Is Tapping on Acupuncture Points an Active Ingredient in Emotional Freedom Techniques? A Systematic Review and Meta-analysis of Comparative Studies

Dawson Church, PhD,* Peta Stapleton, PhD,† Amy Yang, PhD,‡ and Fred Gallo, PhD§

Abstract: Emotional Freedom Techniques (EFTs) combine elements of cognitive restructuring and exposure techniques with acupoint stimulation. Meta-analyses indicate large effect sizes for posttraumatic stress disorder, depression, and anxiety; however, treatment effects may be due to components EFT shares with other therapies. This review hypothesizes whether EFTs acupoint component was an active ingredient. Six studies of adults with diagnosed or self-identified psychological or physical symptoms were compared (n = 403), and three (n = 102) were identified. Pretest vs. posttest EFT treatment showed a large effect size, Cohen’s d = 1.28 (95% confidence interval [CI], 0.56 to 2.00) and Hedges’ g = 1.25 (95% CI, 0.54 to 1.96). Acupressure groups demonstrated moderately stronger outcomes than controls, with weighted posttreatment effect sizes of d = -0.47 (95% CI, -0.94 to 0.0) and g = -0.45 (95% CI, -0.91 to 0.0). Meta-analysis indicated that the acupressure component was an active ingredient and outcomes were not due solely to placebo, nonspecific effects of any therapy, or nonacupuncture components.

Key Words: EFT, Emotional Freedom Techniques, acupressure, acupuncture points, tapping

(Eur Neurol Dis 2018;206: 783–793)

Emotional Freedom Techniques (EFTs) and other “energy psychology” protocols combine elements of established therapies, such as cognitive restructuring and exposure techniques, with the novel ingredient of acupuncture point (acupoint) stimulation (Gallo, 2004). Acupoint stimulation is delivered in the form of percussion with the fingertips (“acupoint tapping”), a type of acupressure. In acupressure, the same discrete points on the body that are used in acupuncture are manually stimulated for therapeutic effects (Au et al., 2015). Although the term “acupressure” evokes an image of continual pressure, tapping on acupoints is another established way of stimulating them and has been used in methods such as qigong and shiatsu for centuries (Jahneke et al., 2010). For a small proportion of individuals who might find tapping on their skin unsettling, such as those who have been physically abused, an alternative way of stimulating the acupoints within an energy psychology framework is to apply light pressure with the fingertips (Diepold, 2000).

While stimulating these acupoints, clients maintain mental focus on a traumatic event. This is facilitated by EFT’s “setup statement,” which combines exposure with cognitive reframing. The client uses the words “even though” before the triggering phrase, and “I deeply and completely accept myself” after it. A typical setup statement might be, “Even though I was so badly hurt in that terrible car crash, I deeply and completely accept myself.” Although tapping on seven prescribed acupoints, the client then uses a “reminder phrase” to facilitate exposure. An example of a reminder phrase is “that terrible car crash.” The purpose of this study is to investigate whether the acupoint component is an active ingredient in the effects measured in studies of EFT.

Empirical Foundation

Psychological Outcomes

Since the publication of a preliminary review of efficacy evidence for acupoint tapping protocols that target psychological concerns (Feinstein, 2008a), more than 50 peer-reviewed randomized controlled trials (RCTs) have shown favorable clinical outcomes (see online research bibliography at research.EFTuniverse.com). Three meta-analyses have found large effect sizes for EFT in the treatment of anxiety (Clond, 2016), depression (Nelms and Castel, 2016), and posttraumatic stress disorder (PTSD) (Sebastian and Nelms, 2017), respectively. A fourth, combining a range of conditions, showed a medium effect size (Gilomen and Lee, 2015). EFT treatment outcomes with PTSD have been particularly encouraging (Church and Feinstein, 2017; Church et al., 2009, 2013, 2016a; Feinstein, 2008b, 2010; Gallo, 2007; Geronilla et al., 2016; Karatzias et al., 2011). Additional studies show rapid improvements for a variety of other psychological challenges such as performance blocks, excessive food cravings, and stress management (see summaries in Church, 2013b; Gallo, 2002, 2004; Mollon, 2007; Schulz, 2009).

Physiological Outcomes

A range of medical diagnoses have also responded favorably to EFT treatment, including fibromyalgia (Brattberg, 2008; Church, 2013c), hypertension (Bach et al., 2016), frozen shoulder (Church and Nelms, 2016), psoriasis (Hodge and Jurgens, 2011), tension headaches (Bougea et al., 2013), pain (Church, 2014; Church and Brooks, 2010; Ormer, 2015), obesity (Stapleton et al., 2012, 2016), traumatic brain injury (Church and Brooks, 2014), and seizure disorders (Swingle, 2010). The demonstrated ability of EFT to treat this disparate range of psychological and physiological conditions has been attributed to the technique’s capacity to rapidly reduce stress levels (Church et al., 2012; Lane, 2009) combined with its postulated strengths for facilitating the adaptive processing of emotional information (Feinstein, 2015).

