

The Oxford Studies: Validating Muscle Response Testing

Part I – Methods Used

When I was practising as a chiropractor in North Queensland, I avidly used a range of muscle-testing-based techniques with my patients – and we loved the results. However, I soon realised that muscle testing was one of the biggest strengths of my practice – but also one of the biggest weaknesses. It was a strength because we could clearly and quickly tune in to the body, ask it what it requires, and focus any therapy on that. It was a weakness because it lacked scientific validity. That is, insufficient (e.g. too little) robust clinical research has been carried out which supports its usefulness – as a result, it is largely thought of as unscientific, meaningless, and even dodgy.

On one hand, I did not necessarily need scientific “proof” that muscle testing “worked” – I saw proof of it in my practice every day. But on the other hand, I was curious. So, I undertook a number of small research projects in my clinic, and well, I did them poorly. I just didn’t have the knowledge I needed to make the results meaningful, and for that reason, I looked for where I could gain this knowledge. This is how I came across [Oxford University’s programme in Evidence-based Health Care](#). It is a programme designed specifically for practitioners who want to learn how to do rigorous clinical research. It was also run alongside Oxford’s Centre for Evidence-based Medicine, giving students access to some of the top clinical researchers in the world. It seemed just what I was looking for – so I applied – and was accepted! So, off I went to England for further tertiary education.

For my research, I had originally planned to study the effectiveness of an emotional healing technique (e.g. HeartSpeak) on those with depression. However, this technique uses muscle response testing (MRT), and as can be imagined, it was met with extreme skepticism within my department (the Department of Primary Health Care Sciences). Before my supervisors would allow me to embark on a large randomised trial, they insisted I demonstrate the efficacy of MRT. This venture then took a life of its own, and the randomised trial on depression was deferred to another time.

The first step that needed to be taken to investigate the validity of MRT was to figure out how to begin. This was not as straight forward as one would think. So, I started by defining what I meant by MRT, and that was: testing one muscle repeatedly as the target condition changed. Having studied Applied Kinesiology (AK) myself, I knew that MRT was jokingly (or not) called “The Arm Push Down Test” – and was often regarded as unsound. However, I also understood that it was used within many different muscle testing technique systems – such as HeartSpeak, Psych-K, Total Body Modification, Contact Reflex Analysis, and dozens of others. So, aside from myself, I knew MRT was used widely around the world. I recognised that MRT was different from the type of muscle testing done in AK, and as such, needed to be considered as distinct.

Reiterating the distinction, in MRT one muscle is tested repeatedly (usually the deltoid) as the target condition changes. That means, one MRT is performed for each target condition (and usually the result of one MRT influences the choice of target condition of the next MRT). A target condition is what one performs the test to detect, and examples in common use include: stress, lies, chiropractic subluxation, meridian imbalance, the need for a particular nutritional supplement, etc. Another important aspect of MRT is that it is a binary test – that is, it has only two possible outcomes, commonly referred to as “strong” and “weak.”

Then, I clearly delineated MRT as being distinct from other forms of manual muscle testing (MMT). For instance, MRT differs from orthopaedic/neurological MMT (ON-MMT) done by many physiotherapists, chiropractors, and osteopaths, in that the target condition for ON-MMT is limited to muscular strength, and the result is not binary, but usually rated on a 0-to-5 scale. As introduced above, MRT differs from Applied Kinesiology style of MMT (AK-MMT) – also a binary test – in that with AK-MMT, any muscle can be tested, and the outcome of the test will have different meanings, dependent upon which muscle was being assessed.

The next step was to determine just how widely used MRT actually is – also called the prevalence of use of MRT. As my advisors argued, if only a handful of people use MRT, then assessing its validity would be impractical. So, I set about this task – which, again, was not simple because those in many different lines of work use MRT – and also those in no particular employment (i.e mothers). For instance, many chiropractors use MRT, but not all, and kinesiologists use MRT but a kinesiologist does different things in different parts of the world. So, interviewing those in particular profession seemed inefficient. Therefore, it was decided that if the various organisations that teach MRT were polled and asked how many they have trained over the years, and if the totals were adjusted for things like attrition, inaccurate accounting, and incompleteness, then a reasonable estimation would be achieved. The results of this polling were interesting for a number of reasons. Firstly, in a painstaking search, only 86 techniques that used MRT were identified; however, undoubtedly the actual number of named techniques probably far exceeds this number. Nevertheless, all teaching organisations were contacted by either telephone or email, with unexpected mixed responses, ranging from extremely helpful to unresponsive to outright hostile. Still, from the data collected, it could be estimated that over 1 million people use MRT worldwide (for the full paper, click [here](#)). This widespread prevalence of use of MRT certainly warranted investigation of its validity.

Next, a thorough literature search had to be carried out, to determine if previous research has already demonstrated MRT to be valid, because PhD research must uncover some new information or insights. I was pleased to learn just how much research on MMT has been done, but did discover that most of it was not associated with MRT specifically. So, yes, my PhD would result in unique research.

