ORIGINAL REPORT

RELIABILITY AND VALIDITY OF THE MEDICAL RESEARCH COUNCIL (MRC) SCALE AND A MODIFIED SCALE FOR TESTING MUSCLE STRENGTH IN PATIENTS WITH RADIAL PALSY

Tatjana Paternostro-Sluga, MD, PhD[1], Martina Grim-Stieger, MD[1], Martin Posch, MD, PhD[2], Othmar Schuhfried, MD[1], Gerda Vacariu, MD[1], Christian Mittermaier, MD[1], Christian Bittner, MD[1] and Veronika Fialka-Moser, MD, PhD[1]

From the [1]Department of Physical Medicine and Rehabilitation and [2]Department of Medical Statistics, Medical University of Vienna, Vienna, Austria

Objective: To assess the inter-rater and intra-rater reliability and validity of the original and a modified Medical Research Council scale for testing muscle strength in radial palsy.
Design: Prospective, randomized validation study
Patients: Thirty-one patients with peripheral paresis of radial innervated forearm muscles were included.
Methods: Wrist extension, finger extension and grip strength were evaluated by manual muscle testing. Dynamometric measurement of grip strength was performed. Pair-wise weighted kappa coefficients were calculated to determine inter-rater and intra-rater reliability. The 2 scores were compared using the signed-rank test. Spearman's correlation coefficients of the maximal relative force measurements with the median (over-raters) Medical Research Council and modified Medical Research Council scores were calculated to determine validity.
Results: Inter-rater agreement of the Medical Research Council scale (finger extension: 0.77; wrist extension: 0.78; grip strength: 0.78) and the modified Medical Research Council scale (finger extension: 0.81; wrist extension: 0.78; grip strength: 0.81) as well as intra-rater agreement of the Medical Research Council scale (finger extension: 0.86; wrist extension: 0.82; grip strength: 0.84) and the modified Medical Research Council scale (finger extension: 0.84, wrist extension: 0.81; grip strength: 0.88) showed almost perfect agreement. Spearman's correlation coefficients of the maximal relative force measurements with the median Medical Research Council and modified Medical Research Council score were both 0.78.
Conclusion: Medical Research Council and modified Medical Research Council scales are measurements with substantial inter-rater and intra-rater reliability in evaluating forearm muscles.

Key words: manual muscle strength testing, Medical Research Council scale, peripheral nerve lesion, radial palsy.

J Rehabil Med 2008; 40: 665–671

Correspondence address: Tatjana Paternostro-Sluga, Department of Physical Medicine and Rehabilitation, Medical University of Vienna/Austria, Waehringer Guertel 18-20, AT-1090 Vienna, Austria. E-mail: tatjana.paternostro-sluga@akhwien.at

INTRODUCTION

For the assessment of muscle strength, quantitative methods using dynamometers (1) and more qualitative methods of manual muscle testing (MMT) are available. Dynamometric testing is not suitable for weak muscles when movement against resistance cannot be performed, as often occurs in the case of peripheral nerve lesions. This is the critical phase of nerve regeneration, when it is not known whether sufficient regeneration will occur. Nerve surgery may be indicated and the decision for or against nerve surgery depends on the clinical course of the disease. Assessment is therefore very important and MMT is the only applicable strength measurement in peripheral nerve lesions with high-grade paresis.

MMT was developed by Lovett and described by Wright in 1912 (2). This technique has been revised, advanced and promoted so that it has resulted in a range of methods from which the investigator may select the most suitable one (3). The scale proposed by the Medical Research Council (MRC) uses the numeral grades 0–5 (4). Kendall & McCreary (5) use percentages, and Daniels & Worthingham (6) use differentiation between Normal, Good, Fair, Poor, Trace and Zero.

The MRC scale is widely accepted and frequently used. Nevertheless, little is known about its reliability and validity in peripheral nerve lesions. Therefore, a major concern of this study was to examine the inter-rater and intra-rater reliability of the MRC scale in patients with peripheral nerve lesions.

Moreover, the MRC scale neither considers the range of motion (ROM) for which a movement can be performed nor defines the strength of resistance against which a movement can be performed (7). These aspects are particularly relevant for grades 3 and 4. Grade 3 of the MRC scale indicates that active movement against gravity is possible; grade 4 denotes that active movement against resistance is possible. To resolve this problem, the guidelines (4) recommend the use of plus and minus subdivisions within grade 4. Grade 4 is subdivided into 3 categories: slight, moderate and strong resistance (8). The problem with this subdivision is that the quantification of resistance is descriptive and that the meaning of "low", "moderate" and "strong" is unclear. The different levels of resistance are highly rater-dependent. Therefore, the modification of resist-

ance for subdivisions of the scale is not an optimum solution. Moreover, no subdivision is provided for grade 3.

