Manual strength testing in 14 upper limb muscles

A study of inter-rater reliability

Jørgen R Jepsen[1], Lise H Laursen[1], Anders I Larsen[2] and Carl-Göran Hagert[3]

[1]Department of Occupational Medicine, Central Hospital, DK-6700 Esbjerg, Denmark, [2]Occupational Health Services, Novozymes, DK-2880 Bagsværd, Denmark, [3]Hand and Upper Extremity Surgery, University of Pittsburgh Medical Center, Pittsburgh, PA 15203, USA
Correspondence JRJ: jrj@ribeamt.dk
Submitted 03-02-24. Accepted 03-11-16

Background Manual muscle testing has been termed a "lost art" and is often considered to be of minor value. The aim of this investigation was to study the inter-rater reliability of manual examination of the maximal voluntary strength in a sample of upper limb muscles.

Patients and methods The material consisted of a series of 41 consecutive patients (82 limbs) who had been referred to a clinic of occupational medicine for various reasons. Two examiners who were blinded as to patient-related information classified 14 muscles in terms of normal or reduced strength. In order to optimize the evaluation, the individual strength was assessed simultaneously on the right and left sides with the limbs in standardized positions that were specific for each muscle. Information on upper limb complaints (pain, weakness and/or numbness/tingling) collected by two other examiners resulted in 38 limbs being classified as symptomatic and 44 as asymptomatic. For each muscle the inter-rater reliability of the assessment of strength into normal or reduced was estimated by κ-statistics. In addition, the odds ratio for the relation to symptoms of the definition in agreement of strength was calculated.

Results The median κ-value for strength in the muscles examined was 0.54 (0.25–0.72). With a median odds ratio of 4.0 (2.5–7.7), reduced strength was significantly associated with the presence of symptoms.

Interpretation This study suggests that manual muscle testing in upper limb disorders has diagnostic potential.

■

Reduced muscle strength may accompany several muscular and neuromuscular conditions, whether or not these have definite patterns of involvement. Even so, evaluation of the strength of individual muscles appears to be a less regular part of the physical examination, which is routinely limited to the evaluation of parameters such as range of motion, grip strength and sensory function. Manual muscle testing has even been termed a "lost art" (Kendall 1991) and current opinion seems to be that it is of limited value in upper limb disorders, except where there is a debilitating degree of weakness (Sapega 1990).

Upper limb patients in occupational medicine and other specialties, in particular orthopedic and hand surgery, often have complaints such as pain, weakness, and/or numbness/tingling. We assessed the inter-rater reliability of strength testing in selected individual upper limb muscles. In addition, we investigated whether there was any association between muscular strength and the complaints of patients who had been referred to an occupational medicine clinic.

Patients and methods

Patients

Over a 5-month period, consecutive patients with or without upper limb disorders from the county of Ribe (population approx. 250 000) were referred to the Department of Occupational Medicine, Esbjerg Central Hospital, by general practitioners and specialists. The main aim was to assess whether the disorder might be work-related, and also to assess the consequences of the disorder regarding ability to work. Prior diagnostic difficulties, no response to prior treatment, or a recurrence of symptoms on resuming work were characteristic of most of the referred patients.

Patients were enrolled in the study regardless of the reason for referral, which might be disorders of the upper limb, lower back, lung, etc. Since a maximum of one study patient per day was allowed into the study, this caused 10 patients to be excluded, all of whom were similar to the study patients with respect to disease pattern and severity. In addition, primary examiners who were not engaged in the muscle strength tests excluded 17 patients in order to assure instructions and blinding during the subsequent physical examination (foreign language patients, patients who had earlier contact with the department, or presenting visible signs suggestive of assignment to the symptomatic or the asymptomatic group, e.g., scars from prior upper limb surgery, or indications of easily recognizable disease such as an antalgic position).

