
http://researchrepository.murdoch.edu.au/26284

Copyright © © 2015 Elsevier Ltd. It is posted here for your personal use. No further distribution is permitted.
The Efficacy of Acupoint Stimulation in the Treatment of Psychological Distress: A Meta-Analysis

Sandro Alfred Gilomen, Christopher William Lee

¹Murdoch University, School of Psychology and Exercise Science, Murdoch, Western Australia, Australia

Correspondence concerning this article should be addressed to Dr Christopher Lee, Murdoch University, School of Psychology and Exercise Science, 90 South Street, Murdoch, Western Australia, Australia. E-mail: Chris.Lee@murdoch.edu.au

*Corresponding author. Tel.: +61 893606838; fax: +61 893606492.

E-mail addresses: chris.lee@murdoch.edu.au (C.W. Lee), Sandro@sgpsychology.com.au (S.A. Gilomen)
Abstract

**Background and objectives:** Emotional Freedom Techniques (EFT) is a type of therapy involving the stimulation of acupuncture points while using a spoken affirmation to target a psychological issue. While some studies cite data indicating EFT is highly efficacious, findings in other studies are unconvincing. The aim of this meta-analysis was to examine the effect of EFT, particular acupoint stimulation, in the treatment of psychological distress.

**Method:** A systematic review of the literature identified 18 randomised control trials published in peer reviewed journals involving a total of 921 participants.

**Results:** A moderate effect size (Hedge’s g = - 0.66; 95% CI: -0.99 to -0.33) and significantly high heterogeneity ($I^2 = 80.78$) across studies was found using a random effects model indicating that EFT, even after removing outliers (decreases in $I^2 = 72.32$ and Hedge’s g = -0.51; 95% CI: -0.78 to -0.23), appears to produce an effect. The analysis involved 12 studies comparing EFT with waitlist controls, 5 with adjuncts and only 1 comparison with an alternate treatment. Meta-regression and subgroup analyses were conducted to examine the effect of moderators on effect size of symptom change following EFT.

**Conclusions:** Due to methodological shortcomings, it was not possible to determine if the effect is due to acupoint stimulation or simply due to treatment elements common with other therapies.

**Keywords**

Acupoint stimulation, Emotional Freedom Techniques, Gold Standard Scale, Meta-Analysis
Emotional Freedom Techniques (EFT) fall into the Energy Psychology modality and have its origin in Thought Field Therapy (TFT) developed by Roger Callahan in the 70s (Callahan, 1985; Callahan & Callahan, 1996; Feinstein, 2012). TFT postulates that thoughts associated with emotional problems are based in a thought field and become perturbations which correspond to meridian points on the body. To eliminate the emotional pathology, the individual must tap the exact meridian point in a precise sequence (Algorithm). It is thought that this process will unblock or balance the energy flow thereby removing the perturbations (Dietrich et al., 2000; Lukoff, 2012). A variant protocol of Callahan’s early formulation was developed by Gary Craig and is used as the standard treatment manual for EFT (Craig, 2011). The protocol requires the person to recall the traumatic event, use a spoken affirmation, for example “Even though I have this fear of dogs, I deeply and completely accept myself”, and tap acupuncture points on the body.

EFT has been used in outpatient settings to treat a range of disorders. This includes addictions (Sparks, 2002), PTSD (Church & Pina, 2011; Karatzias et al., 2011), chronic pain (Bougea et al., 2013; Brattberg, 2008) and anxiety (Irgens et al., 2012). It has also been used to treat habit problems such as weight gain (Sojcher, Gould-Fogerite & Perlman, 2012) The number of EFT sessions delivered ranges from one (Church & Pina, 2011) to eight (Stapleton, 2012).

