Test-Retest-Reliability and Validity of the Kinesiology Muscle Test

R. Lüdtke,[1] B. Kunz,[2] N. Seeber,[2] J. Ring[3]

[1]Karl und Veronica Carstens-Stiftung, Germany [2]Hautklinik, Universitätskrankenhaus Eppendorf, Hamburg, Germany [3]Klinik und Poliklinik für Dermatologie und Allergologie am Biederstein, München, Germany

SUMMARY. Objectives: To assess the test-retest-reliability and validity of the Health Kinesiology muscle test. Patients: Seven patients with clinically and allergologically confirmed wasp venom allergy. Design: Four Health Kinesiology-examiners tested each patient in a random order for 10 verum and 10 placebo bottles. All examiners used the anterior deltoid as indicator muscle. Patients and examiners were completely blinded. Outcome Measures: Weak muscle holds were rated as 'sensitivity' towards the test substance, stable holding as normal (not sensitive). Results: An overall kappa of 0.03 (95%-CI: −0.02–0.07) indicates the test is not reliable. Individual kappas do not substantially vary from examiner to examiner. Sensitivity and specificity were estimated at 40% and 60%. Conclusions: The results suggest that the use of Health Kinesiology as a diagnostic tool is not more useful than random guessing. This should at least be true in patients with insect venom allergy that are tested by examiners with average skills.

Keywords: Applied kinesiology, muscle testing, insect venom allergy, diagnostic tests

INTRODUCTION

Applied Kinesiology (AK) is numbered among the so-called alternative medical procedures. Its fundamental principles were developed in the 60 s by G. Goodheart. AK is increasingly used and spreads more and more around the world.[1,2] One category of AK is 'Health Kinesiology' (HK), developed by J. Scott in the USA.[3] It combines AK, Chinese medicine, and acupunture with psychological knowledge.

AK is mainly (but not exclusively) used for the diagnosis (and therapy) of allergies or food intolerances. Its diagonostic element is a non-invasive muscle test.[4] Proponents of AK state that this muscle test uses a simple body feedback system. The person to be tested is just requested to hold a limb, for example an arm or leg, against a specific light pressure. The result of the test is either the muscle holds strong or is weak, that is it gives way, which is interpreted as an indicator of energetic balance or imbalance, respectively. In addition to purely clinical diagnostics, the muscle test is also used for therapeutic purposes.

Contrary to usual allergy diagnostic tests like radioallergosorbent test for detection of specific IgE antibodies (RAST) or skin tests, HK is a non-invasive method with little stress on the body. This advantage, among others, has led to the fact that proponents of HK recommended HK testing is especially in children.

Despite of the fact that the method is widespread, there is so far little scientific evidence on the value of HK. A scientific assessment seems even more important because the consequences of allergy diagnoses may have very far-reaching consequences on the daily lives of the patients, for example, rejection of protective therapies against life-threatening reactions, long term avoidance of certain foods, changes in living habits, or great investments needed for replacing incriminated furniture.

Our study was done using a model employing IgE mediated allergy to insect (wasp) venom. [5] The

Rainer Lüdtke
Karl und Veronica Carstens-Stiftung, Am Deimelsberg 36, D-45276 Essen. E-mail: r.luedtke@carstens-stiftung.de

aim was to evaluate whether the results of the HK muscle testing are reliable and valid.

SUBJECTS AND METHODS

Patients and examiners

In January 1995, seven patients were recruited from the Allergy Department and Dermatology at the University Hospital in Hamburg-Eppendorf, Germany. All of the test persons had a history of an anaphylactic reaction after insect stings.

Each of the seven patients was tested for wasp allergy using HK muscle testing by four different examiners. Because of a lack of time, only five out of seven patients could be tested by all four examiners. One patient was tested only by one examiner (labelled S) and two patients by two examiners (S and K). For the other patients all examiners participated.

The four examiners varied substantially in their skills and experiences in HK. One examiner (labelled G) had performed the muscle tests for more than eight years in his daily work as a medical doctor and paediatrician. He had passed through at least six courses in HK including those for advanced therapists. Two examiners were medical doctors and had taken standard classes to study HK (K and V). According to the Institute for Applied Kinesiology in Freiburg, Germany, these courses are sufficient enough to perform the muscle tests correctly. Both examiners had had personal experience for at least two years. The last examiner (S) was an absolute beginner who already had learned the muscle test from one of the doctors mentioned previously.

