Intra-tester and inter-tester reliability of the MicroFET 3 hand-held dynamometer

Clarke MN,1 Ní Mhuircheartaigh DA,1 Walsh GM,1 Walsh JM,1,2 Meldrum D1

1 School of Physiotherapy, The Royal College of Surgeons in Ireland, Dublin, Ireland.
2 Department of Physiotherapy, Trinity Centre for Health Science, St. James’s Hospital, Dublin, Ireland.

ABSTRACT

Background: The reliability of the MicroFET 3 has not previously been reported in the literature. The aim of this study was to evaluate intratester and intertester reliability of the MicroFET3 hand-held dynamometer (HHD) in three lower limb muscle groups.

Methods: Maximum voluntary isometric contraction (MVIC) of hip extension, knee extension and ankle plantar-flexion were measured in 38 healthy participants (males=18, females= 20) by two testers on separate days using the MicroFET HHD. The reliability analysis was carried out using intra-class correlation coefficients (ICCs) to measure association and Bland and Altman plots to demonstrate agreement.

Results: The results showed that intratester reliability was moderate to excellent; with associations ranging from ICC 0.56 - 0.92 and higher agreement for knee and ankle than hip measurements was shown. Intertester reliability was lower, with hip and knee associations ranging from ICC 0.60 - 0.66. Ankle measurements inter-tester associations were particularly low (ICC 0.23 and 0.15). These values would not be considered acceptable for clinical use. Bland and Altman plots used to demonstrate agreement between testers displayed a considerable lack of agreement with discrepancies of up to 150N noted in measurements.

Conclusion: The results suggest that the MicroFET3 HHD displayed moderate to excellent intratester reliability and poor to moderate intertester reliability and agreement with discrepancies noted between muscle groups. While use of this instrument can be recommended when consistently used by a single tester, further reliability analysis should be carried out before this instrument could be recommended for use by different testers in the clinical setting.

Keywords: Muscle Strength; Muscle Strength dynamometer; Reliability of Results

Corresponding author: J. M. Walsh, Email: julie.broderick@tcd.ie

INTRODUCTION

Strength testing is one of the cornerstones of physical examination and an increase in strength is the aim of many rehabilitation programs.1 Manual muscle testing (MMT) is the most commonly used method of strength testing in the clinical setting and is a subjective measurement technique whereby the tester applies resistance to a maximum voluntary muscle contraction.2 Among its advantages is its quick and simple procedure providing information that can be useful in differential diagnosis, prognosis and treatment of neuromuscular and musculoskeletal deficits. Various factors must be taken into account when testing muscles including test standardization, appropriate positioning, observation of how the patient performs the test and avoidance of pain or discomfort which may inhibit the participant from performing a maximal contraction.3 There has been much controversy over the reliability of MMT grades4-7 as good visual and palpation skills are necessary to identify changes in muscle grades, a skill that depends on the experience of the examiner 2, 8.

The need for greater objectivity in clinical examination has led to the development of devices such as the hand-held dynamometer (HHD) and isokinetic machines. The isokinetic machine is considered the gold standard in muscle testing with validity8,9 and reliability10 well documented. However, this equipment is time consuming to use, expensive and not readily accessible to most practitioners.12

Dynamometers offer an alternative to the isokinetic machine as they are a portable, time-efficient and relatively inexpensive method of measuring isometric contractions13 and have also been shown to be valid and reliable when compared to the isokinetic machine.12 Also, with the publication of normative data, HHDs offer a simple alternative in clinical practice and provide a more quantitative objective measure than MMT.14

The reliability of HHDs has been previously reported in the literature.14-17 Studies have used a variety of different dynamometers and none of these have compared reliability among the different models. Further to this, reliability of the MicroFET3 has not been established. Before the MicroFET3 can be used clinically or as part of future research, the reliability of this instrument must be determined.18