EFT has been controversial due to its proponent's claims of speed, durability, effectiveness, and mechanism of action, in the treatment of multiple psychological disorders (Callahan, 2001; Coelho, 2007; Devilly, 2005; Feinstein, 2012; Gaudiano, Brown, & Miller, 2012; Lohr, 2001; McCaslin, 2009; Moritz et al., 2011; Pignotti, 2007; Pignotti & Thyer, 2009; Rosen & Davison, 2001). These claims have been made despite questionable empirical support.
The most recent review, by Feinstein (2012), focussed on “Acupoint Stimulation in Treating Psychological Disorders”. Feinstein conducted a literature search, identifying 51 peer-reviewed papers up to April 2012. The studies reviewed investigated clinical outcomes following the tapping of acupuncture points, to address a range of psychological disorders including, depression, anxiety, specific phobias and somatoform disorders. A number of studies in the review also examined the use of EFT to enhance sports performance. Of the 51 studies, a total of 18 met the criteria of randomized controlled trials. Feinstein concluded that all 18 studies reported positive outcomes on at least one clinical measure and that “effect sizes were large in 15 of 16 studies” in which effect size was calculated (Feinstein, 2012). However the review did not test the significance of the effect sizes or use statistical methods to synthesize the data.

The present study is a meta-analysis of all RCTs on EFT identified in our systematic literature search. The aim was to examine the effect of EFT, in particular acupoint stimulation, in the treatment of psychological distress. More specifically, this study seeks to establish if EFT is more efficacious in relieving a person’s psychological distress than comparison treatments or no treatment controls.

1. Method

1.1 Literature Search

Literature searches were conducted in Medline, PsycINFO, PsycArticles, ProQuest, and Science Direct databases. The search was conducted in April 2013 and spanned the previous 30 years. In addition, reference papers from the articles that were retrieved were also
sought. Search terms used were Energy Psychology, Thought Field Therapy, Emotional Freedom Techniques, acupoint tapping, acupoint stimulation, acupuncture point tapping, acupuncture point stimulation, psychological disorder, and randomised outcome or randomized outcome or controlled study. Additional articles were identified by reviewing the research category of the Association for Comprehensive Energy Psychology (ACEP) website and the Energy Psychology Journal. Personal contact was also made with an accredited trainer in Thought Field Therapy to identify any in press articles. Finally a systematic review of the literature by Feinstein (Feinstein, 2012) was used to cross reference any articles identified in the search and to locate any not yet identified that were published up until March 2013.

1.2 Inclusion Criteria

We aimed to assess randomised control trials published in peer reviewed journals that involved treatment of a psychological disorder or psychological distress. The aim of the search was not to restrict the population to any specific type of disorder or distress but to include as many studies as possible that used the intervention of interest. Thus whether the study had participants with a formal diagnosis of PTSD or depression or any disorder they were included. If the study involved psychological distress that was insufficient to meet diagnostic criteria for a formal diagnosis or this was not assessed formally the study was still included. In terms of intervention any study that reported EFT, Thought Field Therapy, acupoint tapping or stimulation, were included. In terms of comparisons the search was not restricted to a comparison intervention of a particular type rather as long as there was a comparison group whether a waitlist control or an active intervention group the study was included. Given the heterogeneity of the studies included, outcome scores were calculated using the measures reported in the included studies.
1.3 Exclusion Criteria

Studies were excluded if they did not randomly assign participants to treatment or if they did not contain a comparison group (either a no treatment control or another type of intervention). Furthermore studies were excluded if the intervention did not target psychological distress or a disorder. Examples of the later was excluding studies where the intervention attempted to improve sport performance

1.4 Quality of included studies

The validity of the included studies was assessed against the Gold Standard (GS) Scale for Post-Traumatic Stress Disorder (PTSD) treatment outcome research (Table 1), adapted from Foa & Meadows (Foa, 1997) by Maxfield & Hyer (Maxfield & Hyer, 2002). This was used because it is an established scale and can be used to study a wide variety of psychological disorders. The included studies were rated on this scale by two Clinical Psychology post graduate students and each study was evaluated and discussed until a consensus was reached.

[Insert Table 1 here]

1.5 Analyses

For each study, Hedge’s g (standardized mean difference) was calculated by subtracting (at post-test) the mean score of the control group from the mean score of the experimental group and dividing the result by the pooled standard deviations of the experimental and control group. Effect sizes of 0.80 and higher are regarded as large, while effect sizes of 0.50-0.80 are moderate, and lower effect sizes are considered small (Cohen,
Hedge’s g is considered to provide a better estimate of effect size compared to Cohen’s d when the sample sizes are small (Grissom, 2010).

