The following full text is a publisher's version.

For additional information about this publication click this link.
http://hdl.handle.net/2066/51143

Please be advised that this information was generated on 2018-02-04 and may be subject to change.
VALIDITY AND REPRODUCIBILITY OF HAND-HELD DYNAMOMETRY IN CHILDREN AGED 4–11 YEARS

Willeke A. van den Beld, MD, MSc1,2,3, Gitty A. C. van der Sanden, MSc1, Rob C. A. Sengers, MD, PhD2, André L. M. Verbeek, MD, PhD3 and Fons J. M. Gabreëls, MD, PhD1

From the 1Department of Rehabilitation Medicine and Interdisciplinary Child Neurology Centre 2Department of Paediatrics 3Department of Epidemiology and Biostatistics, University Medical Centre Nijmegen St Radboud, Nijmegen, The Netherlands

Objective: To evaluate validity and reproducibility of hand-held dynamometry in 11 different muscle groups in children.

Design and patients: Maximum isometric muscle strength was measured with a calibrated hand-held dynamometer in 61 patients aged 4–11 years who had been referred to our specialist centre in the past 3 years because of suspected myopathy. All the patients had had muscle biopsy.

Methods: Validity was assessed by the power to discriminate between patients with and without myopathy, using logistic regression analysis and receiver operating characteristic analysis and specificity at a specifically chosen cut-off point. Reproducibility was evaluated by test-retest reliability in a stratified random sample of 40 patients who returned for re-measurements, using the intraclass correlation coefficient and the standard error of measurement.

Results: In the patients, areas under the receiver operating characteristic curve ranged from 0.66 to 0.88. At a specifically chosen cut-off point, sensitivity varied from 73% to 87%, while specificity varied from 54% to 80%. Intraclass correlation coefficients ranged from 0.73 to 0.91. The standard error of measurement ranged from 3.3 N to 12.2 N.

Conclusion: Performance of hand-held dynamometry varied widely in the 11 muscle groups. Highest performance was observed in the elbow flexors. Test-retest reliability of the mean value of 2 efforts was generally higher than the maximum value.

Key words: hand-held dynamometry, muscle, strength, myopathy, children.

J Rehabil Med 2006; 38: 57–64

Correspondence address: Willeke A. van den Beld, Department of Rehabilitation Medicine (HP 720), University Medical Centre Nijmegen St Radboud, PO Box 9101, 6500 HB Nijmegen, The Netherlands. E-mail: w.vandenbeld@reval.umcn.nl

Submitted February 23, 2005; accepted June 1, 2005

INTRODUCTION

Hand-held dynamometry is frequently used for muscle strength testing in children. However, most validity and reproducibility studies on hand-held dynamometry have been carried out in adults. Too often, adult data, procedures and equipment have been used on children of various ages, with minimal consideration for the differences between children and adults (1). A few reports have been published on paediatric reference values for muscle strength measured by hand-held dynamometry in normal children (2, 3). Unfortunately, these reference values were corrected for only 1 or 2 determinants at a time (for example age, in age groups), whereas isometric muscle strength depends not only on age, but also on other determinants, such as gender and body size (4–6). Reference values corrected for all these determinants (regression prediction equations) are only available for adults (7–10). Regression prediction equations are needed for comparison purposes to interpret isometric muscle strength data obtained from individual children.

Validity and reproducibility are qualities that describe the performance of a test instrument (11, 12). Reproducibility assesses the instrument’s capacity to obtain the same results with repeated measurements (precision). Validity assesses how well an instrument measures what it is intended to measure (accuracy), in the case of hand-held dynamometry: muscle strength. Criterion validity is the instrument’s capacity to predict a particular characteristic associated with the measure. Ideally, criterion validity should be established by comparing the measurement to a golden standard. As loss of muscle strength is a general feature of myopathy, criterion validity of hand-held dynamometry can be assessed by its power to discriminate between patients with and without myopathy. In this study, each patient’s diagnosis on the basis of muscle biopsy and the medical records served as the golden standard (myopathy/no myopathy at the time of hand-held dynamometry).