In order to obtain a more specific clinical picture of a peripheral nerve lesion and its course of motor recovery, a modified MRC (mMRC) scale including ROM was defined. ROM was chosen for the subdivision because this parameter can be quantified more easily than resistance, even in clinical routine.

The aim of the present study was to investigate the inter-rater and intra-rater agreement and validity of the original and mMRC scales for assessment of muscular weakness due to peripheral paresis of radial innervated forearm muscles.

METHODS

Examiners

The study was approved by the local ethics committee and was performed at the department of physical medicine and rehabilitation at the General Hospital, Medical University of Vienna, Vienna, Austria. The 5 examiners were specialists in physical medicine and rehabilitation with 4–10 years of experience in the assessment of muscle strength. The sequence of the examiners was randomized.

Patients

Inclusion criteria. Muscular weakness of more than 3 months' duration in the radial innervated forearm muscles, caused by a peripheral lesion of the radial nerve, a radicular lesion C7 or a lesion of the brachial plexus involving the C7 fibres. In the case of brachial plexus lesions patients could have additional paresis of the median and ulnar innervated muscles of the hand and arm.

Exclusion criteria. ROM less than 40° for the tested movements caused by a contracture of an involved joint or by shrinkage of soft tissue due to scars. Other exclusion criteria were progression of the lesion, and systemic disease of the peripheral nervous system or the central nervous system.

Original and modified MRC scale

The original MRC scale is shown in Table I. The mMRC scale (Table II) was designed as follows: grades 0, 1, 2 and 5 of the mMRC scale are in conformity with the original MRC scale; and grades 3 and 4 are modified by including the active ROM in the grading system.

ROM was measured visually.

Procedure

The strength of wrist extension (extensor carpi ulnaris and radialis muscles), extrinsic finger extension (extensor digitorum muscle) and grip (flexor digitorum superficialis and profundus muscle, intrinsic hand muscles) were evaluated by MMT, graded by the original MRC and mMRC scale. A quantitative muscle testing of grip strength was performed using the Jamar dynamometer (Jamar TEC, Clifton, USA) (9).

Three measurements were taken from the affected and the healthy hand. The testing procedure of inter-rater reliability included 15 min rest between the assessments of the different raters to avoid muscle fatigue.

Table I. *Medical Research Council scale. Aids to examination of the peripheral nervous system. Memorandum no. 45. London: Her Majesty's Stationery Office; 1976*

0	No contraction
1	Flicker or trace contraction
2	Active movement, with gravity eliminated
3	Active movement against gravity
4	Active movement against gravity and resistance
5	Normal power

Table II. *Medical Research Council scale modified according to Paternostro-Sluga et al.*

0	No contraction
1	Flicker or trace contraction
2	Active movement, with gravity eliminated
2–3	Active movement against gravity over less than 50% of the feasible ROM
3	Active movement against gravity over more than 50% of the feasible ROM
3–4	Active movement against resistance over less than 50% of the feasible ROM
4	Active movement against resistance over more than 50% of the feasible ROM
4–5	Active movement against strong resistance over the feasible ROM, but distinctly weaker than the contralateral side
5	Normal power

ROM: range of motion.

All positions and procedures for testing were standardized, strictly defined, and in accordance with the recommendations of the MRC (8).

As a first step, a pilot study comprising 5 patients was performed. The results were used to discuss the problems of clinical testing that arose during the assessment of muscle strength and to estimate the required sample size. Thereafter, the process of clinical strength testing was defined in greater detail and the examiners trained together twice. The pilot patients were not included in the study. Based on the observed standard errors of the pair-wise weighted kappa values in the pilot study, a sample size of at least 30 patients was deemed necessary to achieve weighted kappa estimates with a standard error of less than 0.025 (10).

MMT with the modified MRC scale according to Paternostro-Sluga et al.

Wrist extension, extrinsic finger extension and grip strength were tested. First the feasible passive ROM was evaluated by visual measurement. Movement against gravity was then tested. For this purpose the patient's forearm was pronated. If the movement against gravity amounted to more than 50% of the feasible passive ROM the patient was graded as at least a force grade 3. If an active movement was possible but was less than 50% of the feasible passive ROM the force grade was 2–3.

If movement against gravity was not possible, the forearm was brought into a neutral position between supination and pronation and the wrist into a 0° position. The patient was then asked to perform the movement. The examiner palpated the muscle. If there was no contraction, muscle strength was graded 0. If a contraction was perceived it was graded 1. If a movement could be performed for more than 5° it was graded 2, which meant active movement with gravity eliminated.

If movement against gravity was possible over more than 50% of the feasible passive ROM, testing against resistance followed. The patient's forearm was in pronation. If the movement against resistance could be performed over less than 50% of the feasible passive ROM, muscle strength was graded as grade 3–4. If it was possible to move over more than 50% of the ROM, it was graded 4.

Movement against strong resistance over the entire ROM, but weaker than the contralateral side, was classified as grade 4–5.

