A SYSTEMATIC REVIEW AND META-ANALYSIS OF RANDOMIZED AND NONRANDOMIZED TRIALS OF CLINICAL EMOTIONAL FREEDOM TECHNIQUES (EFT) FOR THE TREATMENT OF DEPRESSION
Jerrod A. Nelms, PhD, MPH*, and Liana Castel, PhD, MSPH

Background: Among a group of therapies collectively known as energy psychology (EP), emotional freedom techniques (EFT) is the most widely practiced. Clinical EFT is an evidence-based practice combining elements of cognitive and exposure therapies with the manual stimulation of acupuncture points (acupoints). Lacking is a recent quantitative meta-analysis that enhances understanding of the variability and clinical significance of outcomes after clinical EFT treatment in reducing depression.

Methods: All studies (2005–2015) evaluating EFT for sufferers of depression were identified by electronic search; these included both outcome studies and randomized controlled trials (RCTs). Our focus was depressive symptoms as measured by a variety of psychometric questionnaires and scales. We used meta-analysis to calculate effect sizes at three time points including posttest, follow-ups less than 90 days, and follow-ups more than 90 days.

Results: In total, 20 studies were qualified for inclusion, 12 RCTs and 8 outcome studies. The number of participants treated with EFT included \(N = 461\) in outcome studies and \(N = 398\) in RCTs. Clinical EFT showed a large effect size in the treatment of depression in RCTs. At posttest, Cohen’s \(d\) for RCTs was 1.85 and for outcome studies was 0.70. Effect sizes for follow-ups less than 90 days were 1.21, and for \(\geq 90\) days were 1.11. EFT were more efficacious than diaphragmatic breathing (DB) and supportive interview (SI) in posttest measurements (\(P = .06\) versus DB, \(P < .001\) versus SI), and sleep hygiene education (SHE) at follow-up (\(P = .036\)). No significant treatment effect difference between EFT and eye movement desensitization and reprocessing (EMDR) was found. EFT were superior to treatment as usual (TAU), and efficacious in treatment time frames ranging from 1 to 10 sessions. The mean of symptom reductions across all studies was \(-41\%\).

Conclusions: The results show that Clinical EFT were highly effective in reducing depressive symptoms in a variety of populations and settings. EFT were equal or superior to TAU and other active treatment controls. The posttest effect size for EFT (\(d = 1.31\)) was larger than that measured in meta-analyses of antidepressant drug trials and psychotherapy studies. EFT produced large treatment effects whether delivered in group or individual format, and participants maintained their gains over time. This meta-analysis extends the existing literature through facilitation of a better understanding of the variability and clinical significance of depression improvement subsequent to EFT treatment.

Key words: meta-analysis, depression, PTSD

(Explore 2016; 12:416-426 © 2016 Elsevier Inc. All rights reserved.)

BACKGROUND
Emotional freedom techniques (EFT) is an evidence-based practice, with more than 100 randomized controlled trials (RCTs), outcome studies, and review articles published in peer-reviewed journals listed in an online research bibliography (Research.EFTuniverse.com). EFT meets the criteria of the American Psychological Association’s Division 12 Task Force on Empirically Validated Treatments (hereafter referred to as APA Division 12 criteria) for a number of psychological conditions, including anxiety, depression, phobias, and PTSD.\(^1\) EFT has also been found to produce improvement in physical conditions such as fibromyalgia,\(^2\) psoriasis,\(^3\) tension headaches,\(^4\) pain,\(^5\) traumatic brain injury,\(^6\) and seizure disorders.\(^7\) In service evaluations performed by the UK National Health Service, EFT has been found to produce an improvement in general physical functioning as well as mental health.\(^8\)–\(^10\) The wide range of conditions for which EFT are effective are usually attributed to the technique’s ability to reduce stress, which is a component of many emotional and physical disorders.\(^11\),\(^12\)

Lucyna Health and Safety Solutions, LLC, Tampa, FL
#Corresponding author.
e-mail: jnelms@lucynasolutions.com