Mechanisms of Physiological Action

Functional magnetic resonance imaging (fMRI) and positron emission tomography scans have investigated neurological shifts after acupoint stimulation. In a 10-year research program conducted at Harvard Medical School, brain imaging experiments showed that the needling or electronic stimulation of specified acupoints consistently produced prominent decreases of activation in the amygdala, hippocampus, and other brain areas associated with fear and pain (Fang et al., 2009; Hui et al., 2005; Napadow et al., 2007). Clinically favorable shifts after acupoint tapping sessions have also been reported for a number of biomarkers, including stress hormones (Church et al., 2012), gene expression (Church et al., 2016b; Feinstein and Church, 2010; Maharaj, 2016), electroencephalogram activity (Lambrout et al., 2003; Swingle et al., 2004),
The Current Study

Because EFT protocols include elements of psychotherapies that have proven to be efficacious (e.g., cognitive reframing and imaginal exposure), what is the evidence that the strong treatment effects found in EFT clinical trials is not due solely to these ingredients, and not to acupressure (Bakker, 2013; McCaslin, 2009)? Is the acupressure component of EFT an instance of the “purple hat fallacy” (Rosen and Davison, 2003), in which a therapist might apply an efficacious therapeutic method while wearing a purple hat and then attribute the treatment success to the hat?

In a critique of the claims of acupoint tapping proponents, Bakker (2014) concluded that empirical research had “not been able to demonstrate an effect beyond nonspecific or placebo effects, or the incorporation of known-effective elements (p 44);” that is, that acupressure does not contribute to EFT’s effectiveness. To address this question, Bakker (2013, 2014) and others (Baker et al., 2009; McCaslin, 2009) have recommended that studies be designed to distinguish any effects of acupressure from those of EFT’s cognitive, exposure, non-specific, placebo, and other therapeutic elements. Studies and meta-analyses of other therapies have addressed similar questions, such as whether the eye movements used in eye movement desensitization and reprocessing are a superfluous purple hat or an active therapeutic ingredient (Lee and Cuipers, 2013).

The current analysis examines this question, “Is acupoint tapping an active ingredient in the manualized EFT protocol (Church, 2013a)?” Studies examining this issue were identified in a literature search and assessed for quality. Those that met predefined quality standards were submitted to meta-analysis. The authors’ hypothesis in conducting this study was that EFT protocols which utilize acupoint tapping produce stronger targeted outcomes than comparable protocols that do not use acupressure.

METHODS

The meta-analysis approach used in this article and the structure of the report conform to the Meta-analysis Reporting Standards criteria described by the American Psychological Association (APA) reporting standards group (APA Publications and Communications Board Working Group on Journal Article Reporting Standards, 2008).

Search Strategy

A systematic search identified studies designed to identify the active ingredients in EFT protocols. Sources searched included Pubmed, APA PsychINFO, Google Scholar, the online research bibliography at Research.EFTuniverse.com, and the references sections of articles that were retrieved. Search terms included “EFT,” “Emotional Freedom Technique” (singular), and “Emotional Freedom Techniques” (plural). Reference lists of retrieved studies were also utilized. Research committees of the Association for Comprehensive Energy Psychology, the National Institute for Integrative Healthcare, and The Energy Medicine Institute were queried about unpublished trials investigating EFT’s active ingredients, although no additional studies were identified. The included studies are all English language articles published in peer-reviewed journals up to November 2016.

The research directors of professional organizations in the field were also contacted personally to ascertain whether or not any file drawer studies, that is, studies with unfavorable results not submitted or published, might exist. They were unaware of any such studies.

Selection Criteria

The active ingredients of a psychotherapeutic approach may be identified by studies using dismantling designs, additive designs, or comparative designs (Wampold and Imel, 2015). Dismantling designs remove components of a clinical protocol, one at a time, and compare outcomes with the full protocol; additive designs compare an existing treatment with the existing treatment plus a new component; comparative designs, when applied to isolating active ingredients, compare the full protocol with a similar protocol that omits the component of interest while replacing it with a substitute activity. EFT studies that compared the full EFT protocol with a similar protocol that did not include acupressure met the criteria for inclusion in this review.

Studies using clinical as well as nonclinical populations were included in the selection process because the question being investigated is whether tapping is an essential ingredient in the positive outcomes that have been reported for both populations. Similarly, the effectiveness of EFT protocols has been shown to generalize among age and sex (e.g., Boath et al., 2013), and no selection restrictions were stipulated for either. Studies with relatively small sample sizes were also included because obtaining statistical significance with a small n requires a large treatment effect. Earlier meta-analyses showed that acupoint studies with small sample sizes have produced outcomes similar to those found in studies using larger samples. A reason for conducting a meta-analysis of studies with small samples is, in fact, to discern patterns among a number of samples by accumulating a larger N.

