

INTEREXAMINER RELIABILITY OF THE DELTOID AND PSOAS MUSCLE TEST

Henry Pollard, DC, PhD,^a Bronwyn Lakay, MChiro,^b Frances Tucker, MChiro,^b Brett Watson, MChiro,^b and Peter Bablis, DC^b

ABSTRACT

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 multidocor assessment/management programs. (J Manipulative Physiol Ther 2005;28:52-56)

Key Indexing Terms: *Chiropractic; Muscle Test; Reliability; Psoas Muscle; Deltoid Muscle*

Muscle testing is widely used by practitioners to determine the strength of a muscle as well as determine its related neurological functioning.¹ In addition, many practitioners use muscle testing as a part of a diagnostic system related to one or more of the following: spinal hypomobility/subluxation/manipulable lesion,²⁻⁴ deficiency in nutritional status,⁵ indicator of neurological,⁶ emotional (neuroemotional technique), and other dysfunction,^{7,8} or as part of systems of treatment such as applied kinesiology (AK).⁹

All the above uses, valid or not, presuppose the reliability of the testing procedure. In a review of the English literature from 1966 to the present using the Medline and CINAHL databases with search key words “muscle test” and “applied kinesiology” and “reliability,” only a small number of papers were found on the topic. Many papers were found on the use of isokinetic machine-based muscle testing systems and their reliability.^{1,10-12} These papers compared isokinetic

muscle testing to muscle testing with and without handheld force transducers.

The literature provides examples of studies that have deemed the reliability of manual muscle testing as either good^{1,13,14} or poor.¹⁵⁻¹⁸ This work further reports the comparison of the manual muscle test with that done by a handheld force transducer concluding that both are reliable.^{19,20} Other research compares the use of isokinetic testing to that of the handheld transducer type, concluding that both are reliable, but the machine-based approach is superior to the handheld approach.¹

In addition, many examples of different muscles being the object of the testing have resulted in different reliability to the procedure of the manual approach.^{1,17} It is apparent that the reliability of the muscle test may depend on the type of muscle being tested,²¹ who is testing it (experienced or novice),^{15,22} in what position it is being tested, and with what procedure.²⁰

The muscle test procedures have been in wide clinical use, and practitioners of varying skill levels use them. Although the outcomes of testing by experienced practitioners or novice practitioners have been determined,^{13,22} no comparative testing of these 2 groups has been performed. The results of such a study could be applied to the multidocor practice that frequently use a combination of both senior and novice clinicians (new graduates) sharing the same patient population. Thus, it is the specific aim of this paper to test the interexaminer reliability of a muscle test as performed by an experienced practitioner and a novice practitioner.

^a Senior Lecturer Macquarie University, Sydney, NSW, Australia.

^b Private Practice of Chiropractic, Cronulla, NSW, Australia.

Sources of financial support: the ONE Research Foundation provided a grant to support this work.

Submit requests for reprints to: Henry Pollard, DC, PhD, The ONE Research Foundation, 1991 Village Park Way, Suite 201-A, Encinitas, CA 92024 (e-mail: hpollard@optushome.com.au).

Paper submitted June 6, 2003; accepted 1 October 2003.

0161-4754/\$30.00

Copyright © 2005 by National University of Health Sciences.

doi:10.1016/j.jmpt.2004.12.008

METHODS

Experimental Procedure

The project included 106 volunteer participants who responded to advertisements that were placed around a university campus on notice boards in various departments and other communal student areas. Of the 106 participants, 50 were men, and 56 were women. The age range of participants was between 19 and 63 years, with a mean age of 26 years. The project received approval from the Macquarie University Human Ethics Committee before commencement. Participants were requested to complete a consent form before the experimental session after the procedure had been completely explained to them. Their participation was entirely voluntary.