416 © 2016 Elsevier Inc. All rights reserved.
ISSN 1550-8307/$36.00
EXPLORE November/December 2016, Vol. 12, No. 6
http://dx.doi.org/10.1016/j.explore.2016.08.001
EFT treatment is standardized via a manual, with most published studies applying the manualized form of the technique. EFT combines acupressure with elements of two other evidence-based psychotherapeutic techniques, namely exposure therapy and cognitive therapy. While vividly recalling a traumatic event, clients are instructed to tap on 12 acupressure points with their fingertips. The client reports the emotional intensity of the event on an 11-point Likert scale both before and after the application of EFT. Studies and clinical reports note rapid reductions in distress after tapping, client safety, and an absence of adverse events after treatment.

Clinical EFT is the manualized evidence-based form of the original EFT methodology.

Clinical EFT is generally considered safe, with few accounts of abreactions or emotional flooding. These benefits extend even to the most highly traumatized populations, such as combat veterans with PTSD, Rwandan genocide orphans, and Haitian earthquake survivors. In a recent critical survey of therapists conducted through Listservs such as acceptance and commitment therapy, the society for the science of clinical psychology, and the association of behavioral and cognitive therapies regarding their use of EFT and similar methods, 42% reported employing these techniques.

Acupoint tapping has been evaluated in several studies using electroencephalogram (EEG) imaging to record brain waves. Swingle et al. used the EEG to compare the brainwave frequencies of automobile accident victims before and after they performed EFT, and noted a reduction of those wave frequencies of automobile accident victims before and after treatment in theta EEG frequencies, which are associated with PTSD. Lambrou et al. used acupressure tapping with clausrophobics, comparing them with a non-clausrophobic group, and found an increase after treatment in theta EEG frequencies, which are associated with relaxation, and a reduction in anxiety symptoms, which persisted on follow-up. Swingle found EFT to be beneficial in the treatment of seizure disorders. Regulation of the autonomic nervous system and the fear response has been found in fMRI studies using acupuncture, and similar effects for acupoint tapping have been proposed.

As EFT uses elements of efficacious therapies, some critics claim that any beneficial effects might be solely due to the merits of these therapies and that acupoint tapping makes no contribution to its success. This is the so-called “purple hat” fallacy, which holds that a practitioner might employ a therapeutic method of demonstrated efficacy while wearing a purple hat and then attribute the success of treatment to the hat. For this reason, dismantling studies that separate acupoint stimulation from the cognitive and exposure portions of the protocol have been recommended, the first of which was performed by Fox. In a randomized controlled trial with university students, the investigator administered the exposure portion of EFT to a control group, though without the cognitive “Setup Statement” that is a component of EFT. In an active control group, mindful breathing was substituted for tapping. The study found that the tapping group showed significantly better results on most measures.

A second dismantling study examined 126 school teachers at risk for burnout. EFT was compared to a control condition that included tapping on a sham location on the body coupled with the exposure component of EFT. Participants were assessed using the Maslach Burnout Inventory, which employs three scales: Emotional Exhaustion, Depersonalization, and Personal Accomplishment. The study reduced the possibility of cross-contamination between the two conditions by not randomizing participants within a single population. Instead, to minimize contact between experimental and control participants, the two samples were taken from different school districts with similar demographic profiles in the same county.

To ensure that the cognitive and exposure portions of both protocols were similar and thus isolate tapping as a viable and measurable technique, Reynolds gave both the EFT and control groups identical lists of troubling situations and cognitions that might contribute to burnout, and they were instructed to focus on them mentally while utilizing the procedure. Instead of using the Setup Statement and prescribed acupressure points of EFT, members of the control group were instructed to tap with an open right hand on the forearm of their left hand. The position of the hand doing the tapping is important, as there are acupoints at the tips of each finger. Thus, this study controlled for the stimulation of those points by tapping with the undersides of the fingers instead of the fingertips, in contrast to an earlier study that unintentionally used fingertip points. The results of the Reynolds study showed that on all three indicators of burnout measured, EFT was superior to the sham tapping employed in the control group.