The next question that had to be answered is *how* to assess the validity of MRT. There are numerous terms that are used to describe tests and measures, such as valid, accurate, precise, reliable, repeatable and so on. Because in colloquial English, the meanings of these words differ from their use in research settings, it became important for me to understand specifically what each term meant. and also, because earlier research studies used these terms. After months of reading, I determined that the place to start was to assess the *accuracy* of muscle testing, and to use the standard protocol for *diagnostic test accuracy studies*, called the STARD Statement.

At first, I was resistant to using the diagnostic test study protocol – since, after all, MRT is not used to *diagnose*, per se. Then I read that a diagnostic test: (1) gains information about a patient, and (2) is used to guide treatment. Since MRT is used for both of these tasks, then this was indeed an appropriate approach.

Since a diagnostic test is used to detect a target condition (e.g. manual blood pressure testing is used to detect hypertension, and a series of blood tests are used to detect diabetes), I had to consider carefully which condition to target in my studies. Since MRT is used to detect a large range of conditions, the pool was extensive – and I knew that this choice would be extremely important.

Furthermore, to assess a diagnostic test, the results of the test in question (called the index test, MRT in this case), must be compared to the results of a reference standard (a test already in use to detect the target condition and already found to be valid). Since the validity of MRT was questionable, I also knew how important it was to select an exceptionally sound reference test. It would be much more convincing if MRT was compared to an established standard rather than to another speculative test.

After much consideration, it was decided that we would use MRT to detect *deceit* (a lie), or put another way, to distinguish false from true spoken statements. Deceit was chosen for a number of specific reasons – primarily because the reference standard would then be the actual verity of the statement, which would be definitively known and could be controlled. Because of this, the reference standard would be a *gold standard*, which would add rigorousness to this series of studies. The paradigm we chose to implement is one very commonly used: the muscle stays strong when a statement is true, and goes weak when a statement is false. While this line of research did not involve explaining why or how this occurs, in the field, it is often explained that lying is a stress, and stress causes the muscle (the body?) to weaken – so it makes sense. Aside it being in common use, this explanation has good face validity and a sound theoretical framework.

Then, we defined *accuracy* of MRT to be the percent correct, and it would be quite straightforward to make this calculation (i.e. the number of MRTs gotten “right” divided by the total number of MRTs performed). Using statistical methods, this number could then be compared to chance to determine if there was a difference. We compared MRT accuracy to chance because, theoretically, in a binary test, the two outcomes (strong and weak in the case of MRT) would be equally likely. However, in practice this may not be the case.

Therefore, we sought to implement a second index test which could be used to compare accuracies. Since it is the opinion of some MRT skeptics that it is not actually MRT making distinctions, but rather that the MRT practitioner is good at “reading” people. With this in mind, the secondary index test we implemented was *intuitive guessing*; that is, without using MRT, but only visual, auditory and kinesthetic clues, the practitioner was asked to guess whether a statement spoken by the patient was true or false. The accuracy, or percent correct, of intuitive guessing could be compared to MRT accuracy to see if there is any difference. If there is no difference, the skeptics’ hypothesis would be correct; however, if there was a difference, then there would be something to MRT after all. This addition of a secondary index test was an important and valuable piece of this research methodology.

The next aspect of the methods that had to be considered was the participant enrolment criteria – that is, defining who we were going to recruit to participate. I wanted to get a very well-rounded view of MRT accuracy, so I wanted a broad sampling of muscle testing practitioners, otherwise known as a heterogeneous sample. We recruited practitioners from any profession, with any amount of experience and any amount of expertise. However, to be included, they had to have had some previous training in some kind of MRT. In contrast, the patients recruited for the first study (the largest one) had to have no previous experience with MRT, and also had to be unknown to the practitioner who was going to test them. For the first study, after doing a sample size calculation, we recruited 48 unique practitioner-patient pairs, meaning a pair could only participate once. For the 3 follow-up studies, 20 practitioner-patient pairs were used.

Blinding is another important aspect of clinical research. In this series of studies, MRT accuracy was calculated using the case when the practitioners were blind to the outcome of the MRT, meaning they did not know if the spoken statement was true or false. In addition, practitioners were not blind to the

paradigm under investigation (i.e. true statements → strong MRT; false statements → weak MRT). Unfortunately, blinding patients was not as straight forward – as they were aware that they were saying true and false statements. To balance this, they were blind to the paradigm being tested – that is, they were not told that their muscle will weaken when they spoke false statements. So, in the end, I believe we attained a fair degree of blinding.

There were a good many factors that we had to consider when designing these studies, and this first part of this 2-part article describes the general methods used. In Part 2, particulars of each of the 6 studies (see Table 1) will be outlined, the results revealed, and their implications discussed.

Table 1 – An outline of this series of studies assessing the validity of MRT

Study 1 – Estimating the Accuracy of MRT

Study 2 – Replication of Study 1

Study 3 – Replacing the Practitioner with Grip Strength Dynamometry

Study 4 – Using Emotionally-arousing Stimuli

Study 5 – Estimating MRT Precision using a Round-robin Format

Dr Anne Jensen, *DC, PGCert, PGDip, MSc, MS, DPhil*