The testing room was arranged so that each of the testers was unable to determine the nature of the others' activities. Subjects were randomly allocated (via drawing a number from a hat) to one of the practitioners. Each practitioner began testing the muscles in reverse order (ie, one practitioner began with the deltoid and the other began with the psoas). The "novice" student researcher (R1) and the "expert" supervisor (S1) each performed the muscle testing of the anterior deltoid and psoas muscles on each participant. The novice practitioner was a final-year (5th year) chiropractic student, and the expert supervisor was a practitioner with 15 years' experience in using the technique. Participants were invited into the testing room and instructed that 2 muscle strength tests would be performed on them, 1 on their shoulder and 1 on their hip.

Description of Muscle Test Procedures

The deltoid muscle strength test was first performed by S1, with the participant sitting facing the tester, the tester standing directly in front of subject. The chosen arm was outstretched in front of the subject at right angles to their body, with their hand in a loose fist, their elbow straight, and their forearm pronated with the hand parallel to the floor. In the manual muscle test, the testers placed their hands at a location approximately 5 cm proximal to the ulnar styloid, in the midline of the forearm. The participant was asked to hold that position against pressure exerted by the tester's hand that was applied in a superior-to-inferior direction. During the test, an algometer (pain-pressure displacement measurement device) was used to determine the amount of load applied to the individual. The same displacement noted in the first test was used for all subsequent attempts at muscle testing to standardize the amount of force used in the testing. Thus, subsequent tests began by placing the algometer at the same location, and the participant was instructed to hold the position once again against pressure. R1 then conducted the upper limb (deltoid) manual muscle strength test in the same manner. The right limb was requested in all cases, unless there was an existing injury to that side. Assessors were again blinded to each other's results.

The participant was then instructed to assume a supine position for testing of psoas muscle strength. The tester lifted the chosen leg and positioned the extended leg to approximately 30° of hip flexion, 10° of abduction from the midline, and external rotation. Pressure was applied in a superior-to-inferior direction, approximately 5 cm proximal to the medial malleolus on the anteromedial border of the tibia for both the manual muscle testing (by S1 and R1) determined once again by the displacement noted with the algometer testing.

Standardization of the force in testing was important. A displacement device was used to indicate the force used on each individual. It was not the purpose of the study to determine the amount of force used in the muscle test, only to use the same amount of force. By using a displacement device (the algometer), a standardized amount of force was required to achieve the same displacement per patient.

As this study was a pragmatic study using procedures commonly used in the clinical setting, a pragmatic approach to the determination of the weak or strong response to the muscle testing was used. Subjects were indicated to have had a positive response to the muscle test if they produced a "large" subjective change (deemed to be greater than 30°) in the depression of the arm (or leg) using the same force as measured by the same degree of deformation of the algometer. The 2 practitioners subjectively noted findings independent of each other, and results were compared.

Analysis

Statistical analysis of data was made using Cohen κ statistic (used to evaluate results of interexaminer agreement). Statistical interpretation of κ statistics is similar to a correlation coefficient (ie, the highest level of agreement is 1.0, whereas the lowest level is 0). Dawson and Trapp²³ have outlined a scale for qualifying the quality of agreement between examiners. They suggest a κ score of 0.93 to 1.0 as an excellent agreement, 0.81 to 0.92 as a very good agreement, 0.61 to 0.80 as a good agreement, 0.41 to 0.60 as a fair agreement, 0.21 to 0.4 as a slight agreement, and 0 to 0.20 as a poor agreement.

RESULTS

Interexaminer Reliability

The interexaminer reliability for the test-retest protocol between the assessors was determined, and significant reliability was attained between testers for both muscles tested. S1 and R1 muscle tests for the deltoid were identical 82.08% of the time (Table 1). Both examiners recorded a 'strong' measure 57 times, and a 'weak' measure 30 times with differing measures totalling 19 times. The κ statistic was 0.62, which showed 'good' agreement on the κ scale.

The measure of both S1 and R1 for the psoas agreed 84.9% of the time (Table 2). Both examiners recorded

Table 1. *Interexaminer deltoid manual muscle testing reliability*

82.08% Agreement		S1	
R1		Strong	Weak
	Strong	57	9
	Weak	10	30

Cohen κ (unweighted): observed agreement = 82.08%; expected agreement = 53.24%; $\kappa = 0.62$ (SE = 0.10); 95% CI 0.43 to 0.81; z (for $k = 0$) = 6.35; 2-sided $P < .0001$; 1-sided $P < .0001$.

