	Research Supporting the Reliability of The Manual Muscle Test
On the reliability and validity of manual muscle testing: a literature review, Cuthbert SC, Goodheart GJ Jr.	*Chiropr Osteopat.* 2007 Mar 6;15(1):4. **ABSTRACT:** A body of basic science and clinical research has been generated on the manual muscle test (MMT) since its first peer-reviewed publication in 1915. The aim of this report is to provide an historical overview, literature review, description, synthesis and critique of the reliability and validity of MMT in the evaluation of the musculoskeletal and nervous systems. **METHODS:** Online resources were searched including Pubmed and CINAHL (each from inception to June 2006). The search terms manual muscle testing or manual muscle test were used. Relevant peer-reviewed studies, commentaries, and reviews were selected. The two reviewers assessed data quality independently, with selection standards based on predefined methodologic criteria. Studies of MMT were categorized by research content type: inter- and intra-examiner reliability studies, and construct, content, concurrent and predictive validity studies. Each study was reviewed in terms of its quality and contribution to knowledge regarding MMT, and its findings presented. **RESULTS:** More than 100 studies related to MMT and the applied kinesiology chiropractic technique (AK) that employs MMT in its methodology were reviewed, including studies on the clinical efficacy of MMT in the diagnosis of patients with symptomatology. With regard to analysis there is evidence for good reliability and validity in the use of MMT for patients with neuromusculoskeletal dysfunction. The observational cohort studies demonstrated good external and internal validity, and the 12 randomized controlled trials (RCTs) that were reviewed show that MMT findings were not dependent upon examiner bias. **CONCLUSION:** The MMT employed by chiropractors, physical therapists, and neurologists was shown to be a clinically useful tool, but its ultimate scientific validation and application requires testing that employs sophisticated research models in the areas of neurophysiology, biomechanics, RCTs, and statistical analysis. **Comment:** This is a landmark study presenting the basic science and clinical research evidence for the reliability and validity of the manual muscle test and applied kinesiology chiropractic technique. The literature review presents the results of more than 100 peer-reviewed studies related to the manual muscle test (MMT) and the applied kinesiology chiropractic technique (AK). **Muscle testing, which is the backbone of AK, now has support for its use in the field of chiropractic to diagnose and treat neuromusculoskeletal dysfunction.**
Interrater reliability and diagnostic accuracy of pelvic girdle pain classification. Cook C, Massa L, Harm-Ernandes I, Segneri R, Adcock J, Kennedy C, Figuers C.	*J Manipulative Physiol Ther.* 2007 May;30(4):252-8. **OBJECTIVE:** The purpose of this study was to measure the reliability of a classification system for pelvic girdle pain (PGP) and diagnostic accuracy of selected examination and clinical special findings for diagnosis of PGP. **METHODS:** The design involved a prospective epidemiological study of **pregnancy-related PGP.** Consecutive subjects were recruited and classified using criteria defined by previous studies. Two clinicians examined the subjects and classified each patient into 1 of 5 classification groups. Clinical examination and clinical special tests were performed on the patients with PGP. Statistical analysis involved interobserver agreement using a kappa statistic and sensitivity and specificity values for the examination and clinical special testing. **RESULTS:** Twenty-one subjects were included in the analyses. Aggregated percentage of agreement for the classification system was 84.6%. The Cohen kappa was 0.78 (CI, 0.64-0.92; P < .0001), which indicated substantial agreement during selection of the classification system. Most clinical examination and clinical special-test findings demonstrated low sensitivity and high specificity, whereas clusters of findings including the lunge, **manual muscle testing of the**

	hip (lower extremities), and passive range of motion of the hip demonstrated the strongest diagnostic value. CONCLUSION: Selected tests and measures are moderately discriminatory in diagnosing PGP. A classification system for diagnosing PGP demonstrates strong agreement and may be useful for clinicians. **Comment:** The MMT method used in this study was the standard one used in AK (Kendall and Kendall's original MMT methods), and the MMT test was found to show the strongest accuracy and sensitivity of all the tests used for evaluation of PGP.
Intra-rater and inter-rater reliability of the 10-point Manual Muscle Test (MMT) of strength in children with juvenile idiopathic inflammatory myopathies (JIIM), Jain M, Smith M, Cintas H, Koziol D, Wesley R, Harris-Love M, Lovell D, Rider LG, Hicks J.	*Phys Occup Ther Pediatr.* 2006;26(3):5-17. **OBJECTIVE:** Children with juvenile idiopathic inflammatory myopathies (JIIM) present with muscle inflammation and decreased strength that may affect their functional abilities. The purpose of this study was to determine the intra-rater and inter-rater reliability of the 0 to 10-point manual muscle testing method for children with JIIM. **METHODS:** For the intra-rater and inter-rater reliability studies, 10 and 9 children with JIIM participated, respectively. For intra-rater reliability, one pediatric therapist completed two assessments in one day with a one-hour break. For inter-rater reliability, four therapists assessed the same child within a single morning. **RESULTS:** Spearman correlations for intra-rater reliability ranged from 0.70 to 1.00. Kendall's W coefficient for inter-rater reliability of groups of muscles (total, proximal, distal, and peripheral) ranged from 0.51 to 0.76. **CONCLUSIONS: The total, proximal, and peripheral Manual Muscle Test (MMT) score, using the 0-10 point scale, has acceptable reliability in JIIM patients.**

Can the Ileocecal Valve Point Predict Low Back Pain Using Manual Muscle Testing? Pollard HP, Bablis P, Bonello R.	*Chiropr Aust* 2006;36:58-62 **Background:** According to some technique groups in chiropractic the ileocecal valve may malfunction and be associated with a large array of health problems that can lead to common chronic health issues prevalent in our society. Many tests commonly used in chiropractic are presumed to identify painful and/or dysfunctional anatomical structures, yet many have undemonstrated reliability. Despite this lack of evidence, they form the basis of many clinical decisions. One cornerstone procedure that is frequently used by chiropractors involves the use of manual muscle testing for diagnostic purposes not considered orthopaedic in nature. A point of the body referred to as the ileocecal valve point is said to indicate the presence of low back pain. This procedure is widely used in Applied Kinesiology (AK) and Neuro-Emotional Technique (NET) chiropractic practice. **Objective:** To determine if correlation of tenderness of the "ileocecal valve point" can predict low back pain in sufferers with and without low back pain. It was the further aim to determine the sensitivity and specificity of the procedure. **Methods:** One hundred (100) subjects with and without low back pain were recruited. Subjects first completed information about their pain status, then the practitioner performed the muscle testing procedure in a separate room. The practitioner provided either a *yes* or *no* response to a research assistant as to whether he had determined if the subject had back pain based on the muscle test procedure. **Results:** Of 67 subjects who reported low back pain, 58 (86.6%) reported a positive test of both low back pain and ICV point test. Of 33 subjects, 32 (97%) with no back pain positively reported no response to ICV point test. Nine (9) subjects (13.4%) reported false negative ICV tests and low back pain, and 1 subject (3%) reported a false positive response for ICV test and no low back pain. **Conclusion:** The majority of subjects with low back pain reported positive ileocecal valve testing, and all but one of the subjects without low back pain reported negative ileocecal valve testing. The application of ileocecal valve testing as a diagnostic measure of low back pain was found to have excellent measures of sensitivity, specificity and diagnostic competency. This study confirms that the use of this test within the limitations of this study is reliably associated with the presence of low back pain. Further testing is required to investigate all aspects of the diagnostic milieu commonly used by proponents of this form of diagnostic testing. **Comment:** In AK, the ileocecal valve dysfunction is not related automatically to low back pain though this is a frequent consequence of the problem. Another interesting research question that might have been posed to the subjects of this study would have been whether they had experiencing any digestive difficulties and its relationship to positive MMT outcomes. The finding of excellent sensitivity and specificity in this research report is noteworthy.
Electromyogram and force patterns in variably timed manual muscle testing of the middle deltoid muscle, Conable K, Corneal J, Hambrick T, Marquina N, Zhang J.	*J Manipulative Physiol Ther.* 2006 May;29(4):305-14. **OBJECTIVE:** The objective of the study was to compare force curves and surface electromyogram from examiners and subjects during manual muscle testing with 3 examiner-identified variants of manual muscle testing (MMT)-examiner-started (ESMMT), patient-started (PSMMT), and undifferentiated/near-simultaneous (NSMMT). **METHODS:** Forty-two volunteer applied kinesiologist doctors tested 36 volunteer students, doctors, and spouses at a professional conference. Start-time difference between examiner and subject muscle contraction, peak force, time of peak force and duration of force was measured. Force and surface electromyogram from examiner and subject were recorded simultaneously during MMT of the middle deltoid muscle for each style of testing used in practice. **RESULTS:** The congruence between examiner label and timing was as follows: ESMMT, 39%; PSMMT 61%; and NSMMT 28% within 50 milliseconds of simultaneous. Mean subject/examiner start-time difference differed significantly between PSMMT (-0.116 seconds) and both ESMMT (-0.018 seconds) and NSMMT (-0.0053 seconds). No clear cutoff between styles was evident. Peak force ranged from 0.548 to 23.615 lb (mean, 8.806 lb; SD, 4.543 lb), and all styles were significantly different.

