an association between fatigability and perceived fatigue is an important goal for clinical research but has proven difficult [37]. Such an association could redefine our understanding of fatigue and how the symptom is treated.

There are two aspects of performance fatigability the cognitive and physical/motor (cf. Figure 1, right). In this paper, we focus on the physical aspect of fatigability. Motor fatigability is typically measured through walking (e.g., 6-minute walking test [29]), handgrip strength [67], or using a knee dynamometer [75]. These approaches have important limitations: 1) requirement of expensive clinical equipment and trained professionals to conduct the tests, and 2) the restriction to a medical facility and, consequently, these assessments are conducted only at a few or single time-points. Recently, Barrios et al. [5] proposed a rapid tapping task on a smartphone as an inexpensive approach to assessing motor fatigability (cf. Figure 1 bottom right). Their controlled lab evaluation demonstrated a high correlation between smartphone tapping and fatigability measurements obtained with a handgrip dynamometer. However, studies on the association between fatigability measured by the tapping task and perceived fatigue are not available. Hence, it is still unknown if 1) a smartphone-based tapping task is a feasible proxy for fatigue, and if 2) the task is valid in uncontrolled environments. In this paper, we seek to investigate both of these questions (cf. Figure 1 bottom left).

This paper's focus is two-fold: (I) To develop a new objective and reliable measure of motor fatigability computed from raw tapping data and demonstrate its usability and validity when performed outside controlled settings and without medical supervision (in-the-wild). We approach this goal by using Barrios et al. [5]'s tapping task and conducting a two-week in-the-wild study with 35 MS patients. Participants performed a 30 s tapping

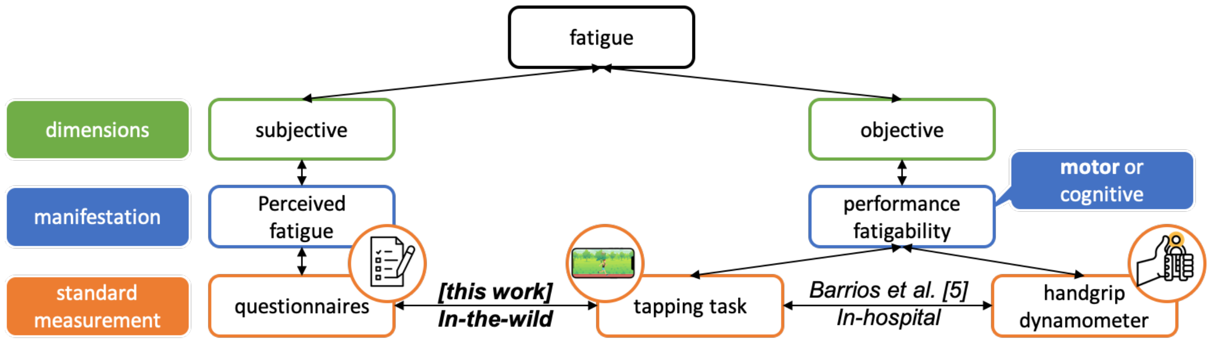

Fig. 1. State of the art fatigue vs. fatigability. Fatigue is currently **only** measured by questionnaires which have several shortcomings like subjective and prone to recall-bias. Motor fatigability can be measured objectively with rapid tapping or a handgrip dynamometer. The association between fatigue and fatigability is not established yet. Hence, we conduct an empirical in-the-wild study to evaluate the feasibility of using rapid tapping on a smartphone as a surrogate for fatigue.

task (i.e., a trial) once per day during the two weeks. Using this data, we introduce a new metric to assess motor fatigability: *tapping frequency*. We show that our new metric is a valid method to assess motor fatigability in-the-wild by comparing it to the previous approaches, specifically *touch duration* introduced by Barrios et al. [5] and strength decline using a handgrip dynamometer. (II) To evaluate the feasibility of establishing an objective and ubiquitous method as a surrogate to quantify perceived fatigue. To this end, we evaluate the performance of our proposed metric, *tapping frequency*, to classify fatigued and non-fatigued patients using *ROC* (receiver operating characteristic) curves and area under the *ROC* curve (AUC_{ROC}). We quantified perceived fatigue during the study with two widely accepted and validated fatigue questionnaires in MS patients: Fatigue Severity Scale (FSS) [44] and Fatigue Scale for Motor and Cognitive Functions (FSMC) [56].

1.1 Contributions

With the results of our in-the-wild study, we make the following core contributions:

- A new smartphone-based metric, *tapping frequency*, to quantify motor fatigability with the tapping task.
- Provide proof of concept of the validity of the tapping task and our metric in uncontrolled environments in-the-wild.
  - The data is consistent across days, with no significant change in tapping frequency.
  - There is a statistical difference between fatigued and non-fatigued groups during the whole study.
  - There is high correlation between our new metric and the handgrip dynamometer during the two-week study.
- The first in-the-wild study evaluating the feasibility of establishing a tapping task as a surrogate for perceived fatigue.
  - Mean tapping frequency ranks motor fatigue according to the FSMC with AUC_{ROC} $\overline{X}$ = .76 ± .05 when averaging thee valid trials.
  - Mean tapping frequency ranks fatigue according to the FSS with AUC_{ROC} $\overline{X}$ = .81 ± .05 when averaging three valid trials.
- Open source our fatigue dataset to the research community, containing data from 35 MS patients[1].

[1] https://www.research-collection.ethz.ch/

Our goal is to develop an objective metric to be used when monitoring patients with fatigue in medical routine or clinical trials, which until now has been hampered by the heterogeneity and subjective nature of questionnaires [27]. We believe our results are an essential step into understanding fatigue. Our smartphone-based tapping technique and evaluation metric have the opportunity to be used regularly by patients outside the clinic and more frequently than currently done in the medical routine. Being an objective method, it also opens the potential for quantifying the direct effects of therapeutic interventions, which is a clear advantage over currently used questionnaires [55].

2 RELATED WORK

2.1 Fatigue vs. Fatigability

The current gold standard to assess MS-related fatigue is through questionnaires [23, 44, 56, 80] such as the Fatigue Severity Scale [44], Fatigue Impact Scale [24], and Modified Fatigue Impact Scale [80] and Fatigue Scale for Motor and Cognitive Functions (FSMC) [56]. The shortcomings of assessing fatigue with questionnaires have been discussed in several studies [9, 37, 39, 65, 83]. Though the current definition of fatigue is to include both perceived fatigue and performance fatigability (cf. Figure 1), standard questionnaires only assess perceived fatigue. Researchers have suggested that perceived fatigue is associated with *fatigability* [20, 48, 74, 82]. However, demonstrating the association has been difficult [37].

Motor fatigability is assessed by quantifying the performance decline rate during a specific task with a fixed duration. Most studies measure motor fatigability through maximal voluntary contractions (MVCs) with a handgrip dynamometer [14, 19, 66, 68, 69, 74]. This method requires specialized equipment and personnel [20, 69]. To overcome this limitation, the latest research is focused on developing alternative fatigability methods. Tanigawa et al. [76] observed motor fatigue during fast tapping with the index finger on a custom button. Boukhvalova et al. [8] hypothesized that tapping with the index finger on a smartphone may measure motor fatigability. However, they did not test their hypothesis. Barrios et al. [5] proposed the use of *touch duration* during a rapid tapping task to quantify motor fatigability using commodity smartphones. The authors collected single tapping sessions per study participant to validate and compare their approach to the commonly accepted technique MVC for 30 s with a handgrip dynamometer within a controlled environment within the hospital (cf. Figure 1 bottom right side). However, they did not explore the relation to fatigue nor tested if their approach would work outside the hospital. Our work seeks to fill this gap. With an empirical study, we evaluate the feasibility of using rapid tapping on a smartphone as a surrogate for fatigue (cf. Figure 1 bottom left side). We base our work on a new metric, *tapping frequency* and compare it to Barrios et al. [5]'s touch duration using our in-the-wild dataset.

2.2 Fatigue Monitoring with mHealth in MS

There are a number mobile applications for managing MS-related fatigue [3, 18, 28, 35]. More Stamina [28] is a mobile application for the self-managing of MS-related fatigue. The app acts as a to-do list where users can input their daily tasks. The user's energy is represented through a visual metaphor (progress bar) and a symbolic unit (Stamina Credits) for quantifying the estimated effort per activity. The app's goal is to facilitate patients' energy management. Jongen et al. [35] introduced MSmonitor, a web-based program for self-management and care of MS patients. Their pilot study data suggests that using MSmonitor led to increased health-related quality of life and helped patients self-manage their fatigue. MS Energize [3] is an iPhone app focused on self-management of fatigue for MS patients. The app works as a coach supporting patients in their fatigue management. Similarly, D'hooghe [18] introduced MS TeleCoach, a mobile application offering telemonitoring of fatigue and telecoaching of physical activity and energy management in persons with MS. Results from their 12-week study indicate an improvement in the fatigue level of the participants measured through the FSMC. Existing mobile application for fatigue monitoring measure fatigue using questionnaires. Tong et al. [78], aimed at predicting MS patients'

FSS score using data from connected devices, background information and daily questions at weekly intervals. Objective metrics to assess fatigue and fatigability are not common in mHealth. Hence, in this paper, we seek to evaluate the feasibility of establishing an objective surrogate for fatigue measurements. Ideally, these new objective fatigue measurements can be incorporated into clinical trials and thus achieve more robust fatigue studies.

2.3 Handgrip Strength and Fatigue

Several ways to quantify fatigue during maximal voluntary contraction with a handgrip dynamometer have been proposed. According to Schwid et al. [66], the simplest method is to compare the maximal strength at the beginning and at the end of the contraction, as suggested by Miller et al. [51]. Bigland–Ritchie et al. [7] found that force declines in a linear manner during sustained muscle contraction at a rate characteristic for each subject. Hence, the slope of the decline indicates the rate of fatigue. Nacul et al. [54] studied handgrip strength as an objective measure of disease status and severity in people with chronic fatigue syndrome (CFS). Their results show that CFS patients had significantly lower mean handgrip strength than healthy controls, suggesting that the mean handgrip could be used as an objective tool for diagnosis and measuring disease severity. Similarly, they found that MS patients have a lower mean handgrip strength than healthy controls.

2.4 Finger Tapping as Disability Metric

Finger tapping quantifies neurological impairment in conditions like Parkinson' Disease (PD) [59, 60, 77]. Finger-to-thumb tapping is a standard task used in PD patients to assess dysfunction of the *extrapyramidal motor system*, which leads to impairment in maintaining alternating movements. However, the task is not unique to PD. Several studies have also shown that MS-related impairment can be measured with finger tapping [1, 13, 64, 71]. Chipchase et al. [13] conducted tapping with the index finger at maximal speed with the participant's hand resting on a surface. Using a counting device, the authors conducted 10 tapping sessions of 10 s each [50]. Their results indicate that the number of taps can differentiate between MS patients and controls, but they found a lack of correlation between finger tapping and fatigue severity. Alusi et al. [1] found a good correlation between the nine-hole peg and tapping a key on a large calculator with the index finger, and thus suggest tapping as a useful objective assessment of upper limb function in tremulous patients with multiple sclerosis. Scherer et al. [64] used alternating left and right index finger tapping to measure tapping speed on a standard PC keyboard (key F1 and key F12). Their task can detect minimal psycho-motor dysfunction in migraine as well as MS-related impairment.

3 METHODS

To analyze the feasibility of establishing smartphone-based objective metrics as a surrogate for fatigue, we conducted a two-week in-the-wild study. We use the FSMC [56] to discriminate between motor-fatigued and non-motor-fatigued participants and the FSS [44] to differentiate fatigued and non-fatigued participants. As a motor fatiguing task, we use the tapping task by Barrios et al. [5] with its original data collection procedure. Participants performed the tapping task with their dominant hand during each day of the two-week study. Through AUC_{ROC}, we evaluate the performance of our smartphone-based metrics to rank fatigued vs. non-fatigued participants in relation to the FSMC and FSS. Participants could exit the study at any point or continue for longer if desired. The local state ethics review board approved this study.

3.1 Participants

We recruited 35 MS patients at a specialized MS clinic (20 female, 15 male), aged 21–53 (M = 36.77, SD = 8.93). All MS patients had a confirmed diagnosis, signed written informed consent, and had Android smartphones. Seven of the 35 MS patients had hand impairments according to the Nine-Hole Peg Test (9-HPT) threshold (cf.

section 3.2.1). The Expanded Disability Status Scale (EDSS) scores ranged from 0 to 6 (M = 2.31, SD = 1.7) and were obtained from the MS clinic at the beginning of the study.

3.2 Tasks and Baselines

Our study started with an on-boarding, during which we explained the study protocol to the participants. We also collected normative outcome measurements using the FSMC, Nine-Hole Peg Test (9-HPT), and handgrip dynamometer. Additionally, we asked participants to install our Android application on their smartphones. Our application included Barrios et al.'s smartphone-based motor fatigability task [5] and sent daily notifications to the participants to remind them to complete the tapping trials during the in-the-wild study as well as to complete the FSS questionnaire once per week directly in the app.

Fig. 2. Study methods: smartphone-based fatigability task on the left, nine-hole peg test centered, and handgrip dynamometer on the right.

3.2.1 Clinical Baseline Methods. We use the 9-HPT to objectify hand function, the FSS and FSMC to categorize fatigue, and a handgrip dynamometer as standard motor fatigability measurement. Neurological impairment was measured using the standard disability rating scale for MS patients (EDSS) [45].

