A Double-Blind, Randomized Study to Assess the Validity of Applied Kinesiology (AK) as a Diagnostic Tool and as a Nonlocal Proximity Effect

Stephan A. Schwartz, [1#] Jessica Utts, PhD,[2] S. James P. Spottiswoode, [3] Christopher W. Shade, PhD,[4] Lisa Tully, PhD,[5] William F. Morris, DO,[6] and Ginette Nachman, MD, PhD[7]

Premise: Applied Kinesiology (AK) is a diagnostic technique widely used within the Integrative Medical community. In essence, it posits that a question can be mentally held in a person's mind, sometimes while they are holding a substance like a vitamin, or a food sample, and by measuring relative muscular weakness an answer as to whether the substance or the condition represented by the question is good for that person can be obtained. This AK is presumed to have a diagnostic capability. That being presumed, this study asks the following questions: (1) Is there a difference in muscular strength when an individual holds a substance that is inimical to life processes (a poison solution), as compared to a substance that is essential for life (normal saline)? (2) Is this effect a transaction involving input from both the person being measured and the kinesiologist doing the measurement or is it only the person being measured? (3) As an extension of question 2, is the result the same when different kinesiologists take the measurement or when no kinesiologist is involved? (4) Does belief, expectation, gender, or time cognition play a role in determining response?

Methodology: To answer these questions, which would help to define the parameters of the AK process, 51 participants were tested during three trials each, first by one kinesiologist, then by another, and finally, with no kinesiologist present by grip strength indicated using a hand dynamometer. Grip strength being a self-administered AK test of relative muscular strength. For each trial, a pair of randomly numbered sealed vials, each pair in a randomly numbered plastic bag, were used as the objects of the trial. In each bag, one vial contained saline solution while the other was filled with a slightly smaller amount of saline solution to which had been added ionic hydroxylamine hydrochloride $(NH_3OH)^+$, producing a toxic solution of 9 mg/ml. Each trial consisted of a separate muscle test for each vial. All present at the trials were blind as to which vial contained the toxin. And all who prepared the vials were blind to the trials. The force used by the kinesiologists in each of their trials was measured via a pressure pad system. The hand dynamometer trials were conducted with no kinesiologist present.

Results: Of the 151 sets of trials, the toxic vial was identified correctly in 80 of them (53%), resulting in a one-tailed exact binomial *P*-value of .258. Results for two of the kinesiologists were almost exactly at chance. For the third kinesiologist, there was a one-tailed exact binomial *P*-value of .18 (unadjusted for multiple testing). Results for the dynamometer were also almost exactly at chance. Testing whether there was a significant difference in proportions for whom the AK test worked based on belief about whether it would work resulted in non-significant χ^2 values of 0.6 ($P = .439$) for the trials with one kinesiologist and 2.222 ($P = .136$) for the hand dynamometer trials. The final variable examined was gender. While there was no significant difference in performance for males and females for the trials of the male kinesiologist or the hand dynamometer, the combined data for the two female kinesiologists did reveal a difference. Of the 33 sessions with females, only 15 were successful (45%), while for the 18 sessions with males, 14 were successful (78%), resulting in a χ^2 statistic of 4.96, $P = .026$. However, given all of the χ^2 tests performed in this section, the results must be interpreted with caution because of multiple testing. Results indicate belief in whether the AK test will work was not significantly related to whether it actually did work. A χ^2 test of the relationship between time perception and correct vial choice showed no significant relationships. The χ^2 statistic for the relationship using the hand dynamometer data was 0.927, $P = .629$.

Conclusion: The data in this study, particularly when seen in the larger context of a review of the literature from the AK field itself by Klinkoski and Leboeuf (1990), which considered 50 papers published between 1981 and 1987 by the International College of Applied Kinesiology, and the survey by Hall, Lewith, Brien, and Little (2008), using standard evaluation

1 Research Associate, Cognitive Sciences Laboratory of the Laboratories for Fundamental Research, Palo Alto, CA.
2 Department of Statistics, University of California—Irvine, Irvine, CA
3 The Resource Group, Washington, DC
4 Chief Scientist, Quicksilver Scientific, Lafayette, CO
5 Quicksilver Scientific, Lafayette, CO
6 Professor and Chair, Osteopathic Medical Manipulation Department, Campbell University School of Osteopathic Medicine, Buies Creek, NC
7 Quintiles Transnational Corporation, Durham, NC

Corresponding Author. Address: Cognitive Sciences Laboratory, P.O. Box 905 Langley, WA 98260
e-mail: saschwartz@earthlink.net

ISSN 1550-8307/$36.00

http://dx.doi.org/10.1016/j.explore.2013.12.002

criteria [quality assessment tool for studies of diagnostic accuracy included in systematic reviews (QUADAS), Standards for Reporting of Diagnostic Studies (STARD), JADAD, and Consolidated Standards of Reporting Trials (CONSORT)], for research methodology, as well as six prior nonclinical studies by Radin (1984), Quintanar and Hill (1988), Braud (1989), Arnett et al. (1999), Ludtke (2001), and Kendler and Keating (2003), all together suggest the following: The research published by the Applied Kinesiology field itself is not to be relied upon, and in the experimental studies that do meet accepted standards of science, Applied Kinesiology has not demonstrated that it is a useful or reliable diagnostic tool upon which health decisions can be based.

Key words: applied kinesiology, nonlocality, consciousness, alternative medicine, diagnostic technique

(Explore 2014; 10:99-108 © 2014 Elsevier Inc. All rights reserved.)