Allergy diagnostics

Allergy diagnostics followed clinically established dermatological standard procedures:[6,7] All patients were carefully examined physically, a thorough personal and family history of allergy was taken with special emphasis in insect venom anaphylaxis. Skin tests (prick and intradermal) were performed in a titration method. Specific IgE antibodis were detected using the CAR-RAST-technique (Pharmacia, Uppsala).

Study design

Each candidate was tested by each examiner with 20 muscle tests. Ten tests were performed with wasp venom, ten with a placebo (NaCl). The order in which the substances were tested was randomised. All substances were packed in small glass bottles which were housed in a cardboard box and handed to the examiner by an assistant. All glass vials were of the same weight, the small cardboard boxes were identical so that all tests were double blind. The examiners were not told how many verum or placebo tests had been planned. Nobody but the statistician and the assistant, who prepared the samples in a separate room, had access to the randomization list.

Kinesiology muscle test

The muscle tests were done in accordance with the rules of HK.[3] First a specialised area over the sternum was tapped for achieving energetic balances. For training purposes an indicator muscle was tested without contact to the allergen. This test was done while standing using the deltoid anterior. The examiner tried to push down the arm of the patient while the patient tried to hold the arm in position without using counter-pressure. To test whether the indicator muscle reacted as expected, two further pretests were performed: First, the participants were requested to say ‘yes’ or ‘no’. A strong reaction is expected for ‘yes’ and a weak one for ‘no’. Second, to confirm this previous result, the indicator muscle was ‘pinched’ and ‘smoothed’. Again, weak and strong reactions are expected. The subsequent tests only were performed when the indicator muscle reacted as expected in both tests.

Further testing was performed while the participants were lying down. First a check of the meridian balance was performed based on the test points for the meridian pairs on the navel. Weak elements were corrected immediately. The correction method was as described by Scott.[3] Testing was begun with the ‘neurovascular points’, when the reaction was strong it was followed by testing of ‘meridian end points’ and then the ‘neurolymphatic points’ and finally the ‘sedation points’. Points which tested weak were corrected correspondingly.

Finally, to finish off the preliminary testing, a navel-balance test (the hand of the participant is placed over the navel) and the testing of allergy test point (a small depression in the bones in front of the ear which the participant must hold during the test) was performed.

For the actual test the prepared cardboard box was put on the so called substance test area below the navel of the participant and the indicator muscle was tested (with simultaneous touching of the allergy test point). An uncertain hold or the arm being depressed without resistance were evaluated as a weak reaction, the stable holding of the arm in the initial position was considered to be a strong reaction. The weak reaction represents an allergy towards the test substance.

Every ten tests, the energy balancing and system testing procedures were repeated. We left at least two hours time before the patient was tested by another examiner. This time gap was chosen to prevent from carry-over effects from examiner to examiner and to give the indicator muscle time to recover.

Table 1 Test-retest-reliability and correctness of Kinesiology muscle test by examiner. Reliability is estimated by Intraclass-Kappa, correctness by the proportion of observations with a correct results

		Examiner				
Substance		**G**	**K**	**S**	**V**	**Total**
Wasp venom	Reliability	−0.07	−0.01	−0.05	−0.05	−0.04
	Correctness	24.0%	45.0%	61.4%	38.0%	43.8%
NaCl	Reliability	0.02	−0.04	0.07	−0.09	0.01
	Correctness	68.0%	60.0%	61.4%	50.0%	60.0%
Overall	Reliability	−0.01	−0.02	0.06	−0.06	0.03
	Correctness	46.0%	52.5%	61.4%	44.0%	51.9%

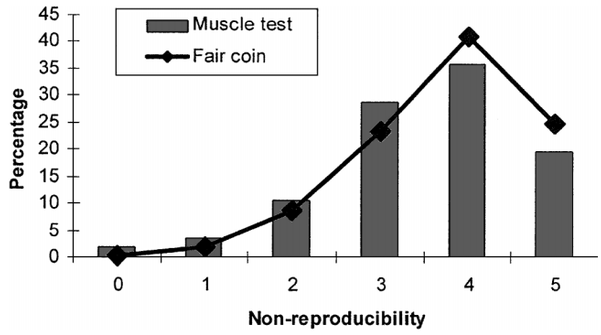

Fig. 1 Non-reproducibility of kinesiology muscle test and tossing of a fair coin. For definition of Non-reproducibility see text.