The main limitation of HHD studies to date is the influence of tester strength, which inherently affects the ability of the tester to stabilize the dynamometer appropriately.2, 12, 17, 19-20 Measurements obtained may therefore vary according to the tester’s strength. There are also inconsistencies in describing different muscle groups when using HHDs18 which is especially evident in larger muscle groups.17 A number of studies have highlighted difficulties in testing lower-limb muscle strength accurately2, 21-22 which may be due to increased force generated by lower-limb muscle groups.21 Brinkman23 concluded that HHDs were not reliable when testing large muscle groups exerting forces greater than 15kg. Agre et al22
carried out a study examining both intra- and inter-tester reliability of muscle strength in upper and lower extremities using a HHD. These authors concluded the instrument was reliable for measuring upper extremities in clinical practice. However, results obtained from lower extremity measurements were less reliable and it was concluded that the HHD was unacceptable for clinical assessment of the lower-limb.22 Consequently the biggest challenge to these devices to date has been accurate measurement of lower limb strength.

A new dynamometer the MicroFET 3 was patented in 2005 (Hoggan Health Industries, West Jordan, UT). No literature has reported reliability measures of this instrument. The reliability of an instrument must be determined in order to eliminate it as a source of error.24,25 The aim of this study therefore was to investigate intra- and inter-tester reliability as well as agreement between testers of the MicroFET3 HHD in three of the lower limb muscle groups.

**METHODOLOGY**

**Participants**

A convenience sample of 38 healthy participants (20 females and 18 males) from a university population volunteered to participate in this study. To be included in the study, participants had to be healthy and between the ages of 18 and 30 years. Participants were excluded if they had a history of lower extremity injury in the three months prior to testing or if they had a history of any medical condition that would preclude exercise participation. Subjects were instructed to refrain from strenuous activity or alcohol for 24 hours prior to testing, and to refrain from eating for at least 3 hours before the test for standardisation.26 Ethical approval was obtained from the Royal College of Surgeons of Ireland (RCSI) research ethics committee and subjects gave written, informed consent prior to participation.

**Study procedure**

The testers involved in this study were three final year female physiotherapy students. Two of these testers were randomly selected for the final analysis. The testers were trained and instructed in the use of the MicroFET3 HHD prior to commencing testing by an experienced physiotherapist familiar with this equipment. Testers undertook a period of training and familiarization in the use of the HHD, to ensure competency and efficiency. In addition, a pilot study was carried out on eight participants prior to commencement of testing. Throughout the testing period, each tester was blinded to the values obtained by the other tester.

Lower limb strength was measured using the MicroFET3 HHD – The MicroFET3 is a battery operated hand-held device which measures peak force. It is a load-cell based strain gauge type of dynamometer whereby a force distorts a strain gauge and converts it to an electrical signal which can be downloaded to accompanying software to support data analysis. The MicroFET3 measures forces in Newtons (N), up to a value of 890N.

Testing was carried out in the RCSI Movement Laboratory on two occasions, one week apart. During the first visit, the study was explained to the participants. Participants warmed up on a stationary bike for five minutes and then performed stretches to the major muscle groups of the lower extremity, holding each stretch for 10 seconds. Following this, the MicroFET3 was used to measure knee extension, hip flexion and ankle plantar-flexion strength on one leg. The leg to be tested was randomly selected by tossing a coin. Testing positions27-29 and stabilizations were standardized and strictly adhered to (Table 1). Verbal instructions for each test were standardized as follows: “Push as hard as you can, as hard as you can, as hard as you can”, in a loud voice. Isometric strength was measured using a make test as this test has been shown to be more reliable than the break test.15 The make test is carried out by the examiner holding the dynamometer stationary while the subject exerts a maximal force against it.10 Participants were instructed to perform three consecutive maximal efforts, lasting five seconds each. In order to avoid fatigue, a rest period of 30 seconds separated each contraction. This procedure was repeated by each tester, giving a total of 18 measurements for each participant. The testing sequence was consistently carried out in the following order; knee extension followed by