Design Quality

Clinical EFT research has been guided by the criteria established by the APA Division 12 Task Force on Empirically Validated Therapies (Chambless et al., 1996, 1998; Chambless and Hollon, 1998). Seven “essential criteria” based on these articles along with material found on the Division 12-maintained Web site for “Research Supported Psychological Treatments” have been identified by Church and colleagues (2014). Although recommendations for updating these criteria have been made (Church et al., 2014; Tolin et al., 2015), they were in effect when the studies analyzed here were being conducted. They include the following:

- That a sufficient sample size be employed to detect statistically significant effects (p < 0.05 or better);
- That valid and reliable assessment tools be employed;
- That clearly defined treatment populations, as determined by assessment tools and/or diagnosis by qualified clinicians, be employed;
- That assignment to the active treatment and a control condition be randomized;
- Blindness to group assignment by interviewers be maintained in studies using interviews for subject selection;
- That treatment manuals are used, or in the case of simple treatments, full descriptions within the study are provided;
- That sufficient information be included so the study’s conclusions can be reviewed for appropriateness, including sample sizes, use of assessment tools that identify targeted outcomes, and the magnitude of statistical significance.

Church et al. (2014) found that more than half of the published EFT studies met these criteria. To be included in the current meta-analysis, a study had to meet all seven criteria. It also had to provide enough information so that effect sizes could be calculated. We will refer to these criteria as the “Design Quality Criteria.”
Selection Procedures

Whether a retrieved study addressed the question of acupoint tapping as an active ingredient in EFT protocols was determined by the principal investigator based on the title of the article and the abstract or full article if the title left any doubt. Study quality determinations based on the Design Quality Criteria were made independently by the first author and an independent colleague and then compared. Although consensus was reached on each of the ratings, the remaining authors would have been consulted to discuss differences of opinion until consensus was attained.

Statistical Methods

The issue of how many studies are required for a meta-analysis is debated by scholars. Although at least two are required, there is not a linear association between the number of studies and the quality of the analysis (Valentine et al., 2010). The reason for this is that a meta-analysis that includes a large number of poorly designed studies will be inferior to one with fewer well-designed studies. Quality control criteria such as those used in this article are therefore appropriate and may eliminate studies, in our case, half of the candidates.

Some have argued that in the case of either a small number of studies or a large number with quality issues, a narrative review should be performed instead of a meta-analysis (Bailar, 1997). This view is decisively rebutted in the most-cited textbook on conducting meta-analyses (Borenstein et al., 2009). Narrative reviews lack the quantitative synthesis provided by a meta-analysis. Although they can evaluate studies, they do not provide calculations of effect size that allow clinicians to determine the effectiveness of an intervention. By aggregating data, meta-analyses can also overcome the limitations of single studies, which are underpowered, but nevertheless be assessing an effective intervention. Finally, because of the transparent statistical methods used in meta-analyses, their results are likely to lead to similar conclusions provided that the differences of opinion that might occur in narrative reviews.

A more graduated interpretation of Cohen’s (1988) three levels of effect size using d or g (small, moderate, and large) is provided by Cicchetti et al. (2011). They advocate the calculation of $r$ as well, and add the categories of “trivial” for studies with values less than 0.10 and “very large” for those greater than 0.70.

Therefore, for direct comparison of the two forms of treatment being investigated (with or without acupoint tapping), we calculated the post/post effect sizes, taking account of group differences at the baseline/pretest as well as the correlation values. Because such baseline differences can cause distortions, the post/post effect sizes were corrected by using the following formula: Effect size corr = Effect size post/post = Effect size pre/pre (described in Becker, 1988; usefully modeled by Seidler and Wagner, 2006). Cohen’s d and Hedges’ g were both used as the means of the effect sizes to facilitate comparisons with other meta-analyses using one or the other. Both effect size numbers are generally interpreted using the convention of small (0.2), medium (0.5), and large (>0.8) effects (Cohen, 1988). The percentage of score change from pretest to posttest was calculated for both the acupoint tapping and nonacupoint conditions. We standardized the scores first by subtracting the lowest possible scores for each measure to make the percentage change comparable between measures.

The next step was the calculation of a weighted effect size for both the pre/post EFT effect and the corrected post/post effect between the EFT and the control treatments. The overall mean effect size for all of the studies combined was weighted by the inverse of the variance of the studies, considering both standard deviations and number of subjects. We also assessed heterogeneity, using “Q” (Borenstein et al., 2009, p 109) to determine whether a randomized or fixed effects model should be used for computing estimates of effect size. All analyses were conducted using “R statistical software” (version 3.2.0).

RESULTS

Retrieved Studies

The literature search identified 249 articles on the general topic of EFTs. Six of them (Church and Nelms, 2016; Fox, 2013; Reynolds, 2015; Rogers and Sears, 2015; Waite and Holder, 2003; Wells et al., 2003) provided information directly applicable to the question of whether acupressure is an active ingredient. All six were comparative studies in which an EFT protocol that included acupoint tapping was compared with a similar protocol in which acupoint tapping was replaced with a substitute activity. No studies using purely dismantling or additive designs were located.

The pooled sample among the six studies included 403 subjects. The study characteristics listed in Table 1 include the conditions treated, the experimental and control treatments, the number of subjects in each treatment group, the assessment instruments used, and the principal findings. All six studies had the experimental group tap on acupoints. Three had the comparative group also tap on sham points (points not used in EFT); two used diaphragmatic breathing (DB) as the comparison; and one used a mindfulness breathing technique (see Fig. 1 for flow of study selection).