The same resistance as that on the contralateral side was rated grade 5.

For grades 0–2, when movement against resistance was not possible, the forearm was held in a neutral position between pronation and supination and the wrist in a 0° position, which was assisted by the examiner.

For testing the other grades the forearm was held in pronation.

Three assessments of each tested movement were made and the best performance was determined. The testing procedure included 15 minutes' rest between the assessments of the different raters in order to avoid muscle fatigue.

Dynamometric measurement

For dynamometric assessment of grip strength a Jamar dynamometer (Jamar TEC, Clifton, USA) (9) was used. The forearm was held in a neutral position between pronation and supination. Three trials were performed and the best trial was used for the evaluation. The dynamometric assessment was perfomed once for all patients and twice within one week for 22 patients, combined with the intra-rater testing.

Inter-rater reliability

To test inter-rater reliability, 5 examiners assessed 31 patients. Patients were permitted a 15-min rest between the 5 evaluations.

Intra-rater reliability

To test intra-rater reliability, one examiner tested 22 patients twice. The median time between the ratings was 7 days.

Validity

To obtain information about the validity of the MMT for grip strength, each patient was measured by MMT as well as a Jamar dynamometer (Jamar TEC, Clifton, USA).

Statistical analysis

For *inter-rater reliability* the pair-wise weighted kappa coefficients for all 5 raters were computed and averaged. To account for the finer classification of the mMRC scale, we assigned scores (0, 1, 2, 3, 4, 5 for the MRC scale and 0, 1, 2, 2.5, 3, 3.5, 4, 4.5, 5 for the mMRC scale) and quantified the amount of disagreement by the difference in the scores. Consequently, a disagreement between, for example, the categories 2 and 3 is weighted equally for both scales. Based on these scores, Cicchetti-Allison kappa coefficient weights were used in the calculation of the weighted kappas (11, p. 554). Kappa values from 0.61 to 0.80 were considered as substantial agreement, kappa values above 0.8 as almost perfect agreement (12).

As an additional measure of agreement, for each proband and scale the maximal deviation of the ratings (maximum score – minimum score) was calculated. A maximal deviation of 0 indicates that the raters agreed perfectly; a maximal deviation of 1 indicates that the lowest rating differs from the highest rating by 1. The resulting (paired) maximal deviations for the MRC and the mMRC scale are compared with the signed-rank test to assess whether the agreement of raters differs between the 2 scales. Additionally, the distribution of the maximal deviations is tabulated. Moreover, pair-wise kappa values for the subsets of patients where the median ratings of the unmodified MRC scores was larger than 0 and lower than 5 were calculated.

For *intra-rater reliability*, pair-wise weighted kappa coefficients for the 2 measurements of one examiner were computed.

From the 3 dynamometric measurements taken at both ratings, first the relative force for each was assessed. This was defined as the ratio between the values for the affected hand and the healthy hand. For each rating the maximum of the 3 ratios was calculated. For these maxima, a variance component analysis (the SAS procedure variance component with the restricted maximum likelihood option) with the independent factors rating (1, 2) and proband was performed. Then the intraclass correlation coefficients defined as (variance between probands) / (sum of all variance components) was calculated. Additionally, the differences in relative force measurements between the first and the second rating were assessed and tested with paired *t*-tests for significant trends.

To obtain information about *validity*, Spearman's correlation coefficient of the maximum of the 3 relative force measurements and the median MRC and the mMRC score over-raters were calculated. For all tests, a 2-sided significance level of 5% was used. Analysis was performed using the statistical software SAS Release 8.2 (SAS Institute, Cary, NC, USA).

RESULTS

Patients

Thirty-one patients with peripheral paresis of the radial innervated forearm muscles were included in the study (16 men and 15 women). The subjects' mean age was 45 years (range 22–84 years), mean height 171 cm (range 155–192 cm) and mean weight 73 kg (range 51–100 kg). The left hand was affected in 13 patients and the right hand in 18. Nineteen patients had a radial nerve lesion, 11 had a lesion of the brachial plexus involving C7 fibres and 1 patient had a radicular lesion C7. The grades of muscle strength rated by the most experienced examiner according to the mMRC scale are shown in Fig. 1.

Inter-rater agreement

As an example, Table III shows the agreement of raters one and 2 for wrist extension for the modified score. For more than half of the patients the ratings agree perfectly. In 1 patient the difference in ratings (quantified by the distance in scores) disagrees by more than one.