The two authors separately calculated effect size data from each study and discrepancies were evaluated and discussed until consensus was reached. Where means and standard deviations were not available in the study (Jones, Thornton, & Andrews, 2010; Schoninger & Hartung, 2010; Waite & Holder, 2003), other statistics (sample size and p-value) were used to calculate the effect size using Comprehensive Meta-Analysis software (CMA) version 2.2057 (Borenstein, Hedges, Higgins, & Rothstein, 2009). In a number of studies standard error (SE) was reported and this was converted to standard deviation (SD) by multiplying the SE by the square root of the sample size minus one.

Mean effect sizes (Hedge’s g) were pooled using CMA. Four studies (Baker & Siegel, 2010; Church, De Asis, & Brooks, 2012; Jain & Rubino, 2012; Waite & Holder, 2003) reported multiple intervention groups. It was decided to use the recommended method of combining the non-EFT intervention groups into a single group. Thus in the study if EFT intervention group was compared to a both an active treatment control (e.g., diaphragmatic breathing) and a no treatment control, then the effects of both these other conditions would be combined to create a single pair-wise comparison as detailed in the Cochrane Collaboration Handbook section 16.5.4 (Higgins & Green, 2011). The alternative strategy of selecting a single pair of interventions (e.g., choosing either ‘diaphragmatic breathing or no intervention as the control) results in a loss of information and is open to results-related choices and is not generally recommended (Higgins & Green, 2011). One study (Salas, 2001) used a within subjects cross-over design. In this instance, rather than discard the study, the mean and standard deviation at the end of the first treatment stage was used as post-test data creating a pair-wise comparison (Higgins, Thompson, Deeks, & Altman, 2003).
In order to generalize to a range of scenarios the random effects model was chosen due to the variation in the studies. Studies differed, as described in Table 2, in the range of independent researchers, sample population (non-clinical, clinical or student), treatment comparisons (EFT compared to EMDR, placebo, treatment as usual or no treatment), dependent variables (Depression, PTSD, Anxiety, Public Speaking, Small Animal Phobia) and outcome measures. By selecting the random effects option in CMA, each study is weighted by the inverse of its variance, in which the variance includes the within-studies variance plus the estimate of the between-studies variance, tau-square (Borenstein, et al., 2009).

To determine the dispersion (true variance and random sampling error) in effect sizes, CMA calculated the I² statistic which is an indicator of heterogeneity or real dispersion between studies and is expressed as a ratio. A value of 0% indicates no observed heterogeneity, and larger values show increasing heterogeneity, with 25% as low, 50% as moderate, and 75% as high heterogeneity (Higgins et al., 2003). Therefore an I² of 75.00 indicates 75% of the effect size is due to true variance between studies while 25% is due to sampling error.

Subgroup analyses were conducted using the mixed effect model recommended by Borenstein et al. (2009). Under this analysis, the random effects model is used for studies within subgroups, and the fixed effects model is used across subgroups. A meta-regression analysis was used to test whether there was a significant relationship between length of treatment time and the effect size, as indicated with a Z-value and an associated p-value. Four subgroup analyses were planned to see if treatment integrity (Maxfield & Hyer, 2002), population characteristics (clinical, non-clinical, student), type of psychological disorder and
treatment comparison (active, non-active controls) moderated the effect size of symptom reduction following EFT.

Publication bias was tested by inspecting the funnel plot on primary outcome measures, and by Duval and Tweedie’s trim and fill procedure (Duval & Tweedie, 2000), which yields an estimate of the effect size after the publication bias has been taken into account (as implemented in Comprehensive Meta-analysis, version 2.2.021).