Studies Excluded From the Meta-analysis

The presence or absence of each of the seven predetermined Design Quality Criteria for each of the six studies is indicated in Table 2. Three of the six studies did not meet all seven criteria, which was required for inclusion in the meta-analysis. They nonetheless provide information pertinent to the effects of acupoint tapping, which are summarized here.

Waite and Holder

The first study to attempt to test whether acupoint tapping is an active ingredient in EFT was conducted by Waite and Holder (2003) with 119 university students, measuring pretreatment and posttreatment self-reported fear ratings. Significant improvements were found on fear ratings for three tapping conditions and no improvement for a nontapping control. Two of the three tapping conditions were intended as placebo treatments: tapping on sham points and tapping on a doll. Because both placebo tapping conditions produced therapeutic effects, the investigators concluded that any effect from EFT was due to distraction, desensitization, or demand characteristics rather than acupoint tapping.

However, in both the sham point and doll tapping groups, the investigators had subjects tapping with the forefinger, inadvertently stimulating an acupuncture point (Large Intestine 1) that is sometimes used in the treatment of “mental restlessness” (Ross, 1995, p 306). Objects such as the doll employed in one control group are noted to possess therapeutic potency in psychotherapeutic methods based on symbolism (Combs and Freedman, 1990). Thus, both comparison conditions that used tapping, which were effective in alleviating fear, were ultimately lending support for the efficacy of acupoint tapping (an argument that has been made by Baker et al., 2009, and by Pasahow, 2010). The study did not, however, meet the Design Quality Criteria for the meta-analysis. The self-report questionnaire had not been standardized for validity or reliability; a rotating group assignment was used instead of randomization; and not enough information was provided to calculate effect sizes. In addition, the study did not adhere to the manualized form of EFT. Instead, participants in the EFT group were instructed to include points not used in EFT (e.g., below the knuckles of the index and middle fingers) while omitting other points (e.g., Triple Warmer 3).
### TABLE 1. Study Characteristics

<table>
<thead>
<tr>
<th>Author (Date)</th>
<th>Condition Treated</th>
<th>Treatments and Controls</th>
<th>n</th>
<th>Assessments</th>
<th>Principal Findings</th>
<th>Statistically Significant Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Church and Nelms (2016)</td>
<td>ROM, pain, anxiety, depression, and seven other psychological conditions</td>
<td>EFT 16</td>
<td></td>
<td>ROM, SUDS, SA-45</td>
<td>No psychological improvement in WL. Posttest improvement in psychological symptoms and pain in DB and EFT groups. On follow-up, both EFT and DB maintained gains for pain, with EFT superior to DB, but only EFT group maintained gains for psychological symptoms.</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DB 18</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>WL 13</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fox (2013)</td>
<td>Study-related emotions</td>
<td>EFT 10</td>
<td></td>
<td>AEQ, PHLMS</td>
<td>EFT group experienced significantly greater increases in enjoyment and hope and significantly greater decreases in anger and shame than did MB control group. No significant change was found for mindfulness.</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MB 10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reynolds (2015)</td>
<td>Professional burnout</td>
<td>EFT 79</td>
<td></td>
<td>MBI</td>
<td>On all three indicators of burnout measured, EFT was significantly superior to the sham tapping control.</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sham 47</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rogers and Sears (2015)</td>
<td>Stress</td>
<td>EFT 26</td>
<td></td>
<td>Stress symptoms</td>
<td>Stress symptoms were significantly reduced in the EFT group compared with the sham tapping control.</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sham 30</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Waite and Holder (2003)</td>
<td>Anxiety and fear</td>
<td>EFT 29</td>
<td></td>
<td>SUDS</td>
<td>EFT, sham, and modeling groups all showed a similar significant decrease in self-report measures for fear, whereas NT control group did not.</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sham 29</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Modeling 32</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>NT 29</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wells et al. (2003)</td>
<td>Specific phobias of small animals, fear</td>
<td>EFT 18</td>
<td></td>
<td>BAT, FQ, SUDS, pulse</td>
<td>EFT produced significantly greater improvement than did DB behaviorally and on three self-report measures, but not on pulse rate. The improvement for EFT was maintained at follow-up.</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DB 17</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

ROM, range of motion; MB, mindful breathing; Sham, sham acupressure points; NT, no treatment; WL, wait list; NA, not applicable; SUDS, subjective units of distress; PHLMS, Philadelphia Mindfulness Scale.

---

**FIGURE 1.** Flow chart of literature search and study selection.
TABLE 2. Studies Included in This Meta-analysis Evaluated on the APA’s Seven Evidence-Based Criteria