PROJECT BACKGROUND AND CONCEPTION

There are literally thousands of health practitioners in complementary and alternative medicine (CAM)—allopathic MDs, homeopathic MDs, osteopathic physicians, naturopaths, chiropractors, nurses, dentists, physiotherapists, and body workers—using a muscle-testing protocol known as Applied Kinesiology (AK) as a diagnostic aid. This technique traces back to the 1940s and an American husband and wife physiotherapist couple, Henry and Florence Kendall,[1] who developed a clinical procedure to ascertain the strength of a muscle and postural procedures allowing them to isolate and test the strength of a specific muscle independent of the surrounding musculature.

Two decades later, a Michigan chiropractor, George J. Goodheart, Jr., picked up this approach and redefined it.[2] Goodheart made a clinical report on a patient who came to him with an immobile shoulder that had troubled him for more than 15 years. Goodheart pressed on small nodules near the origin of the pain and the muscle strength returned to normal as did the range of motion. This led Goodheart to develop a technique for "tugging" on particular muscular trigger points to restore function.[2] As his clinical work continued, he came to believe his muscle-testing technique also had a diagnostic aspect. Goodheart explicitly rejected a "proximity effect," arguing that nutritional testing had to be mediated through lingual receptors and that holding test substances was an aberrant form of AK to be rejected. In spite of his feelings, an informal survey of 47 AK practitioners, who were also chiropractors, done prior to the study to assure that the methodology design was an accurate representation of what really happens in a AK practitioner's clinic revealed that the principal use of AK was as a guide as to how to do the chiropractic adjustment and afterwards, to check whether the treatment had been successful. Closely second, to check for food allergies.

Typically, a patient is asked to hold his or her arm extended out from the shoulder and parallel to the floor in the coronal plane and to resist the downward pressure of the practitioner's hand while the patient is holding or exposed to a test substance (Illustration 1). The relative strength difference, with or without exposure to the substance, supposedly helps the kinesiologist to assess systemic imbalances. For instance, a practitioner might test to see whether a food sensitivity exists. The practitioner might also press on key "trigger points" to find out if they cause muscle weakness. The presumption is that relative weakness is a sign that the substance, or one of the substances, is not good for that individual system. The underlying concept of AK assumes a "proximity effect" in which holding a substance—sometimes in a bottle, or encased in a capsule, or pill—as opposed to actually ingesting it is sufficient to determine the effect of the substance on the individual being tested.

Illustration 1. Trial in progress. Note participant holds vial in left hand while being tested.

Applied Kinesiology has been refined over the past 30 years, both in Europe and in the United States and today is used by many thousands of people as a feedback tool to assist in confirming and gathering psycho-physiological information. Professionally, certification as an Applied Kinesiologist is available from the International College of Applied Kinesiology[3] and from the Touch for Health organization.[4] To reach the highest level of certification, over 300 h of instruction, several proficiency exams, and submission of original research papers are required. But, although there are a wealth of clinical reports[5] and several well-known books, both academic[6] and for the general audience,[7] with muscle testing as their topic, the fundamental question about its diagnostic validity has rarely been directly addressed by properly controlled, randomized, and blinded experimentation.

In a review of the literature from the AK field itself, Klinkoski and Leboeuf[8] considered 50 papers published between 1981 and 1987 by the International College of Applied Kinesiology. Twenty of these papers were classified as research papers, and Klinkoski and Leboeuf evaluated them using seven criteria they felt represented accepted standards

for research methodology. These included the following: identification of sample size, inclusion criteria, blind and naive subjects, and statistical analysis. They concluded, "Although some papers satisfied several of these criteria, none satisfied all seven of them. As none of the papers included adequate statistical analyses, no valid conclusions could be drawn concerning their report of findings."[8]

In 2008, Hall et al. took an even more detailed look at the AK literature. They were explicitly interested in ascertaining whether AK offered "inter-examine reliability" and to determine "whether there was evidence for its therapeutic effectiveness."[9] To do this analysis, they used standard recognized evaluation tools. "Diagnostic accuracy studies were analyzed and scored for methodological quality and quality of reporting using the quality assessment tool for studies of diagnostic accuracy included in systematic reviews (QUADAS) and the Standards for Reporting of Diagnostic Studies (STARD). Clinical studies were analyzed for methodological quality using the JADAD scale and for quality of reporting using the Consolidated Standards of Reporting Trials (CONSORT)."[9]

Twenty-two papers were evaluated by Hall et al. in this way. Their first finding was that overall the methodology of the papers was poor. "QUADAS scored 1–11 out of a possible 14, STARD scores were between 6 and 13 out of 25, JADAD scores were all 0 out of 5 and CONSORT 4–6 out of 22."[9] They felt that with these scores, none of the questions that had prompted their study could be answered. They concluded by stating, "We recommend a pragmatic study of the effectiveness of kinesiology as the most appropriate initial step to determine whether kinesiology has any clinical value."[9]

Six studies by researchers with no vested interest in AK as a clinical technique have also examined the premise of AK in a laboratory setting, and we believe should also be considered. Although none used the accepted kinesiological clinical protocol, indeed were not medical studies per se at all, we cite them because they did meet the standards of methodological rigor, i.e., blindness and randomness, etc., expected in good science, and they did address a core claim of AK.