<table>
<thead>
<tr>
<th>Movement</th>
<th>Position</th>
<th>Strap Position</th>
<th>Dynamometer Position</th>
<th>Instruction</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knee Extension</td>
<td>Sitting in quadriceps bench Hip &amp; knees in 90° flexion Hands on thigh, palmar surface upwards</td>
<td>Lower 1/3 of thigh proximal to knee</td>
<td>Anterior leg proximal to ankle</td>
<td>Straighten out your knee</td>
<td>Wang et al. (2002)27</td>
</tr>
<tr>
<td>Hip Extension</td>
<td>Prone lying on plinth Knee extended Hands by their side palmar surface upwards</td>
<td>Across pelvis</td>
<td>Posterior thigh proximal to knee</td>
<td>Lift your leg off the bed keeping your knee straight</td>
<td>Taylor et al. (2004)28</td>
</tr>
<tr>
<td>Ankle Plantar flexion</td>
<td>Prone lying on plinth Ankle in neutral Hands by their side palmar surface upwards</td>
<td>Lower leg proximal to ankle</td>
<td>Plantar surface of metatarsal heads</td>
<td>Point your toes towards me</td>
<td>Nollet &amp; Beelen (1999)29</td>
</tr>
</tbody>
</table>
hip extension and ankle plantar-flexion and the order in which
the testers performed the procedure was randomised by
tossing a coin. There was a rest period of three minutes
between testers. Measurement 1 recorded by each tester was
regarded as the familiarisation measurement. The maximum
recording from measurements 2 and 3 was documented as the
baseline data against which subsequent data was evaluated
against.31 These results were placed in an envelope, sealed,
and were not consulted until after the second visit. The second
visit involved retesting the participants using the same protocol
and order of testing as during the first visit.

Data analysis
Reliability was evaluated by computing Intra-class Correlation
Coefficients (ICC) which analyse the consistency between two or
more quantitative measures25 and Bland and Altman Plots32
which measured agreement. Intra-class correlation coefficients
were calculated from a single measure chosen as the maximum
value obtained during testing and ICC (3,1), one way random,
was used to evaluate inter-tester reliability while ICC (2,1),
two-way random was used to evaluate inter-tester reliability.33
Correlation or association is greater as the ICC value gets closer
to 1.34 It has been reported in the literature that ‘moderate’
association can be judged by ICC values between 0.50 - 0.75
and ‘good’ to ‘excellent’ association over 0.75.32 However, in
order to consider an instrument suitable for clinical use, the ICC
value must be over 0.90.32 The ICC does not give an indication
of magnitude of disparity between measurements.32 Therefore,
Bland and Altman Plots and standard error of measurement
(SEM) were used to compliment the inter-rater analysis. Bland
and Altman plots provide a visual representation of the degree
of agreement between measurements.32 The SEM is an estimate
of measurement error and the smaller the SEM the greater the
agreement. SEM was calculated using the following formula;
S x (1 ICC) where S corresponds to the pooled standard
deviation and ICC is the reliability coefficient.35 All analysis was
performed using SPSS Version 15 for Windows (SPSS Inc,
Chicago, IL) statistics software.

RESULTS
Thirty-eight healthy participants were recruited to participate in
this study (18 males and 20 females). The mean age was
21.8 ± 2.4 years. Five male participants were excluded from
the knee extension testing protocol. In these instances testers
were unable to match the forces exerted by the participants
and were therefore unable to maintain an isometric
contraction. As a result, the measurements obtained were
deemed unreliable.