	Duration of force ranged from 0.325 to 3.490 seconds (mean, 1.338 seconds; SD, 0.576 seconds), with NSMMT significantly shorter than ESMMT or PSMMT. The shape of force curves did not differ between styles of muscle testing, but differed between facilitated vs. inhibited tests. **CONCLUSIONS:** In this group of doctors, neither start-time difference nor the shape of force curves distinguished styles of MMT. Differences in peak force and test duration may account for differences in outcomes when examiners purposely vary their muscle testing style.
Diagnostic accuracy of the neurological upper limb examination I: inter-rater reproducibility of selected findings and patterns, Jepsen JR, Laursen LH, Hagert CG, Kreiner S, Larsen AI.	*BMC Neurol.* 2006 Feb 16;6:8. **BACKGROUND:** We have previously assessed the reproducibility of manual testing of the strength in 14 individual upper limb muscles in patients with or without upper limb complaints. This investigation aimed at additionally studying sensory disturbances, the mechanosensitivity of nerve trunks, and the occurrence of physical findings in patterns which may potentially reflect a peripheral neuropathy. The reproducibility of this part of the neurological examination has never been reported. **METHODS:** Two blinded examiners performed a semi-quantitative assessment of 82 upper **limbs (strength in 14 individual muscles**, sensibility in 7 homonymous territories, and mechanosensitivity of nerves at 10 locations). Based on the topography of nerves and their muscular and cutaneous innervation we defined 10 neurological patterns each suggesting a focal neuropathy. The individual findings and patterns identified by the two examiners were compared. **RESULTS: Strength, sensibility to touch, pain and vibration, and mechanosensitivity were predominantly assessed with moderate to very good reproducibility (median kappa-values 0.54, 0.69, 0.48, 0.58, and 0.53, respectively). The reproducibility of the defined patterns was fair to excellent (median correlation coefficient = 0.75) and the overall identification of limbs with/without pattern(s) was good (kappa = 0.75). CONCLUSION:** This first part of a study on diagnostic accuracy of a selective neurological examination has demonstrated a promising inter-rater reproducibility of individual neurological items and patterns. Generalization and clinical feasibility require further documentation: 1) Reproducibility in cohorts of other composition, 2) validity with comparison to currently applied standards, and 3) potential benefits that can be attained by the examination.
A method for comparing manual muscle strength measurements with joint moments during walking, Fosang A, Baker R.	*Gait Posture.* 2006 Dec;24(4):406-11. Epub 2006 Jan 18. **Abstract:** This paper describes a protocol for dynamometer assisted manual muscle testing of the major muscle groups of the lower extremity and its application to 11 able-bodied children who also had conventional gait analysis to obtain joint kinetics. Data from the manual muscle testing was processed in such a way that the results for maximum muscle strength (grade 5) and resistance against gravity alone (grade 3) were presented in Nm/kg allowing direct comparison with conventional joint kinetics. The strength measurements of the hip muscles and the knee extensors were between two and three times the moments exerted during normal walking. Those of the knee flexors and dorsiflexors were about five times the joint moments. Measured plantarflexor strength was only just greater than the moment exerted during walking. These results, particularly those for the plantarflexors, question how valid it is to use measures of isometric muscle strength as indicators of muscle function during activity. The study also compares grade 3 muscle strength with both grade 5 strength and the maximum joint moments. For all muscle groups tested grade 3 muscle strength was less than the maximum moment exerted during normal walking. For the plantarflexors it was less than 1% of that moment. The study demonstrates that reliable isometric muscle testing is possible in able-bodied children but requires considerable care and is time consuming. More work is required to understand how measurements made in this way relate to how muscles function during activity. **Comment:** Many studies have compared the findings of MMT with dynamometer tests favorably, however dynamometers are not as sensitive to changes in strength nor to strength

	measurements below 3 during the MMT. The human examiner is the most sensitive of all instruments in relationship to interpreting the MMT.
Reliability of techniques to assess human neuromuscular function in vivo, Clark BC, Cook SB, Ploutz-Snyder LL	*J Electromyogr Kinesiol.* 2006 Jan 18 **Abstract:** The purpose of this study was to comprehensively evaluate the reliability of a large number of commonly utilized experimental tests of in vivo human neuromuscular function separated by 4-weeks. Numerous electrophysiological parameters (i.e., voluntary and evoked electromyogram [EMG] signals), contractile properties (i.e., evoked forces and rates of force development and relaxation), muscle morphology (i.e., MRI-derived cross-sectional area [CSA]) and performance tasks (i.e., steadiness and time to task failure) were assessed from the plantarflexor muscle group in 17 subjects before and following 4-weeks where they maintained their normal lifestyle. The reliability of the measured variables had wide-ranging levels of consistency, with coefficient of variations (CV) ranging from approximately 2% to 20%, and intraclass correlation coefficients (ICC) between 0.53 and 0.99. **Overall, we observed moderate to high-levels of reliability in the vast majority of the variables we assessed** (24 out of the 29 had ICC>0.70 and CV<15%). The variables demonstrating the highest reliability were: CSA (ICC=0.93-0.98), strength (ICC=0.97), an index of nerve conduction velocity (ICC=0.95), and H-reflex amplitude (ICC=0.93). Conversely, the variables demonstrating the lowest reliability were: the amplitude of voluntary EMG signal (ICC=0.53-0.88), and the time to task failure of a sustained submaximal contraction (ICC=0.64). Additionally, relatively little systematic bias (calculated through the limits of agreement) was observed in these measures over the repeat sessions. **In conclusion, while the reliability differed between the various measures, in general it was rather high even when the testing sessions are separated by a relatively long duration of time.**
Intra-rater and inter-rater reliability of the 10-point Manual Muscle Test (MMT) of strength in children with juvenile idiopathic inflammatory myopathies (JIIM), Jain M, Smith M, Cintas H, Koziol D, Wesley R, Harris-Love M, Lovell D, Rider LG, Hicks J.	*Phys Occup Ther Pediatr.* 2006;26(3):5-17. **OBJECTIVE:** Children with juvenile idiopathic inflammatory myopathies (JIIM) present with muscle inflammation and decreased strength that may affect their functional abilities. The purpose of this study was to determine the intra-rater and inter-rater reliability of the 0 to 10-point manual muscle testing method for children with JIIM. **METHODS:** For the intra-rater and inter-rater reliability studies, 10 and 9 children with JIIM participated, respectively. For intra-rater reliability, one pediatric therapist completed two assessments in one day with a one-hour break. For inter-rater reliability, four therapists assessed the same child within a single morning. **RESULTS:** Spearman correlations for intra-rater reliability ranged from 0.70 to 1.00. Kendall's W coefficient for inter-rater reliability of groups of muscles (total, proximal, distal, and peripheral) ranged from 0.51 to 0.76. **CONCLUSIONS: The total, proximal, and peripheral Manual Muscle Test (MMT) score, using the 0-10 point scale, has acceptable reliability in JIIM patients.**
Interexaminer reliability of the deltoid and psoas muscle test, Pollard H, Lakay B, Tucker F, Watson B, Bablis P.	*J Manipulative Physiol Ther*, Jan 2005:28(1):52-6 **Objective:** To determine if 2 practitioners of differing skill levels could reliably agree on the presence of a weak or strong deltoid or psoas muscle. **Study Design:** Interexaminer reliability study of 2 common muscle tests. **Main Outcome Measures:** Cohen κ (unweighted) scores, observer agreement, and 95% confidence intervals (CIs). **Results:** The results showed that an experienced and a novice practitioner have good agreement when using repeated muscle test procedures on the deltoid (κ 0.62) and the psoas (κ 0.67). **Conclusions:** The manual muscle test procedures using the anterior deltoid or psoas showed good interexaminer reliability when used by an experienced and a novice user. These techniques may be used between practitioners in multidoctor assessment/management programs.
Parallel comparison of grip	*Percept Mot Skills.* 2005 Jun;100(3 Pt 1):795-8.