Five of the six studies involved testing whether muscles were stronger or weaker in proximity to sucrose. The results were contradictory. Radin, in 1984, carried out a double-blind experiment with 58 adults testing the hypothesis that proximity to sugar reduces human grip strength. To test this hypothesis, participants were asked to hold bottles identical in appearance and weight that were filled with either sugar or sand. An analysis of variance supported the hypothesis (Illustration 1).[10]

Four years later, Quintanar and Hill carried out a double-blind experiment with 90 undergraduate students in an attempt to replicate Radin's results using a similar protocol. When "additional controls were added to this design to reduce random sampling error, the hypothesis was not supported."[11]

A year later, in 1989, Braud carried out another replication of Radin's work, employing a double-blind experiment with 50 adults. Grip strength of the dominant hand was measured by a chart-recording hand dynamometer for 12 trials. For half of the trials, the subject held an opaque, sealed bottle containing sugar; for the other half of the trials, the subject held a control bottle of equal weight containing sand. Statistical analysis indicated no significant difference in grip strength when holding sugar vs holding sand.[12]

Ten years later, Arnett et al. tested the hypotheses that close physical proximity to sucrose, without its ingestion, affected muscle strength, adding the additional variable of time between meals as potentially influencing the direction of this effect. Twenty-eight college students performed 10 one-arm curls using the dominant arm while holding a bottle containing either sucrose or sand in the free hand, without knowledge of the bottle's contents. Data were first collected following an overnight fast. A week later, data were collected one hour after these students had each consumed a bagel. During the fasting state, the mean maximal isometric strength while holding the bottle of sucrose was significantly greater than when holding a bottle of sand; however, no significant difference was found for the group tested after having consumed a bagel.[13] This study did not counterbalance the order of conditions nor did it control for the participants' amount of experience with AK.

A team of German researchers in 2001, led by Ludtke, carried out a small laboratory pilot study conceptually close to the design of this study. Seven patients with clinically and allergologically confirmed wasp venom allergy were tested by four health Kinesiology examiners, each of whom tested each patient in a random order using 10 venom and 10 placebo bottles. They found no substantial variation amongst practitioners, and concluded, "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."[14]

In 2003, Kendler and Keating carried out the final precursor to the study that is the subject of this paper. Again, the test was the effect of proximity to sucrose. For a sample of 76 participants, a double-blind counterbalanced assessment of grip strength was performed with and without holding a container of sucrose. No significant effect was found for proximity of glucose.[15]

The Ludtke study while closest to this study in concept was still not truly clinical. Of the other five studies, we believe there are two major criticisms, particularly when they are considered as an assessment of Applied Kinesiology. The current study addresses these issues as indicated:

(1) These studies were all grip strength or arm curl studies carried out in non-clinical settings, and not truly tests of AK, as it is practiced in thousands of treatment rooms. With no sacrifice of methodological rigor, this study was designed so that both the set and setting of the trials were equivalent to what takes place in the treatment room of a working health professional.
(2) In most of these studies cited, the central variable governing the participant response was the reaction of individuals to sugar in some form. But, sugar means something quite different to an athletic young individual than it does to a middle-aged, overweight, type II diabetic. None of the studies controlled for this. This study eliminates this as a confounding variable because it is structured to be as unequivocally antipodal in its conditions and the universality of reactions as

possible—normal saline solution vs poison in normal saline solution.

OBJECTIVES

The objectives of this study are to answer the following four questions:

(1) Is there a difference in muscular strength when an individual holds a substance that is inimical to life processes as compared to a substance that is essential for life?
(2) If there is an effect, does the process involve input from both the person being measured and the clinician doing the measurement or is it only the person being measured? This might help us understand the differences in results amongst the four earlier studies.
(3) Is the result the same when different clinicians take the measurement or when no clinician is involved?
(4) Can AK be considered as a reliable diagnostic tool?

This last question deserves a further comment. In a study such as this one, which addresses issues in both basic research and clinical medicine, there were really three possible outcomes: (1) the null hypotheses were supported, (2) statistically significant results occurred but fell short of clinical significance, and (3) the data supported diagnostic levels of confidence.

STATISTICAL SIGNIFICANCE—CLINICAL SIGNIFICANCE

Statistical significance is a mathematical conclusion, which operationally focuses on the null value. Clinical significance in contrast incorporates issues of benefit and harm, and is a matter of ongoing debate. As LeFort[16] observes, "There is general agreement that tests of statistical significance do not provide information about the clinical significance or practical importance of research results." Clinical significance can be defined as the smallest Effect Size (ES) to have a clinically beneficial or harmful value. Hojat and Xu say, "(1) ES is a useful indicator of the practical (clinical) importance of research results that can be operationally defined from being 'negligible' to 'moderate,' to 'important. (2) The ES has two advantages over statistical significance testing: (a) it is independent of the size of the sample; (b) it is a scale-free index. Therefore, ES can be uniformly interpreted in different studies regardless of the sample size and the original scales of the variables."[17]

METHODOLOGY

The fundamental assumption of all AK sessions involving the test of something is that substances that are inimical to life generally, or for the particular individual being tested particularly, in some degree produce weakness in the muscle tested, in comparison with control samples.

This experiment explores the proposition in a deliberately typical clinic setting to increase the clarity of the outcome. In this study, one half of the vials contain a pure saline solution tainted by a notable poison, by definition life degrading, while the other half of the vials serve as controls and contain only the pure saline water solution, by definition essential to life. If the effect is ever going to occur, it should reasonably be expected to occur in a binary experiment in which the two conditions being tested are as antipodal as the differences in this one.

PERSONNEL

Kinesiologists

There were three kinesiologists who conducted the trials in this study. By design, both male and female kinesiologists participated to see whether gender was a factor.

The male kinesiologist originally trained as a civil engineer, and worked in that capacity for a number of years, prior to becoming a chiropractic physician. He completed his chiropractic training at the National University of Health Sciences and then trained under programs of the International College of Applied Kinesiology, completing their standard course in 1992/1993, with subsequent training at regular seminars thereafter. Notably, he received much of this training from Dr. Walter Schmitt, a protégé of AK's founder Dr. Goodheart.