Waite and Holder compared three tapping conditions (EFT points, sham points, and points on a doll) to a non-tapping condition and found that participants in all three tapping groups showed significant improvements, while the non-tapping group did not. From this attempt at a dismantling study, the investigators concluded that EFT owed its efficacy to distraction and desensitization. The same findings have been reinterpreted as supporting EFT, however, as the investigators inadvertently engaged fingertip tapping points in the three groups that improved significantly. Another flaw in Waite and Holder’s study was that it failed to adhere to the manualized form of the method and instead introduced novel variants to the protocol. Its departure from established research norms and its ambiguous results make this study an outlier in EFT research literature.

Another dismantling study performed by Church and Nelms recruited 34 participants with clinically verified “frozen shoulder” consisting of limited range of motion (ROM) measured in five different planes of arm movement. Participants were randomized into a wait list or one of the two treatment groups. ROM, pain, and psychological conditions such as anxiety and depression were assessed before and after a 30-min treatment session, and 30 days later. One treatment group received clinical EFT, while the other received an identical cognitive and exposure protocol but with diaphragmatic breathing (DB) substituted for acupressure. While the DB group showed improvements in pain and psychological distress after treatment, their results were
not as pronounced as those of the tapping group. On follow-up, the participants in the tapping group maintained their gains while those in the DB group did not. As in the Fox and Reynolds studies, these results indicate that acupoint tapping is an active ingredient in the therapeutic results obtained from EFT, and that its efficacy is not solely due to protocols such as exposure and cognitive restructuring that are shared with other therapies.

The fifth dismantling study used a convenience sample of 56 university students randomized into an EFT group and a control group that employed an identical protocol but with sham tapping points. A stress test was administered before and after a single tapping session. The group that had tapped on actual acupressure points exhibited a significantly greater reduction in stress ($p < .0001$). The study had several methodological weaknesses, however, such as the lack of a reliable and valid assessment and the administration of the intervention by one of the investigators.

Several meta-analyses of EFT have been published. The earliest was by Gilomen and Lee and examined studies published through 2012. It identified a moderate treatment effect, but because its time frame did not include the dismantling studies summarized above, it speculated that treatment effects might be due to the non-specific benefits of any therapy. A later meta-analysis of EFT for the treatment of PTSD found very large treatment effects across seven RCTs. It also found equivalence to established therapies and superiority of EFT to treatment as usual. A meta-analysis by Cloninger that did include the dismantling studies found a large effect size in the remission of anxiety symptoms after EFT treatment. This and other evidence points to EFT’s clinically significant effects on a wide variety of psychological symptoms.

**STUDY AIM**

We sought to collect information pertaining to EFT methods most recently used in clinical trials, both randomized and observational, and to examine their relative efficacy in reducing depression in different samples. During the preliminary search of the literature, we organized a review of the studies published in the last 10 years. We chose this time frame as an appropriate span for forming an accurate picture of the state of the research evidence. Further, we postulated that a sufficient number of studies would have been published during that time to allow for meta-analytic calculations. As an easily-taught method that is used both as a self-help technique and in clinical practice, we proposed to identify whether or not there is an empirical basis for EFT’s widespread use for depression.

**METHODS**

**Study Selection**

The overall objective of study selection was to collect reports published in peer-reviewed journals that examined depression symptom levels before and after EFT treatment both in clinical and non-clinical populations. A literature search for English-language articles was performed using MEDLINE/PubMed, PsycINFO, Google Scholar, and references from the retrieved articles. The search is current through June of 2015. Keyword searches included “Emotional Freedom Technique” and “depression,” “Emotional Freedom Techniques” and “depression,” and “EFT” and “depression.” We also included “post-traumatic stress disorder” and “PTSD” as search terms, as we noted that studies of the effect of EFT on PTSD often include an examination of psychological symptoms (including depression) through the use of a standardized questionnaire. Professional organizations in the field, and individual authors, were queried in order to obtain information about unpublished “file drawer” studies as well as studies “in press.”