Statistical methods

As global measures for the test-retest-reliability, intraclass kappa coefficients were used. They were based on calculations of interindividual and intraindividual variances by means of a maximum-likelihood estimation in an ANOVA model incorporating three factors as variance components (patient, examiner, and tested substance).[8] This estimation procedure is valid even if the outcome is not normally but binary scaled.[9] Separate verum or placebo analyses were based on similar variance component models omitting the substance factor. From these models only the intraindividual variances were estimated but interindividual variances were obtained from the whole model. This calculation procedure, in fact, leads to a weighted coefficient with weights chosen proportional to the number of patients that each examiner tested. Intraclass kappa ranges from 1 to 0 where results near 0 indicate a low reliability.

The correctness of the methos was defined as the proportion of observations per examiner and participant having a correct result. A correct result was defined as strong reaction using wasp venom and weak reaction using NaCl.

RESULTS

All seven patients were proven to be allergic to wasp venom with positive skin tests and RAST values.

The global test-retest-reliability of the muscle test is estimated at 0.03 (95%-CI: 0.02 to 0.07). Therefore it can not be distinguished statistically from that of a random number generator, which has by definition a reliability of 0.00.

A subgroup analysis according to examiners and test substances shows that the differences are in each case very small so that no circumstances could be found under which a particularly high or low reliability could be expected (Table 1). Under verum all examiners showed a reliability that was even slightly smaller than random guessing.

The non-reliability of the muscle test can also be illustrated by an alternative approach. For this non-reproducibility was defined simply by counting how often a single result of the muscle tests disagrees with the majority of findings under the same conditions (this means the same examiner, same patient and same substance). For example, a non-reproducibility of 0 means that all ten results agreed, a value of 1 indicates that 9 cut of ten results agreed and a value of 5 means that the muscle was evaluated strong in five cases but weak in the five others.

Figure 1 demonstrates that the distribution of non-reproducibility matched to the distribution for a fair throw of coins. The theoretical mean non-reproducibility for tossing coins is 3.77, the estimated means for wasp venom and NaCl (averaged over each examiner and patient) were not substantially lower at 3.38 and 3.68, respectively. Detailed analyses show that in this approach the inexperienced examiner S performed best of all.

Table 2 Number of positive (weak reaction) and negative (strong reaction) Kiniesiolgoy tests results of examiner G by patients

	Substance	Weak reaction	Strong reaction
Patient 1	Wasp venom	3	7
	NaCl	3	7
Patient 2	Wasp venom	3	7
	NaCl	6	4
Patient 3	Wasp venom	3	7
	NaCl	1	9
Patient 4	Wasp venom	1	9
	NaCl	3	7
Patient 5	Wasp venom	1	9
	NaCl	3	7

The average correctness for all patients and examiners was about 44% for wasp venom and 60% for NaCl. The highest correctness for NaCl was found for the skilled examiner G, at the same time his value for wasp venom is very low at 24%. As Table 2 shows this is a consequence from that he generally more often found strong than weak muscle reactions, regardless of the tested patient or substance. Only examiner S had more than 50% correct decisions using both wasp venom and NaCl (Table 1).

DISCUSSION

Our study indicates that the muscle testing applied according to the principles of Health Kinesiology is not a reliable method for diagnosis of wasp venom allergy. If one assumes that allergy for wasp venom is an adequate and typical model for the evaluation of HK, then the above statement can be generalised for the whole method. This assumption at least can be supposed to be partially correct since Scott and Goss[3] explicitly refer to allergies as a typical und promising indication for HK. Although they do not mention insect venom allergies (the book concentrates on food, pollen, and chemicals), there is no doubt that wasp venom allergies are typical allergies in Scott's sense.