The first female kinesiologist received her chiropractic training at Palmer College of Chiropractic in 1997 and subsequently received her certification for AK from the International College of Applied Kinesiology. She has maintained a chiropractic practice incorporating AK for a decade. This female kinesiologist served as a tester for eight trials. Subsequently, a second female kinesiologist served as the second tester.

The second female kinesiologist is a mechanical engineer who subsequently trained as an Applied Physiology and Touch For Health Certified Instructor. She worked as a parts and dimensional engineer for about 10 years in the large structural investment casting industry supplying jet engine parts to several manufacturers before obtaining instructor certification in Applied Physiology in 2004 from the Touch For Health Organization. She has maintained a private practice for five years, centering on Applied Physiology. Her involvement permitted the study to compare the results of two quite different approaches to muscle testing.

Participants

Over the course of six weeks, before any sessions began, 51 adult volunteers, 33 women, and 18 men, were recruited to take part, for no pay, in what became 151 trials. Each volunteered an average of 2.5 h, approximately what a first visit to a holistic clinician might entail. They were recruited by announcements that the study sought participants to test, and were self-selecting. They ranged in age from their 20s to 70s. The average age was 42 years. All had graduated from high school. In addition, nine had some college and 10 held undergraduate degrees. Eleven held Masters level degrees, two were Registered Nurses (RNs), and four held doctorates. One declined to answer the question on education. All but two participated in all three trial conditions. These two did not take part in the hand dynamometer session because of a

family emergency. Overall, 33 had been previously tested by AK in the course of a clinical visit; 18 had no prior experience with AK.

In addition to basic demographics, the study also included the following measures:

(1) *Time perception*: Since neither the kinesiologist nor the participant knew which vial contained a toxin and no actual direct contact with either the contents of the poison-tainted vials or the pure saline controls occurred, some form of nonlocal awareness must have occurred for AK to work. Schmeidler[18] and Dean et al.[19] in 1974 reported studies in which the Time Metaphor Test of Knapp and Garbutt[20] positively correlated with nonlocal awareness skills. Schwartz and De Mattei[21] found that individuals who self-assign themselves as time Dynamics perform less accurately than either Non-dynamics or Neutrals when they are asked to do a task involving the kind of nonlocal awareness implicit in this study. Participants in this study were asked to select a metaphor that most closely represented their understanding of time: Time is a rushing river (Dynamic). Time is an old woman spinning (Neutral). Time is a still mountain pool (Non-dynamic).
(2) *Belief in whether AK test would work and whether AK test did work*: We considered whether the patient's belief in the AK treatment they were about to receive might affect the outcome. To test this, we used the now classic Sheep/Goat Effect first reported by Schmeidler and McConnell (who coined the terms). This belief effect was chosen because it is now recognized as one of the most consistently determinative variables in the intention research literature.[22] Sheep—those who believe that whatever is being tested in an experiment in which they are taking part—is likely to work generally achieve better success than goats, who are skeptics.

Researchers

There were seven researchers in this study, each carrying out various tasks. Each was blind to the aspects of the study, in which they did not personally participate.

(1) *Principal Investigator*: The Principal Investigator (PI) was present for all the trials that were held in Virginia Beach, Virginia. The PI designed the study and oversaw the conduct of the 151 sessions, including running the computer and the pressure sensor system, the hand dynamometer, and the video camera. The PI transferred the data from the signed affidavit forms signed by each participant at the end of each vial test as well as the affidavits signed by the kinesiologists, when trials involved their participation. This data was then transferred by the PI into a single electronic database, which was sent to Researchers 5 and 6. The PI was blind to all aspects of the vial preparation, as well as their numbering and the numbering of the plastic bags in which the vials were placed.
(2) *Researchers 2 and 3*: Researchers 2 and 3 were in Boulder, Colorado. These researchers prepared the vials, both those with the poison and the controls. They also numbered them. They were blind to which vials were allocated to which trial and the actual trial sessions.
(3) *Researcher 4*: Researcher 4 was in New York, New York, and received the vials from Researchers 2 and 3, as well as the code revealing which were controls and which contained the poison. Researcher 4 first randomized the vials and then put a numbered control vial and a numbered treated vial in small plastic bags and numbered the bags from one to 150, each bag representing one trial. The vial-filled bags were then packed in a box and sent to the PI in Virginia Beach. Researcher 4 retained the codes showing which vials were controls and which contained the poison, and which from each category had been assigned to which bag. Researcher 4 was completely blind to the conduct of the actual trials.
(4) *Researcher 5*: Researcher 5 was in Beverly Hills, California, and carried out the analysis of the pressure pad data, which was then sent to Researcher 6. Researcher 5 was blind to all aspects of the study until all the trial data had been collected.
(5) *Researcher 6*: Researcher 6 was in Davis, California, and received the databases from the PI and Researcher 5. Researcher 6 analyzed these data for correlations with outcomes and demographics. Researcher 6 was blind to all aspects of the study's conduct until all the trial data had been collected.
(6) *Researcher 7*: Researcher 7 was in Durham, North Carolina, and participated in the design of the study and conducted 17 trials over two sessions. Researcher 7 was blind to vial preparation and numbering. Travel distance and the complexities of scheduling made it impossible for Researcher 7 to participate in the study after the second session.

HUMAN EXPERIMENTATION

Institutional Review Board (IRB): Pursuant to the U.S. Code of Federal Regulations—Title 45—Public Welfare—Department of Health and Human Services—National Institutes of Health—Office for Protection from Research Risks—Part 46—Protection of Human Subjects, a conforming Institutional Review Board (IRB) was established.

Informed Consent Form: Pursuant to Title 45 (Revised June 18, 1991, Effective August 19, 1991), participants in the study each signed an Informed Consent Form, which had been approved by the IRB, and which complied with all relevant guidance paragraphs of Title 45, Part 46.