strength measures obtained with a MicroFET 4 and a Jamar dynamometer, Bohannon RW. -- Department of Physical Therapy, School of Allied Health, U-2101, University of Connecticut, Storrs, CT 06269-2101, USA.	**Abstract:** Repeated measures of grip strength obtained bilaterally with a Jamar and a MicroFET 4 dynamometer were compared. Measurements obtained with the MicroFET 4 tended to be slightly (2.2-3.1 lb.) higher but were highly correlated (r > or = .96) with those obtained with the Jamar. Parallel reliability for the two devices was excellent (intraclass correlation coefficient > or = .96). Although clinicians should be cautious about using the devices interchangeably, the MicroFET 4 appears to be a legitimate alternative to the Jamar dynamometer.
Hip muscle weakness and overuse injuries in recreational runners, Niemuth PE, Johnson RJ, Myers MJ, Thieman TJ.	*Clin J Sport Med.* 2005 Jan;15(1):14-21. **OBJECTIVE:** To test for differences in strength of 6 muscle groups of the hip on the involved leg in recreational runners with injuries compared with the uninvolved leg and a control group of noninjured runners. **DESIGN:** Descriptive analysis. **SETTING:** Three outpatient physical therapy clinics in the Minneapolis/St. Paul metropolitan area. **PARTICIPANTS:** Thirty recreational runners (17 female, 13 male) experiencing a single leg overuse injury that presented for treatment between June and September 2002. Thirty noninjured runners (16 female, 14 male) randomly selected from a pool of 46 volunteers from a distance running club served as controls. **MAIN OUTCOME MEASURES:** Self-report demographic information on running habits, leg dominance demonstrated by preferred kicking leg, and injury information. Muscle strength of the 6 major muscle groups of the hip was recorded using a hand-held dynamometer. The highest value of 2 trials was used, and strength values were normalized to body mass(2/3). **RESULTS:** Results comparing the injured and noninjured groups showed that leg dominance did not influence the leg of injury (chi(2)(1) = 0.134; P = 0.71). **Correlations for internal reliability of muscle measurements between trials 1 and 2 with the hand-held dynamometer ranged from 0.80 to 0.90 for the 6 muscle groups measured, and all P values were less than 0.0001.** No significant side-to-side differences in hip group muscle strength were found in the noninjured runners (P = 0.62-0.93). **Among the injured runners, the injured side hip abductor (P = 0.0003) and flexor muscle groups (P = 0.026) were significantly weaker than the noninjured side. In addition, the injured side hip adductor muscle group was significantly stronger (P = 0.010) than the noninjured side.** Duration of symptoms was not a contributing factor to the extent of injury as measured by muscle strength imbalance between injured and uninjured sides. **CONCLUSIONS:** Although no cause-and-effect relationship has been established**,** this is the first study to **show an association between hip abductor, adductor, and flexor muscle group strength imbalance and lower extremity overuse injuries in runners.** Because most running injuries are multifaceted in nature, areas secondary to the site of pain, such as hip muscle groups exhibiting strength imbalances, must also be considered to gain favorable outcomes for injured runners. **The addition of strengthening exercises to specifically identified weak hip muscles may offer better treatment results in patients with running injuries.** **Comment:** This study shows an important evidence-based component of the treatment of hip and pelvic problems in runners as being methods of strengthening the weak muscles around the hip and to improve muscular balance, the *sine qua non* of AK treatment.
Manual strength testing in 14 upper limb muscles: a study of inter-rater reliability, Jepsen, J., Laursen, L., Larsen, A., Hagert, CG. -- Department of Occupational	*Acta Orthop Scand.* 2004 Aug;75(4):442-8. **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

Medicine, Central Hospital, DK-6700 Esbjerg, Denmark. jrj@ribeamt.dk	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 kappa-statistics. In addition, the odds ratio for the relation to symptoms of the definition in agreement of strength was calculated. **RESULTS:** The median kappa-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.
The supine hip extensor manual muscle test: a reliability and validity study, Perry J, Weiss WB, Burnfield JM, Gronley JK. -- Pathokinesiology Laboratory, Rancho Los Amigos National Rehabilitation Center, Downey, CA 90242, USA. pklab@larei.org	*Arch Phys Med Rehabil.* 2004 Aug;85(8):1345-50. **OBJECTIVES:** To define the relative hip extensor muscle strengths values identified by the 4 grades obtained with a supine manual muscle test (MMT) and to compare these values with those indicated by the traditional prone test. **DESIGN:** Comparison of 4 manual supine strength grades with isometric hip extension joint torque; kappa statistic-determined interrater reliability, and analyses of variance identified between grade differences in torque. **SETTING:** Pathokinesiology laboratory. **PARTICIPANTS:** Adult volunteers recruited from local community and outpatient clinics. Reliability testing: 16 adults with postpolio (31 limbs). Validity testing (2 groups): 18 subjects without pathology (18 limbs), and 26 people with clinical signs of hip extensor weakness (51 limbs). **INTERVENTIONS:** Not applicable. **MAIN OUTCOME MEASURES:** Supine hip extensor manual muscle grade and isometric hip extension torque. **RESULTS:** Reliability testing showed excellent agreement (82%). Subjects with pathology had significant differences in mean torque ($P<.01$) for the assigned grade 5 (176 Nm), grade 4 (103 Nm), grade 3 (67 Nm), and grade 2 (19 Nm). Healthy adults showed significant differences between grade 5 (212 Nm) and grade 4 (120 Nm) in mean torque ($P<.05$). **CONCLUSIONS:** The supine MMT is a reliable and valid method with which to assess hip extension strength.
Reliability of hand-held dynamometry in assessment of knee extensor strength after hip fracture, Roy, MA, Doherty, TJ. -- School of Kinesiology, University of Western Ontario, London, Ontario, Canada.	*Am J Phys Med Rehabil.* 2004 Nov;83(11):813-8. **OBJECTIVES:** To examine the reliability of hand-held dynamometry in assessing knee extensor strength in inpatients undergoing rehabilitation after hip fracture and to examine the discriminant validity of this measure. **DESIGN:** A total of 16 subjects (14 women; mean +/- SD, 79 +/- 7 yrs) undergoing inpatient rehabilitation after hip fracture volunteered to participate. Isometric knee extensor strength of the fractured and unfractured sides was determined with a hand-held dynamometer. Subjects were retested 1-2 days after the initial testing session. **RESULTS:** Test-retest intraclass correlation coefficients were high for both the fractured (0.91) and unfractured legs (0.90). A low coefficient of variation was observed for both the fractured (15.3%) and unfractured (14.7%) sides. The maximal knee extensor strength was significantly different when comparing the fractured (7.9 +/- 3 kg) and unfractured (15.6 +/- 4 kg) legs. When comparing test 1 and test 2 mean values for the fractured leg, the scores significantly differed ($t = 3.14$, $P < 0.01$), with 13 of 16 subjects scoring higher on test 2. **CONCLUSIONS:** Hand-held dynamometry is a reliable and valid tool for assessment of knee extensor strength after hip fracture. Reduced knee extensor strength in the fractured leg may be an important component limiting rehabilitation progress in these patients.
Comparison of four tests of quadriceps strength in L3 or L4 radiculopathies, Rainville J, Jouve	*Spine.* 2003 Nov 1;28(21):2466-71. **STUDY DESIGN:** This prospective cohort study evaluated four office tests of quadriceps