When studying the performance of hand-held dynamometry in children, it is necessary to take 2 important epidemiological and statistical concepts into consideration. First, until now, criterion validity of hand-held dynamometry has only been investigated by comparing 2 “extreme” groups of highly selected patients, i.e. children with severe myopathy (severe loss of muscle strength) and healthy children (13–15). However, it is easy to discriminate between these 2 groups, because there are extreme differences in muscle strength between the 2 groups.
and the groups also differ on other aspects (e.g. healthy vs ill) that can influence the measurements. Selection of 2 extreme groups will automatically lead to artificially favourable results. Moreover, to investigate whether hand-held dynamometry can be used as a diagnostic tool to exclude the diagnosis of myopathy to spare more children from painful muscle biopsy, a suspected myopathy population will be most appropriate (16, 17).

Second, reference values are needed to correct the patients’ muscle strength outcomes for age, gender and body size. This is necessary for the above-mentioned clinical interpretation of isometric muscle strength data obtained from individual children and to evaluate validity and reproducibility of hand-held dynamometry. In the evaluation of validity, corrected outcomes make it possible to compare the patients with myopathy to the patients without myopathy, although they differ, for instance, in age: a certain outcome can be good for a 4-year-old patient, but very poor for an 11-year-old patient. In the evaluation of reproducibility, strong correlations between the test/retest and age (gender, body size) can make the results seem more favourable (18). Especially in children aged 4–11 years, in whom muscle strength is very much age and height dependent, the correlations are high (muscle strength is quite different in a 4-year-old compared to an 11-year-old). As, for example, age and height have different effects on reproducibility in each muscle group, reproducibility of the corrected outcomes can be used to compare the muscle groups.

Summarizing, there is a lack of adequate studies on the performance of hand-held dynamometry in children. The aim of the present study was to assess validity and reproducibility of hand-held dynamometry in children aged 4–11 years. To avoid the selection of 2 extreme groups, we investigated a population of patients suspected of having myopathy. Reference values obtained from healthy children were used to correct the patients’ outcomes for age, gender and body size, to avoid any misleading influences from other variables.

**METHODS**

**Subjects**

The staff and children at a primary school were approached and invited to participate in the study. A stratified random sample of 64 healthy primary-school children aged 4–11 years were tested to obtain regression prediction equations. Only children aged 4 years or older were included to ensure that the subjects could follow instructions regarding muscle contraction. The children had no history of medical or neurological problems that could affect muscle strength.

Patients with suspected myopathy were selected if they were aged 4–11 years and had been referred to our specialist centre in the past 3 years. We excluded any patients who were unable to co-operate or could not be motivated because of mental retardation, which was determined at the time of (intended) measurement by the investigators. A total of 68 patients were selected. They were suspected of having myopathy on the basis of their medical history and physical examination. All the patients had undergone muscle biopsy to determine whether muscular disease was present (to distinguish myopathic disease from other paediatric diseases) and to determine the type of muscular disease (if any).

The study was approved by the Research Ethics Committee of the University Medical Centre Nijmegen. Parents gave informed consent for their child to participate. Height (metres) and weight (kg) of each child were measured with bare feet. All measurements were performed by the same investigator. The investigators were blinded against the true diagnosis and clinical course of the patients at the time of testing.

To assess reproducibility, a stratified random sample of 40 patients was asked to return for re-measurements. The potential for recall bias in the outcomes of hand-held dynamometry was minimized by ensuring a delay of at least 2 weeks between measurement and re-measurement. In such a short interval, it was unlikely that muscle strength would have changed to any substantial extent due to disease progression.