- *Nine-hole Peg Test (9-HPT).* The 9-HPT is a standardized, quantitative assessment used to measure finger dexterity [49]. Figure 2 (middle) shows an image of the 9-HPT used in this study. Participants were asked to remove the pegs, one-by-one, from the container to the holes and then to replace them into the container using their dominant hand. As the final score, we used the average of the two trials. Patients with a total time greater than 23.17 s (normative value used at the local hospital) were classified as hand-impaired.
- *Handgrip Dynamometer.* We used sustained handgrip strength as a metric to assess motor fatigability [69]. The test was conducted in an upright position, with both feet on the ground, forearms resting on an armrest. The dynamometer was held with the thumbs facing upwards in line with the forearm, and the grip size was adjusted for comfort. Participants performed maximum voluntary contraction (MVC) for 30 s. The experimenter instructed participants when to start and stop the MVC. Maximum contraction in kilograms was recorded every 3 s for a total of 30 s, resulting in 10 consecutive measurements. Figure 2 (right) depicts the Jamar device that we used.
- *Fatigue Scale for Motor and Cognitive Functions (FSMC) [56].* FSMC is used to assess MS-related cognitive and motor fatigue. The questionnaire consists of ten items that correspond to the cognitive sub-scale and ten items that correspond to fatigue's physical aspects. Participants rated each of the items on a 5-point Likert-type scale, consisting of (1) "Does not apply at all", (2) "Does not apply much", (3) "Slightly applies", (4) "Applies a lot", and (5) "Applies completely". FSMC offers cut-off values that determine fatigue levels

in different aspects (general, cognitive, and physical). With the cut-off values, it is possible to rate the level of fatigue as mild, moderate, or severe. Appendix D, Table 3 shows the different cut-off values for the distinct aspects of fatigue according to FSMC. Participants completed the FSMC before and after the two-week study. We used as the final score the mean of both completed questionnaires. In this study, we only focus on the physical aspect of fatigue as the tapping task is a measurement of motor fatigability. We label participants as non-fatigued if their FSMC physical score is less than 22, otherwise they are considered fatigued. Appendix D, Table 2 shows the items of the FSMC questionnaire.

- *Fatigue Severity Scale (FSS) [44].* FSS is a widely-used questionnaire to assess fatigue in various diseases [44, 81]. The questionnaire consists of 9 items with questions related to how fatigue interferes with the patient's activities. Patients rated the items on a 7-point Likert scale with values ranging from 1 = "strongly disagree" to 7 = "strongly agree". Higher scores indicate greater fatigue severity. The FSS final score is the mean of all items. We classified scores larger than 3.8 as fatigued participants. The FSS has no defined threshold to identify fatigued participants. Thresholds are usually defined depending on the study needs [2, 36, 81]. A score of 4 or higher is commonly used to identify severe fatigue [2, 36]. For the FSMC, we chose to use the lower threshold (mild fatigue). Hence for the FSS, we chose 3.8 as a threshold, representing a more conservative score than the commonly used for severe fatigue. Using this threshold, we identified a correlation of $\rho = 0.85$ ($p < 0.0001$) between the FSS and FSMC scores which is in line with the findings of Penner et al. [56]. Through our mobile application, we reminded participants to complete the FSS questionnaire once per week. We used the mean of the completed questionnaires as the final score. Refer to Appendix A, Table 1 for the complete FSS questionnaire.

3.2.2 Rapid Alternating Finger Tapping [5]. Participants performed rapid finger tapping on their smartphone's screen with their dominant hand. We asked them to keep their hand resting on a flat surface while doing the task with the smartphone set to landscape mode. Figure 2 on the left shows an image of the tapping tasks. The exertion movement required participants to engage the index and middle finger. Participants were asked to complete tapping trials at their maximal performance (maximal speed) for 30 s. This means that, they had to tap as fast as possible without stopping until the app indicated completion. The application was the same as used by Barrios et al. [5]. It did not offer immediate feedback to users if trials were conducted as expected.

3.3 Study Design

Our study included two phases: one in the hospital, and the other in-the-wild, highlighted in orange and blue , respectively, in Figure 3.

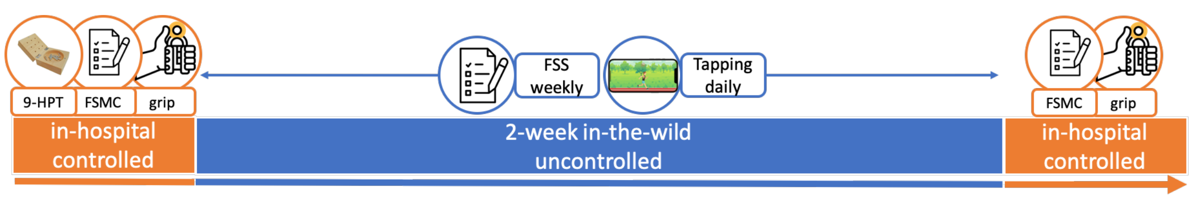

Fig. 3. Study design timeline with two phases: the hospital phase (orange) to gather baseline measurements, and in-the-wild phase (blue), the core of this study. During the in-the-wild phase, participants complete tapping trials daily and the FSS questionnaires once per week. Pre and post in-hospital baselines and questionnaires were average to get the final scores.

3.3.1 In-hospital. We included a pre-and post-study phase, both guided by a healthcare professional. During the pre-phase, participants were briefly introduced to the study and completed the FSMC and a demographic questionnaire. We guided participants through installing and using our Android application on their smartphones. They also received instructions on how to complete the tasks, including a demonstration by the experimenter and a short familiarization session for each of the tasks: handgrip, tapping task, and 9-HPT. Participants completed all tasks with their dominant hand in counterbalanced order. Between tasks, participants rested their arm and hand for three minutes. During the post-phase, as shown in Figure 3, we collected the FSMC scale and handgrip measurements again. Both measures, pre and post, were averaged to obtain the final scores. By combining two measures, we seek to reduce outliers and get more reliable baselines. Previous studies suggest that single MVC handgrip measurements can result in invalid trials [5].

3.3.2 In-the-wild. The in-the-wild experiment started immediately after the initial in-hospital phase. We asked patients to complete tapping trials once a day after receiving the reminder notification. We did not set a specific time to complete the trials. Instead, we allowed notifications to be set randomly during the day to achieve higher fluctuations in the person's energy level. All trials conducted in-the-wild were completed with no supervision. Once per week, participants received a notification for completing the FSS questionnaire. The study duration was two weeks.

3.4 Data Collection

During each 30 s trial of the handgrip dynamometer, we recorded ten samples, which is the maximum sampling rate of the Jamar device we used. For the 9-HPT, we recorded the time (seconds) that participants took to complete the task. Throughout the tapping task, we recorded all touch events on the participant's smartphone using the Android API. We stored all timestamped touch down coordinates and up events, from which we compute touch duration (i.e., how long did the finger touch the screen). Additionally, we computed the *tapping frequency* per second (i.e., number of taps recorded within a 1 s window), as a new feature to quantify fatigability. We define task performance for the tapping task in terms of the tapping frequency. This means tapping frequency will be high for fast tapping (high performance), whereas for slow taps (low performance), tapping frequency decreases. During our analysis, we also incorporate touch duration and its slope following previous fatigability approaches [5]. Touch duration has the opposite behavior of tapping frequency. During a high performance, touch duration decreases, and it increases as performance decays (i.e., taps become slower). FSS scores were stored on the participants' phone.

3.5 Hypotheses

We analyzed the data concerning the following hypotheses:

I We expect to find a comparable decrease in performance between the handgrip task and the in-the-wild tapping trials (tapping frequency), similarly to the findings of Barrios et al. [5] with touch duration.
II We expect to see a difference in tapping performance when comparing fatigued and non-fatigued participants suggesting an association between tapping performance and perceived fatigue.
III We expect the smartphone-based tapping task to be feasible and provide valid results when conducted in unsupervised settings in-the-wild.

To verify H1, we followed the procedure used by Barrios et al. [5]. We compared the tapping performance against the handgrip performance through correlation. However, as this is an in-the-wild study, we compare each tapping trial with mean handgrip of pre-and-post in-hospital phases. We verify H2 by evaluating features derived from the tapping task as a metric to classify fatigued and non-fatigued participants, according to the FSS and FSMC, using *ROC* (receiver operating characteristic) curves and area under the *ROC* curve (AUC_{ROC}). Finally, for H3, we use our validity algorithm to verify the trials in-the-wild.

3.6 Data Processing Pipeline

We use *tapping frequency* as the primary performance metric to assess motor fatigue with the smartphone tapping task. We define tapping frequency as the total number of taps registered during one second. Hence, we compute our feature with a one-second sliding window. We aim to monitor patients reliably in-the-wild. Thus, our data processing pipeline needs to handle noise and invalid tasks. A key difference to Barrios et al. [5] data handling is that their approach did not include gaps verification nor handling. As their study was fully controlled and under supervision, they probably were not affected by these cases as they could easily repeat trials when the experimenter deemed it necessary. However, our study is fully unsupervised. Hence, we see the relevance is verifying the data quality before conducting any analyses. Gaps can occur when a participant gets distracted by an incoming phone notification, call, or external factors. Our data processing pipeline includes three steps: (1) gap removal, (2) task validity, and (3) feature extraction.

3.6.1 Gap Removal. We identify gaps within a tapping trial when no input is recorded on the smartphone's screen for over 843.5 ms. This threshold represents the 0.999th quantile of the time differences between consecutive taps in our in-the-wild dataset. We did not incorporate automatic gap detection in the app, as it would imply that we had a priori knowledge of the tapping frequency of MS patients, which was not the case. Moreover, by setting a threshold without knowing how hand impairment could affect tapping, we could have erroneously stopped trials of participants with motor impairment. From our dataset, we have seen that gaps can occur at any time within a trial. If the gap occurs during the first half of a trial, we move the trial's start to after the gap. If the gap occurs after the second half of a trial, we move the trial's end time to before the gap occurs. We repeat this process recursively until all gaps within a trial have been removed. When removing gaps, we verify that the final trial's length is at least 27 s to ensure enough data for analysis. Shorter tapping trials are classified as invalid.

3.6.2 Tapping Trial Validity. We validate individual tapping trials by verifying that they are completed at maximal performance. First, we derive a continuous time series at a constant sample rate and apply a 2^{nd} order Butterworth low-pass filter with a cutoff frequency of 0.5 Hz. Then, we proceed to find the time of maximal performance (i.e., maximal tapping frequency). We use a low-pass filter to avoid detecting outliers within the tapping frequency. Since the initial 3 seconds of tapping contains inertial behavior [5], we do not consider them when extracting the peak's performance time.

Two conditions must hold to verify sustained maximal performance during a tapping task: (1) The peak of maximal performance should occur during the first half of the task (i.e., the maximal tapping frequency should occur before 15 s). Later peaks in performance indicate that the person failed to start the task at maximal speed. (2) After the peak of maximal performance, we expect a negative slope in the tapping frequency data. To verify this condition, we fit a line to the tapping frequency data, taking the time of maximal peak performance as the task's start time, following, we extract the line's slope. After the maximal peak in performance, a positive slope indicates that the person failed to perform maximal performance from the beginning of the trial. Figure 4 depicts examples of different cases of trial validity: invalid slope (left), invalid maximal performance location (center), valid trial (right).

3.6.3 Feature Extraction. We compute a set of features to evaluate the trial's performance. In particular, we compute the slope of the touch duration [5], slope of the tapping frequency, mean and maximum tapping frequency.

4 RESULTS

In summary, our results show an association between the smartphone-based tapping performance metrics and perceived fatigue measured with the FSMC and FSS. We found a statistically significant difference between fatigued versus non-fatigued participants' performance. The difference between both groups (fatigued and

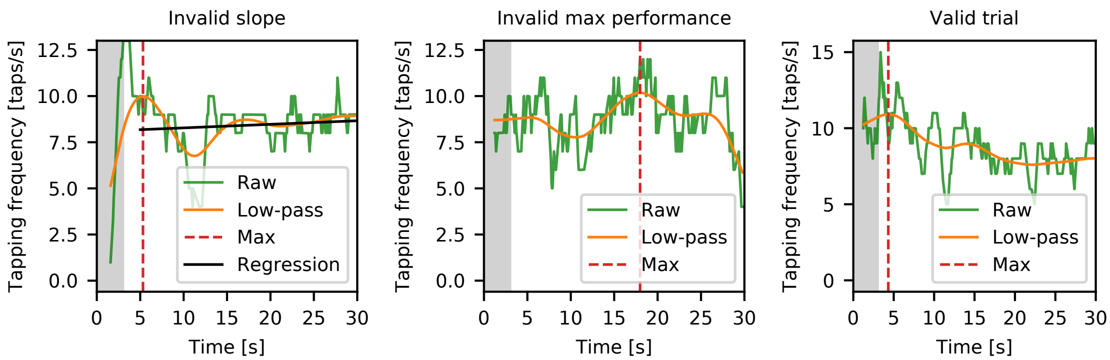

Fig. 4. A trial is invalid when the regression slope of the tapping frequency after the maximum is positive (*left*), or when the maximum of the filtered tapping frequency occurs after 15 s (*center*). Otherwise, the trial is considered valid (*right*), meaning the trial was completed at maximal performance. The first three seconds of the trial, depicted in grey, are discarded to avoid the influence of the initial inertia.

non-fatigued) is significant during the whole study, indicating that the approach is valid in-the-wild and without supervision. Additionally, our new data processing pipeline and core metric, tapping frequency, increase the validity of trials by 19% in comparison to Barrios et al. [5].