<table>
<thead>
<tr>
<th>Author (Date)</th>
<th>Randomization</th>
<th>Defined Population</th>
<th>Sufficient Sample Size</th>
<th>Reliable and Valid Procedures</th>
<th>Clear Treatment Procedures</th>
<th>Sufficient Data for Calculating Effect Sizes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Church and Nelms (2016)</td>
<td>Yes</td>
<td>Participants with clinically verified ROM limitations at three Northern California Integrative Health clinics</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Ye s</td>
<td></td>
<td>Undergraduate students at Liverpool John Moores University</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ye s</td>
<td></td>
<td>Full-time teaching professionals in Northern California</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y es</td>
<td></td>
<td>Students attending Okanagan University College with specific phobias</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fox (2013)</td>
<td>No</td>
<td>Undergraduate students at Liverpool John Moores University</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Ye s</td>
<td></td>
<td>Full-time teaching professionals in Northern California</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reynolds (2015)</td>
<td>No</td>
<td>Participants with spider, mouse, rat, or cockroach phobias</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Ye s</td>
<td></td>
<td>University College with specific phobias</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rogers and Sears (2015)</td>
<td>No</td>
<td>Participants with spider, mouse, rat, or cockroach phobias</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Ye s</td>
<td></td>
<td>University College with specific phobias</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wells et al. (2003)</td>
<td>Yes</td>
<td>Participants with spider, mouse, rat, or cockroach phobias</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Ye s</td>
<td></td>
<td>University College with specific phobias</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* $\text{Effect sizes were} < 0.1$, small; $0.2$, medium; $0.3$, large (Cohen, 1988). Effect sizes: $0.1$, trivial; $0.2$, small; $0.3$, medium; $0.5$, large; $d = 0.7$ and above, very large (Cohen, 1988).

Rogers and Sears
A university study randomized 56 students to an EFT group or a control group (Rogers and Sears, 2015). Both groups were given identical protocols except for the EFT group tapped on the EFT acupoints and the control group tapped on sham points (areas of the skin not identified in the acupuncture literature as treatment points). An inventory of common stress symptoms was administered before and after the two group tapping sessions, which were of 15 to 20 minutes each. Symptoms were reduced by 39.3% in the acupoint tapping group and by 8.1% in the sham tapping group ($p < 0.001$). Again, however, the stress inventory had not been validated, disqualifying the study from inclusion in the meta-analysis. Other weaknesses included the possibility that the sham tapping inadvertently stimulated a traditional acupoint and that one of the investigators conducted the treatments for both groups.

Reynolds
Reynolds (2015) assessed 126 full-time public school teachers (K-12) for burnout risk using the Maslach Burnout Inventory (MBI) (Maslach et al., 1996). EFT was compared with a control group that tapped on sham points on the body in an otherwise identical protocol. To minimize contact and contamination between the treatment and control participants, the two samples were drawn from different school districts but within the same county and with similar demographic profiles. Both the EFT and control groups received identical instructions about situations and cognitions that might contribute to burnout and were asked to focus on them mentally while doing the tapping.

The control group tapped on the left forearm, about an inch above the wrist, with the underside of the fingers of the open right hand. This positioning is important as acupoints at the tips of the fingers will be stimulated if used in tapping, whereas no acupoints are stimulated using the instructed method. This was the first sham point study to correct for Waite and Holder's (2003) having inadvertently stimulated fingertip acupoints. With all other components of the EFT protocol being identical for both groups, the EFT points were superior to the sham points ($p < 0.05$) on the three indicators of burnout being tracked (emotional exhaustion, depersonalization, and personal accomplishment). However, the study was excluded from the meta-analysis because randomization was not used and not enough information to calculate effect sizes was provided.

Studies Included in the Meta-analysis
Three studies ($n = 102$) met the seven predetermined Design Quality Criteria. Each is briefly summarized here.

Wells et al
The first peer-reviewed investigation of EFT used a comparison group that substituted DB for acupoint stimulation (Wells et al., 2003). The only other component of the EFT protocol that was omitted from the comparison condition was use of the setup statement. The outcome on a variety of self-report and behavioral measures was superior for the EFT group and held on follow-up. The study was replicated, with design changes to control carefully for all possible confounding variables such as expectancy effects and nonspecific factors (Baker and Siegel, 2010). This replication corroborated the outcomes identified by Wells et al. (2003).

Fox
A convenience sample of university students was convened and randomly assigned to a treatment session using either the full EFT protocol or a comparison condition in which a mindfulness breathing exercise was substituted for the acupoint tapping (Fox, 2013). The tapping group experienced significantly greater increases in “positive emotions” (e.g., enjoyment, hope, and pride; $p < 0.01$) and greater decreases...
in “negative emotions” (e.g., anger, anxiety, and shame; \( p < 0.01 \)) on a standardized measure than did the comparison group. As with Wells et al. (2003), although the control group maintained mental focus on the traumatic event, they did not use the verbal structure found in EFT’s setup statement.

**Church and Nelms**

Participants diagnosed with frozen shoulder syndrome (limited range of motion in either arm) were recruited and randomly assigned to a wait list or one of two treatment conditions (Church and Nelms, 2016). One treatment group received a 30-minute EFT session that included acupoint tapping. The other received an identical protocol but tapping was replaced by DB (Martarelli et al., 2011). Participants in both active treatment groups showed significant improvements in pain and psychological symptoms posttest. On follow-up, both groups maintained their gains for pain, with EFT being statistically superior to DB. In relation to psychological symptoms, only those in the EFT group maintained their gains on follow-up (\( p < 0.001 \)). The EFT group also showed greater improvement in range of motion than the DB group.