C, Finno M, Limke J. -- The Spine Center, New England Baptist Hospital, Boston, MA 02120, USA. jrainvil@caregroup.harvard.edu	strength in symptomatic adults with radiographic evidence of L3 or L4 nerve root compression. **OBJECTIVE:** The study observed the performance of each test for its ability to detect quadriceps weakness when compared to the asymptomatic side. To determine the potential influence of radicular pain on the performance of the four tests, a control group of patients over the age of 40 with clinical and radiographic L5 or S1 radiculopathies underwent identical testing of quadriceps strength. **SUMMARY OF BACKGROUND DATA:** The L3 and L4 nerve roots innervate the quadriceps; therefore, quadriceps weakness may be a consequence of L3 or L4 radiculopathies. There are no standardized or validated methods to evaluate quadriceps strength in the clinical office setting. This may lead to inconsistent detection by clinicians of quadriceps weakness in cases of L3 or L4 radiculopathy. **METHODS:** Thirty-three consecutive patients with L3 or L4 radiculopathies and 19 with L5 or S1 radiculopathies were studied. The four tests of quadriceps strength included: 1) single leg sit-to-stand test; 2) step-up test; 3) knee-flexed manual muscle testing; and 4) knee-extended manual muscle testing. Results from a second examiner repeating the four tests were used to calculate interrater reliability. **RESULTS:** In L3 and L4 radiculopathies, unilateral quadriceps weakness was detected by the single leg sit-to-stand test in 61%, by knee-flexed manual muscle testing in 42%, by step-up test in 27% and by knee-extended manual muscle testing in 9% of patients. The sit-to-stand test detected weakness in all but one case when weakness was detected by another test. All patients with L5 or S1 radiculopathies could perform the sit-to-stand test. Kappa coefficient was high for sit-to-stand test (0.85), step-up (0.83), and knee-flexed manual muscle testing (0.66), and low for knee-extended manual muscle testing (0.08). **CONCLUSION:** In L3 and L4 radiculopathies, unilateral quadriceps weakness was best detected by a single leg sit-to-stand test. Patients of similar age with radicular pain caused by L5 or S1 radiculopathies could perform this test. As the interrater reliability of the single leg sit-to-stand test is high, clinicians should consider utilizing this test for assessing quadriceps strength in cases of L3 and L4 radiculopathies.
Reliability of 4 outcome measures in pediatric spinal muscular atrophy, Iannaccone ST, Hynan LS, American Spinal Muscular Atrophy Randomized Trials (AmSMART) Group.	*Arch Neurol.* 2003 Aug;60(8):1130-6. **BACKGROUND:** Spinal muscular atrophy is a common neurologic disorder of infants and children with a high mortality rate. Clinical trials have not been attempted in this population until recently. **OBJECTIVE:** To demonstrate that 4 outcome measures are reliable for use in clinical trials in patients with spinal muscular atrophy. **DESIGN, SETTING, PATIENTS**: Thirty-eight children with spinal muscular atrophy who fulfilled inclusion and exclusion criteria were enrolled at 5 pediatric centers for a reliability study. Paired samples statistics were performed comparing results of the qualifying variance visit with a fourth visit. **MAIN OUTCOME MEASURES: Quantitative muscle testing and the Gross Motor Function Measure.** RESULTS: Thirty-four patients and 7 evaluators completed the study. Thirteen patients were aged 2 through 4 years and 21 were 5 through 17 years. The Gross Motor Function Measure was completed by 34 subjects. Six variables for pulmonary function tests were measured in 20 subjects. Quantitative muscle testing was performed on 21 subjects in 8 muscle groups. Thirty-three subjects completed the PedsQL Neuromuscular Module for Parents. **The intraclass correlation coefficient and Bradley-Blackwood procedures indicated a very high level of agreement between measures**. **CONCLUSION: The Gross Motor Function Measure, pulmonary function tests, quantitative muscle testing, and quality of life are reliable outcome measures for clinical trials in pediatric spinal muscular atrophy.**
The reliability of upper- and lower-extremity strength testing in a community survey of older adults, Ottenbacher KJ, Branch LG, Ray L, Gonzales VA, Peek MK, Hinman MR.	*Arch Phys Med Rehabil.* 2002 Oct;83(10):1423-7. **OBJECTIVE:** To examine the stability (test-retest reliability) of strength measures in older adults obtained by nontherapist lay examiners by using a hand-held portable muscle testing device (Nicholas Manual Muscle Tester). **DESIGN:** A prospective relational design was used to collect test-retest data for 1 male subject by using 27 lay raters who completed

-- Division of Rehabilitation Sciences, Sealy Center on Aging, University of Texas Medical Branch, Galveston, TX 77555-1028, USA. kottenbo@utmb.edu	intensive training in manual muscle. **SETTING:** Data were collected from older Mexican-American adults living in the community. **PARTICIPANTS:** Twenty-seven lay raters who completed intensive training in manual muscle testing for a field-based assessment and interview of older adults and 63 Mexican-American subjects completing wave 4 of the Hispanic Established Populations for the Epidemiologic Study of the Elderly. **INTERVENTIONS:** Training involved reviewing a manual describing each testing position followed by approximately 6 hours of instruction and practice supervised by an experienced physical therapist. Lay raters then collected test-retest information on older Mexican-American subjects. **MAIN OUTCOME MEASURE:** Stability (test-retest) for a portable manual muscle testing device. **RESULTS:** Intraclass correlation coefficients (ICCs) were computed for the 27 lay raters examining 1 male subject (2 trials) and 12 lay raters assessing 63 older Mexican-American adults (1 practice and 2 trials recorded). The ICC values for the first 27 lay raters ranged from .74 to.96. The ICC values for the latter 12 lay raters ranged from .87 to.98. No differences were found in ICC values between male or female subjects. **CONCLUSIONS:** Stable and consistent information for upper- and lower-extremity strength was collected from the older adults participating in this study. The results suggest reliable information can be obtained by lay raters using a portable manual muscle testing device if the examiners receive intensive training.
Clinical evaluator reliability for quantitative and manual muscle testing measures of strength in children, Escolar DM, Henricson EK, Mayhew J, Florence J, Leshner R, Patel KM, Clemens PR.	*Muscle Nerve.* 2001 Jun;24(6):787-93. **Abstract:** Measurements of muscle strength in clinical trials of Duchenne muscular dystrophy have relied heavily on manual muscle testing (MMT). The high level of intra- and interrater variability of MMT compromises clinical study results. We compared the reliability of 12 clinical evaluators in performing MMT and quantitative muscle testing (QMT) on 12 children with muscular dystrophy. QMT was reliable, with an interclass correlation coefficient (ICC) of >0.9 for biceps and grip strength, and >0.8 for quadriceps strength. Training of both subjects and evaluators was easily accomplished. MMT was not as reliable, and required repeated training of evaluators to bring all groups to an ICC >0.75 for shoulder abduction, elbow and hip flexion, knee extension, and ankle dorsiflexion. We conclude that QMT shows greater reliability and is easier to implement than MMT. Consequently, QMT will be a superior measure of strength for use in pediatric, neuromuscular, multicenter clinical trials. **Comment:** The I.C.A.K. has always insisted that muscle testing is an art form that is easy to learn but difficult to master. It is the key to diagnostic success in AK. The doctor unable to distinguish the change in muscle strength caused by challenge to the patient's body is reduced in proportion to his ability to diagnose a patient's status. Accurate, consistent, and reproducible MMT is the most important physical talent an AK practitioner will ever develop. This study showed that with training, even previously untrained manual muscle testers could be brought up to a statistically reliable interclass coefficient. Diagnosis via manual muscle testing requires in depth training and consistent application to achieve mastery.
The Clinical Utility of Force/Displacement Analysis of Muscle Testing in Applied Kinesiology, Caruso, W., Leisman, G.	*International Journal of Neuroscience*. 2001; 106:147-157. This study provided a physical record of the phenomena in an AK muscle test. The record allowed the observer to distinguish between conditionally inhibited and conditionally facilitated muscles. This study demonstrates that the difference between these states of muscle function is quantifiable. The authors suggest however that unlike the X-ray of the radiologist and the histological specimen of the clinical pathologist, the objective outcome of an AK muscle test will not be the source of the AK practitioner's judgment; that is, he will continue to rely on his trained perception of the event that produces the record. But the record (conditionally inhibited or conditionally facilitated) will stand after the fact as a piece of objective evidence that others may examine in order to confirm the practitioner's judgment.