4.1 Tapping Frequency as a Valid Motor Fatigability Metric

We propose tapping frequency as a method to quantify motor fatigability using the tapping task on commodity smartphones. We validate our approach by comparing it to two accepted fatigability methods: handgrip dynamometer and Barrios et al.'s touch duration [5] on our in-the-wild dataset. We compute correlations between the mean handgrip and each single tapping trial at the participant level and report the combined correlations. We applied the data processing pipeline and feature generation as described in Section 3.6. To directly compare the handgrip dynamometer's ten measurements, we split the tapping data into ten segments. Next, we discard the first data segment to account for inertia [5]. We perform min-max normalization on the segmented data instead of standardization before computing the segments as done by Barrios et al. [5].

Using our trial validity definition, we classify 87% of the in-the-wild participants trials as valid (cf. Figure 6), which is a 19% increment compared to Barrios et al. [5]'s touch duration using our in-the-wild dataset. The correlation to the handgrip is comparable in both approaches. We used Spearman's correlation and obtained the following values, for Barrios et al. [5] $\bar{\rho} = 0.80$, CI: [0.39, 0.98] ($p < 0.05$ for all except 5 participants), and for our approach $\bar{\rho} = 0.83$, CI: [0.54, 0.99] ($p < 0.05$ for all except 2 participants).

4.2 Fatigue Scores' Distribution

The FSMC and FSS score distributions of our study population are depicted in Figure 5. As tapping is a motor task, we focus our analyses on the physical aspect of the FSMC questionnaire. Following the FSMC cut-off values, we classified 18 patients as fatigued and 17 as non-fatigued. Out of the 35 patients, two did not complete a single FSS questionnaire. The FSS questionnaire was intended to be completed during the in-the-wild phase of the study. From the 32 patients that did complete the FSS survey, we classified 17 patients as fatigued and 15 as non-fatigued. We observe a correlation of $\rho = 0.85$ ($p < 0.0001$) between the FSS and FSMC scores. With the conservative threshold of 4.0 for the FSS, 12 patients would classified as fatigued and 20 as non-fatigued.

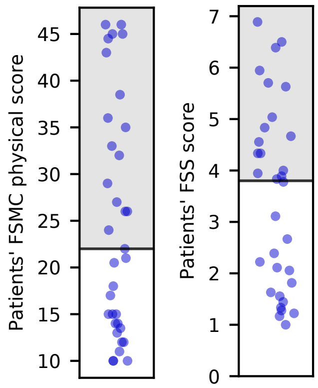

Fig. 5. Average physical FSMC and FSS scores of our study population. In grey, we depict the scores that we considered as *fatigued*.

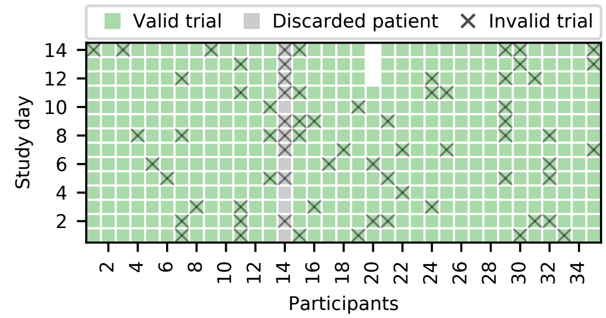

Fig. 6. Total tapping tasks completed per participant over the two week study. One participant was discarded for having more than half of the trials invalids. 87% of trials were labelled as valid.

4.3 Completed Trials and Validity

We collected a total of 487 tapping trials from 35 patients during our in-the-wild study. From those, 70 trials were classified by our validation algorithm as invalid. Figure 6 shows the valid and invalid trials per participant. One participant had more than half of the trials labeled as invalid (cf. Figure 6 "Discarded patient"). We decided to discard data from this participant as the medical examiner noted during the in-hospital session that the participant had very long artificial nails that prevented them from tapping correctly. This resulted in a dataset of 34 patients with 473 tapping trials, of which 61 are labeled as invalid. The rest of the patients completed the study and achieved at least eight valid tapping trials during the whole study. The average validity during the study was 87% (min = 57.0%, max = 100.0%)

4.4 Tapping Frequency Outperforms Handgrip Strength When Analyzing Fatigue

We computed a series of non-parametric Kruskal-Wallis H-tests [22] to identify statistically significant differences between fatigued and non-fatigued participants in terms of mean tapping frequency and handgrip strengths. The results are summarized in Figure 7. Following previous findings, we expect fatigued patients to show lower handgrip strength than non-fatigued patients [54]. Figure 7 (top left) shows the mean tapping frequency distribution of the study population according to the FSMC classification, averaged over all the valid trials of the patients. We observe a statistical significance difference in tapping frequency comparing the fatigued and non-fatigued group with $H = 7.50$ ($p < 0.01$). However, there is no statistically significant difference in terms of the mean handgrip strengths (cf. Figure 7, bottom left).

We found that while mean handgrip is confounded by gender, tapping frequency is not. Mean tapping frequency in female participants shows a significant difference between fatigued and non-fatigued with $H = 8.84$ ($p < 0.01$). However, this is not true within the male participants with $H = 2.72$ ($p = 0.09$). These results are shown in Figure 7 (top center). This may be explained by the smaller sample size compared to the female group. In fact, only six male participants are classified as non-fatigued according to the FSMC scale. In terms of mean handgrip (cf. Figure 7 bottom center), there is a statistically significant difference between genders when analyzing the non-fatigued $H = 11$ ($p < 0.001$) and fatigued $H = 7.25$ ($p < 0.01$) groups. However, handgrip does not show a statistically significant difference between fatigued and non-fatigued participants.

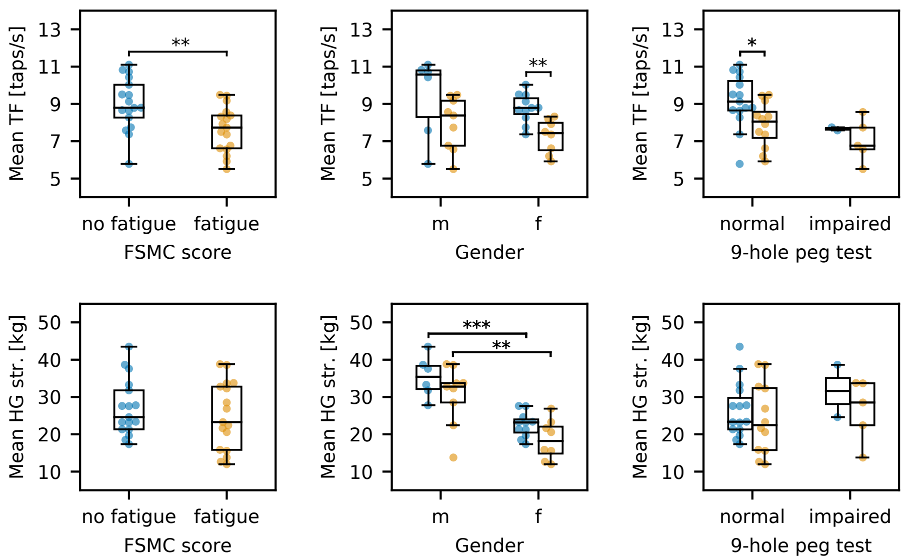

Fig. 7. Mean tapping frequency (*top*) and mean handgrip strength (*bottom*) in function of FSMC motor fatigue, gender, and impairment as defined by the 9-hole peg test. Fatigue is shown in orange and no fatigue in blue.

Analysis on the tapping performance and hand impairment shows statistically significant difference between non-impaired fatigued and non-fatigued participants with $H = 5.72$ ($p < 0.05$). However, there is no significant difference between the fatigued participants of the impaired and non-impaired groups. Only seven participants classified as hand-impaired according to the 9-HPT, and from those only two are non-fatigued. The small amount of non-fatigued and hand-impaired participants does not allow us to calculate whether there is a significant difference among the non-impaired population. In terms of handgrip strength, there is neither a difference between impaired and non-impaired participants, nor a difference within these groups in terms of fatigue. Figure 7 (right) shows the box plots corresponding to this analysis.

We performed a similar analysis to explore mean tapping frequency and handgrip of fatigue and non-fatigued participants in function of the FSS questionnaire, coming to the similar conclusions presented in this section. However, when exploring the influence of gender on the mean tapping frequency and the FSS, we found statistical significant difference for female and male participants. This is opposite to the FSMC where no statistical significance difference was found for male participants. Refer to Appendix C for the box plots related to the FSS scores. Full descriptive statistic on the performed test and additional non-parametric tests are shown on Appendix E.

4.5 Tapping Frequency as a Surrogate for Perceived Fatigue

To explore the association between our metric and perceived fatigue, we computed the predictive power of the *tapping frequency* to rank fatigued participants according to the FSMC and FSS scores. We use *ROC* curves and AUC_{ROC} [34] to evaluate the performance of our metric using it as threshold to classify between fatigued and not fatigued participants. AUC has the advantage that it provides the features' overall classification performance without defining a threshold. Thresholds can be adapted depending on a specific purpose. In some cases, the focus is on high recall, while in others, on accuracy.

4.5.1 Evaluation Setting. To evaluate the robustness of our approach and compute confidence intervals for AUC_{ROC}, we use stratified Monte-Carlo sampling [58] with 1000 iterations and randomly select (without replacement) in each iteration 2/3 of our participants' data (tapping trials) for evaluation. We partitioned the tapping data into six strata, following two partitioning criteria: (a) fatigued as a binary state according to FSMC or FSS, and (b) an age group, which can be one of three: [18,30), [30, 40), and [40,∞). Each participant and all of their data is fully assigned to one of the resulting six strata. Thus, when performing the stratified split, either a participant's data is fully considered or not at all. With this approach, we split at the participant level, ensure class balance, and account for age.

We report the average ($\overline{X}$) AUC_{ROC} with its respective confidence intervals. Additionally, we explore how the predictive power changes when combining more than one tapping trial. Thus, we combine consecutive, valid trials by averaging their features. For visual inspection, we include plots of the ROC curves corresponding to the 1000 splits and averaging three consecutive, valid trials. This section reports results for several features, specifically mean tapping frequency, maximum tapping frequency, and the slope of the tapping frequency. There is no established baseline for this classification task. Nevertheless, we consider the participant's age and the slope of the touch duration as baseline comparisons. Previous research shows that fatigue occurs more frequently in older patients, independently from disease severity [16] , and Barrios et al. [5] proposed touch duration declined rate (slope) as a fatigability metric.

4.5.2 FSMC - Motor Fatigue Ranking According to AUC_{ROC}. Our results show that maximum tapping frequency and mean tapping frequency outperform the other features. When considering a single tapping trial ($t = 1$), tapping frequency ranks fatigue and non-fatigued participants with AUC_{ROC} $\overline{X} = .74 \pm .05$. Furthermore, we observe that the AUC_{ROC} increases when averaging consecutive trials' features. Tapping frequency reaches a maximum when combining three consecutive, valid trials, representing an improvement of 2 percentage points ($p.p$). Figure 8 (right) shows the ROC curves corresponding to the mean and maximum tapping frequency, best-performing features, when averaging three successive valid trials. The slope of the tapping frequency ranks participants with AUC_{ROC} $\overline{X} = .65 \pm .05$ when $t = 3$. Followed by touch duration slope with a AUC_{ROC} $\overline{X} = .60 \pm .05$ when $t = 3$, and age with AUC_{ROC} $\overline{X} = .57 \pm .05$. We computed slopes as features of motor fatigability as suggested in previous research [5, 7]. Similarly, we consider the participant's age as a feature, as previous research has shown that fatigue occurs more frequently in older patients, independently from disease severity [16]. Our suggested metric, tapping frequency, outperforms the baseline touch duration slope [5] and age by 16 $p.p$ and 19 $p.p$ respectively. Our results show that tapping trial performance metrics outperform the motor fatigability metrics for assessing perceived fatigue.

4.5.3 FSS - Fatigue Ranking According to AUC_{ROC}. Fatigue ranking in terms of the FSS questionnaire shows the same behavior as described for the FSMC ranking. Mean tapping frequency and maximum tapping frequency exhibit the best ranking performance. When considering a single tapping trial ($t = 1$), mean tapping frequency ranks fatigue and non-fatigued participants with AUC_{ROC} $\overline{X} = .80 \pm .05$. Furthermore, we observe that the AUC_{ROC} increases when averaging consecutive trials' features. Tapping frequency reaches a maximum when combining three consecutive, valid trials with AUC_{ROC} $\overline{X} = .81 \pm .05$. The next best feature is maximum tapping frequency with AUC_{ROC} $\overline{X} = .77 \pm .05$ when combing three trials ($t = 3$). Following is tapping frequency slope with with AUC_{ROC} $\overline{X} = .61 \pm .05$ when $t = 3$, age with with AUC_{ROC} $\overline{X} = .56 \pm .05$ when $t = 3$, and finally touch duration slope with AUC_{ROC} $\overline{X} = .50 \pm .05$. The touch duration slope shows a random behavior for ranking fatigue according to the FSS. Mean tapping frequency, outperforms the fatigability baseline touch duration slope [5] and age by 31 $p.p$ and 26 $p.p$, respectively (cf. Figure 9). Similar to the FSMC ranking results (cf. Section 4.5.2), tapping trial performance metrics outperform the motor fatigability metrics for assessing perceived motor fatigue.