The final material consisted of a random sample of 41 participants (22 men) with a median age of 44 (25–61) years. Based on presuppositions regarding the distribution of reduced strength and symptoms, this sample size was determined to be adequate to ensure that the statistical calculations were of sufficient power. The study complied with the Helsinki declaration. It was approved by the local Ethics Committee and signed informed consent was obtained from all participants.

Muscles examined

14 upper limb muscles were selected for evaluation of individual strength, and were divided into three groups according to the standard postures of the upper limb during the examination:

1. Posterior deltoid, pectorals, latissimus dorsi (Figure 1);
2. Biceps, triceps, infraspinatus (Figure 2);
3. Extensor carpi radialis brevis (ECRB), flexor carpi radialis (FCR), flexor pollicis longus (FPL), extensor carpi radialis (ECR), abductor pollicis brevis (APB), extensor carpi ulnaris (ECU), flexor digitorum profundus to the 5th digit (FDP V), abductor digiti minimi (ADM) (Figure 3).

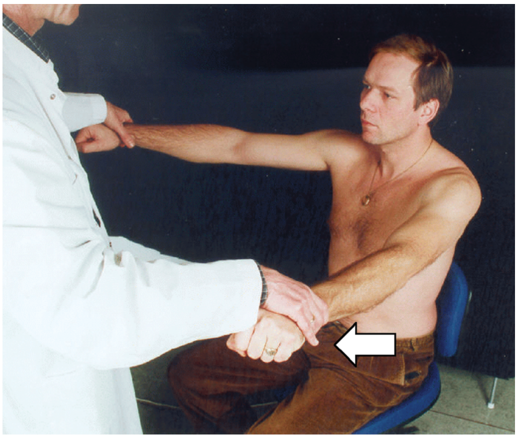

Figure 1. Standard posture I. Testing of the posterior deltoid muscle. The arrow illustrates the direction of the examiner's force against the patient's resistance.

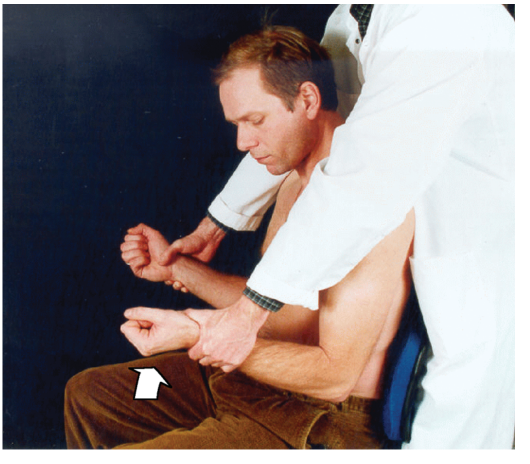

Figure 2. Standard posture II. Testing of the triceps muscle. The arrow illustrates the direction of the examiner's force against the patient's resistance.

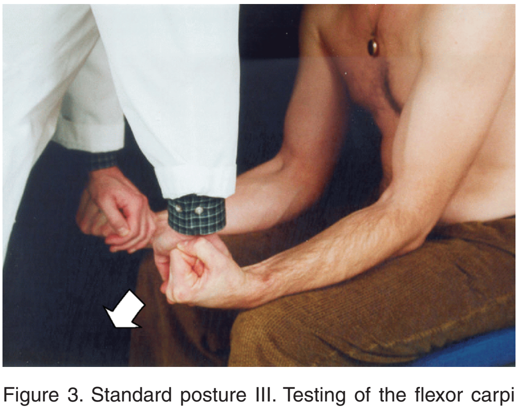

Figure 3. Standard posture III. Testing of the flexor carpi radialis muscle. The arrow illustrates the direction of the examiner's force against the patient's resistance.

Methods

Anamnestic information. The patients were interviewed about upper limb complaints by two examiners who were not involved in the muscle tests. Limbs with any of the following complaints, pain, weakness, and/or numbness/tingling, were defined as being symptomatic by these two examiners irrespective of the duration and severity of the complaints. Limbs without any of these complaints were defined as being asymptomatic.