Though some researchers (e.g., Cochrane Collaborative) view the RCT as the only acceptable evidence for treatment outcome, many systematic reviews are indeterminate because they include insufficient numbers of RCTs while rejecting large numbers of non-randomized controlled studies. We decided to include all studies published and relevant to our aim(s), independent of their research design, in order to increase the sample size and provide clinically informative results.

Of the RCTs, only those that assessed depression as an outcome measure and met the APA Division 12 criteria were included. To meet the Division 12 efficacy criteria as a
“well-established treatment,” two independent research teams must have demonstrated in a peer-reviewed report that the treatment is either (1) more effective than a placebo or (2) that it is equivalent to another treatment whose efficacy has been established.46,49 In order to be designated as “efficacious” method, the treatment must also be standardized (which is usually taken to mean that the practitioner delivering the treatment adhered to manualized procedures), and the client population must be clearly specified so that the study's generalizability can be inferred. To achieve a “probably efficacious treatment” designation, only one study meeting the above criteria is required. The “probably efficacious” criteria can be met by two studies in which the control condition is a wait-list group instead of a placebo condition or an established treatment. The two studies may be by the same investigator or different investigators. The Division 12 criteria were therefore used as a convenient published standard allowing us to screen for the quality of RCTs.

Study Eligibility

The first author (J.N.) screened the abstracts of all publications obtained by the search strategy. Studies meeting the following inclusion criteria were selected for the meta-analysis: (1) at least one session of EFT, (2) reporting of interval or ratio data, (3) the use of valid and reliable psychometric questionnaires; (4) pre-treatment and post-treatment depression level data presented, and (5) sufficient reporting of study results (e.g., means and standard deviations) to allow for effect size computation. Some studies utilized both repeated measures design (before and after EFT), as well as comparisons (EFT versus active treatment or wait list). The absence of a control group was not an exclusion criterion because calculation of effect size can be performed on pre–post changes. However, given that between-group and within-group analyses are methodologically different,50 two separate analyses were conducted. In their review, Shrier et al.50 argue for the inclusion of non-RCTs in meta-analyses. They conclude that the benefits outweigh the disadvantages and point out that the effect sizes found in outcome studies are generally similar to those in RCTs. We did not require an independent observer-rated diagnosis of depression since we believe that self-reported depressive symptoms are more clinically significant to depression sufferers than observer-rated measures, and the discrepancies between the two are anyway usually minor.51

One study by Nemiro and Papworth52 assessed general mental health, with depression and anxiety measured in sub-scales. The investigators only provided the general mental health scale, and despite their depression results being requested several times for this meta-analysis, they did not provide these data and we were therefore unable to include it. We also excluded several studies which re-analyzed data from studies already included. Excluded studies were the studies of Church12; which was a detailed analysis of the interaction of pain, anxiety, and depression using data extracted from Church et al.13; Hartung and Stein,60 which compared office to telephone sessions, and Stein and Brooks (2011), which compared EFT delivered by life coaches versus licensed mental health professionals.

Data Coding

For all articles selected, the full articles were obtained and inspected to assess their relevance based on the preplanned criteria for inclusion. Data were extracted using a predesigned data collection form: (1) number of subjects, (2) geographic origin of the study, (3) treatment delivery format, (4) description of population, (5) mean age, (6) percentage of women, (7) assessment measures, (8) homework, (9) protocol length, (10) trial context, and (11) summary statistics required for computation of effect sizes. Data were then reviewed by the first and the second authors (J.N. and L.C.) to reach an agreement. Any disagreements were resolved by an outside reviewer.