EXPERIMENTAL SET AND SETTING

Experimental Area

Because one of the purposes of this study was to approximate as closely as possible the set and setting in which Applied Kinesiology is actually carried out, and because the majority of health practitioners utilizing AK are doctors of chiropractic, the trials for this study were all conducted in a treatment room of a working chiropractic clinic in Virginia Beach, Virginia. Participants filled out their paperwork, just as they would on their first visit to the clinic. They waited in the

waiting room, just as they would as patients, and were called in, one-by-one, just as patients would be in a clinical practice.

VIAL PREPARATION

Researchers 2 and 3 prepared the 300 vials at the laboratories of Quicksilver Scientific, LLC in Lafayette, Colorado. We specifically used glass for the vials, instead of plastic to obviate any criticism about leaching from the container that might influence the outcome. It is worth noting that AK practitioners routinely test substances in both glass and plastic (Illustration 2).

Illustration 2. A set of vials, one control, one with poison, each randomly numbered, and randomly placed in the plastic bag that was, then, randomly given a trial number.

Isotonic Saline Solution

In order to obviate any criticism that commercially produced normal saline might arguably contain some unaccounted-for trace substance, the researchers prepared the saline solution in their lab from naturally produced (sun-dried, unprocessed) sea salt (light-grey Celtic Sea salt). In order to assure uniformity across vials, the saline solution was prepared in one batch from sea salt dissolved in 18-MΩ deionized water to a ratio of 9 mg/ml (0.9% m/v) and then filter sterilized through a 0.22-μm nylon membrane filter. Overall, 150 of the test vials were filled with 6.0 cc of the saline solution and were uniquely numbered.

Isotonic Poison Solution

In contrast to the earlier studies that compared sugar with sand, a potentially compromised variable, since different people react quite differently to sugar, this study was designed to produce the starkest most antipodal options. $NH_2OH{\bullet}HCl$ ionizes in solution to NH_3OH^+ and Cl^-. The LD_{50} of $NH_2OH{\bullet}HCl$ for mice is 408 mg/kg and for rats is 120 mg/kg. A total of 150 test vials were filled with 6.0 cc of isotonic (0.9% m/v) hydroxylamine hydrochloride ($NH_2OH{\bullet}HCl$, ACS Plus Grade, Fisher Scientific) solution and randomly and uniquely numbered.

This toxin was chosen for several reasons: (1) it is unequivocally a poison to any human system; (2) it would dissolve in water without changing the water's sensorial characteristics—no visual change, no smell, etc.—so that nothing would reveal which vial contained the toxin; (3) it could be disposed of with relative ease; and (4) even if an accident were to occur, it would not be an unmanageable threat. Although highly toxic, in the form and potency used, as an ionic $(NH_3OH)^+$ solution at a pH of ~3.2, little absorption through skin would have happened even if a vial had broken, so there was a very small margin of risk if exposure occurred. Only actually drinking the contents of the vial would have exposed participants to any significant risk, and at the level of the solution used and the quantity involved, even that would have presented a nonlethal threat. However, from the premise of AK, it clearly allowed a sharply contrasting condition when compared with the untainted saline solution.

Numbering

The vials were randomly numbered as one lot, using an Arizona State University Random Number Generator program RandomNum v 2.1,[23] then the vials with the toxin were boxed in one box and those with just the untainted saline solution in another. The two boxes were sent to Researcher 4 in New York, where one from each box was randomly taken and placed in a plastic bag. When this was completed, the bags were randomly numbered with the trial for which they were intended. This produced 150 bags, each containing one toxin and one untainted vial.

PRESSURE SENSOR

Tact Array System

To continuously measure the force being exerted by the kinesiologists throughout a trial, a tactile pressure sensor system manufactured by PPS of Los Angeles, California., was used (Illustration 3). The system is stretchable and fits snugly over the palm of the hand. It is USB powered for Windows PCs and supported by proprietary PC-resident software which displays and tracks time-series, average, and peak pressures.

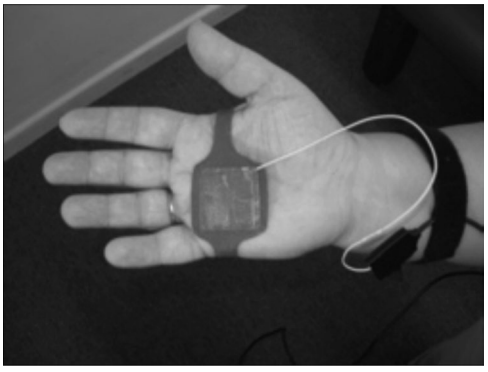


Illustration 3. The pressure pad installed prior to trial.

HAND DYNAMOMETER

Baseline Hand Dynamometer

The Jamar Hydraulic Hand Dynamometer gives accurate grip strength readings without the subject being able to "feel" the handle move. The results are reliably consistent with published

Baseline and Jamar® studies. The internationally accepted design assures reliability, user convenience, and measurement repeat-ability. The five position-adjustable handle can accommodate any hand size. The maximum reading is recorded by a slave hand that moves with the reading hand and remains in place until the unit is reset to zero. The strength reading can be viewed as pounds or kilograms. This permitted measurement of any grip strength differences between the vials (Illustration 4).

Illustration 4. Jamar hydraulic hand dynamometer.

VIDEO

To preserve a permanent record of all trials, each was recorded in its entirety on a Sony Handycam model DCR-HC38 Digital Video Camera Recorder.

SESSION METHODOLOGY

In order to approximate the experience of a patient in a clinic, the trials were designed for groups of three to nine individuals per day. In that way, the experience was very much like a first visit to a new physician. There were nine sessions in all, spread across parts of four months, with each day's session running from two to three hours.