Blinded physical examination of upper limb muscles. Independently of each other, and blinded as to any patient characteristics, two examiners (JRJ and LHL) performed identical manual physical examinations of the maximal voluntary strength in each of the 14 muscles on both sides. These examinations were done in rapid succession. For all muscles except ECU (for technical reasons), the two sides were examined simultaneously in order to reveal most accurately any discrepancy in the individual muscle strength between right and left side. The patients were asked not to communicate with the examiner during or after the testing, and to provide maximal muscle effort on both sides for each muscle tested, despite any potential discomfort.

In order to stabilize the limb and to minimize discomfort during muscle strength testing, the postures employed were modifications by one of the authors (C-GH) of previously presented techniques (The Editorial Committee for the Guarantors of Brain 1986, Lister 1993). The criteria proposed for optimization of muscle strength testing (Kendall 1991) were met. The patient was positioned comfortably in a chair. Upper limb postures were standardized for each muscle, with the part to be tested positioned with stabilization of the part proximal or adjacent to the tested part. During testing it was aimed to maximize the length of the lever arm. Each test moment was defined from the functional anatomy of the individual muscle. The aim was to optimize the position to facilitate the exertion of maximal strength by each muscle. A uniform placement of the hand for applying pressure was defined, and the pressure was applied directly and gradually, opposite to the line of pull of the muscle being tested. The intent to assess the peak strength as well as the ability of the individual to hold the force at a constant level during testing (endurance) demanded elements of isometric testing to be combined with an evaluation of eccentric dynamic resistance. Testing was performed up to three times.

The strength in each muscle was graded into 6 levels (grades 0–5). Grade 5 represents normal (i.e. powerful) strength. Grade 4 was subdivided into 4–, 4, and 4+, representing slight, moderate, and strong resistance, respectively (Editorial Committee for the Guarantors of Brain 1986). Assessment of strength in each muscle was based on a comparison of the intact contralateral muscle or—with bilaterally reduced strength in the particular muscle—to other muscles of the individual in which the strength was assessed as intact. The latter interpretation was based on former clinical experience from examination of patients with and without upper limb complaints, and taking sex, age and general constitution into account.

The individual muscles were evaluated from proximal to distal with three standard postures of the upper limbs:

I. The patient's arms were elevated horizontally forward, with the elbows kept fully extended, the forearms pronated, the wrists kept at neutral and the hand clenched. Standing in front of the patient, the arm adduction (pectoral muscles) and abduction (posterior deltoid) were tested by applying force against the patient's wrists from inward-out and from outward-in, respectively (Figure 1). The preferred exit position for the posterior deltoid is to have the patient keep the arms 30 degrees outward. The patient then lowered the arms with the elbows still fully extended, but with the forearms now at neutral position and the clenched hands pressed toward the knees at the same time as the examiner was gripping the wrist and lifting the arms upward (latissimus dorsi);

II. The patient's upper arms were now kept along the sides of the chest, the elbows flexed at a right angle with the forearms pointing forward and kept at neutral position, the wrists kept at neutral and the hands clenched. Standing in front of the patient, the examiner leaned forward toward the patient's wrists, asking the patient to "carry" the examiner (elbow flexion, defined as biceps). Resisted by the patient, the examiner then pressed the patient's clenched hands inward (infraspinatus). For this test, the patient's upper arms were rotated 30 degrees in the outward direction. Finally, standing behind the patient, the examiner lifted the patient's wrists upward (triceps) against the patient's resistance (Figure 2);

III. The patient leaned forward, resting the forearms on the thighs with the wrists just distal to the knees. There were three positions for the forearms.

The patient's forearms fully supinated (Figure 3). With the patient's hands clenched and the wrists slightly flexed, the examiner leaned forward, pressing toward the proximal interphalangeal joint knuckles of the index and long fingers to extend the wrists of the patient (FCR). The patient's hands now opened with the dorsum of each hand resting on the knee, and when a) bending the distal joint of the thumbs the examiner tried to straighten them out (FPL), and b) straighten the flexed distal joint of the small fingers (FDP V). c) The patient then had the small fingers abducted. The examiner applied pressure at the tip of the fingers in the radial direction toward the ring finger (ADM). d) The patient brought the thumbs into opposition and the examiner pressed them down toward the palm (APB).