2. Results

2.1 Inclusion of Studies

The electronic searches, cross referencing with websites, personal communication with authors of some studies, an accredited trainer in the field and the review paper resulted in 82 unique studies (see figure 1). Of these, 42 were excluded because they were either discussion papers (Devilly, 2005; Pignotti & Thyer, 2009), reviews (Feinstein, 2012), reports on case studies (McCarty, 2006) or theoretical framework papers (Church, 2010). A further 4 were excluded as they were either an unpublished Doctor of Psychology dissertation (Connais, 2009; Haynes, 2010; Kruse, 2011) or a Master’s thesis (Salas, 2001). Of the remaining 36 studies, a further 3 were excluded. One was a conference presentation (Dinter & Church, 2009), one was a follow up paper of a previous study (Stapleton, 2012) and the last paper used electrical acupoint stimulation (Zhang, Feng, Xie, Xu, & Chen, 2011). Of the remaining 33 published research papers, 13 were not RCTs and 2 papers (Church, 2009; Llewellyn-Edwards & Llewellyn-Edwards, 2012) investigated sports performance only. In total, 18 studies were located that satisfied the inclusion criteria.

[Insert Figure 1 here]
2.2 Description of Included Studies

The 18 studies examining the impact of Emotional Focused Techniques (EFT treatment) on psychological distress included a total of 921 respondents (475 in the EFT conditions, and 446 in the comparison conditions). Study characteristics are presented in Table 2.

[Insert table 2 here]

In eight of the studies, all or most participants met criteria for a clinical diagnosis. In six studies, participants were students who reported various levels of distress in relation to a specific disorder. In four studies, participants were non-clinical volunteers reporting various levels of distress. Four studies examined PTSD; a further four examined Specific Phobias, while five studies examined an anxiety-related disorder. Two studies examined pain-related disorders, one examined food cravings and one examined non-specific psychological distress. Only one study examined the impact of EFT on depression. All outcome measures used to assess psychological variables in the 18 studies were self-report and specific to the population group they were treating. For example, studies relating to specific phobias (Baker & Siegel, 2010; Salas, 2001; Waite & Holder, 2003; Wells, Polglase, Andrews, Carrington, & Baker, 2003) used the Fear Questionnaire, the Fear of Specific Animal Questionnaire, the Subjective Units of Distress Scale, the Behavioural Approach Task and the Beck Anxiety Inventory. Studies addressing anxiety-related disorders (Irgens, Dammen, Nysæter, & Hoffart, 2012; Jones, et al., 2010; Sachin & Rubino, 2012; Schoninger & Hartung, 2010; Sezgin & Ozcan, 2009) used a variety of measures such as the Symptom Checklist 90 Revised, the Personal Report of Communication Stress, the Personal Report of
Communication Anxiety and the State-Trait Anxiety Inventory. Studies examining PTSD (Church et al., 2013; Church, Piña, Reategui, & Brooks, 2011; Connolly & Sakai, 2011; Karatzias et al., 2011) all used different measures relevant to their populations. Church, Piña, Reategui, & Brooks (2011) used the Impact of Events Scale for adolescent males. Karatzias et al. (2011) used the Clinical Administered PTSD Scale, the PTSD Checklist, the Hospital Anxiety and Depression Scale and the Satisfaction with Life Scale for an adult population (See Table 2 and Table 3 for more details).

2.3 Quality of Included Studies by Randomisation, Blind Assessors and Specific Diagnosis

Only six of the eighteen studies specifically stated that they used an adequate means of randomisation (e.g. by referring to randomizor.org). The remaining twelve studies only stated they used randomization, and in one study (Waite & Holder, 2003) the authors were contacted for clarification of the randomization process. Blinding was also problematic in many of the studies. When assessors are not blind to treatment or independent, both expectations and therapeutic alliance can lead to bias (Maxfield & Hyer, 2002). Only four studies reported that assessors were blind to treatment conditions. Seven studies reported specification of diagnosis using the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV-TR) while the remaining eleven studies reported either using participant self report or screening participants using the experimental measures.

In terms of assessor reliability, thirteen studies did not clearly report whether the assessors were trained, or if the interviews, observations or measurement processes were standardized. The use of manualized treatment was reported by thirteen studies and only four studies reported adequately assessing treatment fidelity. Table 3 summarises these studies in terms of their methodological qualities based on the GS scale.
2.4 Effect Sizes

The results indicating the difference between EFT treatment and controls are presented in Table 4. The effect sizes and 95% confidence intervals of the individual studies are plotted in Fig. 2. The mean effect size indicating the difference between EFT treatment and controls was moderate with Hedge’s $g = -0.66$ (random effects model: 95% CI: -0.99 to -0.33 with $\mu$ (population mean) $=-0.65$) and high heterogeneity ($I^2 = 80.78$) indicating high variability across studies.