Concerning the inter-rater agreement of the original MRC scale as well as the mMRC scale, the average weighted pair-
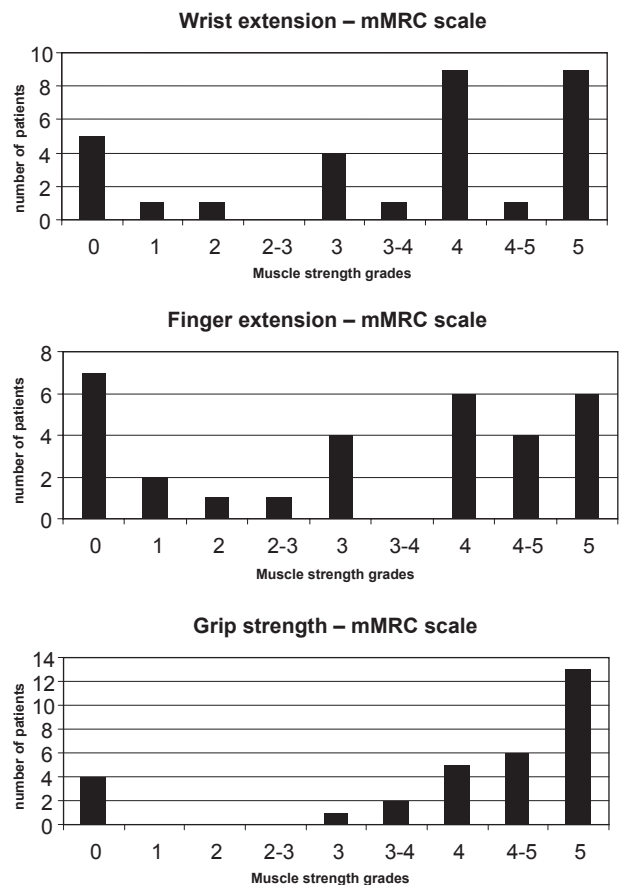

Fig. 1. Grades of muscle strength for all subjects, rated by the most experienced examiner according to the modified Medical Research Council (mMRC) scale for: (a) wrist extension; (b) finger extension; and (c) grip strength.

Table III. *Frequency table of agreements of raters 1 and 2 for wrist extension. The shaded fields indicate perfect agreement. For example, value 2 in row "4", column "3–4" indicates that 2 patients were given a rating of 4 by rater 1 but a rating of 3–4 by rater 2.*

	Rater 2								
Rater 1	0	1	2	2–3	3	3–4	4	4–5	5
0	3	1	1						
1			1						
2		1							
2–3									
3				1	2	1			
3–4						1			
4				1	2	4	1	1	
4–5								1	
5							2	7	

wise kappas showed substantial agreement for all tested muscles (average weighted pair-wise kappas: MRC scale: wrist extension 0.78, finger extension 0.77, grip strength 0.78; mMRC scale: wrist extension 0.78, finger extension 0.81, grip strength 0.81, see Table IV (a and b). The asymptotic standard error estimates for the pair-wise weighted kappas ranged from 0.03 to 0.18 over all scores.

The maximal inter-rater deviations of the MRC and mMRC ratings for each patient are shown in Table V. None of the differences in the inter-rater deviations between the MRC and the mMRC score were significant (2-sided signed-rank test, all $p > 0.05$).

Table IVa. *The weighted pair-wise kappa coefficients for all 5 raters and the averaged kappa coefficients over raters for the assessment of inter-rater agreement for wrist extension with the modified Medical Research Council (mMRC) scale. The asymptotic standard error estimates for the pair-wise weighted kappas ranged from 0.05 to 0.10 over all scores*

	mMRC scale				
Rater	1	2	3	4	5
1		0.75	0.81	0.77	0.69
2	0.75		0.79	0.79	0.78
3	0.81	0.79		0.83	0.89
4	0.77	0.79	0.83		0.70
5	0.69	0.78	0.89	0.70	
Average	0.76	0.78	0.83	0.77	0.77

Table IVb. *The average weighted pair-wise kappa coefficients averaged over all pairs of raters. The asymptotic standard error estimates for the pair-wise weighted kappas ranged from 0.03 to 0.18 over all scores*

	MRC scale		mMRC scale	
	Mean		Mean	
Type	kappa	Min–Max	kappa	Min–Max
Wrist extension	0.78	0.67–0.90	0.78	0.69–0.89
Finger extension	0.77	0.64–0.93	0.81	0.72–0.92
Grip	0.78	0.64–0.88	0.81	0.74–0.86

Min: minimum; Max: maximum; MRC: Medical Research Council; mMRC: modified Medical Research Council.

Table V. *Frequencies of the maximal inter-rater deviations for the Medical Research Council (MRC) and the modified MRC (mMRC) scale for wrist extension, finger extension and grip strength. Where the modified scale has additional categories, the distance between adjacent categories was set to 0.5*

	Wrist extension		Finger extension		Grip strength	
Deviation	MRC	mMRC	MRC	mMRC	MRC	mMRC
0	13	8	10	6	17	14
0.5		10		12		10
1	16	9	19	10	14	5
1.5		3		1		2
≥2	2	1	2	2	0	0

If patients with grade 0 and grade 5 are omitted, the average kappa values decrease substantially (MRC scale: wrist extension 0.62, finger extension 0.50, grip strength 0.26; modified MRC scale: wrist extension 0.61, finger extension 0.61, grip strength 0.42)

Intra-rater agreement

Concerning the intra-rater agreement the results were also notable: the weighted kappa coefficients of the original MRC as well as the mMRC scale were all above 0.8 and thus indicated nearly perfect agreement (MRC scale: wrist extension 0.82, finger extension 0.86, grip strength 0.84; mMRC scale: wrist extension 0.81, finger extension 0.84, grip strength 0.88). The asymptotic standard errors of the kappa values did not exceed 0.12 for the MRC and 0.08 for the mMRC.