The patient's forearms in neutral. The patient kept the thumbs fully extended and the examiner pressed down at the tip of each thumb (EPL).

The patient's forearms fully pronated. The patient kept the hands open and wrists extended while the examiner leaned forward, pressing against the knuckles of the index and long fingers to flex the patient's wrists (ECRB). Finally, the distal part of the patient's forearm was firmly held by the examiner's one hand while pressing the ulnar-deviated wrist in radial direction (ECU).

Statistics

The grading of strength was redefined into two categories: normal (grade 5) or reduced (grade 4+ or less) (Editorial Committee for the Guarantors of Brain 1986).

Reliability. Cohen's κ statistics, a means of testing whether agreement between raters of categorical data exceeds chance levels, was used for the analyses of the inter-rater variation (same as above) of strength in each individual muscle: $\kappa = (p_o - p_e)/(1 - p_e)$ where p_o is the proportion of observed agreement, and p_e is the proportion of agreement expected by chance. Analyses were performed for all patients, for patients with unilateral complaints, and for symptomatic and asymptomatic limbs, respectively. The κ coefficient has a maximum value of 1.0 and is interpreted as $\kappa < 0.2$ = poor, κ: 0.21–0.40 = fair, κ: 0.41–0.60 = moderate, κ: 0.61–0.80 = good, κ: 0.81–1.00 = very good (Altman 1994).

Association of strength reduction defined by both examiners to symptoms. The odds ratios (OR) were calculated for each muscle. OR = (a/c)/(b/d) with a representing the number of symptomatic limbs with reduced strength, b representing the number of asymptomatic limbs with reduced strength, c the number of symptomatic limbs with normal strength, and d the number of asymptomatic limbs with normal strength.

Results

Anamnestic findings

22 patients were referred due to complaints (pain, weakness and/or numbness/tingling) from one upper limb and 5 patients were referred due to similar complaints from both upper limbs. Among patients who were referred for reasons other than upper limb complaints, 6 also had complaints pertaining to one of the upper limbs. Although only 8 patients had no upper limb symptoms, the composition of the sample with 38 symptomatic and 44 asymptomatic limbs allowed a balanced distribution.

Inter-rater reliability

Individual muscles. The median relative inter-rater agreement of normal or reduced muscle strength was 81 (72–87)%. The median κ-value was 0.54 (0.25–0.72), and 0.57 (0.32–0.82) when calculations were restricted to the 28 subjects with unilateral complaints (Table).

Subdivision of the sample into symptomatic and asymptomatic limbs reduced the κ-values. Despite the high level of agreement between the two examiners with regard to the 44 asymptomatic limbs, the small number of muscles with reduced strength resulted in a median κ of 0.32 (0.17–0.56) while the more balanced distribution of normal or reduced strength for the 38 symptomatic limbs resulted in a moderate median κ of 0.43 (-0.15–0.63) The inter-rater reliability for the latissimus dorsi, infraspinatus, and APB muscles was poor to fair (Table).

Association of symptoms (presence or absence of pain, weakness, and/or numbness/tingling) to muscle strength (reduced or normal) defined by the blinded examiners. With a median odds ratio of 4.0 (2.5–7.7) and the confidence intervals for all 14 muscles well above 1.0, the definition in agreement of reduced strength was significantly associated with presence of complaints for all individual muscles (Table).

Discussion

In spite of their high prevalence and serious effects on life quality and work capacity, our ability to diagnose, manage and prevent upper limb disorders occurring in an occupational context has made only slow progress in our experience. Pain, weakness, and/or numbness/tingling are common complaints in these patients, but often the pathology responsible is not located and characterized by the conventional physical examination.