Table 2. *Interexaminer psoas manual muscle testing reliability*

84.90% Agreement		S1	
R1		Strong	Weak
	Strong	30	9
	Weak	7	60

Cohen κ (unweighted): observed agreement = 84.91%; expected agreement = 53.99%; $\kappa = 0.67$ (SE = 0.10); 95% CI 0.48 to 0.86; z (for $k = 0$) = 6.92; 2-sided $P < .0001$; 1-sided $P < .0001$.

a ‘strong’ measure 30 times, and both recorded a ‘weak’ measure 60 times with different measures recorded 16 times. The κ statistic was 0.67, which once again showed ‘good’ agreement.

DISCUSSION

The use of muscle testing for grading the response to injury is widespread and is accepted for such a purpose.¹ By contrast, the use of muscle testing as a vehicle to detect neurological and musculoskeletal functioning elsewhere other than the specific strength testing throughout the peripheral and central nervous systems is less well established.^{6,15,24,25} Much controversy surrounds the use of these procedures for such purposes.^{2,5,13,26,27} Although such a debate is a healthy call to arms for the commencement of a research agenda to investigate such procedures, it remains a lofty goal for future investigations.²⁶

Before establishing the applied use of such procedures, investigation of the basic functions of this muscle testing approach is needed. Thus, it was the purpose of this paper to investigate the base-level agreement of 2 practitioners with differing skill levels. The determination of the reliability of these commonly used muscle test procedures is important because of a need for practitioners of varying skill levels that use them. The results of this study support the use of these procedures in the multidocor practice. Such clinics frequently use combinations of experienced and inexperienced clinicians sharing the same patient population. Such a scenario is frequently encountered with experienced practitioners employing new graduates.

These skill levels were defined by the use of an intern practitioner versus an established musculoskeletal practitioner (15 years of experience). These practitioners were tested using 2 of the most common muscle testing procedures: the anterior deltoid and psoas muscle tests. These tests are instrumental to AK and other diagnostic system approaches used by manual and other therapists.⁹ Because of the prominent role that muscle testing has in the technique of AK, muscle testing reliability is often considered interchangeable with the diagnostic protocol that is AK. Such extrapolation is inappropriate and unfortunate because it leads to confusion within the literature examining both muscle testing and AK protocol.

Our 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 psoas (κ 0.67). Other researchers have examined the muscle test reliability in these muscles and have found them to have good or better level of interexaminer reliability.^{1,14} Frese et al¹⁶ found the interrater reliability of the middle trapezius and gluteus medius to be low and concluded that their use in muscle testing should be considered questionable. Ludtke et al¹⁵ when attempting to use the anterior deltoid to test subject reactions to insect venom allergy found that the muscle test was an unreliable indicator in patients with known allergy. Also, an older investigation by Grossi²⁷ determined that the muscle spindle technique used by AK practitioners (as it pertains to the isometric quadriceps strength) was also unreliable.

Thus, the results of the various studies can be described as mixed at best. Much crossover has occurred between the reliability of the testing of the actual muscle test and the validity of the application of the muscle test to diagnostic approaches. Many studies have questioned some of the diagnostic uses to which muscle testing has been put.^{4,5,8,26} Kenney et al⁵ have concluded that AK is unreliable to test nutrient status, whereas Schmitt and Leisman²⁵ in a more recent pilot study have concluded the opposite in a hypoallergenic food hypersensitivity response test using muscle testing.

In a review of chiropractic diagnostic methods, Hestbaek and Leboeuf-Yde²⁶ concluded that the manual muscle testing used by sacro-occipital technique is reliable but not the diagnostic procedure to which it is applied (as no studies had been performed). They concluded that very little documentation was available at that time to validate the practice of applied kinesiology and its use of muscle testing as a whole. The conclusion by Hestbaek and Leboeuf-Yde²⁶ is similar to a previous publication, which concluded that during the period 1981 to 1987, the published research on AK was limited by inadequately performed research that prevented the drawing of any clear conclusion as to the efficacy of the procedures.²⁴

Despite these repeated calls for quality research into what is one of the most prolific chiropractic procedures, recent researches by Caruso and Leisman^{13,14} have potentially shown a way for the production of research into the muscle

test. They may help some of the reliability and validity issues associated with the use of the muscle test.