The frequency of intra-rater agreements for grades 4, 4–5 and 5 are shown in Table VI.

The maximum of the 3 relative force measurements with the dynamometer measured at the 2 time-points resulted in a very high intraclass correlation coefficient of 0.98.

None of the differences in the maximum of the second and the maximum of the first rating were significantly different from zero.

Dynamometer measurements

Grip strength was rated grade 5 in 20 patients. For these patients, the maximum muscle strength in the affected hand (over the 3 short-term repetitions) was 28.58 (standard deviation (SD) 19.96) kg and ranged from 7.26 to 77.11 kg. In the healthy hand the maximum muscle strength was 46.27 (SD 14.51) kg and ranged from 18.14 to 74.84 kg. For these patients the median ratio between the affected hand and the healthy hand was 0.65, which indicates that the affected hand had 65% of the muscle strength of the healthy hand. Seventy-five percent of these patients had a force ratio between 33% and 95%.

The median value of 6 patients who were assigned grade 4 was 0.12, which indicates that the affected hand had 12% of the muscle strength of the healthy hand. 75% of these patients had a force ratio between 5% and 21%.

Four patients with grade 0 and the patient with grade 3 had a force measurement of 0 (see Fig. 2).

Table VI. *Frequency table of intra-rater agreements of all patients that had a score of 4 or more at the first measurement in the Medical Research Council (MRC) and modified MRC (mMRC). For example, in the wrist mMRC sub-table the value 1 in row "4", column "4–5" indicates that 1 patient was rated 4 at the first measurement but 4–5 at the second measurement*

Measurement 1	Measurement 2															
	Wrist					Finger extension						Grip strength				
	MRC		mMRC			MRC		mMRC				MRC		mMRC		
	4	5	4	4–5	5	4	5	3–4	4	4–5	5	4	5	4	4–5	5
4	10	1	7	1	1	7		1	4			9	2	3	1	0
4–5			1						2					1	3	1
5		5			5		5				4		9			9

Validity

Concerning validity, Spearman's correlation coefficient of median grip strength measured by the original MRC scale with the maximal relative force measurements was 0.78.

The correlation of the mMRC scale with the maximal relative force measurements was also 0.78.

DISCUSSION

In the present study, the reliability of the original MRC scale for radial palsy was tested and was shown to be substantially good for wrist and finger extension. This is important as the MRC scale is frequently used in clinical routine as well as scientific studies (13–21).

A weakness of the original MRC scale is that it does not consider clinically relevant changes in the strength range of grade 3 and 4 in the recovery process after lesions of the peripheral nervous system. The original MRC scale does not include the ROM for which a movement can be performed. From the clinical point of view, this is an important parameter to follow the regeneration process. If a patient after, for example, traumatic nerve lesion can move against gravity by 5° and 6 weeks later he can move by 60°, then the examiners know that a further improvement has occurred. If, after 6 weeks, movement against gravity is still only 15°, than there might be an obstacle to the regeneration process. Without the modification to the scale the patient would have been scored grade 3 both times. Moreover, it is assumed that the functional relevance of whether a movement can be executed over 15° or 60° is significant. Therefore, a new mMRC scale was designed as an instrument with more grades, in order to represent better the clinical changes that occur in the motor recovery process after peripheral nerve lesions. It was decided to use sub-divisions based on ROM rather than resistance, because ROM is easier to measure and quantify than resistance.

After defining the mMRC scale according to Paternostro-Sluga et al., its reliability was tested and was shown to be as good as the reliability of the original scale. Moreover, it could be shown that the margin of deviation of the mMRC scale was no worse than the margin of deviation of the original MRC scale.

The reliability and validity of various MMT techniques have been tested in patients with poliomyelitis (22–24) and muscular dystrophy (3, 25–27). Florence et al. (3) tested the intra-rater reliability of a modified MRC scale that differentiated between movement against maximal (grade 4+), moderate (grade 4), minimal (grade 4–) and transient (grade 3) resistance. These grades were less reliable than those given in positions in which the factors of gravity and resistance had been eliminated. Moreover, in that study the intra-rater reliability for distal muscles were not as consistent (e.g. weighted kappa for wrist extensors 0.69) as that for the proximal muscles (e.g. weighted kappa for hip flexors 0.9) (3).