Reliability. This study is the first to demonstrate in a sample of symptomatic or asymptomatic limbs the inter-rater reliability of the assessment of 14 muscles of normal strength, or with mostly minor strength reduction. The muscles studied were selected to be representative of the whole limb from the shoulder region all the way out to the hand, with 6 muscles being extensors and the remaining 8 being flexors. Taking the inter-rater agreement in the interval of 72%–87% into consideration, κ-values that were only fair for the latissimus dorsi, infraspinatus, and APB muscles can be partly explained by a smaller proportion of reduced strength in these muscles. However, challenges from bilateral symmetrical disorder may also be an issue. Our findings are supported by the very few former studies on the reliability of manual muscle testing which were, however, based on a smaller number of patients or muscles (Iddings and Smith 1961, Silver et al. 1970, Viikari-Juntura 1987).

The large variability in strength between individuals in terms of absolute figures restrains the comparison of the measured strength with normative values, and gravity cannot be used as an objective standard when strength is not severely reduced. This variability challenges the assessment particularly with bilateral strength reductions. For each muscle, the strength measured is related to normal power (The Editorial Committee for the Guarantors of Brain 1986) which, however, is a term relating to the presuppositions of the examiner with regard to the physical condition of each individual. In spite of being blinded as to patient-related information, the examiners appraised the normal strength of individuals from their age, sex and general physical constitution. In addition to technical skills, this subjective aspect of manual muscle testing demands of the examiner clinical experience that cannot be standardized in quantitative terms. Despite these difficulties, the strength was assessed with reasonable reliability, but better with unilateral than with bilateral complaints (Table).

In contrast to quantitative dynamometric measurements, manual muscle strength testing enables the examiner to guide the tested part into the exact test position and to give a precise amount of pressure to determine the strength. Previous recommendations concerning the increased use of manual muscle testing in clinical practice (Kendall 1991, Marx et al. 1999) are supported by this study which even demonstrated a comparable or better reliability than other diagnostic tests in common use, including the brachial plexus tension and shoulder abduction relief tests (Viikari-Juntura 1987), passive cervical spine motion (Viikari-Juntura 1987, Smedmark et al. 2000), trigger point palpation (Viikari-Juntura 1987), and tendon reflexes (Manschot et al. 2000), as well as the Babinski sign for the lower limb (Maher et al. 1992).

Combining symptoms and physical findings is fundamental to the diagnostic process, but also infers a risk of bias. The blinded and independent physical examination in this study has aimed at eliminating such bias. Even so, there may have been inconsistencies in measurement resulting from sources of variability relating to the patient and the clinician.

The patients investigated included those for whom the assessment of strength is intended.

Inter-rater reliability and relation to complaints of manual testing of strength in 14 muscles in 82 limbs

	Muscles	All limbs (n=82) κ (95% CI)	Unilateral complaints (n=56) κ (95% CI)	Symptomatic limbs (n=38) NS [a]	RS [a]	κ	Asymptomatic limbs (n=44) NS [a]	RS [a]	κ	Odds ratios (95% CI) [b]
	Posterior deltoid	0.60 (0.42–0.77)	0.69 (0.49–0.88)	3	32	0.63	22	9	0.36	6.5 (2.2–19.)
I	Pectorals	0.55 (0.34–0.76)	0.55 (0.30–0.79)	16	12	0.47	41	0	-	3.6 (2.4–5.3)
	Latissimus dorsi	0.37 (0.19–0.56)	0.32 (0.11–0.53)	11	12	0.27	35	1	0.17	3.9 (2.3–6.6)
	Biceps	0.57 (0.40–0.75)	0.63 (0.43–0.84)	7	20	0.40	33	5	0.56	4.6 (2.3–9.2)
II	Triceps	0.72 (0.57–0.88)	0.82 (0.68–0.97)	5	25	0.42	38	3	0.63	7.7 (3.3–18)
	Infraspinatus	0.25 (0.05–0.46)	0.36 (0.14–0.58)	4	13	-0.15	34	3	0.28	7.7 (3.0–20)
	ECRB	0.69 (0.53–0.85)	0.79 (0.63–0.94)	8	24	0.62	36	2	0.34	5.1 (2.7–9.6)
	FCR	0.46 (0.25–0.66)	0.52 (0.29–0.76)	8	14	0.19	40	1	0.37	5.6 (2.9–11)
	FPL	0.51 (0.30–0.73)	0.65 (0.43–0.87)	15	12	0.42	39	1	0.29	3.3 (2.1–5.3)
III	EPL	0.52 (0.33–0.71)	0.67 (0.48–0.87)	11	20	0.61	28	4	0.22	3.0 (1.7–5.0)
	APB	0.33 (0.11–0.55)	0.48 (0.22–0.73)	10	10	0.09	39	4	0.29	3.5 (1.8–6.7)
	ECU	0.56 (0.37–0.76)	0.54 (0.31–0.78)	11	15	0.38	39	2	0.54	4.0 (2.3–7.0)
	FDP V	0.54 (0.31–0.78)	0.59 (0.32–0.86)	19	9	0.44	43	0	-	3.3 (2.2–4.7)
	ADM	0.46 (0.22–0.70)	0.37 (0.08–0.66)	21	8	0.46	38	1	0.23	2.5 (1.7–3.8)