2.5 Publication Bias

Examination of the funnel plot (see figure 3) suggested that two studies were possible outliers (greater than 2 standard deviations from the pooled mean of the studies), and their 95% confidence intervals fell outside the 95% confidence interval of the pooled effect size for all studies. After removal of these two studies (Church, et al., 2012; Church, et al., 2011) from the sample the mean effect size was -0.51 (95% CI: -0.78 to -0.23) with a small decrease in heterogeneity though remaining high ($P = 72.32$). Further inspection of the funnel plot and Duval and Tweedie’s (2000) trim and fill procedure indicated a significant publication bias. After adjustment for the bias (2 studies filled to the right) the effect size was -0.51 with an expected slight increase in the confidence interval (random effect model: 95% CI: -0.88 to -0.14).
2.6 Subgroup Analyses

An a priori decision was made to conduct subgroup analyses to examine the possible effect of moderators on the effect size of symptom change following EFT. The data presented above requires examination of the variability across studies as the $I^2$ statistic indicates a high degree of true variance (non-sampling error) between study effect sizes.

[Insert Table 4 here]

The results of the subgroup analyses are presented in Table 4. There were no indications for significant differences between studies that did or did not define symptoms, between studies that did or did not use blind assessors or between studies that used trained assessors. Equally there were no significant differences identified between studies in the way they used or adhered to treatment manuals/protocols or to treatment fidelity. We found no significant differences between studies with active controls such as EMDR or progressive muscle relaxation and non-active controls of waitlist and no treatment. There were also no significant differences between studies in terms of whether the population consisted of clinical, community volunteers, or student participants. Further analysis of moderators found no significant differences between studies aimed at psychological disorders of anxiety, depression, food cravings, post-traumatic stress disorder (PTSD), somatoform or specific phobias.

2.7 Meta-Regression Analysis-Time in Treatment

There was no identified significant relationship between the time spent in treatment and effect size indicating that the amount of treatment time (10 minutes or 360 minutes) was not a moderator. Statistically, the time ratio did not relate to the EFT effect size ($p = 0.70$)
with the point estimate of the slope at -0.03 (95% CL: -0.20 -0.12). The findings of a non-significant relationship must be interpreted with caution due to the very high heterogeneity between studies.

3. **Discussion**

The purpose of the study was to conduct a meta-analysis to investigate whether or not treatment outcomes for psychological distress were efficacious using acupoint stimulation/EFT. In that way, the conclusion that EFT has large effect sizes reported by Feinstein (2012) could be tested. The data was not consistent with a large effect size, the 18 randomized control studies identified resulted in a moderate effect size in favour of EFT using a random effects model. Heterogeneity was high in all calculations even after the removal of outliers.

In treatment outcome studies, it is important to demonstrate that an intervention is better than no intervention (Stevens, Hynan, & Allen, 2000). The findings in this study appear consistent with Feinstein’s conclusions of a positive effect for EFT, i.e. the majority of studies (12) indicate EFT is better than no treatment or waitlist controls. However, study design, treatment comparisons and multiple subgroup analyses (particularly the disorders subgroup analysis) raise major concerns for any conclusive support. Only one of the studies compared EFT to a complete treatment protocol with an evidence base. Five of the remaining studies compared EFT with adjuncts to established minimal interventions such as Diaphragmatic Breathing (DB) or Progressive Muscle Relaxation (PMR). Neither of these minimal interventions is supported as a stand-alone treatment (Conrad, Isaac, & Roth, 2008; Deacon et al., 2012; Stevens, Hynan, Allen, Braun, & McCart, 2007). Only one study, Karatzias (2012) compared EFT with a recognised treatment, Eye Movement Desensitization
and Reprocessing (EMDR), and found both interventions produced significant therapeutic gains with slightly non-significant larger effect size in favour of EMDR.