Barr et al. (26) assessed the reliability of an mMRC scale that uses plus and minus sub-divisions (4). Perfect agreement was seen 35.9% of the time, consistency within one consecutive strength grade was found 66.5% of the time, and within 2 consecutive steps 84.7% of the time. The agreement for measures of proximal muscle strength ($r = 0.80$) was found to be more consistent than that for measures of distal muscle strength ($r = 0.58$) (26). Other studies addressed the reliability of a composite score, weighted by a factor that assessed muscle bulk rather than assigning grades to individual muscle groups or individual grades within a particular score (22–24). Some
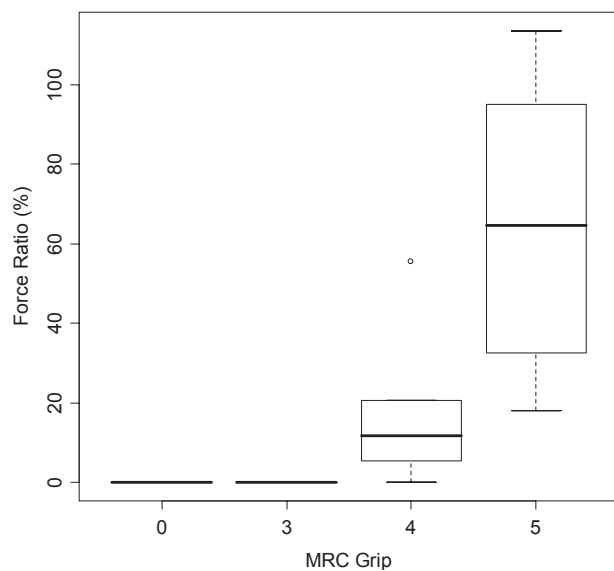

Fig. 2. Dynamometric measurements (Jamar TEC, Clifton, USA) of grip strength are shown as the maximum relative force, defined as the ratio of values for the affected hand and the healthy hand. MRC: Medical Research Council.

studies that addressed inter-rater reliability used a sum score of various muscles rather than analysing reliability for individual muscle groups (27, 28). Escolar et al. (27) determined a sum score of the mMRC scale and compared the reliability of MMT and quantitative muscle testing. MMT was not as reliable and required repeated training of evaluators to bring all groups to a correlation coefficient > 0.75 (27). Kleyweg et al. (28) registered nearly perfect inter-observer agreement of a sum score of various muscles tested with the MRC scale in patients with Guillain-Barré syndrome. The MMT method described by Daniels & Worthingham (6) was shown to be reliable (29, 30). A more recently published study that addressed inter-rater reliability of MMT only differentiated between "normal" or "reduced" power (31), which is too approximate for assessment of motor recovery after peripheral nerve lesion.

Brandsama et al. (32) tested the reliability of the 6-point MRC scale of intrinsic muscles of the hand. They suggested testing specific movements rather than selective muscles because it is difficult to isolate, and hence grade, most of the intrinsic muscles of the hand (32). They also introduced a mMRC scale (33), which includes the description of ROM as well as resistance into the 6-point original MRC scale. In their mMRC scale, grade 3 has to have normal ROM. In our modified scale, grade 3 has to have more than 50% of the feasible ROM and additional 3 grades (grade 2–3, 3–4 and 4–5) were included, which was assumed to represent better the clinical course of nerve regeneration.

The testing procedure of inter-rater reliability included 15 min rest between the assessments of the different raters to avoid muscle fatigue.

All examiners were specialists in physical medicine and rehabilitation, with different levels of experience. Open points concerning the MMT procedure were discussed during the pilot phase and the examiners were trained to carry out the procedure. One of the issues was to improve the clinical differentiation between intrinsic and extrinsic finger extension in the presence of extrinsic finger extension paresis.

The median time between the ratings for testing the intra-rater reliability was one week. This time span was selected because the rater would probably not remember the result of the first rating and the subject's clinical condition would remain largely unchanged. Only patients with chronic paresis were included. Thus, a constant muscle force during the entire examination procedure (one week) can be assumed. A further indicator for constant muscle force, at least for grip strength, was that the maximum of the 3 relative force measurements with the dynamometer measured at the 2 time points resulted in a very high intraclass correlation coefficient of 0.98.

Three assessments of each tested movement were made. The best result was determined in order to allow a learning effect and exclude false low muscle grading due to rapid fatigue in weak muscles.

Patients with paresis of the radial innervated forearm muscles were chosen for examination because wrist extension and extrinsic finger extension are hardly influenced by co-activation of muscles innervated by other nerves. In this context it has to be considered that the effect of gravity on the wrist and fingers

is much less than on the leg or proximal upper limb and this might limit the generalization of the results.

Excluding patients with grade 0 and grade 5 decreases the reliability level. This shows that the assessment of the different grades of paresis is much more difficult than the assessment of a muscle that has no contraction at all or is evaluated as normal. This emphasizes the importance of training within the team.