Measuring knee extensor muscle strength, Bohannon RW. -- Department of Physical Therapy, School of Allied Health, University of Connecticut, Storrs 06269-2101, USA.	*Am J Phys Med Rehabil.* 2001 Jan;80(1):13-8. **OBJECTIVE:** To compare manual muscle test with hand-held dynamometer measurements of knee extension strength. A secondary analysis of measurements (n = 256 knees) from 128 acute rehabilitation patients was performed. **DESIGN:** Knee extensor muscle testing was conducted according to the technique of Hislop and Montgomery; 0 to 5 grades were converted to an expanded 0 to 12 scale. Dynamometry was used to measure the isometric knee extension force with 'gravity eliminated.' **RESULTS:** Manual muscle test and dynamometer measures were highly correlated (r = 0.768; P < 0.001); the correlation was higher when the quadratic nature of the relationship was taken into account (R = 0.887; P < 0.001). Although the dynamometer forces that were associated with different manual muscle test grades differed overall (F = 67.736; P < 0.001), the forces associated with some of the higher grades did not differ statistically. **CONCLUSIONS:** These findings reinforce the convergent construct validity of the manual muscle test and dynamometry measurements but challenge the discriminant construct validity of manual muscle testing. An alternative manual muscle testing grading scheme is suggested that provides for discriminant validity and retains convergent validity.
A Force/Displacement Analysis of Muscle Testing, Caruso, B., Leisman, G.	*Perceptual and Motor Skills.* 2000; 91:683-692. Using a force transducer developed by Dr. Caruso, this study demonstrated the difference between muscles that the examiners perceived to be "weak" or inhibited, and those perceived to be "strong" or facilitated. This study also demonstrated that the muscle tests of examiners with over five years of clinical experience using AK procedures had reliability and reproducibility when their outcomes were compared. Also, the perception of inhibition or facilitation made by the examiner was corroborated by test pressure analysis using the instrumentation developed.
Muscle force measured using "break" testing with a hand-held myometer in normal subjects aged 20 to 69 years, Phillips BA, Lo SK, Mastaglia FL. -- Centre for Neuromuscular and Neurological Disorders, University of Western Australia, Australian Neuromuscular Research Institute, Perth.	*Arch Phys Med Rehabil.* 2000 Oct;81(10):1442-3. **OBJECTIVE:** To measure the strength of 17 muscle groups in the upper and lower extremities in a large group of healthy subjects using "break" testing with a hand-held myometer, and to examine the intrasession and intersession reliability of the testing protocol. **SUBJECTS AND INSTRUMENTATION:** A convenience sample of 20 men and 20 women in each decade of age from 20 to 69 years (n = 200) was tested using a Penny & Giles hand-held myometer. **RESULTS:** Reliability coefficients were >.85 for both intrasession and intersession reliability, except for the ankle dorsiflexors. Men exerted a significantly greater force than women for all muscle groups. Age, weight, and side of testing were significant predictors of force in the majority of muscle groups. The fifth percentile values, as the lower limit of normal, are reported separately for gender and side of testing for each decade of age. **CONCLUSION:** Using the testing protocol specified in this study, data from patients with various neuromuscular diseases may be compared with the appropriate gender- and age-matched normal data to accurately identify the presence of weakness.
Hand-held dynamometry reliability in persons with neuropathic weakness, Kilmer DD, McCrory MA, Wright NC, Rosko RA, Kim HR, Aitkens SG. -- Department of Physical Medicine and Rehabilitation,	*Arch Phys Med Rehabil.* 2000 Nov;81(11):1538-9. **OBJECTIVE:** To determine test-retest reliability of hand-held dynamometry (HHD) in measuring strength of persons with neuropathic weakness. **DESIGN:** Intratester and intertester reliability of HHD-measured strength over a 7- to 10-day period. In addition, HHD knee strength was compared with criterion standard of fixed dynamometry (FD). **SETTING:** Human performance laboratory of a university. PARTICIPANTS: Convenience sample of ambulatory outpatients with Hereditary Motor and Sensory Neuropathy, Type I (HMSN) (n = 10) and able-bodied controls (CTL) (n = 11). **MAIN OUTCOME MEASURE:** Maximal isometric torque. **RESULTS:** Intratester intraclass

School of Medicine, University of California, Davis, USA.	correlation coefficients (ICCs) were high, generally ranging from .82 to .96 for HHD- and FD-measured strength for both HMSN and CTL groups. There were no significant differences between sessions for HHD-measured strength, while FD-measured strength was only significantly different for knee extension ($p < .01$). Intertester reliability was generally good for both HHD- and FD-measured strength, with ICCs ranging from .72 to .97 for HMSN and CTL groups. Exceptions were knee extensors and ankle dorsiflexors for the CTL group. Knee extensor strength was significantly lower measured by HHD compared with FD ($p < .01$), but knee flexor strength was similar for the two methods. **CONCLUSION:** HHD appears to be a reliable method to measure maximal isometric strength in persons with neurogenic weakness, and may be useful to quickly and objectively evaluate strength in the clinical setting.
Interexaminer Agreement for Applied Kinesiology Manual Muscle Testing, Lawson, A., Calderon, L.	*Perceptual and Motor Skills.* 1997; 84:539-546. This study demonstrated significant interexaminer reliability for individual tests of the pectoralis major and piriformis muscles, but not for the tensor fascia lata or hamstring, which are essentially tests of groups of muscles at once. The primary importance of this study is that it demonstrates the reliability and reproducibility of muscle testing as a clinical tool, while also highlighting the need for clinicians to be aware of potential inaccuracies involved with the testing of some muscle groups.
Grade 4 in manual muscle testing: the problem with submaximal strength assessment, Dvir Z.	*Clin Rehabil.* 1997 Feb;11(1):36-41. **OBJECTIVE:** To compare the static moment of force required for a muscle group to support a limb segment against gravity with the maximal dynamic moment it can generate. **DESIGN:** Based on anthropometric measures of both sexes and theoretical calculations, the estimated anti-gravity static muscular moments (MGM) at the shoulder, elbow, hip and knee joints were compared with published data relating to the isokinetic strength (MIM) of the same muscle groups. RESULTS: The ratio of static to dynamic moment, MGM/MIM, was drastically higher in muscles operating on the proximal compared with the more distal joints. In women, the values of this ratio in the shoulder, hip, elbow and knee muscles were 7-27%, 5-65%, 7% and 5-10% respectively. The corresponding figures in men were 7-21%, 4-44%, 8-10% and 5-8%. The ratios relating to the abductors, flexors and extensors of the hip joint were substantially higher in women than in men. **CONCLUSIONS:** Since MGM and MIM correspond to grades 3 and 5 in manual muscle testing, ***the findings of this theoretical analysis indicate that elbow and knee muscles assessed as having grade 4 may generate as low as 10% of their maximal strength. With regard to shoulder and hip muscles the corresponding values are typically around 20% and 30-40%*** Coupled with the very limited human precision in sensing of force, these findings indicate that where quantitative targets in muscle strength conditioning are set or when an accurate measure of impairment is being sought, grade 4 cannot and should not serve as a valid criterion.
The manual muscle examination for rotator cuff strength. An electromyographic investigation, Kelly BT, Kadrmas WR, Speer KP.	*Am J Sports Med.* 1996 Sep-Oct;24(5):581-8. **Abstract:** The electromyographic activity of eight muscles of the rotator cuff and shoulder girdle (supraspinatus, infraspinatus, subscapularis, pectoralis, latissimus dorsi, and the anterior, middle, and posterior deltoid) was measured from the nondominant shoulders of 11 subjects during a series of 29 isometric contractions. The contractions simulated different positions used for strength testing of the rotator cuff and involved elevation, external rotation, and internal rotation at three degrees of initial humeral rotation (-45 degrees of internal rotation, 0 degree, +45 degrees of external rotation) and scapular elevation (0 degree, 45 degrees, 90 degrees). Isolation of the supraspinatus muscle was best achieved with the test position of elevation at 90 degrees of scapular elevation and +45 degrees (external rotation) of humeral rotation. Isolation of the infraspinatus muscle was best achieved with external rotation at 0 degree of scapular elevation and -45 degrees

	(internal rotation) of humeral rotation. Isolation of the subscapularis muscle was best achieved with the Gerber push-off test. This study used four criteria for identifying the optimal manual muscle test for each rotator cuff muscle: 1) maximal activation of the cuff muscle, 2) minimal contribution from involved shoulder synergists, 3) minimal provocation of pain, and 4) good test-retest reliability. Based on the results of this study and known painful arcs of motion, an objective identification of the optimal tests for the manual muscle testing of the cuff was elucidated.
A preliminary inquiry into manual muscle testing response in phobic and control subjects exposed to threatening stimuli	*J Manipulative Physiol Ther.* 1996 Jun;19(5):310-6. **Objective:** To determine phobic and non-phobic subject response to a provocative threat stimulus and to determine variables that confound the response. **Design:** Randomized blind examiner test-retest of randomized phobic and control subjects with qualitative, semistructured, information postintervention interview. **Setting:** Private chiropractic clinic. Subjects: Thirteen phobic individuals, as determined by the Diagnostic and Statistical Manual of Mental Disorders, Third Edition – Revised (DSM-III-R), and 14 control volunteer subjects. **Intervention:** Manual muscle testing was performed while each subject viewed a threat stimulus (i.e., a cue word on a printed card). The results were recorded as "weak" or "strong." **Results:** The analysis of the data demonstrates poor inter- (K = -0.19) and intraexaminer reliability (K = -0.14- +0.29). **The test for independence for valid muscle testing was strong for both examiners (p = .462, p = 1.00) When confounding variables were corrected for, the validity of muscle testing increased to 91%. Conclusion:** This preliminary inquiry demonstrates the need for musculoskeletal, attentional and presensitized subject variables to be controlled to ascertain if muscle testing can be reliably used as a tool to identify emotional arousal.
Comparison of a hand-held and fixed dynamometer in measuring strength of patients with neuromuscular disease, Brinkmann JR.	*J Orthop Sports Phys Ther.* 1994 Feb;19(2):100-4. **Abstract:** While numerous studies report acceptable reliability of hand-held dynamometers, very little information is available on factors affecting measurements and comparisons with other force measurement systems. A hand-held dynamometer was compared to a fixed dynamometer to determine if the two systems of force measurement yielded comparable results. Twenty-one patients with neuromuscular disease were measured for maximal isometric strength of 12 muscle groups with both force measurement systems using standardized positioning and stabilization procedures. Only one of the 12 muscle groups tested demonstrated significantly different force measurements between the two systems. Good association was found between both systems in force measurements, with Pearson correlation coefficients ranging from .76 to .90. We conclude that a hand-held dynamometer and a fixed dynamometer yield comparable results in patients with neuromuscular disease, provided that testing is limited to muscle groups producing relatively low forces.
Isokinetic Muscle Testing: Is It Clinically Useful?, Almekinders LC, Oman J.	*J Am Acad Orthop Surg*, 1994 Jul;2(4):221-225. **Abstract**: The use of computer-driven muscle-testing devices has become increasingly popular during the past two decades. This expensive equipment allows evaluation of muscles and muscle groups in an isokinetic manner. Isokinetic muscle testing is performed with a constant speed of angular motion but variable resistance. Isokinetic dynamometers have been shown to produce relatively reliable data when testing simple, uniaxial joints, such as the knee, as well as when testing the spine in flexion and extension. Isokinetic strength data are generally not helpful in the diagnosis of orthopedic abnormalities. Isokinetic testing can be helpful during the rehabilitation of orthopedic patients, since it allows easy monitoring of progress. It also enables the patient to work on muscle rehabilitation in a controlled manner at higher speeds than are possible with more conventional exercise equipment. An isokinetic rehabilitation program can be easily