<table>
<thead>
<tr>
<th>Muscle Group</th>
<th>ICC (95% CI)</th>
<th>SEM (N)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tester 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Knee (n=33)</td>
<td>0.86 (0.74, 0.93)</td>
<td>19.38</td>
</tr>
<tr>
<td>Hip (n=38)</td>
<td>0.57 (0.3, 0.75)</td>
<td>33.64</td>
</tr>
<tr>
<td>Ankle (n=38)</td>
<td>0.56 (0.29, 0.74)</td>
<td>21.89</td>
</tr>
<tr>
<td>Tester 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Knee</td>
<td>-13.4</td>
<td>30.0</td>
</tr>
<tr>
<td>Hip</td>
<td>-28.2</td>
<td>23.8</td>
</tr>
<tr>
<td>Ankle</td>
<td>-8.2</td>
<td>26.7</td>
</tr>
</tbody>
</table>

Table 2 Intra-tester reliability and agreement

<table>
<thead>
<tr>
<th>Movement</th>
<th>d (N)</th>
<th>SD diff</th>
<th>95% Limits of agreement (N)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tester 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Knee</td>
<td>-9.7</td>
<td>27.4</td>
<td>-64.5 - 45.1</td>
</tr>
<tr>
<td>Hip</td>
<td>-21.4</td>
<td>35.4</td>
<td>-92.7 - 49.3</td>
</tr>
<tr>
<td>Ankle</td>
<td>10.3</td>
<td>31.0</td>
<td>-51.7 - 72.4</td>
</tr>
<tr>
<td>Tester 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Knee</td>
<td>-13.4</td>
<td>30.0</td>
<td>-73.5 - 46.7</td>
</tr>
<tr>
<td>Hip</td>
<td>-28.2</td>
<td>23.8</td>
<td>-75.8 - 19.3</td>
</tr>
<tr>
<td>Ankle</td>
<td>-8.2</td>
<td>26.7</td>
<td>-61.7 - 45.3</td>
</tr>
</tbody>
</table>

ICC= Intraclass correlation co-efficients, 95% CI = 95% confidence interval, SEM = standard error of measurement,
D=difference, N=Newtons, SD diff= Standard deviation of difference
Intra-tester reliability and agreement

Intra-tester reliability results for participants are presented in Tables 2. ICC values for the knee testing protocol were good to excellent (0.86 and 0.88). ICC values obtained by tester 1 for the hip and ankle demonstrated moderate association (0.56 and 0.57). However results obtained by tester 2 were good to excellent (0.88 and 0.92). The standard error of measurements (SEM) was higher in tester 1 than tester 2 except in the knee testing protocol (Table 2). The mean differences for both knee and ankle measurements were closer to zero indicating better agreement. Hip measurements displayed a wider range of values indicating weaker agreement (Table 2).

Intra-tester reliability and agreement

ICCs for inter-tester reliability for both days are shown in Table 3. When all subjects were analyzed together moderate association was demonstrated for the hip and knee (ICCs ranged from .60 - .66). Ankle measurement ICCs showed poor association on both days (ICCs .23 and 0.15). A selection of Bland and Altman plots to demonstrate agreement between testers are shown in Figs. 1, 2, and 3 displaying a considerable lack of agreement with discrepancies of up to 150N noted in measurements.

DISCUSSION

The aim of this paper was to investigate the intra-tester and inter-tester reliability of the MicroFET3 HHD. Our results indicated moderate to excellent intra-tester reliability and lower inter-tester reliability based on maximal isometric contractions.

It has been reported that ‘moderate’ association can be judged by ICC values between 0.50 to 0.75 and ‘good’ to ‘excellent’ association over 0.75.32 However in order to consider an instrument suitable for clinical use, the ICC value must be over 0.90.32 Values calculated in this study reflect those reported by other authors who examined the same muscle groups.20, 27, 29, 36
The high force levels produced by the male participants may have posed difficulty for the female testers in maintaining the resistance for the five-second duration. However, this should not have affected association and agreement when analysed. In addition, there was very little evidence of a difference in reliability when the sexes were analysed separately.