In their recent papers, Caruso and Leisman^{13,14} investigated the force displacement data from an AK muscle test procedure. They were able to determine that experienced practitioners were able to discriminate between a weak and strong muscle by a simple force displacement analysis. They found that experienced researchers gained most of their sensory information from the leading-edge force pulse of the impulse they applied to the muscle they were testing, and the impulse could be visualized by plotting the slope of the force displacement curve (dx/dF). They determined that a weak muscle showed a large slope, whereas a strong muscle showed a small slope. In both related publications, they reasoned that the inability to document or effectively test the dx/dF of the various muscles being tested could explain (at least in part) the variability of the responses found by various researchers. In addition, findings by other researchers, such as Mulroy et al,²² have added to a potential explanation by suggesting that the outcome testing of large muscles such as the quadriceps may be determined by the strength of the assessor, and not just the target muscle or subject.

Additional to the mechanical factors affecting muscle test performance, mental factors may also affect strength testing. Monti et al²⁸ have concluded that a belief in a certain statement may affect the output of the muscle being tested. They tested the effect on muscle testing strength while the subject responds to a congruent (true) or incongruent (false) statement and found that incongruent statements results in a weaker muscle test performance than that which used congruent statements. Leisman et al²⁹ also previously showed that practice can increase the efficiency of subjective muscle testing and that a fatiguing muscle responds differently to strong or weak muscle as measured using integrated electromyography.

The results of our present investigation and that of the various studies reviewed above suggest that the use of the anterior deltoid and psoas muscles for the purpose of muscle testing may be reliable using examiners of differing skill levels. These results may support the use of muscle testing of the deltoid and psoas muscles by different practitioners on a similar population, a scenario that might be encountered in the multidocor clinic; however, more research is needed.

What cannot be concluded is the reliability of other muscles other than the 2 target muscles of this investigation, although we see no cause as to why other muscles may not be reliably tested when similar care is used. In addition, no conclusion can be made of the reliability or validity of any diagnostic system that may use these reliable muscle tests. Thus, we call for further outcomes-based research into the practice of AK and other diagnostic protocols that use the manual muscle test as a component part of their diagnostic application.

CONCLUSION

The manual muscle testing procedure using the anterior deltoid or the psoas muscle shows good test-retest reliability when used by the experienced and novice users in this study. We find that this investigation can make no claim on the validity of those diagnostic procedures used by practitioners for functions other than basic muscle testing. We suggest further research be undertaken to address the noted shortcomings in the literature on the validity of use of these procedures in conjunction with diagnostic protocols to which they are so often used.