	tailored with concentric and eccentric components that closely resemble muscle actions during occupational and sports activities.
The manual muscle examination for rotator cuff strength. An electromyographic investigation, Kelly BT, Kadrmas WR, Speer KP.	*Am J Sports Med.* 1996 Sep-Oct;24(5):581-8. **Abstract:** The electromyographic activity of eight muscles of the rotator cuff and shoulder girdle (supraspinatus, infraspinatus, subscapularis, pectoralis, latissimus dorsi, and the anterior, middle, and posterior deltoid) was measured from the nondominant shoulders of 11 subjects during a series of 29 isometric contractions. The contractions simulated different positions used for strength testing of the rotator cuff and involved elevation, external rotation, and internal rotation at three degrees of initial humeral rotation (-45 degrees of internal rotation, 0 degree, +45 degrees of external rotation) and scapular elevation (0 degree, 45 degrees, 90 degrees). Isolation of the supraspinatus muscle was best achieved with the test position of elevation at 90 degrees of scapular elevation and +45 degrees (external rotation) of humeral rotation. Isolation of the infraspinatus muscle was best achieved with external rotation at 0 degree of scapular elevation and -45 degrees (internal rotation) of humeral rotation. Isolation of the subscapularis muscle was best achieved with the Gerber push-off test. This study used four criteria for identifying the optimal manual muscle test for each rotator cuff muscle: 1) maximal activation of the cuff muscle, 2) minimal contribution from involved shoulder synergists, 3) minimal provocation of pain, and 4) good test-retest reliability. Based on the results of this study and known painful arcs of motion, an objective identification of the optimal tests for the manual muscle testing of the cuff was elucidated.
Intrarater reliability of manual muscle test (Medical Research Council scale) grades in Duchenne's muscular dystrophy, Florence JM, Pandya S, King WM, Robison JD, Baty J, Miller JP, Schierbecker J, Signore LC.	*Phys Ther.* 1992 Feb;72(2):115-22; discussion 122-6 **Abstract: The purpose of this study was to document the intrarater reliability of manual muscle test (MMT) grades in assessing muscle strength in patients with Duchenne's muscular dystrophy (DMD).** Subjects were 102 boys, aged 5 to 15 years, who were participating in a double-blind, multicenter trial to document the effects of prednisone on muscle strength in patients with DMD. Four physical therapists participated in the study. Two identical (duplicate) evaluations were performed within 5 days of each other by the same examiner initially and after 6 and 12 months of treatment. A total of 18 muscle groups were tested on each patient, 16 of them bilaterally, using a modification of the Medical Research Council scale. Reliability of muscle strength grades obtained for individual muscle groups and of individual muscle strength grades was analyzed using Cohen's weighted Kappa. **The reliability of grades for individual muscle groups ranged from .65 to .93, with the proximal muscles having the higher reliability values. The reliability of individual muscle strength grades ranged from .80 to .99, with those in the gravity-eliminated range scoring the highest. We conclude the MMT grades are reliable for assessing muscle strength in boys with DMD when consecutive evaluations are performed by the same physical therapist.**
Relationship between two measures of upper extremity strength: manual muscle test compared to hand-held myometry, Schwartz S, Cohen ME, Herbison GJ, Shah A.	*Arch Phys Med Rehabil.* 1992 Nov;73(11):1063-8. **Abstract:** One hundred and twenty-two individuals with spinal cord injuries at levels C4-6, Frankel classifications A through D, were evaluated to determine the relationship between the manual muscle test (MMT) and hand-held myometry as accurate methods for measuring muscular strength. More specifically, this study attempted to define a range of myometry scores that could be correlated with discrete MMT grades. It also investigated which of the two modalities (MMT or hand-held myometry) is the best reflection of improvement in muscle strength over time. Sequential motor strength examinations using both modalities were performed at 72 hours, one week, and two weeks post SCI and then one, two, three, four, six, 12, 18, and 24 months post injury. The data analyses included calculations of Spearman ranked correlations, analyses of variance, and linear regressions.

	Results showed that 22 of 24 correlations between MMT and myometry were significant at p values less than .001. The range of myometry measurements for a particular MMT grade appears to be most specific for MMT scores less than 4 (i.e., poor-plus to good), and less specific for MMT scores greater than or equal to 4. The results of this study also indicate that myometry measurements detect increases in strength over time, which are not reflected by changes in MMT scores.
Objective Measurement of Proprioceptive Technique Consequences on Muscular Maximal Voluntary Contraction During Manual Muscle Testing, Perot, C., Meldener, R., Gouble, F. -- Departement de genie biologique, URA CNRS 858, Universite de technologie, Compiegne.	*Agressologie*. 1991; 32(10):471-474. This study measured the electrical activity in muscles. It established that there was a significant difference in electrical activity in the muscle, and that this corresponded with the difference found between “strong” versus “weak” muscle testing outcomes by AK practitioners. It further established that these outcomes were not attributable to increased or decreased testing force from the doctor during the tests. In addition, the study showed that manual treatment methods used by AK practitioners to reduce the level of tone of spindle cells in the muscle are in fact capable of creating a reduction in tone of the muscle, as had been observed clinically. Response of Tibialis anterior muscle to a "proprioceptive technique" used in applied kinesiology was investigated during manual muscle testing using a graphical registration of both mechanical and electromyographic parameters. Experiments were conducted blind on ten subjects. Each subject was tested ten times, five as reference, five after proprioceptive technique application reputed to be inhibitory. Results indicated that when examiner-subject coordination was good an inhibition was easily registered. Therefore reliability of the proposed procedure is mostly dependent upon satisfactory subject-examiner coordination that is also necessary in standard clinical manual muscle testing.
Reliability of testing measures in Duchenne or Becker muscular dystrophy, Barr AE, Diamond BE, Wade CK, Harashima T, Pecorella WA, Potts CC, Rosenthal H, Fleiss JL, McMahon DJ.	*Arch Phys Med Rehabil.* 1991 Apr;72(5):315-9. **Abstract:** In a multiinstitutional collaborative study, we ascertained the interevaluator and intraevaluator reliability of six physical therapists who performed assessment measures on 36 boys (11.7 +/- 3.9 years) with Duchenne or Becker muscular dystrophy. Upper and lower extremities were evaluated by manual muscle testing for function, range of motion, and strength. The data were analyzed using intraclass correlation coefficients (ICCs). For the interevaluator phase, ICCs were as follows: **average muscle strength, .90; range of motion, .76; and upper extremity functional performance, .58. For the intraevaluator phase, corresponding ICCs were .80 to .96; .33 to .97; .34 to 1.00. Our results confirm and extend observations by others that these assessment measures are sufficiently reliable for use in a multiinstitutional collaborative effort. Such results can be used to design clinical trials that have sufficient statistical power to detect changes in the rate of disease progression.** Investigators planning clinical trials in a multiinstitutional collaborative setting should first standardize the assessment methods, provide evaluator training, and document reliability.
Reliability of Manual Muscle Testing with a Computerized Dynamometer, Hsieh, C.Y., Phillips, R.B.	*Journal of Manipulative and Physiological Therapeutics*. 1990; 13:72-82. **Abstract:** The purpose of this study was to investigate the reliability of manual dynamometry. Three testers participated and performed the doctor-and-patient-initiated testing methods as described in the applied kinesiology literature. Three muscles from each