**Hand-held dynamometry**

Maximum isometric muscle strength was measured with the Micro-FET2, a calibrated digital hand-held dynamometer (Hoggan Health Industries Inc., USA). This hand-held dynamometer displays force measurements to a maximum of 440 Newton (N). Eleven different muscle groups were tested bilaterally according to the order and procedures described in Table I, derived from Bohannon (19). Van der

**Table I. Standard protocol for testing muscle group**

<table>
<thead>
<tr>
<th>Muscle group</th>
<th>Position</th>
<th>Limb/joint positions</th>
<th>Dynamometer placement</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Elbow flexors</td>
<td>Supine</td>
<td>Shoulder 30° abducted, elbow 90° flexed, forearm supinated</td>
<td>Just proximal to wrist on flexor surface of forearm</td>
</tr>
<tr>
<td>2. Elbow extensors</td>
<td>Supine</td>
<td>As for elbow flexors</td>
<td>Just proximal to wrist on extensor surface of forearm</td>
</tr>
<tr>
<td>3. Shoulder extensors</td>
<td>Supine</td>
<td>Shoulder 90° anteflexed, elbow extended</td>
<td>Just proximal to elbow on extensor surface of arm</td>
</tr>
<tr>
<td>4. Shoulder abductors</td>
<td>Supine</td>
<td>Shoulder 45° abducted, elbow extended</td>
<td>Just proximal to lateral epicondyle of humerus</td>
</tr>
<tr>
<td>5. Wrist extensors</td>
<td>Sitting</td>
<td>Elbow 90° flexed, forearm supported and pronated, wrist in neutral position, fingers flexed</td>
<td>Just proximal to third metacarpal head</td>
</tr>
<tr>
<td>6. Hip flexors</td>
<td>Supine</td>
<td>Hip 90° flexed, knee relaxed, ankle supported by investigator</td>
<td>Just proximal to knee on anterior surface of thigh</td>
</tr>
<tr>
<td>7. Hip extensors</td>
<td>Supine</td>
<td>Hip 90° flexed, knee relaxed</td>
<td>Just proximal to knee on posterior surface of thigh</td>
</tr>
<tr>
<td>8. Hip abductors</td>
<td>Supine</td>
<td>Hips 45° flexed, knees 90° flexed, contralateral knee supported by chest of investigator</td>
<td>Lateral epicondyle of knee</td>
</tr>
<tr>
<td>9. Knee flexors</td>
<td>Sitting</td>
<td>Knee 90° flexed</td>
<td>Just proximal to ankle on posterior surface of leg</td>
</tr>
<tr>
<td>10. Knee extensors</td>
<td>Sitting</td>
<td>Knee 90° flexed</td>
<td>Just proximal to ankle on anterior surface of leg</td>
</tr>
<tr>
<td>11. Foot dorsiflexors</td>
<td>Sitting</td>
<td>Knee 90° flexed, foot in neutral position</td>
<td>Just proximal to metatarsophalangeal joints on dorsal surface of foot</td>
</tr>
</tbody>
</table>
Ploeg et al. (20) and Bäckman et al. (2). Most comfortable and stable sometimes the mean value was used and sometimes the maximum value. Measuring error, in comparison with the use of a single measurement analyses, as this is an effective and important method to reduce error. The mean of the 2 measurements (efforts) on each side was used in the test. 2 efforts were measured on each side. Hand dominance was determined by asking the child to write down his or her name, or to draw a picture. Lower-limb dominance was determined based on hand dominance and confirmed by observing which lower limb was preferred for hopping (22). The dominant limb was tested first. Hip abductors were tested on 1 side only, because in the position hand dominance and confirmed by observing which lower limb non-dominant sides, we tested 64 healthy primary-school children aged 4 – 11 years. Regression prediction equations were constructed on the basis of stepwise multiple regression analyses (12, 24). By inserting the age, gender, height, weight and/or body mass index (weight/height2) of a patient into the equations, their predicted maximum isometric muscle strength could be calculated. The patients’ observed outcomes could then be compared to their predicted outcomes and a value assigned to describe the proportion of deficit (8). The corrected outcomes used in the analyses were calculated as follows:

\[
\text{Corrected outcome} = \frac{\text{Observed outcome} - \text{Predicted outcome}}{\text{Predicted outcome}}
\]