Prior to beginning, participants were all briefed, their questions were answered, and they were asked to read and sign the informed consent form and to fill out a questionnaire. The video camera was started, the pressure sensor was calibrated, and the treatment room readied to begin. By arrangement, as this was happening, the first kinesiologist arrived in the building and was wired with the pressure sensor system by the Principal Investigator, whereupon the trials began in the treatment room with the first participant.

The kinesiologist tested the individual's muscle strength, using their AK techniques, until satisfied that the participant was in a condition to proceed with the test. Prior to each trial, the pressure sensor would be zeroed out and allowed to run for a few seconds showing a zero baseline. When this was established, the PI told the kinesiologist to begin. At the end of the trial, for each vial, the PI, the kinesiologist, and the participant, each independently filled out, dated, and signed a form attesting to which was used, and which produced the weaker response. At the end of each trial, the vials were returned to their plastic bag, all paperwork was collected, and that trial was at an end.

The first participant would leave the treatment room, the next participant was asked to enter, and the same process proceeded until the last participant had been tested. When the entire group for the session had been tested by the first kinesiologist, that kinesiologist left the building, and the second kinesiologist, by prior arrangement, entered. There would be a break while the new kinesiologist was wired for the pressure sensor. Then the trial cycle began and the patients, in the same sequence rotation, went through a second time.

Then the second kinesiologist left the building, there was another short break, and the third trial cycle, this one with just the hand dynamometer, began and proceeded in the same sequence. On the hand dynamometer trials, the participants held the vials, asking themselves "Is this harmful to me?" pointing the dynamometer dial toward the video camera, and when they felt it was appropriate, squeezed the grip. The strength in pounds was recorded, both on video and on the trial form, the instrument was reset, and the second vial was tested in the same way and the grip strength recorded.

HYPOTHESES

Affirmative hypothesis 1: There will be a significant difference in the level of weakness associated with the presence of the toxin as compared to the presence of the normal saline solution. There will be a statistically significant correct perception of the vials with the toxin in which weakness is associated with the presence of the toxin.

Affirmative hypothesis 2: The effect will occur under all three conditions of testing.

Affirmative hypothesis 3: We hypothesize that if Applied Kinesiology is a viable diagnostic test then there should be no statistically significant difference in outcome regardless of which kinesiologist performs the test.

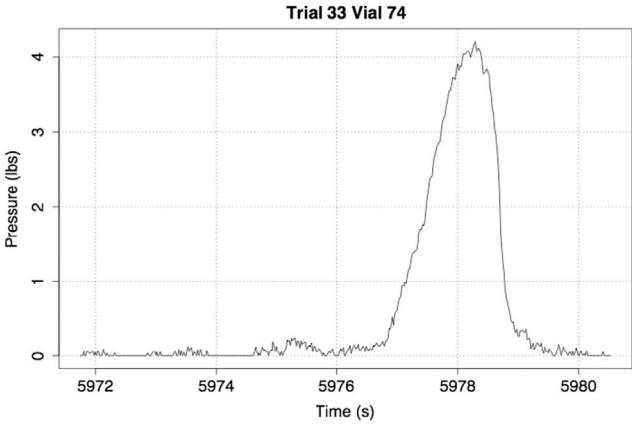
Affirmative hypothesis 4: We hypothesize that for Applied Kinesiology to be a viable diagnostic test, there will be no statistically significant difference arising from gender, belief, or time cognition.

DATA ANALYSIS AND RESULTS

Data Collected

For each of the 51 participants, the demographic and belief categorical variables listed in the section on "Participants" were measured. During the experiment, 49 of the 51 participants contributed results for three pairs of vials, with one toxic vial and one non-toxic vial in each pair, assigned in random order. All 51 participants worked with the male kinesiologist for one pair of vials and with either female kinesiologist 1 or female kinesiologist 2 for another pair. Overall, 49 of the 51 participants worked with the hand dynamometer for the third pair of vials. Therefore, there were a total of 150 pairs of vials tested. For each pair of vials, it was determined whether the participant correctly identified the toxic vial by noting if the trial with that vial produced a weaker response than the trial with the pure saline vial.

In addition, for the two sets of trials with the kinesiologists, continuous pressure data was measured. Four measurements



Graph 1. Representative Session Pressure Graph.

were derived from each of these trials: maximum pressure, mean pressure, median pressure, and integral pressure (the area under the pressure curve). These results can be used to determine if the kinesiologist, perhaps unconsciously, detected which was the toxic vial because, if so, more pressure should have been applied with the toxic vial than with the non-toxic vial (see Graph 1 above for an example of continuous pressure data measurement).

RESULTS FOR PRESSURE DATA

Under the null hypothesis that it is not possible to detect which is the toxic vial, the 151 sets of trials represent a binomial experiment with $n = 151$ and $P = ½$, where P is the probability of correctly selecting the toxic vial.[24] Under the alternative hypothesis, the probability of selecting the toxic vial is higher, thus the alternative hypothesis is $P > ½$. Of the 151 sets of trials, the toxic vial was identified correctly in 80 of them (53%), resulting in a one-tailed exact binomial P-value of .258. Although the result is in the correct direction, it is not statistically significant. Results for the male kinesiologist and female kinesiologist 1 were almost exactly at chance, with 26 out of 51 correct and four out of eight correct, respectively. Results for the dynamometer were also almost exactly at chance, with 25 correct out of 49 tests. For female kinesiologist 2, there were 25 correct guesses out of 43 tries, resulting in a one-tailed exact binomial P-value of .18 (unadjusted for multiple testing).