Grip strength was also measured in order to obtain information about validity. The limitation of the validity testing is the fact that strength of the radial innervated forearm muscles was not directly assessed. Grip strength may be weak in the presence of radial palsy alone, as wrist dorsal extension cannot be performed, which is an important prerequisite for a strong grip. Some patients also had additional paresis of median or ulnar innervated forearm muscles, resulting in reduced grip strength. A high percentage of patients had a strong grip, which may have improved the correlation coefficient.

Hand-held dynamometers are described for wrist extension (34, 35) and for wrist and metacarpophalangeal joint extension (36). These instruments were not available in the present test setting and therefore grip strength was selected as the parameter for validation.

Measurement with the Jamar dynamometer showed a distinct difference between the left and the right sides and a wide variance of measurements for patients assigned MRC grade 5. Clinicians have to be aware that there is a wide range of strength levels summarized under grade 5. It is necessary to differentiate grade 5 dynamometry. Beasley (37) reported in post-polio children that MMT classified as "normal" were those whose knee extension force was only 50% of normal. The overestimation of the extent to which a patient is "normal" by MMT was also described by Bohannon (38). Moreover, it was shown (39) that, by comparing MMT with hand-held dynamometry, strength differences and deficits in strength were missed at least 25% of the time by MMT in acute rehabilitation patients (39).

For MMT with the mMRC scale, the ROM was measured visually. In former studies it was shown that visual and goniometric measurements of the ROM of the knee joint were equally reliable (40). Moreover, visual measurement can be applied easily and rapidly in the clinical setting. The idea was to create a scale that can be used unmodified in everyday clinical practice. However, measurement with a goniometer might have improved reliability.

In conclusion, the MRC as well as the mMRC scale are manual measurements of muscle strength in peripheral nerve palsy of forearm muscles with substantial inter-rater and intra-rater reliability and strong validity and can be recommended for clinical use. To ensure equal test conditions it is advisable to train the evaluators in advance.

REFERENCES

1. Beasley WC. Instrumentation and equipment for quantitative clinical muscle testing. Arch Phys Med Rehabil 1956; 37: 604–621.
2. Wright W. Muscle training in the treatment of infantile paralysis. Boston Med Surg J 1912; 167: 567.

3. Florence JM, Pandya S, King WM, Robison JD, Baty J, Miller JP, et al. Intrarater reliability of manual muscle test (Medical Research Council scale) grades in Duchenne's muscular dystrophy. Phys Ther 1992; 72: 115–122.

4. Medical Research Council. Aids to examination of the peripheral nervous system. Memorandum no. 45. London: Her Majesty's Stationary Office; 1976.

5. Kendall F, McCreary E, editors. Muscle testing and function, 3rd edn. Baltimore, MD: Williams & Wilkins; 1983.

6. Daniels L, Worthingham C, editors. Muscle testing: technique of manual examination, 5th edn. Philadelphia: WB Saunders Co.; 1986.

7. Bohannon RW. Manual muscle testing of the limbs: considerations, limitations, and alternatives. Phys Ther Pract 1992; 2: 11–21.

8. Brain. Aids to the examination of peripheral nervous system, 4th edn. Edinburgh; London; New York; Philadelphia; St:Louis; Sydney; Toronto: WB Saunders Co., 2000.

9. Mathiowetz V, Weber K, Volland G, Kashman N. Reliability and validity of grip and pinch strength evaluations. J Hand Surg [Am] 1984; 9: 222–226.

10. Fleiss JL, Cohen J, Everitt BS. Large-sample standard errors of kappa and weighted kappa. Psychol Bull 1969; 72: 323–327.

11. SAS Institute Inc., SAS procedures guide, vers 8. Cary, NC: SAS Institute Inc; 1999.

12. Landis IR, Koch GG. The measurement of observer agreement for categorial data. SAS/STAT® user's guide, version 8, Cary, NC: SAS Institute Inc.; 1999.

13. Slawek J, Bogucki A, Reclawowicz D. Botulinum toxin type A for upper limb spasticity following stroke: an open-label study with individualised, flexible injection regimens. Neurol Sci 2005; 26: 32–39.

14. Hsu SP, Shih YH, Huang MC, Chuang TY, Huang WC, Wu HM, et al. Repair of multiple cervical root avulsion with sural nerve graft. Injury 2004; 35: 896–907.

15. Van Empelen R, Jennekens-Schinkel A, Buskens E, Helders PJ, van Nieuwenhuizen O; Dutch Collaborative Epilepsy Surgery Programme. Functional consequences of hemispherectomy. Brain 2004; 127: 2071–2079.

16. De Carvalho M, Scotto M, Lopes A, Swash M. Clinical and neurophysiological evaluation of progression in amyotrophic lateral sclerosis. Muscle Nerve 2003; 28: 630–633.