Treatment effects were calculated using Cohen’s \( d \). EFT showed a large effect size for pain (\( d = 0.9 \)), anxiety (\( d = 0.9 \)), and depression (\( d = 1.1 \)). The study was explicitly designed to meet both the CONSORT (Wells et al., 2003) and the results were as follows:

- **Table 1**: Includes measures for calculating effect sizes: a) Symptom Assessment-45 Questionnaire anxiety subscale; b) positive symptom total (PST); c) Achievement Emotions Questionnaire anxiety and negative emotion subscales; d) MBI emotional exhaustion subscale; e) stress inventory; f) Behavioral Approach Task (BAT); and g) Fear Questionnaire (FQ) (see abbreviation legend in Table 1).

For all the measures except the BAT, a lower score indicated improvement. The positive effect sizes ranged from 0.49 to 3.14 for these measures, with improvement being based on a decreased score from pretest to posttest. For BAT, a higher score represents improvement, thus an effect size of \(-1.1\) translates to improvement at posttest. We reversed the sign for effect size of BAT before the analysis so that positive values represent improvement for all measures.

The weighted effect sizes for the three studies were high (\( >0.8 \)), with a Cohen’s \( d \) of 1.28 (95% confidence interval [CI], 0.56 to 2.00) and a Hedges’ \( g \) of 1.25 (95% CI, 0.54 to 1.96).

**Measures for Calculating Effect Sizes**

**Heterogeneity**

Heterogeneity, assessed using “Q” (Borenstein et al., 2009), was high within the pre/post effect sizes in the EFT group as well as within the post/test effect sizes between EFT and the controls (from 17.09 to 24.02; \( p < 0.005 \)). This was expected given the small number of studies included in the analysis. Based on this high heterogeneity, we selected a random effects model for the meta-analysis.

**EFT vs. Comparison Condition Effect Sizes**

Comparisons of the full EFT protocol with similar protocols that substituted acupoint tapping with an alternative activity (tapping on sham points, DB, or a meditative breathing technique) were made to directly address the question of whether acupoint tapping is an active or inert component in EFT. Table 4 shows these comparisons. Each study found a significant posttreatment improvement for participants in the EFT group (\( p < 0.02 \)). The protocols that used a substitute for acupoint tapping showed a significant positive effect in two of the studies (Church and Nelms, 2016; Wells et al., 2003), although the acupoint tapping condition produced significantly stronger effects in each. Again, appropriate adjustments were made for the BAT due to a higher rather than lower score indicating improvement.

Table 4 compares effect sizes at pretest to determine the baseline differences between the groups (pre/pre), at posttest for both groups (post/post), and posttest corrected for baseline differences using the formula described in the Statistical Methods section (post/post corrected). The weighted post/post corrected effect sizes are 0.47 (95% CI, −0.94 to 0.0) for Cohen’s \( d \) and 0.45 (95% CI, −0.91 to 0.0) for Hedges’ \( g \). Using the convention that 0.5 (which may be positive or negative) indicates a medium treatment effect, the calculations demonstrate a moderately superior treatment result from EFT with acupressure when compared with control treatments. The bivariate correlations between the pre and post differences ranged from weak to strong (see Table 4).

We wondered whether the inclusion of RCTs that did not meet the Design Quality Criteria would have changed the size of the treatment effect. The only excluded study for which enough information was presented to determine effect size was Rogers and Sears (2015), and we performed an additional calculation that included it. This did change the overall effect sizes, but only slightly, to \( d = 1.25 \) (95% CI, 0.66 to 1.84), as contrasted with the weighted effect sizes for the three qualifying studies of \( d = 1.28 \) (95% CI, 0.56 to 2.00). These findings, whether including or excluding the study with the weaker design, indicate that the full EFT protocol has a stronger effect than the protocols that substituted acupoint tapping with an active control.

**DISCUSSION**

Six studies of varying quality were retrieved in a literature search, which compared the full EFT protocol with a similar protocol that substituted acupoint tapping with either a breathing technique or

---

**TABLE 3. Effect Sizes for Pre/Post EFT for Those Studies Meeting APA Criteria**

<table>
<thead>
<tr>
<th>Author (Date)</th>
<th>No. Treatment</th>
<th>Measure</th>
<th>Pretest Mean (SD)</th>
<th>Posttest Mean (SD)</th>
<th>Cohen’s d (SE)</th>
<th>Hedges’ g (SE)</th>
<th>r*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Church and Nelms (2016)</td>
<td>One 30-minute session</td>
<td>Anxiety subscale</td>
<td>58 (9)</td>
<td>52 (7)</td>
<td>0.74 (0.37)</td>
<td>0.73 (0.37)</td>
<td>0.35</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PST</td>
<td>56 (0)</td>
<td>50 (9)</td>
<td>0.94 (0.37)</td>
<td>0.92 (0.37)</td>
<td>0.42</td>
</tr>
<tr>
<td>Fox (2013)</td>
<td>One 45-minute session</td>
<td>Anxiety subscale</td>
<td>2.63 (0.66)</td>
<td>2.34 (0.61)</td>
<td>0.52 (0.45)</td>
<td>0.49 (0.45)</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Negative emotion</td>
<td>2.35 (0.49)</td>
<td>2.04 (0.42)</td>
<td>0.68 (0.46)</td>
<td>0.65 (0.46)</td>
<td>0.32</td>
</tr>
<tr>
<td>Wells et al. (2003)</td>
<td>One 30-minute session</td>
<td>BAT</td>
<td>4.5 (2.5)</td>
<td>6.8 (1.7)</td>
<td>−1.1 (0.36)</td>
<td>−1.1 (0.36)</td>
<td>0.47</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fear questionnaire</td>
<td>31.9 (7.3)</td>
<td>15.1 (2.0)</td>
<td>3.14 (0.50)</td>
<td>3.07 (0.49)</td>
<td>0.84</td>
</tr>
</tbody>
</table>