Validity was evaluated using logistic regression analysis (24). Predictions were made of whether a patient had myopathy based on the outcomes of the muscle group (muscle strength). The discriminative power (how well hand-held dynamometry could distinguish between patients with and without myopathy) was assessed by drawing a receiver operating characteristic (ROC) curve for each muscle group. Sensitivity (Se, true-positive fraction) at every possible cut-off point was plotted against the false-positive fractions or 100% minus specificity (Sp) (25, 26). With the trapezium method, the area under the curve (AUC) could be calculated as a measure of the discriminative power. A non-discriminative test will have an ROC curve that coincides with the diagonal and an AUC of 0.5, whereas a perfect test will have an ROC curve in the upper-left hand corner of the diagram and an AUC of 1.0. A combination of Se and Sp at a specifically chosen cut-off point was used to assess the predictive validity of hand-held dynamometry for myopathy. If hand-held dynamometry is used for diagnostic purposes to spare more children from painful muscle biopsy, it must be high, so that as few myopathy patients as possible are missed. Therefore, Se is most important at this specifically chosen cut-off point. For the same reason, Sp is also displayed at the Se = 100% cut-off point.

Reproducibility was assessed by evaluating the test-retest reliability of the measurements and re-measurements. Test-retest reliability is an instrument’s capacity to obtain the same measurement values on different occasions (12). Test-retest reliability was estimated by the intraclass correlation coefficient (ICC [21]) (23, 27 – 31). The ICC was used as a measure of agreement between the test and retest values obtained from each individual (class). It can range from 0 to 1.00; an ICC of zero means there is no agreement between the test and retest, whereas an ICC of 1.0 means perfect test-retest reliability. The ICCs of the observed outcomes and the ICCs of the corrected outcomes were calculated. In addition, the standard error of measurement (SEM) was calculated in each muscle group (32 – 36). Statistical data analysis was performed using the SAS package (37).

RESULTS

In the group of 68 patients, 3 could not be motivated, 2 were afraid of the hand-held dynamometer and 2 were unable to perform the movements in a correct and standardized way. Thus, 61 patients remained for testing and analysis. The population comprised 40 boys and 21 girls. The age range was 4 – 11 years in boys as well as in girls. Mean age was 7 years 10 months (SD 2 years 3 months) in boys and 7 years 5 months (SD 2 years) in girls. All the patients had had muscle biopsy. Most of the 61 patients had undergone needle biopsy (a few had undergone open biopsy). The myopathies in the patient group included different diagnoses, for example congenital myopathies, muscular dystrophies and inflammatory myopathies. Muscle biopsy was abnormal in 18 out of the 61 patients; 3 of them had myositis which, according to the medical records, had cleared up by the time hand-held dynamometry was performed. Thus, in our analyses, 15 patients were considered to have myopathy and 46 were considered myopathy-free.

In the healthy primary-school children, stepwise multiple regression analyses showed that age, height, gender and body mass index contributed to predicting the maximum isometric muscle strength in children. In contrast with the body mass index, weight did not make any independent contribution to the prediction of muscle strength in the separate muscle groups. Height played an important role in many regression prediction equations. Thus height explained a large part of the maximum isometric muscle strength in children aged 4 – 11 years.

The validity results are presented in Table I by means of the AUCs, Se and Sp. AUCs ranged from 0.66 to 0.88. Examples of ROC curves of the elbow flexors, elbow extensors, shoulder abductors and hip flexors (mean of 2 efforts, dominant side) are shown in Fig. 1. The AUC of the mean value of 2 efforts was similar to the AUC of the maximum value of 2 efforts (Table II). Therefore, predictive validity was assessed using the mean value. At a specifically chosen cut-off point, Se varied from 73% to 87%, while Sp varied from 54% to 80%. At the Se = 100% cut-off point, Sp varied from 0% to 48%. The ROC curves also display the specifically chosen cut-off points (Fig. 1a – d).