The mechanisms of action for acupoint stimulation are unclear and several authors have attempted to develop formulations to link the body mind connection as an explanation for reported study effects. Commons (2000) considered EFT treatment to replace the conditioned responses with an unconditioned response reducing the intensity of emotional response elicited by previously conditioned stimuli associated with trauma. Isaacs (2004) suggested that “dual attention” produced by sensory stimuli, tapping and memory recall create a deconditioning response and may be an underlying mechanism in EFT. Feinstein (2012) proposed that the arousal of the amygdala and other brain structures in the stress response are desensitised using EFT. These explanations remain controversial and require further studies to better understand the mechanism of action.

3.1 Directions for future research

No study, identified to date, has conducted a comprehensive analysis of tapping versus no tapping. This, therefore, fails to address the fundamental question “Is acupoint stimulation and EFT an effective treatment for psychological distress?” Only one study, Waite & Holder (2003), attempted to compare EFT with tapping non acupuncture points, tapping a doll and making a paper doll as the control condition. The study produced a small non-significant effect size indicating no significant differences among the first 3 conditions, in reducing reported fear ratings, whereas the control condition showed no decrease. Criticisms of this study were raised in the energy psychology literature reporting fingertips contain meridian endpoints, which are purportedly among the most responsive points on a meridian and the mere use of the fingers to tap “anything” results in the stimulation of these meridians contributing to the effect of EFT on the person(Baker & Carrington, 2005). The
field also called for careful dismantling studies that look at EFT components and operations separately and as yet no studies have been identified.

If EFT is to claim positive clinical outcomes following the tapping of acupuncture points, it is recommended that studies compare tapping with no tapping and conduct the dismantling studies called for since 2005. It is curious why, in nearly a decade, no one in the field has addressed this fundamental concern, with an appropriately designed study. It may be beneficial to focus on one psychological disorder and conduct sufficient rigorous replicable studies to be able to adequately determine the effect size with a greater degree of understanding about the limitations of the data and conversely a greater understanding of the benefits of the treatment (Cumming, 2013). It may be possible that clinicians are motivated by a desire to use treatment protocols that purport to deliver effective outcomes in a short period of time enabling them to service more people. If this assumption has any validity it is incumbent upon clinicians to ensure professional adherence to evidence based practice (American Psychological Association, 2006). This is important in terms of finding a reliable understanding of the mechanisms of action for EFT.

Another possible area for future research is to assess the efficacy of adding EFT to a treatment that is already evidence based and assess if this provides any benefit. For example does a tapping task add anything to prolonged exposure therapy when treating PTSD.

This study has several limitations. Most concerning is that the quality of the studies only met the GS scale at the low to medium quality range with an average score of 4.58 (maximum rating of 7) for all studies. The average rating per item was 0.68 with a range of 1.0 to 0.19, the mode was 3.5 with a 4.5 median score. Other than stating that randomization was used, only six studies described adequate sequence generation with three studies adequately conducting procedures which concealed allocation to respondents. Only five
studies adequately reported assessor training with performance supervision, or reliability checks. The role of the therapist is crucial and therapist competency should be assessed in all treatment outcome studies (Maxfield & Hyer, 2002), yet only four studies met the adequate rating for implementing treatment fidelity procedures. The number of studies included was small, limiting the ability to conduct in depth analysis of possible moderator variables, especially in the disorders subgroup analysis. Heterogeneity was high and possibly due to the fact that there were insufficient numbers of studies addressing the same dependent variables. Under the random-effects model high heterogeneity and a small number of studies limits the potential power of the meta-analysis (Borenstein et al., 2009).

These limitations not with-standing, EFT appears to produce an effect. This finding, however, must be interpreted in light of the methodological shortcomings of the studies which were included in the meta-analysis. While it seems clear that EFT produces an effect, it remains unclear as to whether this effect is due to the claimed therapeutic properties of acupoint stimulation, or if in fact the positive effects are attributable to elements of EFT that are common to long standing evidence based therapies such as CBT. Furthermore the improvement could be due to other so-called general factors such as the therapeutic alliance (Goldfried & Davila, 2005; Price & Jones, 1998). The efficacy of EFT beyond such general therapy effects has yet to be tested.

**Disclosure of Interest.** The second author receives payments for CBT and EMDR trainings.
References


Callahan, R., & Callahan, J. (1996). *Thought Field Therapy (TFT) and Trauma: Treatment and Theory*. Indian Wells, California Callahan Techniques.