Besides computing the total average effect size of EFT intervention for depression, we also calculated effect sizes across multiple time points. We did this by structuring an estimating equation with weights assigned to each study based on the number of participants used in the analysis. Time points analyzed included posttest, less than 90 days, and 90 days or greater. Ambiguity concerning study sample sizes (i.e., unspecified attrition) was solved by the conservative approach of using the smallest number for which there was clear documentation.

Depression was the dependent variable. From the studies selected for the meta-analysis, we extracted all available psychometric data, which were available from a variety of questionnaires: Symptom Assessment-4553 (SA-45), Beck Depression Inventory II54 (BDI II), Depression Anxiety Stress Scales55 (DASS), Hospital Anxiety and Depression Scale56 (HADS), and the Profile of Mood States57 (POMS) depression subscale.

Statistical Methods

To assess clinically relevant treatment effects, the percentage change in depressive symptoms before and after treatment was calculated using normed T-scores or whatever other scoring method was specified in the relevant assessments. Effect sizes were measured using Cohen’s d. When the necessary data were available, all effect sizes were calculated directly using the following formula: \( d = (M_1 - M_2)/S \), where \( M_1 \) is the pretest mean, \( M_2 \) the posttest or follow-up mean, and \( S \) the standard deviation for the pooled sample. We used Cohen's \( d \) rather than Hedge's \( g \) because it is more commonly used in meta-analyses and therefore allowed us to compare results with meta-analyses of other treatments such as psychopharmacology and psychotherapy. For RCTs, the effect size was first calculated only for the EFT treatment group. For within-group outcome studies without a control group, baseline scores instead of control group scores were used in the formula. If these data were not provided, \( d \) was estimated using conversion equations for significance tests. If an RCT had an active control, such as EMDR, a comparison was then made between the treatment effectiveness of the two therapies. The overall mean effect size for all of the studies combined was weighted by the variance of the studies, considering both standard deviations and number of subjects. Effect sizes are usually interpreted using Cohen's original convention of small (<0.2), medium (0.5), and large (>0.8) treatment effects.
We identified 20 studies for inclusion in the meta-analysis. Eight studies investigated EFT outcomes only; the other 12 were RCTs. The pooled sample for outcome studies comprised 461 subjects, while the pooled RCT sample consisted of 398 participants who received EFT treatment. Study characteristics are listed in Table 1. Tables 2 and 3 outline the reported results for outcome studies and RCTs, respectively. Several types of psychological surveys were employed to measure depressive symptoms in the studies included in this meta-analysis.

We first calculated effect sizes as Cohen’s $d$ comparing depression scores at time points after EFT intervention with pre-intervention scores. We then stratified time points into