[a] NS = normal strength, RS = reduced strength. Strength as defined by agreement between the two blinded examiners.
[b] Odds ratios (95% CI) for the relation of strength (see above) to complaints.
n = numbers of limbs. ECRB = extensor carpi radialis brevis, FCR = flexor carpi radialis, FPL = flexor pollicis longus, ECR = extensor carpi radialis, APB = abductor pollicis brevis, ECU = extensor carpi ulnaris, FDP V = flexor digitorum profundus to the fifth digit, ADM = abductor digiti minimi.

The spectrum of disease was sufficiently broad, ranging from patients with healthy limbs to those with limbs variably afflicted in one or both sides. This was one advantage of the study and suggests the diagnostic feasibility of muscle strength testing in clinical settings in which a wide variability in presentations and severity of upper limb disorders is to be expected.

Clinical variability may have resulted from the physical examination and its interpretation. We have attempted to reduce inconsistencies in measurement by optimized and standardized testing conditions defined to match the clinical setting. Each muscle was tested in a position aimed at stabilizing the limb and minimizing discomfort, and the grading of strength was converted into two well-defined categories only (normal and reduced). Technical skills in the manual muscle strength testing procedure and its interpretation are other crucial issues. Some experience is needed in performing the examination correctly, and in interpreting perceived strength as normal or reduced, especially when only slight. Both examiners had learned the examination technique a short time before the study. One (JRJ) had used it for two years prior to the study, and the other (LHL) had used it for only 2 months following an update on upper limb anatomy and also supervised examination of about 20 patients.

Despite the fact that no patient-related information was disclosed to the examiners and that there was no oral communication during the examination, it can be argued that the examiners may have been biased from the patients' symptom status being apparent from nonverbal communication. A strength reduction at or below 3 may show as impaired active motion or no motion at all. In the sample of patients studied, however, the strength reductions encountered were minor. Only two and three muscles were graded as 3 by the two examiners, respectively, out of a total of 14 muscles examined in 82 limbs. No muscles were graded below 3. The intact or only slightly reduced (4–, 4, or 4+) strength in the remaining muscles was not apparent from visible atrophy, abnormal movement pattern or rhythm, or reduced active range of motion. Consequently, almost all limbs including those with reduced strength would look completely normal, and facial expressions or withdrawal indicating simultaneous pain did not accompany the testing.

The reliability of findings and of their association with complaints in the sample of patients studied may deviate from another sample with different severity and frequency of reduced strength. Accordingly, no conclusions relating to other settings can be drawn from this study.

Association with symptoms. Despite the significant correlation between reduced strength of individual muscles and patient complaints from the upper limb in terms of pain, weakness, numbness and/or tingling, the clinical relevance of our findings was not analyzed further in this study. However, in this context it is of interest to refer to a previous study (Stål et al. 1998) in which reduced individual muscle strength confined to forearm muscles innervated by the median nerve was a reliable diagnostic sign of median nerve compression at the elbow level, the so-called pronator syndrome.