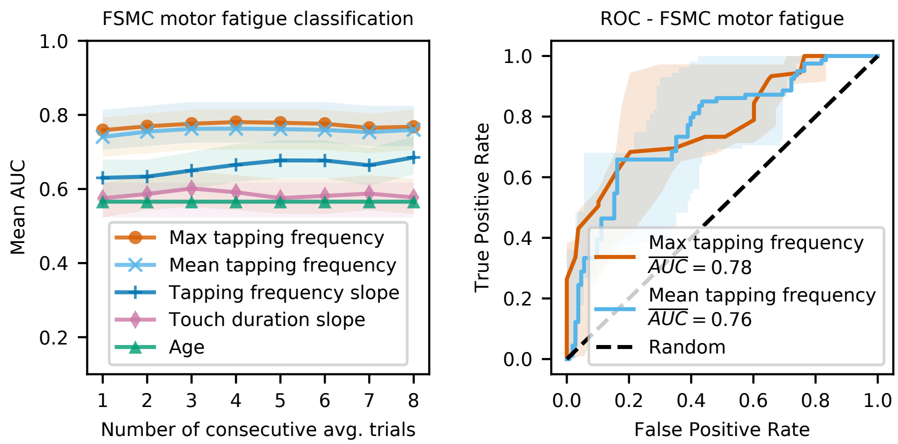

Fig. 8. Mean AUC_{ROC} when ranking motor fatigue according to FSMC of all participants (N=34) on the left. Mean tapping frequency shows the best performance in comparison to the other features. Also, reliability increases when averaging the features of consecutive valid trials (t). ROC curves for mean and maximum tapping frequency with $t = 3$ are displayed on the right. Data generated using Monte-Carlo simulation with 1000 iterations.

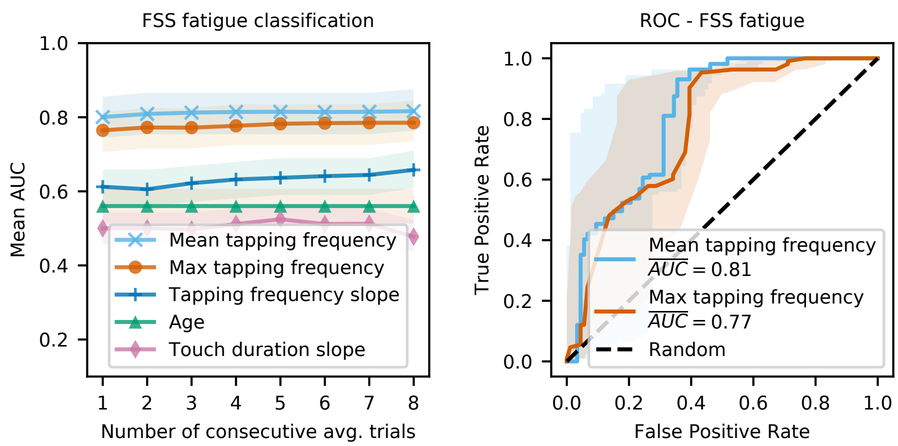

Fig. 9. Mean AUC_{ROC} for fatigue according to FSS of all participants (N=32) on the left. Mean tapping frequency shows the best performance in comparison to the other features. Also, reliability increases when averaging the features of consecutive valid trials (t). ROC curves for mean and maximum tapping frequency with $t = 3$ are displayed on the right. Data generated using Monte-Carlo simulation with 1000 iterations.

4.6 Participants' Adherence – Temporal Analysis

To verify that our approach is valid in-the-wild, we analyzed how the metric outcomes varied over the two-week study. We compute a sliding window and average the mean tapping frequency over three consecutive, valid trials. Afterward, we compute a series of Kruskal-Wallis H-tests to verify if the statistically significant difference between fatigued and non-fatigued patients according to the FSMC and FSS held along the two-weeks. Figure 10 and Figure 11 offer an overview of the results corresponding to the FSMC and FSS respectively. The results show that the statistically significant difference between the fatigued and non-fatigued participants in terms of the mean tapping frequency holds in-the-wild for both questionnaires. This confirms that our metric is valid in unsupervised settings and that the approach is suitable for monitoring fatigue remotely.

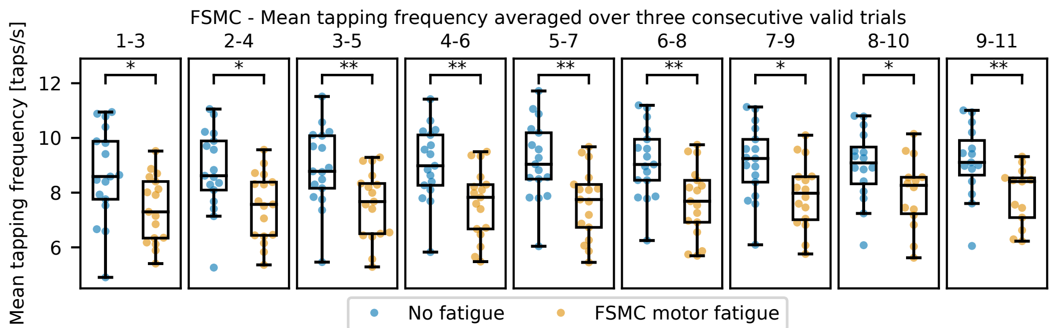

Fig. 10. Mean tapping frequency of three averaged valid asks during the course of the study grouped by motor fatigue as defined by the FSMC questionnaire.

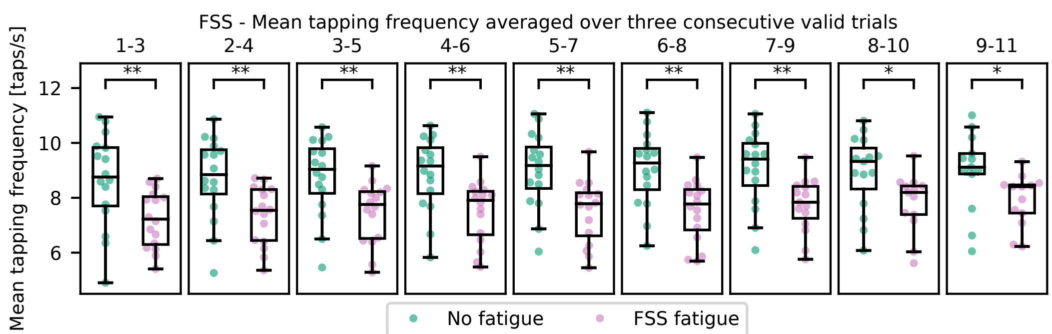

Fig. 11. Mean tapping frequency of three averaged valid asks during the course of the study grouped by fatigue as defined by the FSS questionnaire.

5 DISCUSSION

In this section, we discuss the implications of our findings, the limitations of our work, and possible directions for future research.

5.1 Implication of Subjective/Objective Measurements of Fatigue

There is a clear unmet medical need to develop an objective measure to assess both motor and cognitive fatigue in MS patients. Availability of such a tool would be an essential component to develop new therapies and improve routine medical care by helping to assess the effect of an intervention and to differentiate from various confounding symptoms, e.g., sleepiness, mood alterations, and others. Despite being a debilitating symptom affecting 90% of all MS patients throughout the disease, there is still is no approved therapy available. Different compounds have been tested in randomized placebo-controlled clinical trials (or are being used as off-label treatments). The results of these trials were inconsistent; some reported good efficacy of the therapies, whereas others did not show a benefit of the therapies [10, 15, 31, 43, 46, 47, 53, 55, 62, 73]. The outcome measures in all these trials were subjective questionnaires. It is well accepted that the magnitude of the placebo effect is an important reason for the variability in efficacy [55, 61, 70]. Hence, an objective measure would overcome this limitation for treatment development and provide a useful medical decision-making tool.

The smartphone-based tapping task is easy to administer, and because of its game-like character, we believe it could potentially have a higher acceptance than standard questionnaires. While the tapping task takes less than a minute, the FSMC questionnaire involves 20 items that have to be answered by deciding on five given choices of a Likert scale. However, user acceptance needs to be assessed in future studies.

Further, the tapping task provides a direct (to the point) measurement. It could be administered several times a day to quantify fluctuations in performance fatigability, typical of fatigue. Questionnaires evaluate the symptom only retrospectively, usually for two weeks, and are therefore less precise and not sensitive to fluctuations or short-term changes in the severity of the symptom (i.e., following physical/medical interventions). Hence, a more continuous (higher frequency) assessment is an advantage for following patients over time and assessing new interventions' effects. However, one has to consider that fatigability relates to a specific task, while fatigue questionnaires cover a general feeling, which affects the person as a whole. Thus, our approach complements existing fatigue quantification methods.

Our study provides a proof-of-concept for an association of motor fatigability, assessed by the tapping task, with subjective motor fatigue, assessed by the FSMC, which has been developed and validated in MS patients. Furthermore, the association between smartphone-based motor fatigability and perceived fatigue has also been confirmed with an independently validated fatigue questionnaire, FSS. Hence, the study provides early evidence for an association of the objective smartphone-based motor fatigability measurement and perceived fatigue in MS patients. Nevertheless, further and more extensive studies are needed to establish the predictive value of the tapping task to subjective fatigue.

5.2 Tapping Frequency as Reliable Smartphone-based Motor Fatigability Metric

We believe that our proposed method is less prone to outliers in comparison to Barrios et al. [5]. Touch duration could have an erroneous representation of the tapping task performance, given that the metric fails to account for the time when fingers perform their air motion. An example of this behavior is when the person is fast at lifting the fingers from the smartphone screen, but their finger's air motion is slow. Our metric, tapping frequency, does not suffer from this phenomenon, as it reflects the full dynamics of the tapping task. Additionally, with our gap removal, we seek to have a more flexible approach.

5.2.1 Gap Removal for In-the-wild Studies. The gap removal intends to gain as much value from the data as possible while avoiding discarding complete trials, a key feature for in-the-wild studies. We believe this is particularly useful for unsupervised settings where the person may get distracted while performing a trial. Distractions could be caused by phone notifications, calls or external factors. Additionally, we noticed the utility of our validation algorithm when it detected problems with one patient. Later we learned that the patient had very long artificial nails that caused unreliable tapping. In summary, our new method makes fewer assumptions, increased validity by 19%, and shows a comparable correlation to the clinical baseline (handgrip).

5.3 Tapping Frequency – Difference Between Fatigued and Non-fatigued Patients

From the tapping frequency, we learned that non-fatigued participants delivered a higher mean tapping frequency than fatigued participants and that this difference is statistically significant. Patients defined as non-fatigued according to the FSS and FSMC questionnaires achieved higher maximum tapping frequencies. In contrast, we notice no statistically significant difference between fatigued and non-fatigued patients when using the handgrip dynamometer. Moreover, tapping frequency is independent of gender, while handgrip dynamometer is not. Hence, our approach shows advantages and outperforms the commonly used handgrip dynamometer for monitoring motor fatigue.

5.4 Participants' Adherence to the Study Protocol

Through our experiment, we examined participants' adherence to the study protocol over two weeks. Adherence during the study was good. Analysis of the participants' two-week behavior shows no significant change in tapping frequency over time. All patients completed the two-week protocol, and the number of invalid trials did not show a particular pattern. Using our validity algorithm, we analyzed the completed tapping trials and found out that only a small percentage was invalid. Our analysis shows that combining several tapping trials is advisable to achieve higher confidence in the results. We show that the average of three tapping trials is sufficient to classify fatigue.

5.5 External Validity of the Results

There is no standard objective method to measure overall fatigue, particularly perceived fatigue, other than standard questionnaires. Hence, to develop a new approach, one has to rely on these validated questionnaires as a reference. Therefore, as part of this study, we aimed to assess the association of motor fatigability, assessed with the tapping task, with perceived fatigue rated by standard questionnaires. The following steps have been taken to ensure the validity of the results. First, we validate tapping frequency as an objective measure of motor fatigability against a standard reference method (handgrip dynamometer). Second, the validity of an unsupervised assessment of the smartphone-based task has been confirmed in an in-the-wild study in MS patients. Third, we use the in-the-wild data to assess whether the results of the tapping task can be used as a surrogate for subjective fatigue, being classified using two different questionnaires, both validated in MS patients. Overall, the results provide early evidence for using the smartphone-based tapping task as a surrogate for perceived fatigue. However, more extensive and independent studies are needed to confirm the results and establish an objective task of motor fatigability as a surrogate for subjective fatigue.

5.6 User-interface, Interaction and Design Improvements

Informal feedback from the participants suggests that performing daily tasks can produce a lack of motivation and boredom. This can be addressed in further studies by introducing a gamification mechanism to keep the participant engaged and motivated. To achieve better results and avoid demotivating the users, we recommend combining three tapping trials. However, most importantly, we do not advise conducting the tapping task daily for prolonged periods. An alternative approach would be to require tapping trials for three consecutive days every 1-2 weeks. Further studies are necessary to estimate a suitable periodicity for the tapping task.

5.6.1 Immediate Validity Feedback. We only applied our validity algorithm during a post-processing phase. In future task design improvements, we recommend incorporating immediate feedback to the user to further reduce the total percentage of invalid trials. Trials can be automatically stopped when gaps exceeded a defined threshold of 1 s. When this occurs, users can be notified of the specific problem (large gap) and can be asked to re-start the tapping trial from the beginning.

5.6.2 Maximum Tapping Frequency and Shorter Trials. Our results indicate that maximal tapping frequency could also be a suitable surrogate for fatigue. This has important implications as it would mean that our proposed validity algorithm would change, and potentially fewer trials will be discarded. Additionally, this would imply that the tapping trials could be shorter than 30 s. However, further studies are needed to evaluate the full implications of such changes. Further analysis suggests that the mean tapping frequency measured during only 15 s of a tapping trial produces comparable results, indicating that a shorter task may be viable. However, further studies are needed to confirm this hypothesis. In addition, we do not know how patients' behavior and intrinsic motivation will change when performing the task in a shorter time frame. Based on our observations, we expect 20 seconds of tapping to be a suitable compromise. We do not recommend shorter trials as we know the initial 3 s of tapping account for task inertia and momentum [5]. Moreover, applying the gap removal algorithm also reduces the effective trial length, but trials need to be sufficiently long to quantify fatigue.