	subject were tested. Fifteen normal volunteer adults had their muscles tested by the doctor-initiated method and another and another 15 had their muscles tested by the patient-initiated method. Each tester took two observations per muscle. The testing procedures were repeated 7 days later. The results showed that the intratester reliability coefficients were 0.55, 0.75 and 0.76 for testers 1, 2 and 3, respectively, when the doctor-initiated method was used; 0.96, 0.99 and 0.97 when the patient-initiated method was used. The intertester reliability coefficients were 0.77 and 0.59 on day 1 and day 2, respectively, for the doctor-initiated method; 0.95 and 0.96 for the patient-initiated method. It is concluded that manual dynamometry is an acceptable procedure for the patient-initiated method and is not acceptable for the doctor-initiated method. **Comment:** Numerous hand-held dynamometers have been developed to quantify the manual muscle test. These are units that are interposed between the examiner's hand and the subject's limb being tested. There is a constant effort to upgrade the hand-held units. These units can measure many aspects of the manual muscle test, but none of them has all of the measuring capacities that are in each of them separately. To date none of these units have been capable of measuring the manual muscle test as used in applied kinesiology with consistent reliability. This study does show a correlation, but it does not validate the complete system of manual muscle testing as used in AK. It appears that the major difference between testing against fixed transducers – whether isometric or concentric – is that the muscle is required to simply produce power; in manual muscle testing, the muscle is required to adapt to the changing pressure of the examiner's force. This requires effective functioning in the gamma system adjusting the neuromuscular spindle cell, and proper interpretation of its afferent supply and response by the nervous system.
Manual muscle testing, Mendell JR, Florence J.	*Muscle Nerve.* 1990;13 Suppl:S16-20
Somatosensory Evoked Potential Changes During Muscle Testing, Leisman, G., Shambaugh, P., Ferentz, A.	*International Journal of Neuroscience.* 1989; 45:143-151. This study measured the way the central nervous system is functioning when muscles test strong versus when they test weak. Clear, consistent and predictable differences were identified in the central nervous system between weak and strong muscle test outcomes. This supports the idea that manual muscle testing outcome changes reflect changes in the central nervous system.
Reliability of quantitative muscle testing in healthy children and in children with Duchenne muscular dystrophy using a hand-held dynamometer, Stuberg WA, Metcalf WK.	*Phys Ther.* 1988 Jun;68(6):977-82. **Abstract:** The purpose of this study was to examine intratester and test-retest reliability using a hand-held dynamometer for the measurement of isometric muscle strength in 28 healthy children and children with Duchenne muscular dystrophy. The Dystrophic Group consisted of 14 children diagnosed with Duchenne muscular dystrophy, and the Healthy Group consisted of 14 age-matched children with no history of orthopedic or neuromuscular disorders. One physical therapist tested hip and knee extension, elbow flexion, and shoulder abduction in each child bilaterally. A two-way analysis of variance for repeated measures was used to analyze differences between measurements taken within and across the testing sessions. Pearson product-moment correlation coefficients were determined on mean values across the testing sessions for each variable. No significant differences (p greater than .05) between measurements taken within or across testing sessions were found in either the Dystrophic Group or the Healthy Group. Correlation coefficients for the Dystrophic Group ranged from .83 to .99 for the variables tested. Correlation coefficients for the Healthy Group ranged from .74 to .99. The results suggest that the hand-held dynamometer can be used as a reliable instrument in measuring the isometric strength of selected muscles in children.
Intrarater reliability of manual	*Phys Ther,* 1987 Sep;67(9):1342-7.

muscle testing and hand-held dynametric muscle testing, Wadsworth CT, Krishnan R, Sear M, Harrold J, Nielsen DH.	**Abstract:** Physical therapists require an accurate, reliable method for measuring muscle strength. They often use manual muscle testing or hand-held dynametric muscle testing (DMT), but few studies document the reliability of MMT or compare the reliability of the two types of testing. We designed this study to determine the intrarater reliability of MMT and DMT. A physical therapist performed manual and dynametric strength tests of the same five muscle groups on 11 patients and then repeated the tests two days later. The correlation coefficients were high and significantly different from zero for four muscle groups tested dynametrically and for two muscle groups tested manually. The test-retest reliability coefficients for two muscle groups tested manually could not be calculated because the values between subjects were identical. We concluded that both MMT and DMT are reliable testing methods, given the conditions described in this study. Both testing methods have specific applications and limitations, which we discuss.
Clinical Reliability of Manual Muscle Testing, Frese, E., Brown, M., Norton, B.J. .	*Phys Ther.* 1987; 67:1072-1076. **Abstract:** The purposes of this study were to develop a protocol to examine the reliability of manual muscle testing in a clinical setting and to use that protocol to assess the interrater reliability of manually testing the strength of the middle trapezius and gluteus medius muscles. One hundred ten patients with various diagnoses participated as subjects, and 11 physical therapists participated as examiners in this study. The results showed that interrater reliability for right and left middle trapezius and gluteus medius muscles were low. The percentage of therapists obtaining a rating of the same grade or within one third of a grade ranged from 50% to 60% for the four muscles. This study indicates that using manual muscle testing to make accurate clinical assessments of patient status is of questionable value. **Comment:** This study demonstrates that manual muscle testing as a diagnostic tool is only as good as the operator conducting the test. The study notes that the 11 physical therapists who were the examiners were recent graduates of a physical therapy program, with only an average of 2.3 years of clinical experience. Two different types of manual muscle testing procedures were also employed during this study, thereby increasing the variability of outcomes. The methods of manual muscle testing used in this study were not those taught by the I.C.A.K., and the numerous variables in a manual muscle test (patient positioning, accuracy of timing during the test, and consistency of the type of testing done on the patients by the examiners) were not accounted for, any one of which may influence the perception of strength or weakness on testing.
Finger flexion function in rheumatoid arthritis: the reliability of eight simple tests, Armstrong R, Horrocks A, Rickman S, Heinrich I, Kay A, Gibson T.	*Br J Rheumatol* 1987;26:118-122. The inter- and intra-observer errors of eight tests of finger flexion function were estimated from the results obtained by three observers assessing 10 patients with rheumatoid hand involvement. Measurements of finger flexion and muscle power involved both conventional and novel techniques using simple and easily constructed apparatus. For each test, measurements were in agreement between observers and were reproducible on three occasions. These tests may now be used with confidence by other investigators.
Reliability of lumbar paravertebral EMG assessment in chronic low back pain, Ahern DK, Follick MJ, Council JR, Laser-Wolston N.	*Arch Phys Med Rehabil.* 1986 Oct;67(10):762-5. **Abstract:** The reliability of lumbar paravertebral EMG assessment was investigated in a sample of 70 patients with chronic low back pain, (CLBP). Dual-site EMG monitoring was employed during both static postures and movements. Flexion and rotation indices were divided to assess the reliability of patterning of paravertebral EMG during movement. Within-session reliabilities computed for the full sample ranged from 0.66 to 0.97, and between-session reliabilities, computed on a subset of 29 patients retested after varying intervals, ranged from 0.26 to 0.92. Average EMG levels, flexion, and rotation indices