The continuous pressure data was used to test whether the kinesiologists are likely to apply more pressure when the participant is holding a toxic vial than when holding a non-toxic vial. Although a complex model could be used to combine results for all three kinesiologists, the most direct way to answer the questions of interest is to conduct separate paired t-tests for each pressure measure and each kinesiologist. For each test, the null hypothesis is that the population mean difference in the pressure measurement is 0. The alternative hypothesis is that the population mean difference is greater than 0, indicating that the pressure is greater when the participant holds a toxic vial than with a non-toxic vial (Tables 1 and 2).

The results of the 12 t-tests are shown in Table 3. For kinesiologist JE, all results were in the opposite direction from the alternative hypothesis, with more pressure applied with the non-toxic vials than with the toxic ones. If a one-tailed test in the opposite direction had been used, the t-test for maximum pressure alone would have been statistically significant. However the sample size was so small ($n = 6$) that no conclusions should be made about what would happen if a larger sample were available. For the male kinesiologist and female kinesiologist 2, the mean and median pressure results are suggestive, but because so many tests were done, the P-values need to take multiple testing into account. For these two practitioners, it appears that the maximum pressure applied over the time period was very similar, on average, for the toxic and non-toxic vials.

RESULTS FOR BELIEF DATA

Participants were asked whether they thought the experiment would work, and 33 said yes while 18 said they were not sure. No one said no. Table 4 shows whether the experiment did work for that participant, for the session with the male kinesiologist and the session with the hand dynamometer. Testing whether there is a significant difference in proportions for whom it worked based on belief about whether it would work resulted in non-significant χ^2 values of 0.6 ($P = .439$) for the trials with the male kinesiologist and 2.222 ($P = .136$) for the hand dynamometer trials. There were too few trials with the other kinesiologists to do a meaningful χ^2 test. These results indicate that belief in whether the experiment will work is not significantly related to whether it actually does work. Results were similar for the question about whether kinesiology could work in general.

RESULTS FOR TIME PERCEPTION

As described in the section on participants, they were asked about their time perception, categorized as Dynamics, Neutrals, and Non-dynamics. A χ^2 test of the relationship between time perception and correct vial choice showed no significant relationships. For example, the χ^2 statistic for the relationship using the hand dynamometer data was 0.927, $P = .629$.

Table 1. One-Tailed t-Test Results for Pressure Measurements

Kinesiologist	Mean Pressure	Median Pressure	Max Pressure	Integral Pressure
Male ($n = 50$)	$t = 1.63$, $P = .055$	$t = 2.14$, $P = .019$	$t = 0.00$, $P = .50$	$t = 0.94$, $P = .176$
Female 1 ($n = 6$)	$t = -0.97$, $P = .812$	$t = -0.03$, $P = .513$	$t = -2.42$, $P = .970$	$t = -1.37$, $P = .885$
Female 2 ($n = 43$)	$t = 1.30$, $P = .100$	$t = 1.33$, $P = .095$	$t = -0.35$, $P = .635$	$t = -0.59$, $P = .720$

Table 2. Sheep/Goat Results

AK Can Work			AK Will Work in My Trial		
Not Sure	No	Yes	Not Sure	No	Yes
18	0	33	9	1	41
35%	0%	65%	18%	2%	80%

AK, applied kinesiology.

GENDER VARIABLE

The final variable examined was gender. While there was no significant difference in performance for males and females for the trials with the male kinesiologist or the hand dynamometer, the combined data for the two female kinesiologists did reveal a difference. Of the 33 sessions with females, only 15 were successful (45%), while for the 18 sessions with males, 14 were successful (78%), resulting in a χ^2 statistic of 4.96, $P = .026$. However, given all of the χ^2 tests performed in this section, the results must be interpreted with caution because of multiple testing.

DISCUSSION

Since this is a study in which the participants never actually touch the toxic solution within the vials, one must ask whether we are measuring a physical proximity effect? Collaterally, one might ask that if we were measuring a physical proximity effect, would the proximity of the two vials in the plastic bag constitute a confounding variable? There is within physical chemistry nothing to suggest this. Therefore, we must look elsewhere. One obvious place would be in the homeopathic literature, where substances diluted beyond the Avogadro number (6.022×10^{23}) where there is no longer any physical presence of the potentized remedy, yet clinically such remedies are still effective, and must involve some field effect since the result cannot be chemical. A search of this literature yields two insights: (1) the remedies do cancel one another if two open vials are place proximate to one another, even when they are beyond Avogadro dilution and (2) there is no problem in storing remedies in close proximity, as long as they remain sealed. Indeed, that is how they are stored in homeopathic pharmacies. We believe that the only rationale of AK that makes sense is that an AK reaction represents nonlocal acquisition of information. Therefore, one can conclude that because the vials were never opened, physical proximity in the plastic envelope is not a confounding variable.

Table 3. Belief in Whether AK Test Would Work and Whether AK Test Did Work

	Male Kinesiologist		Hand Dynamometer	
Will Work?	Did Work	Did Not Work	Did Work	Did Not Work
Not sure	4	6	3	7
Yes	22	19	22	17

Table 4. Time Perception Results

Dynamics	Neutrals	Non-Dynamics
26	16	8
52%	32%	16%

Note: One participant could not make a decision as to how he viewed time.

Is it possible that if one holds a vial and then a control vial, a kind of linger effect might be a confounding variable? This argument moves the locus of action from the vial to the person being tested, again, as the result of physical proximity. This fails for the same reason as the vial-to-vial scenario. It is also a materialist argument. First of all, since there are an equal number of vials in the two conditions, the linger effect would presumably move from treated to control, which would occur 50% of the time.