5.7 Limitations and Future Challenges

5.7.1 Tapping and Impairment. A larger study population is needed for evaluating the reliability of our metric in MS patients with hand impairment. Only two of seven hand-impaired patients were non-fatigued. Hence, at this point, we cannot conclude if there is a statistically significant difference between fatigued and non-fatigued patients within this specific population. However, we do not see this as a significant drawback of our approach. Our results show that our tapping task is feasible and valid in our MS cohort and is, therefore, a promising tool for patients with other disease entities, such as post-COVID19 syndrome, which is not associated with hand impairment. Future studies should include larger numbers of MS patients combining the whole spectrum of disabilities and further expand on other diseases, particularly those that do not entail hand impairment.

5.7.2 Recognizing Different Fatigue Levels. In this study, we used the FSMC as a 2-level assessment tool. However, the FSMC offers thresholds for the different fatigue levels: "mild," "moderate," "severe." We used the FSMC for binary classification and considered patients fatigued once they exceeded the lowest threshold (mild fatigue). During future work, we plan to explore using our approach for classifying the multiple fatigue levels. A larger study population is needed for assessing the feasibility of this approach.

5.7.3 Recommendation for Future Trials. First, single tapping task measurements are usually not reliable as they could be classified as invalid. Averaging values of several trials lead to the best results when analyzing fatigue. The frequency of the measurements is also an important point that should be taken into account. Even though we did not conduct specific interviews to get feedback about the usability of the task and study design, some patients gave informal feedback indicating that frequent testing may become tedious or tiresome.

6 CONCLUSIONS

We introduced a new metric as a proxy to objectively quantify perceived fatigue. Our metric, mean tapping frequency, is derived from a simple tapping task performed on commodity smartphones. The validity of the metric has been confirmed by a significant correlation with handgrip strength measurements, which is the current standard procedure in measuring motor fatigability. Additionally, we demonstrate that our approach is comparable to touch duration, which has been recently reported as a motor fatigability metric of the tapping task [5]. Our two-week in-the-wild study, in 35 MS patients, shows that mean tapping frequency can rank fatigued and non-fatigued with AUC_{ROC} $\overline{X} = .76 \pm 5$ according to the FSMC, and with AUC_{ROC} $\overline{X} = .81 \pm .05$ according to the FSS, indicating an association between fatigue and our smartphone-based assessment metric.

In summary, our results show that: (1) Tapping frequency is a valid motor fatigability metric. (2) Our data processing pipeline maintains task validity with an increase of 19% over Barrios et al.'s method [5]. (3) Mean tapping frequency can discriminate fatigue rated by two clinical fatigue scales (FSS and FSMC). (4) Mean tapping frequency as an objective fatigue metric is valid in-the-wild. (5) Combining several trials improves the reliability of fatigue prediction. Future studies in MS patients with hand impairment are needed to establish the validity of our metric in this population. Furthermore, future longitudinal studies are needed to establish optimal time intervals between tapping trials and verify if our metric can be established as a surrogate for perceived fatigue.

Our goal was to study the feasibility of establishing an objective metric as a surrogate for perceived fatigue. We are confident that our work is a step towards the ubiquitous and objective quantification of the symptom. Our simple model provides good interpretability and a higher chance of being adopted in clinical practice. Providing a novel tool to continuously follow patients with fatigue meets an important unmet medical need in MS and in many areas of medicine, where fatigue is a prevalent condition. An objective and reliable measure as a surrogate for fatigue facilitates further research on this devastating symptom, particularly the development of novel therapies. Additionally, the ability to monitor patients over time and independently from medical facilities (i.e., in-the-wild) provides an important advantage in assessing the effects of therapeutic interventions.

REFERENCES

[1] S H Alusi, J Worthington, S Glickman, L J Findley, and P G Bain. 2000. Evaluation of three different ways of assessing tremor in multiple sclerosis. Journal of Neurology, Neurosurgery & Psychiatry 68, 6 (2000), 756–760. https://doi.org/10.1136/jnnp.68.6.756

[2] Kadriye Armutlu, Nilufer Cetisli Korkmaz, Ilke Keser, Vildan Sumbuloglu, Derya Irem Akbiyik, Zafer Guney, and Rana Karabudak. 2007. The validity and reliability of the Fatigue Severity Scale in Turkish multiple sclerosis patients. International Journal of Rehabilitation Research 30, 1 (2007). https://doi.org/10.1097/MRR.0b013e3280146ec4

[3] Duncan R. Babbage, Kirsten van Kessel, Juliet Drown, Sarah Thomas, Ann Sezier, Peter Thomas, and Paula Kersten. 2019. MS Energize: Field trial of an app for self-management of fatigue for people with multiple sclerosis. Internet Interventions 18 (2019), 100291. https://doi.org/10.1016/j.invent.2019.100291

[4] Rohit Bakshi. 2003. Fatigue associated with multiple sclerosis: diagnosis, impact and management. Multiple Sclerosis Journal 9, 3 (2003), 219–227. https://doi.org/10.1191/1352458503ms904oa

[5] Liliana Barrios, Pietro Oldrati, David Lindlbauer, Marc Hilty, Helen Hayward-Koennecke, Christian Holz, and Andreas Lutterotti. 2020. A Rapid Tapping Task on Commodity Smartphones to Assess Motor Fatigability. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20). Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/3313831.3376588

[6] B L Belza. 1995. Comparison of self-reported fatigue in rheumatoid arthritis and controls. The Journal of rheumatology 22, 4 (1995).

[7] B. Bigland-Ritchie, R. Johansson, O. C. Lippold, and J. J. Woods. 1983. Contractile speed and EMG changes during fatigue of sustained maximal voluntary contractions. Journal of Neurophysiology 50, 1 (1983), 313–324. https://doi.org/10.1152/jn.1983.50.1.313

[8] Alexandra K. Boukhvalova, Emily Kowalczyk, Thomas Harris, Peter Kosa, Alison Wichman, Mary A. Sandford, Atif Memon, and Bibiana Bielekova. 2018. Identifying and Quantifying Neurological Disability via Smartphone. Frontiers in Neurology 9 (2018), 740. https://doi.org/10.3389/fneur.2018.00740

[9] Tiffany J. Braley and Ronald D. Chervin. 2010. Fatigue in Multiple Sclerosis: Mechanisms, Evaluation, and Treatment. Sleep 33, 8 (08 2010), 1061–1067. https://doi.org/10.1093/sleep/33.8.1061

[10] A. Brioschi, S. Gramigna, E. Werth, F. Staub, C. Ruffieux, C. Bassetti, M. Schluep, and J. M. Annoni. 2009. Effect of Modafinil on Subjective Fatigue in Multiple Sclerosis and Stroke Patients. European Neurology 62, 4 (2009), 243–249.

[11] J.M. Casillas, S. Damak, J.C. Chauvet-Gelinier, G. Deley, and P. Ornetti. 2006. Fatigue in patients with cardiovascular disease. Annales de Réadaptation et de Médecine Physique 49, 6 (2006), 392 – 402. https://doi.org/10.1016/j.annrmp.2006.04.003

[12] Abhijit Chaudhuri and Peter O Behan. 2004. Fatigue in neurological disorders. The Lancet 363, 9413 (2004), 978 – 988. https://doi.org/10.1016/S0140-6736(04)15794-2

[13] SY Chipchase, NB Lincoln, and KA Radford. 2003. Measuring fatigue in people with Multiple Sclerosis. Disability and Rehabilitation 25, 14 (2003), 778–784. https://doi.org/10.1080/0963828031000093477

[14] David H. Clarke. 1986. Sex Differences in Strength and Fatigability. Research Quarterly for Exercise and Sport 57, 2 (1986), 144–149. https://doi.org/10.1080/02701367.1986.10762190

[15] Ronald A. Cohen and Marc Fisher. 1989. Amantadine Treatment of Fatigue Associated With Multiple Sclerosis. Archives of Neurology 46, 6 (06 1989), 676–680. https://doi.org/10.1001/archneur.1989.00520420096030

[16] C. Colosimo, E. Millefiorini, M. G. Grasso, F. Vinci, M. Fiorelli, T. Koudriavtseva, and C. Pozzilli. 1995. Fatigue in MS is associated with specific clinical features. Acta Neurologica Scandinavica 92, 5 (1995), 353–355. https://doi.org/10.1111/j.1600-0404.1995.tb00145.x

[17] Carlos del Rio and Preeti N. Malani. 2020. COVID-19—New Insights on a Rapidly Changing Epidemic. JAMA 323 (04 2020). https://doi.org/10.1001/jama.2020.3072

[18] Marie D'hooghe, Geert Van Gassen, Daphne Kos, Olivier Bouquiaux, Melissa Cambron, Danny Decoo, Andreas Lysandropoulos, Bart Van Wijmeersch, Barbara Willekens, Iris-Katharina Penner, and Guy Nagels. 2018. Improving fatigue in multiple sclerosis by

smartphone-supported energy management: The MS TeleCoach feasibility study. Multiple Sclerosis and Related Disorders 22 (2018), 90 – 96. https://doi.org/10.1016/j.msard.2018.03.020

[19] Ruth Djaldetti, Ilan Ziv, Anat Achiron, and Eldad Melamed. 1996. Fatigue in multiple sclerosis compared with chronic fatigue syndrome. Neurology 46, 3 (1996), 632–635. https://doi.org/10.1212/WNL.46.3.632 arXiv:https://n.neurology.org/content/46/3/632.full.pdf

[20] Bruce H. Dobkin. 2008. Fatigue Versus Activity-Dependent Fatigability in Patients With Central or Peripheral Motor Impairments. Neurorehabilitation and Neural Repair 22, 2 (2008), 105–110. https://doi.org/10.1177/1545968308315046

[21] Kristin Falk, Karl Swedberg, Fannie Gaston-Johansson, and Inger Ekman. 2007. Fatigue is a Prevalent and Severe Symptom Associated with Uncertainty and Sense of Coherence in Patients with Chronic Heart Failure. European Journal of Cardiovascular Nursing 6, 2 (2007), 99–104. https://doi.org/10.1016/j.ejcnurse.2006.05.004

[22] Andy Field and Graham J Hole. 2003. How to Design and Report Experiments. SAGE Publications Ltd.

[23] John D. Fisk, Paul G. Ritvo, Lynn Ross, David A. Haase, Thomas J. Marrie, and Walter F. Schlech. 1994. Measuring the Functional Impact of Fatigue: Initial Validation of the Fatigue Impact Scale. Clinical Infectious Diseases 18, Supplement 1 (1994), S79–S83. https://doi.org/10.1093/clinids/18.supplement_1.s79

[24] John D. Fisk, Paul G. Ritvo, Lynn Ross, David A. Haase, Thomas J. Marrie, and Walter F. Schlech. 1994. Measuring the Functional Impact of Fatigue: Initial Validation of the Fatigue Impact Scale. Clinical Infectious Diseases 18, Supplement 1 (01 1994), S79–S83.

[25] JE Freal, GH Kraft, and JK Coryell. 1984. Symptomatic fatigue in multiple sclerosis. Archives of physical medicine and rehabilitation 65, 3 (March 1984), 135—138. http://europepmc.org/abstract/MED/6703889

[26] J Friedman and H Friedman. 1993. Fatigue in Parkinson's disease. Neurology 43, 10 (1993), 2016–2016.

[27] Joseph H. Friedman, Guido Alves, Peter Hagell, Johan Marinus, Laura Marsh, Pablo Martinez-Martin, Christopher G. Goetz, Werner Poewe, Olivier Rascol, Cristina Sampaio, Glenn Stebbins, and Anette Schrag. 2010. Fatigue rating scales critique and recommendations by the Movement Disorders Society task force on rating scales for Parkinson's disease. Movement Disorders 25, 7 (2010), 805–822. https://doi.org/10.1002/mds.22989

[28] Guido Giunti, Octavio Rivera-Romero, Jan Kool, Jens Bansi, Jose Luis Sevillano, Anabel Granja-Dominguez, Guillermo Izquierdo-Ayuso, and Diego Giunta. 2020. Evaluation of More Stamina, a Mobile App for Fatigue Management in Persons with Multiple Sclerosis: Protocol for a Feasibility, Acceptability, and Usability Study. JMIR Res Protoc 9, 8 (2020). https://doi.org/10.2196/18196

[29] Myla D. Goldman, Ruth Ann Marrie, , and Jeffrey A. Cohen. 2008. Evaluation of the six-minute walk in multiple sclerosis subjects and healthy controls. Multiple Sclerosis Journal 14, 3 (2008), 383–390. https://doi.org/10.1177/1352458507082607

[30] Multiple Sclerosis Clinical Practice Guidelines. 1998. Fatigue and multiple sclerosis: evidence-based management strategies for fatigue in multiple sclerosis. Paralyzed Veterans of America.

[31] W. Hader, P. Duquette, A. Auty Winnipeg, S. Hashimoto, J. Noseworthy, G. Sawa, D. Brunet, R. Nelson, T. Gray, G. Klein, G. Francis, Y. Lapierre, B. Weinshenker, W. Barkas, S. Philips, M. Girard, and T. J. Murray. 1987. A Randomized Controlled Trial of Amantadine in Fatigue Associated With Multiple Sclerosis. Canadian Journal of Neurological Sciences 14, 3 (Aug. 1987), 273–278. https://doi.org/10.1017/S0317167100026603 Copyright: Copyright 2016 Elsevier B.V., All rights reserved.