Intra-tester reliability data obtained for Tester 1 and Tester 2 were consistent for knee measurements. However data for the hip and ankle varied considerably. Testers 1 and 2 demonstrated ‘moderate to excellent’ association for knee measurements (0.86 - 0.88), however there was less correlation between values calculated for the hip and ankle (0.57 - 0.92 and 0.56 - 0.88, respectively). It is also important to note that although the mean difference of the ankle measurements were less than that of the hip and knee for Tester 2, the degree of variation relative to the actual torque produced by the ankle is quite high, adding further to the poorer reliability measures at the ankle. Testers reported difficulty maintaining the position of the HHD on the metatarsal heads due to the shape of the transducer head and the high force levels exerted by the plantar-flexors. This is consistent with findings of Agre et al\textsuperscript{22} and Colombo et al\textsuperscript{37} who reported that “off centre” loading, i.e. force applied to the muscle at an angle other than 90\textdegree, resulted in inaccurate readings on the HHD, and therefore poorer results obtained for the lower limb muscle group.\textsuperscript{22} Testers also reported difficulty aligning their arms in the direction of these high force levels due to inability to elevate the bed to an ergonomically advantageous position. A high level of localized tenderness was reported by participants where the dynamometer was placed during the testing protocol. Studies investigating the reliability of the HHD in measuring hip musculature, have reported that participants occasionally complained of discomfort due to positioning of the HHD, which may have limited the generation of a maximal contraction and may have influenced the validity of these measurements.\textsuperscript{21} It is possible that these were contributing factors to the inconsistent figures.

Inter-tester reliability as estimated by the ICC (2,1) and Bland and Altman plots was lower than that of intra-tester. Although hip and knee values ranged from 0.60 - 0.66 these would not be considered acceptable for use in the clinical setting. Ankle ICCs were very low on both days (0.23 and 0.15) indicating poor association between testers regardless of day tested. As the testers in this study were final year physiotherapy students, lack of experience in using the dynamometer clinically may have contributed to the low inter-tester reliability.

As reported elsewhere in the literature,\textsuperscript{2, 21} it was also noted that intra-tester reliability tends to be greater than inter-tester reliability. This may be due to less inherent variability, tester strength\textsuperscript{17} and subconscious participant inhibition with perceived weaker testers.\textsuperscript{34} It is also possible that the tester, in anticipation of impending force levels exerted by the participants, may have applied a pressure to the limb evoking a myostatic stretch reflex which could result in greater force production.\textsuperscript{22}

This study employed Bland and Altman plots to assess the agreement of inter-tester strength measurements. In addition to allowing for a more thorough analysis, they provide visualisation of differences in the measures and any outliers present. The values for the mean difference for knee and ankle were close to zero indicating good agreement. The mean difference for the hip measurements ranged from -21.4N to -28.2N, indicating weaker agreement. The 95\% limits of agreement for the mean difference included zero indicating minimal bias between the two measures. These results need to be interpreted with caution however, as the sample size in this study was less than 50. Research indicates that 95\% limits of agreement will be wide if a sample size of at least 50 is not adhered to.\textsuperscript{34}

\section*{Conclusion}

The results of this study suggest that the MicroFET3 HHD has moderate to excellent intra-tester reliability and poor to moderate inter-tester reliability based on maximal isometric contractions of the hip extensors, knee extensors and ankle plantar-flexors. Further research is required to investigate whether tester strength influences reliability of the MicroFET3 HHD and whether improved intra and inter tester reliability can be obtained with modification of the testing procedure.

Finally, not withstanding these limitations, our results show that the MicroFET3 HHD can be considered a reliable instrument for testing muscle strength of knee extensors, hip extensors and ankle plantar-flexors, provided the measurements are consistently carried out by the same tester. This preliminary study may therefore guide other research in the area in the possible future use of the MicroFET3 HHD for clinical and research purposes.

\section*{Declaration of interest}

The authors report no conflict of interest.

\section*{References}


10. Bember MG, Grump KJ, Massey BH: Assessment of technical