AK plausibly can only occur if the muscular reaction results from the acquisition of nonlocal information.

The objectives of this study were to answer the following four questions:

(1) Is there a difference in muscular strength when an individual holds a substance that is inimical to life processes, as compared to a substance that is essential for life?
(2) If there is an effect, does the process involve input from both the person being measured and the clinician doing the measurement or is it only the person being measured? This might help us understand the differences in results amongst the four earlier studies.
(3) Is the result the same when different clinicians take the measurement or when no clinician is involved?
(4) Can AK be considered as a reliable diagnostic tool?

The task proposed by Hall et al.,[9] "We recommend a pragmatic study of the effectiveness of kinesiology as the most appropriate initial step to determine whether kinesiology has any clinical value," appears to have been accomplished, and the result is that it lacks clinical value.

The data in this study, particularly when considered in the larger context of the five prior non-clinical studies of Radin (1984), Quintanar and Hill (1988), Braud (1989), Arnett et al. (1999), and Kendler and Keating (2003), and the even larger horizon of the 50 papers assessed by Klinkoski and Leboeuf (1990) and the 22 papers examined by Hall et al. (2008), all seem to suggest one common conclusion: Applied Kinesiology does not presently provide a useful or reliable diagnostic tool upon which health decisions can be based.

Acknowledgments

The authors wish to thank the BIAL foundation for its gracious underwriting of the major costs of this research under Bursary 167/06, and Atlantic University for funding augmenting BIAL's support. We also thank Christine Fallwell, D.C. for making her chiropractic clinic available for the trials.

REFERENCES

1. Kendall H, Muscles Kendall F. *Testing and Function.* Baltimore, MD: Williams & Wilkins Co; 1949.
2. Gin RH, Green BN. George Goodheart Jr., D.C., and a history of applied kinesiology. *J Manipulative Physiol Ther.* 1997;20 (5):331–337. PMID: 9200049.
3. Available at: http://www.icak.com/. Accessed 16.05.2008.
4. Available at: http://www.touch4health.com/index.html. Accessed 05.06.2008.
5. Haas M, Peterson D, Hoyer D, et al. Muscle testing response to provocative vertebral challenge and spinal manipulation: a randomized controlled trial of construct validity. *J Manipulative Physiol Ther.* 1994;17:141–148. PMID: 8006528;
   Daniels L, Worthingham C. *Muscle Testing: Techniques of Manual Examination. Anatomical Drawings by Sigal L. Test and Gait Illus by Black H, Levering C.* Philadelphia, PA: Saunders; 1972.
6. Hislop H. *Daniels and Worthingham's Muscle Testing: Techniques of Manual Examination. Contrib. Connolly B.* Philadelphia, PA: W.B. Saunders; 1995.
7. Hawkins DR. *Power vs. Force: The Hidden Determinants of Human Behavior.* Carlsbad, CA: Hay House; 2002.
8. 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;13:190–194. PMID: 2351880.
9. Hall S, Lewith G, Brien S, et al. A review of the literature in applied and specialized kinesiology. *Forsch Komplement Med.* 2008;15(1):40–46. PMID: 18334813.
10. Radin DI. A possible proximity effect on human grip strength. *Percept Mot Skills.* 1984;58(3):887–888. PMID: 6473035.
11. Quintanar AF, Hill TV. Sugar proximity and human grip strength. *Percept Mot Skills.* 1988;67(3):853–854. PMID: 3226837.
12. Braud WG. A possible proximity effect on human grip strength: an attempted replication. *Percept Mot Skills.* 1989;68(1):157–158. PMID: 2928038.
13. Arnett MG, Friedenberg J, Kendler BS. Double-blind study of possible proximity effect of sucrose on skeletal muscle strength. *Percept Mot Skills.* 1999;89(3 Pt1):966–968. PMID: 10665031.
14. Lüdtke R, Kunz B, Seeber N, et al. Test–retest-reliability and validity of the kinesiology muscle test. *Complement Ther Med.* 2001;9(3):141–145. PMID: 11926427.
15. Kendler BS, Keating TM. Re-evaluation of a possible proximity effect on muscular strength. *Percept Mot Skills.* 2003;97 (2):371–374. PMID: 14620219.
16. LeFort S. The statistical versus clinical significance debate. *J Nurs Scholarsh.* 2007;25(1):57–62.
17. Hojat M, Xu G. A visitor's guide to effect sizes. Statistical significance versus practical (clinical) importance of research findings. *Adv Health Sci Educ.* 2004;9(3):241–249.
18. Schmeidler G. An experiment on precognitive clairvoyance: precognition scores related to the subject's ways of viewing time. *J Parapsychol.* 1964;28:93–101.
19. Dean D, Mihalasky J, et al. *Executive ESP.* Englewood Cliffs, NJ: Prentice-Hall; 1974.
20. Knapp AH, Garbutt JT. Time imagery and the achievement motive. *J Pers.* 1958;26.
21. Schwartz SA, De Mattei RD. Proceedings of the Parapsychology Association 1982, and Research in Parapsychology 1982. Metuchen, NJ: Scarecrow; 1983.
22. Schmeidler G, McConnell R. *ESP and Personality Patterns.* New Haven, Conn: Yale University Press; 1958.
23. RandomNum 2.1. Available at: http://www.public.asu.edu/~mwwatkin..
24. By mistake, the vials for trial 101 were used twice with different participants, several weeks apart in time. Bag 101 was the last trial in session, using the hand dynamometer in Session 5, and the first trial in Session 6 involving the female kinesiologist 2. All conditions of blindness obtained in both sessions. This duplication makes no meaningful difference to the outcome statistics, and the authors felt it was important to report exactly what happened. Thus the references elsewhere to "151 trials.".