[32] Martin Heine, Lizanne Eva van den Akker, Lyan Blikman, Trynke Hoekstra, Erik van Munster, Olaf Verschuren, Anne Visser-Meily, Gert Kwakkel, V. de Groot, H. Beckerman, A. Malekzadeh, L. E. van den Akker, M. Looijmans, S. A. Sanches, J. Dekker, E. H. Collette, B. W. van Oosten, C. E. Teunissen, M. A. Blankenstein, I. C. J. M. Eijssen, M. Rietberg, M. Heine, O. Verschuren, G. Kwakkel, J. M. A. Visser-Meily, I. G. L. van de Port, E. Lindeman, L. J. M. Blikman, J. van Meeteren, J. B. J. Bussmann, H. J. Stam, R. Q. Hintzen, H. G. A. Hacking, E. L. Hoogervorst, S. T. F. M. Frequin, H. Knoop, B. A. de Jong, G. Bleijenberg, F. A. J. de Laat, M. C. Verhulsdonck, E. Th. L. van Munster, C. J. Oosterwijk, and G. J. Aarts. 2016. Real-Time Assessment of Fatigue in Patients With Multiple Sclerosis: How Does It Relate to Commonly Used Self-Report Fatigue Questionnaires? Archives of Physical Medicine and Rehabilitation 97, 11 (2021/02/03 2016), 1887–1894.e1. https://doi.org/10.1016/j.apmr.2016.04.019

[33] Sarah Hewlett, Zoë Cockshott, Margaret Byron, Karen Kitchen, Sue Tipler, Denise Pope, and Maggie Hehir. 2005. Patients' perceptions of fatigue in rheumatoid arthritis: Overwhelming, uncontrollable, ignored. Arthritis Care & Research 53, 5 (2005), 697–702. https://doi.org/10.1002/art.21450

[34] Jin Huang and C. X. Ling. 2005. Using AUC and accuracy in evaluating learning algorithms. IEEE Transactions on Knowledge and Data Engineering 17, 3 (2005), 299–310.

[35] Peter Joseph Jongen, Ludovicus G Sinnige, Björn M van Geel, Freek Verheul, Wim I Verhagen, Ruud A van der Kruijk, Reinoud Haverkamp, Hans M Schrijver, J Coby Baart, Leo H Visser, Edo P Arnoldus, H Jacobus Gilhuis, Paul Pop, Monique Booy, Wim Lemmens, Rogier Donders, Anton Kool, and Esther van Noort. 2015. The interactive web-based program MSmonitor for self-management and multidisciplinary care in multiple sclerosis: concept, content, and pilot results. Patient preference and adherence 9 (12 2015), 1741–1750. https://doi.org/10.2147/PPA.S93783

[36] H. Kaynak, A. Altintaş, D. Kaynak, Ö. Uyanik, S. Saip, J. Ağaoğlu, G. Önder, and A. Siva. 2006. Fatigue and sleep disturbance in multiple sclerosis. European Journal of Neurology 13, 12 (2006), 1333–1339. https://doi.org/10.1111/j.1468-1331.2006.01499.x

[37] Benzi M Kluger, Lauren B Krupp, and Roger M Enoka. 2013. Fatigue and fatigability in neurologic inllnesses. Neurology 80 (2013), 409–416. https://doi.org/10.1212/WNL.0b013e31827f07be

[38] Gisela Kobelt, Alan Thompson, Jenny Berg, Mia Gannedahl, and Jennifer Eriksson. 2017. New insights into the burden and costs of multiple sclerosis in Europe. Multiple Sclerosis Journal 23, 8 (2017), 1123–1136. https://doi.org/10.1177/1352458517694432

[39] D. Kos, E. Kerckhofs, G. Nagels, M. B. D'hooghe, and S. Ilsbroukx. 2007. Origin of Fatigue in Multiple Sclerosis: Review of the Literature. Neurorehabilitation and Neural Repair 22, 1 (2007), 91–100. https://doi.org/10.1177/1545968306298934

[40] Lauren Krupp. 2006. Editorial. Multiple Sclerosis Journal 12, 4 (2006), 367–368. https://doi.org/10.1191/135248506ms1373ed

[41] Lauren B. Krupp. 2003. Fatigue in Multiple Sclerosis. CNS Drugs 17, 4 (2003), 225–234. https://doi.org/10.2165/00023210-200317040-00002

[42] Lauren B. Krupp, Luis A. Alvarez, Nicholas G. LaRocca, and Labe C. Scheinberg. 1988. Fatigue in Multiple Sclerosis. Archives of Neurology 45, 4 (04 1988), 435–437. https://doi.org/10.1001/archneur.1988.00520280085020

[43] L. B. Krupp, P. K. Coyle, C. Doscher, A. Miller, A. H. Cross, L. Jandorf, J. Halper, B. Johnson, L. Morgante, and R. Grimson. 1995. Fatigue therapy in multiple sclerosis. Neurology 45, 11 (1995), 1956–1961. https://doi.org/10.1212/WNL.45.11.1956

[44] Lauren B. Krupp, Nicholas G. LaRocca, Joanne Muir-Nash, and Alfred D. Steinberg. 1989. The Fatigue Severity Scale: Application to Patients With Multiple Sclerosis and Systemic Lupus Erythematosus. Archives of Neurology 46, 10 (10 1989), 1121–1123. https://doi.org/10.1001/archneur.1989.00520460115022

[45] John F. Kurtzke. 1983. Rating neurologic impairment in multiple sclerosis. Neurology 33, 11 (1983), 1444–1444.

[46] Rüdiger Lange, Marek Volkmer, Christoph Heesen, and Joachim Liepert. 2009. Modafinil effects in multiple sclerosis patients with fatigue. Journal of Neurology 256, 4 (2009), 645–650.

[47] Alenka Horvat Ledinek, Mojca Cizek Sajko, and Uros Rot. 2013. Evaluating the effects of amantadin, modafinil and acetyl-l-carnitine on fatigue in multiple sclerosis – result of a pilot randomized, blind study. Clinical Neurology and Neurosurgery 115 (2013), S86–S89. https://doi.org/10.1016/j.clineuro.2013.09.029 Proceedings of the 6th Dubrovnik International Conference on Multiple Sclerosis, Dubrovnik, Croatia, 10-12 May 2012.

[48] Bryan D. Loy, Ruby L. Taylor, Brett W. Fling, and Fay B. Horak. 2017. Relationship between perceived fatigue and performance fatigability in people with multiple sclerosis: A systematic review and meta-analysis. Journal of Psychosomatic Research 100 (2017), 1–7. https://doi.org/10.1016/j.jpsychores.2017.06.017

[49] Virgil Mathiowetz, Karen Weber, Nancy Kashman, and Gloria Volland. 1985. Adult Norms for the Nine Hole Peg Test of Finger Dexterity. The Occupational Therapy Journal of Research 5, 1 (1985), 24–38. https://doi.org/10.1177/153944928500500102

[50] Anna Mazur-Mosiewicz and Raymond S. Dean. 2011. Halstead-Reitan Neuropsychological Test Battery. Springer US, Boston, MA, 727–731. https://doi.org/10.1007/978-0-387-79061-9_1311

[51] R. G. Miller, R. S. Moussavi, A. T. Green, P. J. Carson, and M. W. Weiner. 1993. The fatigue of rapid repetitive movements. Neurology 43, 4 (1993), 755–755. https://doi.org/10.1212/WNL.43.4.755

[52] Sandra A. Mitchell. 2010. Cancer-related fatigue: state of the science. Clinical Translation 2 (5) (2010), 364–383.

[53] F Möller, J Poettgen, F Broemel, A Neuhaus, M Daumer, and C Heesen. 2011. HAGIL (Hamburg Vigil Study): a randomized placebo-controlled double-blind study with modafinil for treatment of fatigue in patients with multiple sclerosis. Multiple Sclerosis Journal 17, 8 (2011), 1002–1009.

[54] Luis Carlos Nacul, Kathleen Mudie, Caroline C. Kingdon, Taane G. Clark, and Eliana Mattos Lacerda. 2018. Hand Grip Strength as a Clinical Biomarker for ME/CFS and Disease Severity. Frontiers in Neurology 9 (2018), 992. https://doi.org/10.3389/fneur.2018.00992

[55] Bardia Nourbakhsh, Nisha Revirajan, Bridget Morris, Christian Cordano, Jennifer Creasman, Michael Manguinao, Kristen Krysko, Alice Rutatangwa, Caroline Auvray, Salman Aljarallah, Chengshi Jin, Ellen Mowry, Charles McCulloch, and Emmanuelle Waubant. 2021. Safety and efficacy of amantadine, modafinil, and methylphenidate for fatigue in multiple sclerosis: a randomised, placebo-controlled, crossover, double-blind trial. The Lancet Neurology 20, 1 (2021), 38–48. https://doi.org/10.1016/S1474-4422(20)30354-9

[56] Iris-Katharina Penner, Chiara Raselli, Markus Stöcklin, Klaus Opwis, Ludwig Kappos, and Pasquale Calabrese. 2009. The Fatigue Scale for Motor and Cognitive Functions (FSMC): validation of a new instrument to assess multiple sclerosis-related fatigue. Multiple Sclerosis Journal 15, 12 (2020/07/30 2009), 1509–1517. https://doi.org/10.1177/1352458509348519

[57] Ray Perrin, Lisa Riste, Mark Hann, Andreas Walther, Annice Mukherjee, and Adrian Heald. 2020. Into the looking glass: Post-viral syndrome post COVID-19. Medical hypotheses 144 (11 2020), 110055–110055.

[58] Kristopher J. Preacher and James P. Selig. 2012. Advantages of Monte Carlo Confidence Intervals for Indirect Effects. Communication Methods and Measures 6, 2 (2012), 77–98. https://doi.org/10.1080/19312458.2012.679848

[59] John Prince, Siddharth Arora, and Maarten de Vos. 2018. Big data in Parkinson's disease: using smartphones to remotely detect longitudinal disease phenotypes. Physiological Measurement 39, 4 (2018), 044005. https://doi.org/10.1088/1361-6579/aab512

[60] Blake P. Printy, Lindsey M. Renken, John P. Herrmann, Isac Lee, Bryant Johnson, Emily Knight, Georgeta Varga, and D. Diane Whitmer. 2014. Smartphone application for classification of motor impairment severity in Parkinson's disease. In 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. https://doi.org/10.1109/EMBC.2014.6944176

[61] E Pucci, P Branãs, R D'Amico, G Giuliani, A Solari, and C Taus. 2007. Amantadine for fatigue in multiple sclerosis. The Cochrane database of systematic reviews 2007, 1 (01 2007), CD002818–CD002818.

[62] K W Rammohan, J H Rosenberg, D J Lynn, A M Blumenfeld, C P Pollak, and H N Nagaraja. 2002. Efficacy and safety of modafinil (Provigil) for the treatment of fatigue in multiple sclerosis: a two centre phase 2 study. Journal of neurology, neurosurgery, and psychiatry 72, 2