17. Tzvetanov P, Rousseff RT. Median SSEP changes in hemiplegic stroke: long-term predictive values regarding ADL recovery. NeuroRehabil 2003; 18: 317–324.

18. Van den Berg-Vos RM, Franssen H, Wokke JH, Van den Berg LH. Multifocal motor neuropathy: long-term clinical and electrophysiological assessment of intravenous immunoglobulin maintenance treatment. Brain 2002; 125: 1875–1886.

19. Kalita J, Misra UK, Bansal R. Central motor conduction studies in patients with Guillain Barre syndrome. Electromyogr Clin Neurophysiol 2001; 41: 243–246.

20. Vlak M, van der Kooi E, Angelini C. Correlation of clinical function and muscle CT scan images in limb-girdle muscular dystrophy. Neurol Sci 2000; 21: S975–S977.

21. Swash M, Brown MM, Thakkar C. CT muscle imaging and the clinical assessment of neuromuscular disease. Muscle Nerve 1995; 18: 708–714.

22. Smith LK, Iddings DM, Spencer WA, Harrington PR. Muscle testing, part I: description of a numerical index for clinical research. Phys Ther Rev 1961; 41: 99–105.

23. Iddings DM, Smith LK, Spencer WA. Muscle testing. 2. Reliability in clinical use. Phys Ther Rev 1961; 41: 249–256.

24. Lilienfeld AM, Jacobs M, Willis M. A study of the reproducibility of muscle testing and certain other aspects of muscle scoring. Phys Ther Rev 1954; 34: 279–289.

25. Florence JM, Pandya S, King WM, Robison JD, Signore LC, Wentzell M, Province MA. Clinical trials in Duchenne dystrophy. Standardization and reliability of evaluation procedures. Phys Ther 1984; 64: 41–45.

26. Barr AE, Diamond BE, Wade CK, Harashima T, Pecorella WA, Potts CC, et al. Reliability of testing measures in Duchenne or Becker muscular dystrophy. Arch Phys Med Rehabil 1991; 72: 315–319.

27. Escolar DM, Henricson EK, Mayhew J, Florence J, Leshner R, Patel KM, Clemens PR. Clinical evaluator reliability for quantitative and manual muscle testing measures of strength in children. Muscle Nerve 2001; 24: 787–793.

28. Kleyweg RP, van der Meché FG, Schmitz PI. Interobserver agreement in the assessment of muscle strength and functional abilities in Guillain-Barré Syndrome. Muscle Nerve 1991; 14: 1103–1109.

29. Silver M, McElroy A, Morrow L, Heafner BK. Further standardization of manual muscle test for clinical study: applied in chronic renal disease. Phys Ther 1970; 50: 1456–1466.

30. Nitz JC, Burns YR, Jackson RV. A longitudinal physical profile assessment of skeletal muscle manifestation in myotonic dystrophy. Clin Rehabil 1999; 13: 64–73.

31. Jepsen J, Laursen L, Larsen A, Hagert CG. Manual strength testing in 14 upper limb muscles: a study of inter-rater reliability. Acta Orthop Scand 2004; 75: 442–448.

32. Brandsma JW, Schreuders TA. Sensible manual muscle strength testing to evaluate and monitor strength of the intrinsic muscles of the hand: a commentary. J Hand Ther 2001; 14: 273–278.

33. Brandsma JW, Schreuders TA, Birke JA, Piefer A, Oostendorp R. Manual Muscle Strength testing: intraobserver and interobserver reliabilities for the intrinsic muscles of the hand. J Hand Ther 1995; 8: 185–190.

34. Bohannon RW. Test-retest reliability of hand-held dynamometry during a single session of strength assessment. Phys Ther 1986; 66: 206–209.

35. Bohannon RW. Upper extremity strength and strength relationships among young women. J Orthop Sports Phys Ther 1986; 8: 128–133.

36. Richards RR, Gordon R, Beaton D. Measurement of wrist, metacarpophalangeal joint, and thumb extension strength in a normal population. J Hand Surg 1993; 18A: 253–261.

37. Beasley WC. Influence of method on estimates of normal knee extensor force among normal and postpolio children. Phys Ther Rev 1956; 36: 21–41.

38. Bohannon RW. Manual muscle test scores and dynamometer test scores of knee extension strength. Arch Phys Med Rehabil 1986; 67: 390–392.

39. Bohannon RW. Manual muscle testing: does it meet the standards of an adequate screening test? Clin Rehabil 2005; 19: 662–667.

40. Käfer W, Fraitzl CR, Kinkel S, Clessienne CB, Puhl W, Kessler S. Outcome-Messung in der Knieendoprothetik: Ist die klinische Bestimmung der Gelenkbeweglichkeit eine zuverlässig messbare Ergebnisgröße? Z Orthop Ihre Grenzgeb 2005; 143: 25–29 (in German).