Table III shows an example of the classification of patients according to muscle strength at a specifically chosen cut-off point and the presence of myopathy. The ICCs of test-retest reliability are presented in Table IV: when we used the observed outcomes, the ICCs ranged from 0.83 to 0.95, while with the corrected outcomes, the ICCs ranged from 0.73 to 0.91. The SEMs ranged from 3.3 N to 12.2 N (Table V). There was a tendency for the ICC of the mean value to be higher than the ICC of the maximum value and accordingly, there was a

---

1 The regression prediction equations can be obtained on request from the first author.
tendency for the SEM of the mean value to be lower than the SEM of the maximum value.

Our analyses on predictive validity showed that in all the muscle groups, a reasonably high Se was associated with a sizeable false-positive fraction (Sp ≤ 80%) (Table II). As Se is very important (see Data analysis), we explored cut-off points with higher Se. However, in all the muscle groups, Sp decreased out of all proportion compared to the increase in Se. The cut-off points with an Se of 100% were associated with high false-positive fractions (substantial consequential decrease in Sp, Table II). For example, in the elbow flexors (dominant side), at the Se = 100% cut-off point, Sp was only 39%.

**DISCUSSION**

In this study on hand-held dynamometry, major improvements in methodology comprised a suspected myopathy population, the standardized and blinded measurements, the use of AUCs to analyse discriminative power and the use of cut-off points of muscle strength to obtain predictive information. In addition, we corrected the observed outcomes for age, gender and body size using regression prediction equations calculated on data obtained from 64 healthy primary-school children. If the regression prediction equations had been slightly different, e.g. on the basis of more children, this would not have had any substantial consequences on the results of this study. Regression prediction equations were used for correction purposes, for example to exclude the strong influences of e.g. age and height on test/retest, so that better comparisons could be made of test-retest reliability in the different muscle groups. Thus, for comparison purposes, the ICCs of the corrected outcomes can be used. Table IV confirms the strong influence of e.g. age and height on test-retest reliability in children: the ICCs were much lower after the outcomes had been corrected for these variables. For example, without correction, the ICCs of the elbow flexors (mean of 2 efforts) on the dominant side and non-dominant side were as high as 0.94 and 0.95, whereas after correction, these values were 0.85 and 0.90, respectively. Similarly, without correction, the ICCs of the shoulder abductors were as high as 0.94 and 0.91, whereas after correction, these values were 0.82 and 0.77, respectively. Extremely high ICCs would have also occurred if hand-held dynamometry had been investigated in 2 extreme groups of children, instead of in this suspected myopathy population, because of higher variation in outcomes between 2 extreme groups (38, 39).

The tendency for the ICC of the mean value of 2 efforts to be higher than the ICC of the maximum value and accordingly, the tendency for the SEM of the mean value to be lower than the SEM of the maximum value can be explained by the fact that measurement error was reduced by using the mean of 2 measurements in comparison with using the maximum (a single measurement) (23). Thus, using the maximum value of 2 efforts
was of no advantage to the validity or test-retest reliability of hand-held dynamometry in children. Owing to generally higher test-retest reliability, it appeared to be preferable to use the mean value of 2 efforts.

An AUC of 0.5 (i.e. that reflects a non-discriminating test) on all muscle groups falls outside the 95% confidence interval, which means that each muscle group had discriminative power. However, hand-held dynamometry performance showed wide variation in the 11 different muscle groups and thus helped to indicate which muscle groups can be used most reliably in children aged 4–11 years. Hand-held dynamometry of the elbow flexors showed the highest performance (Tables II and IV, Fig. 1). Relatively low performance of hand-held dynamometry was observed on the shoulder abductors, hip flexors and hip extensors. Some of the children found it difficult to understand the muscle strength measurement of the shoulder abductors and to perform the test in a correct and standardized way. The same applied to the hip flexors and hip extensors, as the children performed the movements in different ways. These issues may have affected validity and reproducibility. Thus, in children aged