(02 2002), 179–183.
[63] Thorsten Rudroff, John H. Kindred, and Nathaniel B. Ketelhut. 2016. Fatigue in Multiple Sclerosis: Misconceptions and Future Research Directions. Frontiers in Neurology 7 (2016), 122.
[64] P. Scherer, H. Bauer, and K. Baum. 1997. Alternate finger tapping test in patients with migraine. Acta Neurologica Scandinavica 96, 6 (1997), 392–396. https://doi.org/10.1111/j.1600-0404.1997.tb00304.x
[65] Steven R Schwid, Melissa Covington, Benjamin M Segal, and Andrew D Goodman. 2002. Fatigue in multiple sclerosis: current understanding and future directions. Journal of rehabilitation research and development 39 2 (2002), 211–24.
[66] S. R. Schwid, C. A. Thornton, S. Pandya, K. L. Manzur, M. Sanjak, M. D. Petrie, M. P. McDermott, and A. D. Goodman. 1999. Quantitative assessment of motor fatigue and strength in MS. Neurology 53, 4 (1999), 743. https://doi.org/10.1212/WNL.53.4.743
[67] Deborah Severijns, Ilse Lamers, Lore Kerkhofs, and Peter Feys. 2014. Hand Grip Fatigability in Persons with Multiple Sclerosis According to Hand Dominance and Disease Progression. Journal of rehabilitation medicine 47 (09 2014). https://doi.org/10.2340/16501977-1897
[68] Deborah Severijns, Ilse Lamers, Lore Kerkhofs, and Peter Feys. 2015. Hand grip fatigability in persons with multiple sclerosis according to hand dominance and disease progression. J Rehabil Med. (2015), 154–160. https://doi.org/10.2340/16501977-1897
[69] Deborah Severijns, Inge Zijdewind, Ulrik Dalgas, Ilse Lamers, Caroline Lismont, and Peter Feys. 2017. The Assessment of Motor Fatigability in Persons With Multiple Sclerosis: A Systematic Review. Neurorehabilitation and Neural Repair 31, 5 (2017), 413–431.
[70] Ping Sheng, Lijun Hou, Xiang Wang, Xiaowen Wang, Chengguang Huang, Mingkun Yu, Xi Han, and Yan Dong. 2013. Efficacy of modafinil on fatigue and excessive daytime sleepiness associated with neurological disorders: a systematic review and meta-analysis. PloS one 8, 12 (12 2013), e81802–e81802.
[71] Afsaneh Shirani, Braeden D. Newton, and Darin T. Okuda. 2017. Finger tapping impairments are highly sensitive for evaluating upper motor neuron lesions. BMC Neurology 17, 1 (2017), 55. https://doi.org/10.1186/s12883-017-0829-y
[72] E. R. Spelten, J. H. A. M. Verbeek, A. L. J. Uitterhoeve, A. C. Ansink, J. van der Lelie, T. M. de Reijke, M. Kammeijer, J. C. J. M. de Haes, and M. A. G. Sprangers. 2003. Cancer, fatigue and the return of patients to work—a prospective cohort study. European Journal of Cancer 39, 11 (2003), 1562–1567. https://doi.org/10.1016/S0959-8049(03)00364-2
[73] B. Stankoff, E. Waubant, C. Confavreux, G. Edan, M. Debouverie, L. Rumbach, T. Moreau, J. Pelletier, C. Lubetzki, and M. Clanet. 2005. Modafinil for fatigue in MS. Neurology 64, 7 (2005), 1139–1143. https://doi.org/10.1212/01.WNL.0000158272.27070.6A
[74] Anneke Steens, Astrid de Vries, Jolien Hemmen, Thea Heersema, Marco Heerings, Natasha Maurits, and Inge Zijdewind. 2012. Fatigue Perceived by Multiple Sclerosis Patients Is Associated With Muscle Fatigue. Neurorehabilitation and Neural Repair 26, 1 (2012), 48–57.
[75] Jukka Surakka, Anders Romberg, Juhani Ruutiainen, Arja Virtanen, Sirkka Aunola, and Kari Mäentaka. 2004. Assessment of muscle strength and motor fatigue with a knee dynamometer in subjects with multiple sclerosis: a new fatigue index. Clinical Rehabilitation 18, 6 (2004), 652–659. https://doi.org/10.1191/0269215504cr781oa
[76] Makoto Tanigawa, Jason Stein, John Park, Peter Kosa, Irene Cortese, and Bibiana Bielekova. 2017. Finger and foot tapping as alternative outcomes of upper and lower extremity function in multiple sclerosis. Multiple sclerosis journal - experimental, translational and clinical 3, 1 (2017). https://doi.org/10.1177/2055217316688930
[77] Ana Lisa Taylor Tavares, Gregory S.X.E Jefferis, Mandy Koop, Bruce C. Hill, Trevor Hastie, Gary Heit, and Helen M. Bronte-Stewart. 2005. Quantitative measurements of alternating finger tapping in Parkinson's disease correlate with UPDRS motor disability and reveal the improvement in fine motor control from medication and deep brain stimulation. Movement Disorders 20, 10 (2005), 1286–1298.
[78] Catherine Tong, Matthew Craner, Matthieu Vegreville, and Nicholas D. Lane. 2019. Tracking Fatigue and Health State in Multiple Sclerosis Patients Using Connnected Wellness Devices. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 3, 3, Article 106 (Sept. 2019), 19 pages. https://doi.org/10.1145/3351264
[79] Liam Townsend, Adam H. Dyer, Karen Jones, Jean Dunne, Aoife Mooney, Fiona Gaffney, Laura O'Connor, Deirdre Leavy, Kate O'Brien, Joanne Dowds, Jamie A. Sugrue, David Hopkins, Ignacio Martin-Loeches, Cliona Ni Cheallaigh, Parthiban Nadarajan, Anne Marie McLaughlin, Nollaig M. Bourke, Colm Bergin, Cliona O'Farrelly, Ciaran Bannan, and Niall Conlon. 2020. Persistent fatigue following SARS-CoV-2 infection is common and independent of severity of initial infection. PLOS ONE 15, 11 (11 2020), 1–12.
[80] Nieves Téllez, Jordi Río, Mar Tintoré, Carlos Nos, Inés Galán, and Xavier Montalban. 2005. Does the Modified Fatigue Impact Scale offer a more comprehensive assessment of fatigue in MS? Multiple Sclerosis Journal 11, 2 (2005), 198–202.
[81] Philipp O Valko, Claudio L Bassetti, Konrad E Bloch, Ulrike Held, and Christian R Baumann. 2008. Validation of the fatigue severity scale in a Swiss cohort. Sleep 31, 11 (2008), 1601–1607. https://doi.org/10.1093/sleep/31.11.1601
[82] Ria Wolkorte, Dorothea J. Heersema, and Inge Zijdewind. 2015. Muscle Fatigability During a Sustained Index Finger Abduction and Depression Scores Are Associated With Perceived Fatigue in Patients With Relapsing-Remitting Multiple Sclerosis. Neurorehabilitation and Neural Repair 29, 8 (2015), 796–802. https://doi.org/10.1177/1545968314567151
[83] Inge Zijdewind, Roeland Prak, and Ria Wolkorte. 2016. Fatigue and Fatigability in Persons With Multiple Sclerosis. Exercise and Sport Sciences Reviews 80 (2016), 123–128. https://doi.org/10.1249/JES.0000000000000088

A FATIGUE SEVERITY SCALE (FSS) [44, 81]

Table 1. Fatigue Severity Scale (FSS) [44, 81].

	Strongly disagree (1) -> Strongly agree (7)
1. My motivation is lower when I am fatigued.	
2. Exercise brings on my fatigue.	
3. I am easily fatigued.	
4. Fatigue interferes with my physical functioning.	
5. Fatigue causes frequent problems for me.	
6. My fatigue prevents sustained physical functioning.	
7. Fatigue interferes with carrying out certain duties and responsibilities.	
8. Fatigue is among my most disabling symptoms.	
9. Fatigue interferes with my work, family, or social life.	

B MAXIMUM TAPPING FREQUENCY VS. MAXIMUM HANDGRIP

As depicted in Figure 12 (top left), there is a significant difference between the maximum tapping frequency of patients that do not have motor fatigue and those who are classified as motor fatigued using the FSMC questionnaire, with Kruskal-Wallis $H = 8.67$ ($p < 0.01$). However, there is no statistically significant difference between the same groups using the maximum handgrip strength (cf. Figure 12 bottom left).

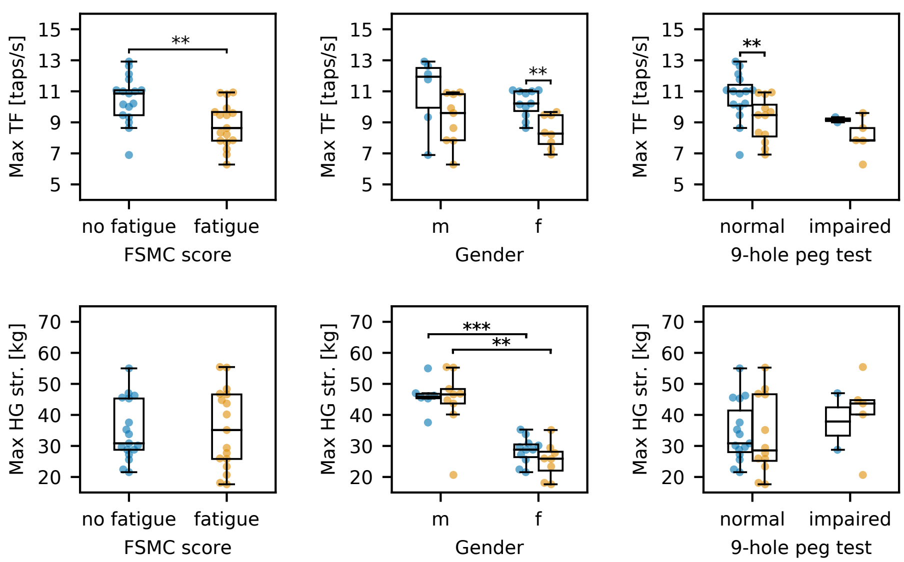

Fig. 12. Maximum tapping frequency (*top*) and maximum handgrip strength (*bottom*) in function of FSMC motor fatigue, gender, and impairment as defined by the 9-hole peg test.

When grouping by gender (Figure 12 center), there is a significant difference between non-fatigued and motor fatigued females as defined by the FSMC questionnaire, with $H = 8.36$ ($p < 0.01$), while no significant difference is found in males, where we have a smaller sample size. The handgrip shows no difference between non-fatigued and fatigued patients within the gender groups, but it shows a significant difference between genders, with $H = 11.0$ ($p < 0.001$) and $H = 8.33$ ($p < 0.01$) for non-fatigued and fatigued patients, respectively.

Figure 12 (right) shows the boxplots when grouping by impairment as defined by the 9-HPT. There is a significant difference between the maximum tapping frequency of non-fatigued and motor fatigued patients that are not hand impaired, with $H = 6.69$ ($p < 0.01$), while no significant difference is found in impaired participants, where we have a very small sample size. The max handgrip strength shows no difference between and within the groups.

C MEAN TAPPING FREQUENCY VS. MEAN HANDGRIP ACCORDING TO FSS

When grouping by gender (Figure 13 center), there is a significant difference between non-fatigued and motor fatigued females as defined by the FSS questionnaire, with $H = 4.93$ ($p < 0.05$), a difference is also found in males, with $H = 4.82$ ($p < 0.05$). The handgrip shows no difference between non-fatigued and fatigued patients within the gender groups, but it shows a significant difference between genders, with $H = 10.58$ ($p < 0.01$) and $H = 6.35$ ($p < 0.01$) for non-fatigued and fatigued patients, respectively.

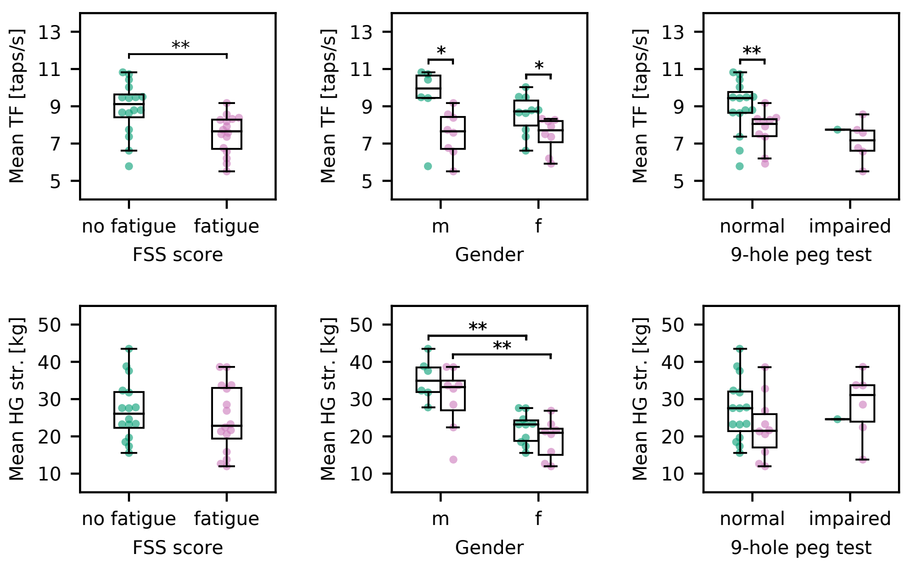

Fig. 13. Mean tapping frequency (*top*) and mean handgrip strength (*bottom*) in function of FSS fatigue, gender, and impairment as defined by the 9-hole peg test.

Figure 13 (right) shows the boxplots when grouping by impairment as defined by the 9-HPT. There is a significant difference between the mean tapping frequency of non-fatigued and motor fatigued patients that are not hand impaired, with $H = 6.5$ ($p < 0.01$), while no significant difference is found in impaired participants, where we have a very small sample size. The mean handgrip strength shows no difference between and within the groups.

D FATIGUE SCALE FOR MOTOR AND COGNITIVE FUNCTIONS (FSMC) [56]

Table 2. Fatigue Scale for Motor and Cognitive Functions (FSMC) [56].

		Does not apply at all	Does not apply much	Slightly applies	Applies a lot	Applies completely
1.	When I concentrate for a long time, I get exhausted sooner than other people of my age.					
2.	When I am experiencing episodes of exhaustion, my movements become noticeably clumsier and less coordinated.					
3.	Because of my episodes of exhaustion, I now need more frequent and/or longer rests during physical activity than I used to.					
4.	When I am experiencing episodes of exhaustion, I am incapable of making decisions.					
5.	When faced with stressful situations, I now find that I get physically exhausted quicker than I used to.					
6.	Because of my episodes of exhaustion, I now have considerably less social contact than I used to.					
7.	Because of my episodes of exhaustion, I now find it more difficult to learn new things than I used to.					
8.	The demands of my work exhaust me mentally more quickly than they used to.					
9.	I feel the episodes of exhaustion particularly strongly in my muscles.					
10.	I no longer have the stamina for long periods of physical activity that I used to have.					
11.	My powers of concentration decrease considerably when I'm under stress.					
12.	When I am experiencing episodes of exhaustion, I am less motivated than others to start activities that involve physical effort.					
13.	My thinking gets increasingly slow when it is hot.					
14.	When I am experiencing an episode of exhaustion, my movements become noticeably slower.					
15.	Because of my episodes of exhaustion, I now feel less like doing things which require concentration.					

Table 2 cont. Fatigue Scale for Motor and Cognitive Functions (FSMC) [56].

16.	When an episode of exhaustion comes on, I am simply no longer able to react quickly.					
17.	When I am experiencing episodes of exhaustion, certain words simply escape me.					
18.	When I am experiencing episodes of exhaustion, I lose concentration considerably quicker than I used to.					
19.	When it is hot, my main feeling is one of extreme physical weakness and lack of energy.					
20.	During episodes of exhaustion, I am noticeably more forgetful.					