	showed no statistically significant differences between surgical (n = 40) and nonsurgical patients (n = 30), although EMG variability was consistently greater for surgical patients across the postures and movements. **These results indicate that lumbar paravertebral EMG can be reliably measured and therefore has potential utility as an assessment and treatment variable in CLBP.**
Clinical trials in Duchenne dystrophy. Standardization and reliability of evaluation procedures, Florence JM, Pandya S, King WM, Robison JD, Signore LC, Wentzell M, Province MA.	*Phys Ther.* 1984 Jan;64(1):41-5. **Abstract:** A multiclinic, collaborative study has been designed to assess the natural progression and efficacy of treatment of Duchenne muscular dystrophy. This article describes the protocol for the evaluation technique and the method used to establish within (intraobserver) and between (interobserver) reliability of the protocol evaluation procedures. Standardized patient evaluations were used, and consistency of evaluation was monitored by a computer. The reliability of the measures was analyzed 1) within observers by comparing the results of each of the first three tests done by each evaluator for all patients and 2) between observers by comparing, at multicenter group meetings, the results of each of the four evaluators' tests of the same patient. We have demonstrated reliability for an evaluation method that will provide an objective foundation on which to claim a drug or therapeutic procedure does or does not have an effect in treating Duchenne muscular dystrophy. **Comment:** This paper showed that there was significant improvement in the degree of consistency of a given examiner's MMT scores when the examiner had more clinical experience and training in MMT. Many other researchers of MMT have discussed the importance of considering the examiner's training upon the outcomes of studies that assess strength via MMT.
The efficacy of manual assessment of muscle strength using a new device, Marino M, Nicholas JA, Gleim GW, Rosenthal P, Nicholas SJ.	*Am J Sports Med.* 1982 Nov-Dec;10(6):360-4 **Abstract:** The purpose of this study was to compare the manual assessment of muscle strength with a small, handheld (by the examiner) force-measuring device developed by the Institute of Sports Medicine and Athletic Trauma (ISMAT). One hundred twenty-eight patients presented with a known lower extremity orthopaedic pathology. All patients were clinically evaluated for hip abductor and hip flexor weakness in standard positions using the "break test" technique. All 128 patients were then evaluated with the ISMAT Manual Muscle Tester, a small, hand-held device which recorded the peak force (kg) required to break a muscle contraction. Three bilateral measures of hip abduction and hip flexion were recorded, averaged, and compared to the subjective clinical evaluation using a chisquare analysis. **Bilateral values which were within 5% of each other were not considered significant and therefore not included in the calculations. The average hip abduction and hip flexion scores measured by the ISMAT tester were consistent with the examiner's perception of muscle weakness (P less than 0.001). The results demonstrate consistent detection of muscle weakness by the ISMAT Manual Muscle tester over a broad range of testing conditions.**
Applied kinesiology: an experimental evaluation by double blind methodology, Jacobs, G.	*J Manipulative Physiol Ther,* 1981;4:141-145. **Abstract:** The object was to develop a double blind experiment for testing the premise of a muscle testing procedure referred to as Applied Kinesiology (A.K.). In a non-blind test there was a significant difference (P<.05) in muscle response to lump sugar versus the same amount in a 10 ml solution of distilled H_20. It was not possible to demonstrate significant differences in response to sweet and non-sweet sugar solutions of various concentrations, to fresh sesame oil and heated, or to fresh corn oil and aged corn oil when stimuli were applied in a double blind experiment. It was concluded on the basis of the parameters of this study that the AK response generally expected did not occur. **Comment:** This study demonstrated a random muscle weakening on gustatory stimulation

	with refined sugar. The “AK response generally expected” was that muscles would weaken upon gustatory stimulation with refined sugar. This is an example of experimental bias against sugar. In some patients sugar will cause improved muscle function depending on the patient’s physiological status at the time. None of the literature from the I.C.A.K. suggests that all individuals weaken on oral stimulation with refined sugars. **This double-blind study did demonstrate, however, an 81.9% agreement between two testers, indicating good inter-examiner reliability.**
An Experimental Evaluation of Kinesiology in Allergy and Deficiency Disease Diagnosis, Scopp, A.	*Journal of Orthomolecular Psychiatry*. 1979; 7(2):137-8. **Abstract:** An accurate, reliable, and quick method for determining vitamin-mineral imbalances and food allergies is critical in establishing a balanced behavioral biochemistry. Dr. George Goodheart has clinically observed that whenever a patient was deficient in a specific vitamin or food factor, a specific muscle which he found to be associated with the nutrient was almost always unilaterally weak (weak on right or left side of the body) when the muscle strength was tested using kinesiology procedures (Goodheart, 1976). Goodheart and coworkers have established an association between approximately 50 vitamins and minerals on the one hand, and 50 associated muscles which are weak when the nutrient is deficient. According to Goodheart, chewing or ingestion of the critical food factor results in dramatic restoration of muscle strength within 10 seconds. The purpose of this study was to test some of Goodheart’s observations in a more formal manner and to adapt his principles to cerebral allergy testing. 10 naïve subjects were given 10 muscles tests by 6 trained testers. **Pearson Product-Moment Correlation between testers was .91, suggesting that muscle testing is reliable between testers.** Subjects with unilateral weak muscles were then given either a placebo or the nutrient which Goodheart believes to be associated with the unilateral muscle. The increase in muscle tone measured approximately 10 seconds after ingestion was 21% for the nutrient group and was statistically significant ($p < .05$) increase in comparison with the placebo group. Muscle tone was measured by a Jaymar dynamometer with the muscle tested according to kinesiology procedures described by Kendall and Kendall. In the cerebral allergy testing part of the study, a 15% decrease in muscle tone of the pectoralis major clavicular was used as the criterion for cerebral allergy. The muscle testing method was then compared to results obtained by a Philpott-type fast with progressive reintroduction of foods**. Correlation between foods identified as provocative by muscle testing and by the fast was .81. Observation of clinical results obtained with muscle testing suggests the method has substantial clinical utility.**
Factors influencing manual muscle tests in physical therapy, Nicholas JA, Sapega A, Kraus H, Webb JN.	*J Bone Joint Surg Am.* 1978 Mar;60(2):186-90 **Abstract:** To determine whether it is the amount or the duration of the force applied manually by the tester, or both, that determines the tester's perception of the strength of the hip flexor or abductor muscles, an electromechanical device was designed which was placed between the tester's hand and the subject's limb. With the device we measured the force applied to the limb, the time interval during which it was applied, and the angular position of the limb during the entire test. In 240 such tests, the testers' ratings of the differences in strength between the right and left sides were correlated with seven variables involving force and time. It was found statistically that the impluse--that is, the duration of the tester's effort multiplied by the average applied force during each test--was the factor that most influenced the tester in the ratings. **Comment:** The skills of the examiner related to conducting tests and interpreting the derived information will affect the usefulness of muscle performance data. The examiner is obliged to follow a standardized protocol that specifies patient position, verbal instructions or demonstration to the patient, alignment of the muscle and direction of examiner

	resistance to insure precise, repeatable, and reliable MMT results. When a muscle is tested in voluntary isometric contraction, EMG testing reveals that additional muscle fibers contract at low forces; when the force increases, the rate of firing becomes the predominant mechanism to increase strength. Tension, velocity, and electrical activity are interdependent and indicate the importance of proper neurologic control for the muscle to meet the changing pressure demands of the MMT. This requires effective function of the gamma system adjusting the neuromuscular spindle cell, and proper interpretation of its afferent supply by the central nervous system. Thus it is patient or more precisely the patient's neuromuscular adaptive capacity that is being examined during a proper MMT.
Experimental Characterization of The Reactive Muscle Phenomenon, Triano J, Davis B.	*Chiro Econ*, Sept/Oct 1976:44-50 **Abstract:** As the authors recognize, much discussion and doubt has been cast towards the clinical impressions and teachings of applied kinesiology since its original inception. Though today more and more practitioners are utilizing these techniques, there has of yet been no clear definition of these techniques, as well as other chiropractic techniques, by scientific investigation. Muscle testing provides an excellent approach to verification of physiologic effects of chiropractic techniques through the widely accepted media of electromyography. A group of 40 subjects were analyzed and found to include 10 individuals susceptible to the reactive muscle phenomenon of Goodheart. These subjects were then studied in order to characterize the reactive phenomenon. This study demonstrated that the reactive muscle phenomenon is, in fact, a physiologic imbalance of muscle and that it is not simply a psychologic suggestion or an overpowering of the tested arm by brute force. These data have provided the first quantitative evidence, to our knowledge, that in the case of the deltoid-rhomboid interactions the clinically observed "reactive muscle" represents a real physiological phenomenon. Further studies to detect its mechanism and clarify its treatment are under way.
Further standardization of manual muscle test for clinical study: applied in chronic renal disease, Silver M, McElroy A, Morrow L, Heafner BK.	*Phys Ther.* 1970 Oct;50(10):1456-66. This study demonstrated an interexaminer reliability above 90% and a test-retest reliability above 95%.
Muscle testing. 2. Reliability in clinical use, Iddings DM, Smith LK, Spencer, WA.	*Phys Ther Rev.* 1961 Apr;41:249-56. This study demonstrated an interexaminer reliability above 90% and a test-retest reliability above 95%.
A study of the reproducibility of muscle testing and certain other aspects of muscle scoring, Lilienfeld AM, Jacobs M, Willis M.	*Phys Ther Rev.* 1954 Jun;34(6):279-89. This study demonstrated an interexaminer reliability above 90% and a test-retest reliability above 95%.