Our clear conclusion is in agreement with the majority of studies that deal with test-retest-reliability of AK muscle testing. Pothmann et al.[10] found non-reproducible results within a project to evaluate AK for food intolerances. There is no indication that the test-retest reliability for double-blind tests is more than accidental. Similarly, Haas et al.[11] and Peterson[12] demonstrated non-reproducibility within healthy persons. In contrast, Hsieh and Phillips[13] state a high test-retest-reliability. But a close examination of their results shows that they actually measured the reliability of the computerized dynamometer which was used to objectify the muscle tests.

Reliability is a basic requirement for validity. It is therefore surprising that Peterson[12] did not prove test-retest reliability but found a high sensitivity in diagnosing phobias. Note, that his validity results are possibly biased as they were obtained after breaking the code. Other studies found contradictory results on validity as well. Kenney et al.[14] could not link muscle testing to biochemical tests assessing nutritional deficiencies, but Jacobs et al.[15] concluded that 'AK enhanced but did not replace clinical/laboratory diagnosis of thyroid dysfunction'. Other studies found statistically significant correlations between AK muscle testing and objective neurophysiologic measures (for an overview see Motyka and Yanuck[16]).

Results on inter-examiner-reliability are mostly negative as well: Kenney et al.[15] showed that different examiners do not come to the same conclusions about the nutritional deficiencies (thiamine, zinc, vitamins A and C). Peterson[12] estimated an inter-examiner-reliability that was slightly smaller than random guessing. Positive results were presented by Jacobs et al.[16] They found excellent inter-examiner-reliability but used an inappropriate statistical analysis which make results difficult to compare. Lawson and Calderon[17] found inconsistent results showing huge differences in inter-examiner-reliability from muscle to muscle. These differences may be one reason why our results are negative: possibly we chose the wrong indicator muscle.

All previous studies investigated AK but not explicity HK. As HK uses specialised technique their results cannot be simply transferred. Nevertheless, they give hints that the fundamental concept of Kinesiology muscle testing is questionable.

Proponents of AK often do criticise the above mentioned studies because they did not follow standard AK procedures.[16] This cannot be applied to our study. We followed in detail the procedures that were described by Scott and Goss.[3] Thus, some described phenomena that could invalidate our results (such as 'switching'[18]) were unlikely to occur.

Nevertheless, our negative results may be due to some general drawbacks of the study. The first problem obviously is its setting. Scott and Goss mention that stress may invalidate the muscle test. Although none of the patients or examiners reported any problems one may argue that the study setting itself and especially the blinding of ratings puts heavy stress on examiners and patients. The data of Peterson partly support this hypothesis: the sensitivity of the muscle tests increases when situations were excluded that describe instabilities of the patient-examiner-relationship.

Small patient numbers and corresponding high probabilities of false decisions are obviously not a problem of this study. The small confidence intervals around the estimated kappas do confirm our negative conclusions. However, it has to be

mentioned that our study patients were not a random sample of all allergic patients and therefore may not be representative. However, there are no hints that we systematically selected patients who were unapproachable by HK.

In our study each patient was tested up to 80 times a day, including the pre-testing procedures. Therefore the indicator muscle may have got tired and have produced false negative results. For several reasons we still believe that our results are unbiased and valid: first, between each series of 20 examinations there was a time gap of several hours which should give the muscle enough time to recover. Second, this number of tests is not unusual and reflects daily practice of HK. Third, additional analyses show no improvement of results when only the first 10 examinations of each series were included.

Motyka and Yanuck[16] state that the level of training of the examiners is critical to the assessment of AK muscle testing. This is not true in our study where the results were bad independently from the skills and experiences of the examiners. Moreover, this argument contradicts basic AK publications,[3] where Scott and Goss unambiguously state that the HK muscle testing can be done even by lay staff that have not joined any course on HK. Nevertheless, we cannot reject the hypothesis that results are much better when only highly skilled HK specialists are under study.

Another theoretical drawback of our study is the estimation of specificity. Although we used the word 'correctness' instead of validity, we implicitly identify 'correct' results for wasp venom with sensitivity, and in the NaCl case with specificity. The latter relies on the assumption that no person is allergic to common salt. Moreover, we assumed that there are no effects of the paper box where the substances were placed in order to guarantee blinding. These assumptions seem reasonable. In any case, this argument may invalidate the results on validity but not on reliability.