Table III. Classification of patients according to muscle strength in the elbow extensors (dominant side) and presence of myopathy

<table>
<thead>
<tr>
<th>Muscle strength</th>
<th>Myopathy</th>
<th>No myopathy</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>12*</td>
<td>10†</td>
<td>22</td>
</tr>
<tr>
<td>Negative</td>
<td>3‡</td>
<td>36§</td>
<td>39</td>
</tr>
<tr>
<td>Total</td>
<td>15</td>
<td>46</td>
<td>61</td>
</tr>
</tbody>
</table>

*True positive; †False positive; ‡False negative; §True negative. Sensitivity (Se) = percentage or proportion of patients with myopathy who had a positive muscle strength outcome for myopathy = 12/15 * 100% = 80.0%.
Specificity (Sp) = percentage or proportion of patients without myopathy who had a negative muscle strength outcome for myopathy = 36/46 * 100% = 78.3%.
False-negative rate (100% - Se) = 3/15 * 100% = 20.0%.
False-positive rate (100% - Sp) = 10/46 * 100% = 21.7%.
Positive predictive value (+PV) = probability that myopathy was present in a patient with a positive muscle strength outcome = 12/22 * 100% = 54.5%.
Negative predictive value (-PV) = probability that myopathy was not present in a patient with a negative muscle strength outcome = 36/39 * 100% = 92.3%.
4–11 years, hand-held dynamometry was more reliable in certain muscle groups than in others. In the diagnosis of paediatric myopathies, technical examinations (biochemical, electrophysiological, DNA examinations and imaging techniques) show normal results in a substantial proportion of patients with myopathy (false-negative fraction). Therefore, a painful muscle biopsy is necessary in children with suspected myopathy to diagnose or exclude myopathy (40). In our study population, all the patients were suspected of having myopathy, so the predictive validity of hand-held dynamometry could be assessed (16). Although some muscle groups showed better discriminative power than others, the predictive validity of hand-held dynamometry was not sufficient to serve as an indicator on which to base the decision of whether or not to...
perform muscle biopsy (diagnostic tool). There were no cut-off points in any of the muscle groups with a high Se as well as an acceptable Sp. Thus, in our opinion, negative muscle strength outcomes for myopathy (according to a specific cut-off point) are unable to conclusively exclude the diagnosis of myopathy and spare more children from muscle biopsy.

Test-retest reliability of a test is very important, not only in relation with validity (a single measurement with poor test-retest reliability has low validity (11)), but also because hand-held dynamometry is used to monitor muscle strength in children. If, for example, hand-held dynamometry is used as an outcome measure for treatment intervention in clinical trials, it should have good sensitivity to change (responsiveness) in children. However, no results have been published on the sensitivity of hand-held dynamometry to change in children. As high test-retest reliability is crucial for good sensitivity to change, muscle groups with high test-retest reliability should be chosen when using hand-held dynamometry for monitoring purposes.

ACKNOWLEDGEMENTS

We thank Stichting Kinderpostzegels Nederland (Foundation for Children’s Welfare Stamps Netherlands) and Johanna KinderFonds (Johanna Children’s Fund) for their financial support, as well as Vereniging Spierziekten Nederland (Dutch Neuromuscular Diseases Association) for their moral support. Special thanks to all the children, their parents, Mr APM Bouman, head of the Nijmeegse Schoolvereniging School II and all the teachers for their cooperation. Mr H ter Laak is acknowledged for the muscle biopsy assessments. We thank Mr T Feuth and Mr MA van’t Hof for their help with the statistical analyses. We also thank Mr AJM van den Beld for his constructive comments on the manuscript.

Financial support: Stichting Kinderpostzegels Nederland (Foundation for Children’s Welfare Stamps Netherlands) and Johanna KinderFonds (Johanna Children’s Fund).

REFERENCES