*r value: 0.1, small; 0.3, medium; 0.5, large (Cohen, 1988). Effect size: <0.1, trivial; 0.2, small; 0.3, medium; d = 0.5 large; d = 0.7 and above, very large (Cicchetti, 2008).
tapping on sham points. One of the studies (Waite and Holder, 2003), which was designed to have participants tap on sham points, inadvertently used traditional acupuncture points as the comparison condition. The other five studies found that the protocol that used acupoint tapping produced outcomes that were superior to those of the comparison treatment. The three studies that met quality criteria and were submitted to meta-analysis showed pretreatment to posttreatment effect sizes on at least one measure that were medium (Fox, 2013) or high (Church and Nelms, 2016; Wells et al., 2003). Meta-analysis showed a moderate aggregate effect when the acupoint tapping condition was compared with the comparison condition, indicating that acupressure made a contribution to the treatment effects measured. The treatment effect sizes measured in the meta-analysis were in broad agreement with the results of the more than 100 clinical trials of EFT found in the literature.

These findings are consistent with evidence from the broader acupressure field. A systematic review of 66 studies (n = 7265) comparing traditional point acupressure with sham point acupressure found both to have a clinical effect, but the traditional points had a significantly stronger effect than the sham points (Tan et al., 2015). For example, in one of the reviewed studies, paramedics were taught to administer acupressure, using either points that are typically employed for pain and anxiety or sham points (Kober et al., 2002). Brief nontapping acupressure treatments were provided to patients for whom an ambulance had been called but who did not require an immediate medical intervention. Those receiving acupressure with the true points had significantly less pain and anxiety upon reaching the hospital than patients receiving sham point treatments (p < 0.01). A partial replication yielded similar findings (Lang et al., 2007).

### Limitations and Justifications

Studies investigating whether acupoint tapping is an active ingredient in the clinical outcomes produced by EFT are limited in number and vary in their design, rigor, sample size, and the conditions treated. The three studies meeting the preidentified quality criteria were each based on a single treatment session, the populations treated were highly heterogeneous (from those with self-identified fears of small animals to individuals presenting with shoulder pain to an opportunity sample of college undergraduates), and the studies had relatively small sample sizes (20, 35, and 47, respectively).

Although a full course of treatment would provide information that a single treatment session does not, the focus of this study is whether acupoint tapping protocols are superior to similar protocols that do not include acupoint tapping. If a difference is found after a single session, the question has been addressed in a manner that isolates as many variables as possible. Similarly, finding the same effects with a range of populations suggests that the active mechanism operates independent of the population, a premise about EFT that was affirmed in an explicit test of EFT's generalizability by Boath et al. (2013). Of course, studies that include multiple treatment sessions for a variety of diagnosed mental conditions would provide useful information, but existing evidence suggests that they would be consistent with the findings of the investigations retrieved for this review.

Studies with small sample sizes run the risk of failing to detect a true effect, of limiting the degree to which the study’s findings can be generalized to a target population, and may lend presumed evidence to a false premise (Button et al., 2013; Faber and Fonseca, 2014). Although all three studies in the meta-analysis met the Design Quality Criteria requiring that the sample size was capable of detecting statistically significant effects, each sample size was relatively small. Small sample sizes, however, necessitate larger effects to achieve statistical significance (which is the reason small sample sizes are prone to not detect existing effects).

The use of meta-analysis also reveals trends common to the individual studies. In the case of the current investigation, each of the
studies for which effect sizes could be calculated found the influence to be in the same direction, with acupoint tapping having a stronger effect than the comparison condition. Viewing these findings in the context of the extensive acupressure investigations showing that traditional points are more effective than sham points (Tan et al., 2015) further corroborates the findings of the EFT comparison studies.

Most meta-analyses include more than the three studies that were available for this review, and therefore, one cannot ignore that the individual studies may have had potential differences. Meta-analyses can, however, be used to detect early trends in a body of research (Davey et al., 2011), particularly in a new area of investigation or when focusing on an isolated question such as active ingredients. The issue examined here is important for the field, and all available evidence was considered for the analysis. In posing the question “How many studies do you need to do a meta-analysis,” Valentine et al. (2010) argued that useful information can be deduced from just two studies “because all other synthesis techniques are less transparent and/or less likely to be valid” (p 245). In this case, the meta-analysis findings were consistent with the conclusions reached by the investigators in the individual studies.