REFERENCES

1. Escolar DM, Henricson EK, Mayhew J, Florence J, Leshner R, Patel KM, et al. Clinical evaluator reliability for quantitative and manual muscle testing measures of strength in children. *Muscle Nerve* 2001;24:787-93.
2. Motyka TM, Yanuck SF. Expanding the neurological examination using functional neurologic assessment: part I methodological considerations. *Int J Neurosci* 1999;97:61-76.
3. Schmitt WH, Yanuck SF. Expanding the neurological examination using functional neurologic assessment: part II neurologic basis of applied kinesiology. *Int J Neurosci* 1999;97:77-108.
4. Haas M, Peterson D, Hoyer D, Ross G. Muscle testing response to provocative vertebral challenge and spinal manipulation: a randomized controlled trial of construct validity. *J Manipulative Physiol Ther* 1994;17:141-8.
5. Kenney JJ, Clemens R, Forsythe KD. Applied kinesiology unreliable for assessing nutrient status. *J Am Diet Assoc* 1988;88:698-704.
6. Klein MG, Whyte J, Keenan MA, Esquenazi A, Polansky M. Changes in strength over time among polio survivors. *Arch Phys Med Rehabil* 2000;81:1059-64.
7. Jacobs GE, Franks TL, Gilman PG. Diagnosis of thyroid dysfunction: applied kinesiology compared to clinical observations and laboratory tests. *J Manipulative Physiol Ther* 1984;7:99-104.
8. Schissel MJ, Dodes JE. Dentistry and alternative therapy. *N Y State Dent J* 1997;63:32-7.
9. Walther D. Applied kinesiology: basic procedures and muscle testing. Volume 1. USA: System DC, 1981. p. 1-40.
10. Almekinders LC, Oman J. Isokinetic muscle testing: is it clinically useful? *J Am Acad Orthop Surg* 1994;2:221-5.
11. Sloan C. Review of the reliability and validity of myometry with children. *Phys Occup Ther Pediatr* 2002;22:79-93.
12. Kannus P. Isokinetic evaluation of muscular performance: implications for muscle testing and rehabilitation. *Int J Sports Med* 1994;15(Suppl. 1):S11-8.
13. Caruso W, Leisman G. A force/displacement analysis of muscle testing. *Percept Mot Skills* 2000;91:683-92.
14. Caruso W, Leisman G. The clinical utility of force/displacement analysis of muscle testing in applied kinesiology. *Int J Neurosci* 2001;106:147-57.
15. Ludtke R, Kunz B, Seeber N, Ring J. Test-retest-reliability and validity of the kinesiology muscle test. *Complement Ther Med* 2001;9:141-5.
16. Frese E, Brown M, Norton BJ. Clinical reliability of manual muscle testing. Middle trapezius and gluteus medius muscles. *Phys Ther* 1987;67:1072-6.

17. Lawson A, Calderon L. Interexaminer agreement for applied kinesiology manual muscle testing. *Percept Mot Skills* 1997;84:539-46.
18. Hsieh CY, Phillips RB. Reliability of manual muscle testing with a computerized dynamometer. *J Manipulative Physiol Ther* 1990;13:72-82.
19. Aitkens S, Lord J, Bernauer E, Fowler WM, Lieberman JS, Berck P. Relationship of manual muscle testing to objective strength measurements. *Muscle Nerve* 1989;12:173-7.
20. Wadsworth CT, Krishnan R, Sear M, Harrold J, Nielsen DH. Intrarater reliability of manual muscle testing and hand-held dynamometric muscle testing. *Phys Ther* 1987;67:1342-7.
21. Bohannon RW, Corrigan D. A broad range of forces is encompassed by the maximum manual muscle test grade of five. *Percept Mot Skills* 2000;90:747-50.
22. Mulroy SJ, Lassen KD, Chambers SH, Perry J. The ability of male and female clinicians to effectively test knee extension strength using manual muscle testing. *J Orthop Sports Phys Ther* 1997;26:192-9.
23. Dawson B, Trapp RG. *Basic & clinical biostatistics*. 3rd ed. Boston: McGraw-Hill, 2001. p. 334-43.
24. Klinkoski B, Leboeuf C. A review of the research papers published by the International College of Applied Kinesiology from 1981 to 1987. *J Manipulative Physiol Ther* 1990;13:190-4.
25. Schmitt Jr WH, Leisman G. Correlation of applied kinesiology muscle testing findings with serum immunoglobulin levels for food allergies. *Int J Neurosci* 1998;96:237-44.
26. Hestbaek L, Leboeuf-Yde C. Are chiropractic tests for the lumbo-pelvic spine reliable and valid? A systematic critical literature review. *J Manipulative Physiol Ther* 2001;24:145-6.
27. Grossi JA. Effects of an applied kinesiology technique on quadriceps femoris muscle isometric strength. *Phys Ther* 1981;61:1011-6.
28. Monti DA, Sinnott J, Marchese M, Kunkel EJ, Greeson JM. Muscle test comparisons of congruent and incongruent self-referential statements. *Percept Mot Skills* 1999;88:1019-28.
29. Leisman G, Zenhausern R, Ferentz A, Tefera T, Zemcov A. Electromyographic effects of fatigue and task repetition on the validity of estimates of strong and weak muscles in applied kinesiological muscle-testing procedures. *Percept Mot Skills* 1995;80:963-77.