From our study, we can only conclude that strength reductions occurring in one or more of the 14 muscles we examined were common in patients with upper limb complaints, whereas the reason for this has not been defined yet, but will be investigated further.

We have chosen to study individual muscle strength as opposed to the assessment of strength in groups of muscles, because the latter would not identify a differential involvement of muscles, and the diagnostic potential would consequently be reduced. Reduced grip strength, for example, can accompany a number of orthopedic conditions as well as systemic disorders, but it is of limited significance for the differential diagnosis. In contrast, provided that strength in representative individual muscles can be reliably assessed, the identification or exclusion of selectively reduced strength in representative muscles would improve the diagnostic options, with some focal conditions being more likely and others less likely.

To summarize, we have found that the inter-rater reliability of manual testing of voluntary strength was moderate to good in 11 of 14 representative upper limb muscles. A reliability that was only "fair" for the remaining three muscles may appear to be disappointing, but it is nevertheless comparable to that of other physical examinations in common use. For all muscles, the strength reductions were significantly associated with the presence of upper limb symptoms in terms of pain, weakness numbness and/or tingling.

The authors wish to thank Professor Gisela Sjøgaard (National Institute of Occupational Health, Copenhagen), Dr. Børge Balle (retired, Hirtshals) and Dr. Per Sabro Nielsen (Central Hospital, Esbjerg) for valuable advice during execution and publication of the study. Financial support was received from Statens Sundhedsvidenskabelige Forskningsråd, Copenhagen (Grant no. 9702593), Den Samfundsvidenskabelige Forskningsfond, Ringkøbing (Grant no. 2-44-4-18-97), and Lida & Oskar Nielsens Fond, Esbjerg.

No competing interests declared.

Altman D G. Practical statistics for medical research. Chapman & Hall, London 1994: 403-9.

Iddings D M, Smith L K. Muscle testing: Part 2. Reliability in clinical use. Phys Ther Rev 1961; 41 (4): 249-56.

Kendall F P. Manual muscle testing: there is no substitute. J Hand Ther 1991; 4: 159-61.

Lister G. Appendix: Muscle testing. In: The hand. Diagnosis and indications 3 ed. (ed. Lister G). Churchill Livingstone, Edinburgh 1993: 513-26.

Maher J, Reilly M, Daly L, Hutchinson M. Plantar power: reproducibility of the plantar response. Br Med J 1992; 304: 482.

Manschot S, van Passel L, Buskens E, Algra A, van Gijn J. Mayo and NINDS scales for assessment of tendon reflexes: between observer agreement and implications for communication. J Neurol Neurosurg Psychiatry 2000; 64 (2): 253-5.

Marx R G, Bombardier C, Wright J G. What do we know about the reliability and validity of physical examination tests used to examine the upper extremity? J Hand Surg (Am) 1999; 24 (1): 185-93.

Sapega A A. Muscle performance evaluation in orthopaedic practice. J Bone Joint Surg (Am) 1990; 72 (10): 1562-74.

Silver M, McElroy A, Morrow L, Heafner B K. Further standardization of manual muscle test for clinical study: applied in chronic renal disease. Phys Ther 1970; 50 (10): 1456-66.

Smedmark V, Wallin M, Arvidsson I. Inter-examiner reliability in assessing passive intervertebral motion of the cervical spine. Man Ther 2000; 5 (2): 97-101.

Stål M, Hagert C-G, Moritz U. Upper extremity nerve involvement in Swedish female machine milkers. Am J Ind Med 1998; 33: 551-9.

The Editorial Committee for the Guarantors of Brain. Aids to the examination of the peripheral nervous system. London: Ballière Tindall 1986.

Viikari-Juntura E. Interexaminer reliability of observations in physical examinations of the neck. Phys Ther 1987; 67 (10): 1526-32.