The estimation procedure of group-specific reliabilities is somewhat unusual since intraindividual variances were calculated for each group separately but interindividual variance estimates included both groups. Note, that this approach ensures correct interindividual variances but restricting on verum (or placebo) results would have not. The estimation of overall test-retest-reliabilities ignores dependencies within patients.

More detailed results, especially on the various pre-tests that we had made, may be taken from Kunz et al.[19] These pre-tests support our conclusion that HK is neither reliable nor valid when used in patients with insect venom allergies that are tested by examiners with average skills in day by day situations. Moreover, they show no hint that HK can be helpful in diagnosing other allergies (for example grass pollen).

REFERENCES

1. Christenson MG, Morgan DRD. Job Analysis of Chiropractic: a Report, Survey Analysis, and Summary of the Practice of Chiropractic within the United States. Greeley: National Board of Chiropractic Examiners, 1993.
2. Christenson MG, Morgan DRD. Job Analysis of Chiropractic in Australia and New Zealand: a Report, Survey Analysis, and Summary of the Practice of Chiropractic within Australia and New Zealand. Greeley: National Board of Chiropractic Examiners, 1993.
3. Scott J, Goss K. Allergie und der Weg, sich in wenigen Minuten davon zu befreien. Freiburg: Verlag für angewandte Kinesiologie, 1990.
4. La Tourelle M, Courtenay A. Was ist angewandte Kinesiologie. Freiburg: Verlag für angewandte Kinesiologie, 1992.
5. Müller U. Insektenstichallergie. Stuttgart: Fischer, 1998.
6. Ring J. Angewandte Allergologie, 2nd ed. München: MMV-Medizin, 1991.
7. Middleton E, Ellis D, Reed C, eds. Principles and practice of allergy 4th ed. St Louis: Mosby, 1999.
8. Shrout PE, Fleiss JL. Intraclass Correlations: Uses in Asssing Rater Reliability. Psych Bulletin 1979; 86: 420–428.
9. Kraemer HC. Measurement of reliability for categorical data in medical research. Stat Meth Med Res 1992; 1: 183–199.
10. Pothmann R, v. Frankenberg S, Hoicke C, Lüdtke R, Weingarten H. Evaluation der klinisch angewandten Kinesiologie bei Nahrungsmittelunvertäglichkeiten im Kindesalter. In: Albrecht H, Frühwald M eds. Jahrbuch der karl und Veronica Carstens-Stiftung 4. Stuttgart: Hippokrates 1998; 116–130.
11. Hass M, Peterson D, Hoyer D, Ross G. The reliability of muscle testing response to a provocative vertebral challenge. Chiro Technique 1993; 5: 95–100.
12. Peterson KB. A preliminary inquiry into manual muscle testing response in phobic and control subjects exposed to threatening stimuli. J Manipulative Physiol Ther 1996; 19: 310–316.
13. Hsieh CY, Phillips RB. Reliability of manual muscle testing with a computerized dynamometer. J Manipulative Physiol Ther 1990; 13: 72–82.
14. Kenney JJ, Clemens R, Forsythe KD. Applied kinesiology unreliable for assessing nutrient nutrient status. J AM Diet Assoc 1998; 88: 698–704.
15. Jacobs GE, Franks TL, Gilman PG. Diagnosis of thyroid dysfunction: Applid Kinesiology compared to clinical observations and laboratory tests. J Manipulative Physiol Ther 1984; 7: 99–104.
16. Motyka TM, Yanuck SF. Expanding the Neurological examination using functional neurologic assessment Part I: Methodological considerations. Intern J Neuroscience 1999; 97: 61–76.
17. Lawson A, Calderon L. Interexaminer agreement for applied kinesiology manual muscle testing. Percept Mot Skills 1997; 84: 539–546.
18. Von der Rosen J. Verschiedene wichtige Teile der Kinesiologie in der täglichen Praxis. Erfahrungsheilkunde 2000; 49: 347–348.
19. Kunz B, Seeber N, Lüdtke R, Ring J. Kinesiologie in der Allergiediagnostik. In: Garbe C, Rassner G, eds. Dermatologie - Leitlinien und Qualitätssicherung für Diagnostik und Therapie. Berlin Heidelberg: Springer-Verlag 1998; 136–139.