Table 3. FSMC cut-off values [56]. We focus our study in the motor aspect of fatigue and classify as motor fatigued participants with FSMC physical score ≥ 22; otherwise, we consider them non-fatigued.

	Cut-off	Classification
FSMC sum score	≥ 43	Mild fatigue
	≥ 53	Moderate fatigue
	≥ 63	Severe fatigue
FSMC cognitive score	≥ 22	Mild cognitive fatigue
	≥ 28	Moderate cognitive fatigue
	≥ 34	Severe cognitive fatigue
FSMC physical score	≥ 22	Mild motor fatigue
	≥ 27	Moderate motor fatigue
	≥ 32	Severe motor fatigue

E DESCRIPTIVE STATISTICS AND NON-PARAMETRIC TEST RESULTS

Table 4. FSMC motor fatigued vs. non-fatigued differences. Non-parametric hypotheses tests with dependent variable *Metric* (mean tapping frequency or mean handgrip strengths) and independent variable motor fatigue classification. Test conducted with IBM SPSS Statistics Version 27. Kruskal-Wallis H^*, Mann-Whitney U^*, Kolmogorov-Smirnov Z^*, Median Test $Median^*$.

Metric	Case	N	FSMC fatigued	FSMC Non-fatigued	Test	$p < 0.05$
mean TF [taps/sec]	all	34 $M = 8.30$ $SD = 1.47$	17 $M = 7.62$ $SD = 1.22$	17 $M = 8.99$ $SD = 1.40$	$H = 7.498$ $U = 65.000$ $Z = 1.543$ $Median = 8.35$	✓$p = .006$ ✓$p = .006$ ✓$p = .017$ ✓$p = .016$
	male	15 $M = 8.53$ $SD = 1.84$	9 $M = 7.96$ $SD = 1.41$	6 $M = 9.41$ $SD = 2.19$	$H = 2.347$ $U = 14.000$ $Z = 1.265$ $Median = 8.56$	$p = .126$ $p = .126$ $p = .082$ $p = .315$
	female	19 $M = 8.13$ $SD = 1.12$	8 $M = 7.25$ $SD = 0.91$	11 $M = 8.76$ $SD = 0.77$	$H = 8.84$ $U = 8.00$ $Z = 1.565$ $Median = 8.27$	✓$p = .003$ ✓$p = .003$ ✓$p = .015$ ✓$p = .02$
	impaired	7 $M = 7.21$ $SD = 1.00$	5 $M = 7.03$ $SD = 1.17$	2 $M = 7.66$ $SD = 0.11$	$H = .600$ $U = 3.000$ $Z = .717$ $Median = 7.58$	$p = .439$ $p = .439$ $p = .683$ $p = 1$
	non-impaired	27 $M = 8.59$ $SD = 1.45$	12 $M = 7.87$ $SD = 1.20$	15 $M = 9.17$ $SD = 1.41$	$H = 5.717$ $U = 41.000$ $Z = 1.42$ $Median = 8.67$	✓$p = .017$ ✓$p = .017$ ✓$p = .035$ $p = .054$
mean HG [kg]	all	34 $M = 25.97$ $SD = 8.30$	17 $M = 24.88$ $SD = 9.07$	17 $M = 27.07$ $SD = 7.56$	$H = .406$ $U = 126.00$ $Z = .857$ $Median = 23.988$	$p = .524$ $p = .540$ $p = .454$ $p = .732$
	male	15 $M = 32.47$ $SD = 7.34$	9 $M = 30.51$ $SD = 8.00$	6 $M = 35.41$ $SD = 5.59$	$H = .681$ $U = 20.00$ $Z = .527$ $Median = 33.27$	$p = .409$ $p = .409$ $p = .944$ $p = 1.0$
	female	19 $M = 20.84$ $SD = 4.62$	8 $M = 18.55$ $SD = 5.34$	11 $M = 22.51$ $SD = 3.33$	$H = 2.727$ $U = 24.00$ $Z = 1.076$ $Median = 21.32$	$p = .099$ $p = .0993$ $p = .197$ $p = .650$
	impaired	7 $M = 27.91$ $SD = 8.40$	5 $M = 26.43$ $SD = 8.47$	2 $M = 31.62$ $SD = 9.92$	$H = .600$ $U = 3.000$ $Z = .598$ $Median = 28.521$	$p = .439$ $p = .439$ $p = .867$ $p = 1$
	non-impaired	27 $M = 25.47$ $SD = 8.36$	12 $M = 24.23$ $SD = 9.600$	15 $M = 26.46$ $SD = 7.41$	$H = 5.36$ $U = 75.000$ $Z = .861$ $Median = 23.26$	$p = .464$ $p = .464$ $p = .449$ $p = .704$

Table 5. FSS Fatigued vs. non-fatigued differences. Non-parametric hypotheses tests with dependent variable *Metric* (mean tapping frequency or mean handgrip strengths) and independent variable fatigue classification according to FSS. Test conducted with IBM SPSS Statistics Version 27. Kruskal-Wallis H^*, Mann-Whitney U^*, Kolmogorov-Smirnov Z^*, Median Test $Median^*$.

Metric	Case	N	FSS fatigued	FSS Non-fatigued	Test	$p < 0.05$
mean TF [taps/sec]	all	32	16	16	$H = 9.091$	✓ $p = .003$
		$M = 8.19$	$M = 7.50$	$M = 8.89$	$U = 48.000$	✓ $p = .003$
		$SD = 1.42$	$SD = 1.04$	$SD = 1.43$	$Z = 1.945$	✓ $p = .001$
					$Median = 8.29$	✓ $p = .005$
	male	14	8	6	$H = 4.817$	✓ $p = .028$
		$M = 8.35$	$M = 7.53$	$M = 9.45$	$U = 7.000$	✓ $p = .028$
		$SD = 1.76$	$SD = 1.20$	$SD = 1.89$	$Z = 1.543$	✓ $p = .017$
					$Median = 8.47$	$p = .103$
	female	18	8	10	$H = 4.934$	✓ $p = .026$
		$M = 8.07$	$M = 7.46$	$M = 8.56$	$U = 15.00$	✓ $p = .026$
		$SD = 1.12$	$SD = 0.93$	$SD = 1.04$	$Z = 1.467$	✓ $p = .026$
					$Median = 8.22$	$p = .152$
	impaired	7	6	1	-	-
		$M = 7.21$	$M = 7.11$	$M = 7.746$	-	-
		$SD = 1.00$	$SD = 1.07$		-	-
					-	-
	non-impaired	25	10	15	$H = 6.511$	✓ $p = .011$
		$M = 8.47$	$M = 7.72$	$M = 8.97$	$U = 29.000$	✓ $p = .011$
		$SD = 1.41$	$SD = 1.01$	$SD = 1.44$	$Z = 1.715$	✓ $p = .006$
					$Median = 8.632$	✓ $p = .004$
mean HG [kg]	all	32	16	16	$H = .513$	$p = .474$
		$M = 25.89$	$M = 24.77$	$M = 27.01$	$U = 109.000$	$p = .474$
		$SD = 8.41$	$SD = 8.88$	$SD = 8.05$	$Z = .707$	$p = .699$
					$Median = 23.988$	$p = .480$
	male	14	8	6	$H = .600$	$p = .439$
		$M = 32.41$	$M = 30.26$	$M = 35.28$	$U = 18.000$	$p = .439$
		$SD = 7.61$	$SD = 8.49$	$SD = 5.70$	$Z = .617$	$p = .841$
					$Median = 33.22$	$p = 1$
	female	18	8	10	$H = 1.334$	$p = .248$
		$M = 20.81$	$M = 19.27$	$M = 22.05$	$U = 27.00$	$p = .248$
		$SD = 4.75$	$SD = 5.27$	$SD = 4.14$	$Z = .738$	$p = .648$
					$Median = 21.48$	$p = .637$
	impaired	7	6	1	-	-
		$M = 27.91$	$M = 28.46$	$M = 24.60$	-	-
		$SD = 8.40$	$SD = 9.06$		-	-
					-	-
	non-impaired	25	10	15	$H = 1.772$	$p = .183$
		$M = 25.33$	$M = 22.55$	$M = 27.18$	$U = 51.000$	$p = .183$
		$SD = 8.50$	$SD = 8.44$	$SD = 8.31$	$Z = .816$	$p = .518$
					$Median = 23.26$	$p = .226$

Table 6. Female vs. male differences. Non-parametric hypotheses tests with dependent variable *Metric* (mean tapping frequency or mean handgrip strengths) and independent variable gender (male or female). Data set corresponding to FSMC motor fatigue classification. Test conducted with IBM SPSS Statistics Version 27. Kruskal-Wallis H^*, Mann-Whitney U^*, Kolmogorov-Smirnov Z^*, Median Test $Median^*$.

Metric	Case	N	Male	Female	Test	$p < 0.05$
mean TF [taps/sec]	all	34	15	19	$H = .556$	$p = .456$
		$M = 8.31$	$M = 8.54$	$M = 8.13$	$U = 121.000$	$p = .456$
		$SD = 1.47$	$SD = 1.84$	$SD = 1.12$	$Z = .894$	$p = .401$
					$Median = 8.35$	$p = .300$
	fatigued	17	9	8	$H = 1.815$	$p = .178$
		$M = 7.62$	$M = 7.96$	$M = 7.25$	$U = 22.000$	$p = .178$
		$SD = 1.22$	$SD = 1.41$	$SD = 0.91$	$Z = 1.143$	$p = .146$
					$Median = 7.733$	$p = .637$
	non-fatigued	17	6	11	$H = 1.455$	$p = .228$
		$M = 8.99$	$M = 9.40$	$M = 8.76$	$U = 21.000$	$p = .228$
		$SD = 1.41$	$SD = 2.20$	$SD = 0.78$	$Z = 1.31$	$p = .063$
					$Median = 1.455$	$p = .228$
mean HG [kg]	all	34	15	19	$H = 16.609$	$\checkmark p = .000$
		$M = 25.97$	$M = 32.47$	$M = 20.84$	$U = 25.000$	$\checkmark p = .000$
		$SD = 8.30$	$SD = 7.34$	$SD = 4.61$	$Z = 2.509$	$\checkmark p = .000$
					$Median = 23.98$	$\checkmark p = .000$
	fatigued	17	9	8	$H = 7.259$	$\checkmark p = .007$
		$M = 24.88$	$M = 30.50$	$M = 18.54$	$U = 8.000$	$\checkmark p = .007$
		$SD = 9.07$	$SD = 8.00$	$SD = 5.34$	$Z = 1.601$	$\checkmark p = .012$
					$Median = 23.26$	$\checkmark p = .015$
	non-fatigued	17	6	11	$H = 11.000$	$\checkmark p = .001$
		$M = 27.06$	$M = 35.42$	$M = 22.51$	$U = 0.000$	$\checkmark p = .001$
		$SD = 7.56$	$SD = 5.59$	$SD = 3.33$	$Z = 1.970$	$\checkmark p = .001$
					$Median = 24.602$	$\checkmark p = .002$

Table 7. Female vs. male differences. Non-parametric hypotheses tests with dependent variable *Metric* (mean tapping frequency or mean handgrip strengths) and independent variable gender (male or female). Data set corresponding to FSS fatigue classification. Test conducted with IBM SPSS Statistics Version 27. Kruskal-Wallis H^*, Mann-Whitney U^*, Kolmogorov-Smirnov Z^*, Median Test $Median^*$.

Metric	Case	N	Male	Female	Test	$p < 0.05$
mean TF [taps/sec]	all	32 $M = 8.19$ $SD = 1.42$	14 $M = 8.35$ $SD = 1.77$	18 $M = 8.07$ $SD = 1.12$	$H = .244$ $U = 113.000$ $Z = .735$ $Median = 8.29$	$p = .621$ $p = .6216$ $p = .653$ $p = .476$
	fatigued	16 $M = 7.50$ $SD = 1.04$	8 $M = 7.53$ $SD = 1.20$	8 $M = 7.46$ $SD = 0.93$	$H = .176$ $U = 28.000$ $Z = .750$ $Median = 7.65$	$p = .674$ $p = .674$ $p = .627$ $p = .1$
	non-fatigued	16 $M = 8.89$ $SD = 1.43$	6 $M = 9.45$ $SD = 1.89$	10 $M = 8.56$ $SD = 1.05$	$H = 2.647$ $U = 15.000$ $Z = 1.033$ $Median = 9.11$	$p = .104$ $p = .104$ $p = .236$ $p = .119$
mean HG [kg]	all	32 $M = 25.89$ $SD = 8.41$	14 $M = 32.41$ $SD = 7.62$	18 $M = 20.81$ $SD = 4.74$	$H = 15.013$ $U = 24.000$ $Z = 2.405$ $Median = 23.98$	✓$p = .000$ ✓$p = .000$ ✓$p = .000$ ✓$p = .000$
	fatigued	16 $M = 24.77$ $SD = 8.89$	8 $M = 30.26$ $SD = 8.50$	8 $M = 19.27$ $SD = 5.27$	$H = 6.353$ $U = 8.000$ $Z = 1.5$ $Median = 22.84$	✓$p = .012$ ✓$p = .012$ ✓$p = .022$ $p = .132$
	non-fatigued	16 $M = 27.02$ $SD = 8.05$	6 $M = 35.28$ $SD = 5.70$	10 $M = 22.06$ $SD = 4.14$	$H = 10.588$ $U = 0.000$ $Z = 1.936$ $Median = 26.078$	✓$p = .001$ ✓$p = .001$ ✓$p = .001$ ✓$p = .007$