Limitations of the meta-analytic technique itself (Hofmann and Smits, 2008) also, of course, apply. Other limitations of the current meta-analysis are that two of the qualifying studies lacked features of more robust designs—such as diagnostic pretreatment and posttreatment interviews and stringent measures to insure compliance with the protocols being tested—shortcomings that should be corrected in future investigations. The principal investigators were also proponents of EFT, a source of potential bias.

Comparison studies that substitute a component of the full protocol with an activity that is not believed to be as therapeutic may be biased by the therapist's allegiance to or faith in the full protocol (Wampold and Imel, 2015). The therapist providing the comparison treatments for Church and Nelms (2016) was selected because he regularly used DB, the substitute activity, in his practice, and was not an advocate of EFT. The articles presenting the other two studies did not report any attempts to control for therapist allegiance. Finally, although we searched for “grey literature” by asking the research committees of relevant professional organizations whether unpublished trials might exist, it is possible that unknown studies with null results are extant, reflecting publication bias.

Although the studies reviewed here suffer from the above limitations, each provides evidence bearing upon the question of whether tapping is an essential ingredient in EFT protocols, and each corroborates the essential findings of each of the others for which effect sizes could be calculated. In all instances, the protocols that included acupoint tapping produced significantly stronger effects than the protocols that did not.

Clinical Implications

Studies attempting to identify the essential components of various psychotherapies have been largely unsuccessful, with less than one percent of the variability in therapy outcome being attributable to the treatment method (Wampold and Imel, 2015). Further, Wampold and Imel (2015) distinguish between dismantling, additive, and comparative experimental designs. The studies examined in this meta-analysis were comparative studies; they included a substitute activity that was either an active control or a sham treatment. By using a substitute activity, comparative designs control for the placebo effect, which dismantling and additive designs do not. Differences in additive and dismantling designs might be due partly or solely to the placebo effect produced by the added component, whereas comparative designs control for this confounding variable with a plausible placebo or active treatment. The studies examined, therefore, controlled for the placebo effect.

Each of the EFT comparison studies found the protocols that included acupoint tapping to produce significantly stronger outcomes than those that did not, unlike the very small differences found by Wampold and Imel (2015). One possible explanation for the effects of acupoint tapping being so strong goes beyond EFT. Advantages of somatic interventions, particularly in the treatment of trauma-based conditions, have been gaining increasing recognition since the publication of Van der Kolk (1994, 2015). Somatic interventions are qualitatively different from other components of psychotherapy (Church, 2016; May, 2005; Rothschild, 2000), and acupoint tapping is a body-based approach that is easily learned and integrated into psychological frameworks. In fact, a recent survey of the membership of the Association for Comprehensive Energy Psychology showed that the majority of licensed therapists adopting acupoint tapping combine it with other clinical methods rather than using as a stand-alone treatment (Feinstein, 2016).

Although the introduction of acupoint tapping into modern clinical practice is relatively recent (Gallo, 2004), the use of acupressure for emotional disorders extends back at least 5000 years (Gach and Henning, 2004). Acupressure, used independent of EFT, has been shown to be effective with a range of physical and psychological conditions, particularly pain. Meta-analyses of acupressure studies for pain and nausea have found highly significant ($p < 0.0001$) treatment effects (Chen et al., 2013; Helmreich et al., 2006). If acupoint tapping done simultaneously with established mental health interventions appreciably increases therapeutic impact, the implications for clinical practice could be substantial.

Suggestions for Further Research

The evidence evaluated here, demonstrating that acupoint tapping is an active ingredient in EFT protocols, can be reinforced with studies using larger sample sizes, multisession treatments, and more stringent designs. Further research into the mechanisms that allow the rapid responses reported in clinical trials will give clinicians greater confidence in applying the technique. Meanwhile, with growing numbers of clinicians introducing EFT into their practices, research on the integration of acupoint tapping with conventional treatment approaches, such as cognitive behavior therapy and psychodynamic psychotherapy, will provide guidance on a variety of practical clinical questions.

CONCLUSIONS

In the six studies that have examined the active ingredients of EFT, investigators have come to divergent conclusions. The authors of five of the articles reported that their findings lend support to the premise that acupoint tapping is an active ingredient; the authors of the remaining study (Waite and Holder, 2003) speculated that other factors, such as “systematic desensitization and distraction,” may be “mediators of EFT’s apparent effectiveness” (p 24). The current article sought to address this fundamental discrepancy. All six studies were reviewed, and a meta-analysis of the three that met the quality criteria published by the APA Division 12 Task Force on Empirically Validated Treatments was conducted. Although EFT borrows components from other treatment methods, especially cognitive therapy and imaginative exposure, the evidence analyzed in this report supports the hypothesis that acupoint tapping is an active ingredient in EFTs efficacy and that the large treatment effects measured are not due solely to the components EFT shares with other methods.

DISCLOSURE

Dawson Church, Peta Stapleton, and Fred Gallo may provide services and trainings using the methods reviewed in this article.

Funding for the data analysis was provided by the National Institute for Integrative Healthcare (NIIH.org).
REFERENCES


Rosen GM, Davison GC (2003) Psychology should list empirically supported principles of change (ESPs) and not credential trademarked therapies or other treatment packages. *Behav Modif.* 27:300–312.