<table>
<thead>
<tr>
<th>Number</th>
<th>Study</th>
<th>Mean Age</th>
<th>% Women</th>
<th>Country</th>
<th>Treatment</th>
<th>Population</th>
<th>Instrument</th>
<th>Delivery Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Baker and Hoffman (2015)</td>
<td>55</td>
<td>100</td>
<td>UK</td>
<td>EFT shortcut basic recipe</td>
<td>Women with breast cancer taking hormones</td>
<td>POMS</td>
<td>Group</td>
</tr>
<tr>
<td>2</td>
<td>Church and Brooks (2013)</td>
<td>54</td>
<td>72</td>
<td>US</td>
<td>Clinical EFT</td>
<td>Workshop participants with addiction issues</td>
<td>SA-45</td>
<td>Group</td>
</tr>
<tr>
<td>3</td>
<td>Church and Brooks (2010)</td>
<td>48</td>
<td>76</td>
<td>US</td>
<td>Clinical EFT</td>
<td>Healthcare personnel</td>
<td>SA-45</td>
<td>Group</td>
</tr>
<tr>
<td>4</td>
<td>Church et al. (2009)</td>
<td>NR</td>
<td>43</td>
<td>US</td>
<td>Clinical EFT</td>
<td>Veterans</td>
<td>SA-45</td>
<td>Group</td>
</tr>
<tr>
<td>5</td>
<td>Rowe (2005)</td>
<td>53</td>
<td>74</td>
<td>US</td>
<td>Clinical EFT</td>
<td>Convenience sample of workshop participants</td>
<td>SA-45</td>
<td>Group</td>
</tr>
<tr>
<td>6</td>
<td>Stapleton et al. (2014)</td>
<td>NR</td>
<td>82</td>
<td>Australia</td>
<td>Clinical EFT</td>
<td>Major Depressive Disorder clients at a university clinic</td>
<td>BDI II; SA-45</td>
<td>Group</td>
</tr>
<tr>
<td>7</td>
<td>Stewart et al.</td>
<td>45</td>
<td>77</td>
<td>UK</td>
<td>Clinical EFT</td>
<td>National Health Service hospital patients</td>
<td>HADS</td>
<td>Group</td>
</tr>
<tr>
<td>9</td>
<td>Brattberg</td>
<td>44</td>
<td>100</td>
<td>Sweden</td>
<td>Clinical EFT</td>
<td>Fibromyalgia patients on sick leave</td>
<td>HADS</td>
<td>Group</td>
</tr>
<tr>
<td>10</td>
<td>Church et al.</td>
<td>52</td>
<td>10</td>
<td>US</td>
<td>Clinical EFT</td>
<td>Veterans with PTSD</td>
<td>SA-45</td>
<td>Individual</td>
</tr>
<tr>
<td>11</td>
<td>Church et al. (2014)</td>
<td>56</td>
<td>44</td>
<td>US</td>
<td>Clinical EFT</td>
<td>Veterans at risk of PTSD</td>
<td>SA-45</td>
<td>Individual</td>
</tr>
<tr>
<td>12</td>
<td>Church et al. (2012)</td>
<td>49</td>
<td>82</td>
<td>US</td>
<td>Clinical EFT</td>
<td>Nonclinical subjects</td>
<td>SA-45</td>
<td>Individual</td>
</tr>
<tr>
<td>13</td>
<td>Stapleton et al. (2011)</td>
<td>NR</td>
<td>89</td>
<td>Australia</td>
<td>Clinical EFT</td>
<td>Patients with major depressive disorder</td>
<td>SA-45</td>
<td>Group</td>
</tr>
<tr>
<td>14</td>
<td>Karatzias et al. (2011)</td>
<td>40</td>
<td>52</td>
<td>UK</td>
<td>EFT Movie Technique</td>
<td>National Health Service psychotherapy referrals</td>
<td>HADS</td>
<td>Individual</td>
</tr>
<tr>
<td>15</td>
<td>Lee et al. (2015)</td>
<td>80</td>
<td>100</td>
<td>South Korea</td>
<td>Clinical EFT</td>
<td>Senior insomnia patients</td>
<td>GDS-K</td>
<td>Group</td>
</tr>
<tr>
<td>16</td>
<td>Church et al. (2012)</td>
<td>17</td>
<td>80</td>
<td>Philippines</td>
<td>Clinical EFT</td>
<td>Psychology students</td>
<td>BDI</td>
<td>Group</td>
</tr>
<tr>
<td>17</td>
<td>Geronilla et al. (2014)</td>
<td>50</td>
<td>19</td>
<td>US</td>
<td>Clinical EFT</td>
<td>Veterans with PTSD</td>
<td>SA-45</td>
<td>Individual</td>
</tr>
<tr>
<td>18</td>
<td>Church and Nelms (2015)</td>
<td>54</td>
<td>63</td>
<td>US</td>
<td>Clinical EFT</td>
<td>Adults with frozen shoulder</td>
<td>SA-45</td>
<td>Individual</td>
</tr>
<tr>
<td>19</td>
<td>Church et al.</td>
<td>60</td>
<td>31</td>
<td>US</td>
<td>Clinical EFT</td>
<td>Veterans</td>
<td>SA-45</td>
<td>Individual</td>
</tr>
<tr>
<td>20</td>
<td>Stapleton et al. (2013)</td>
<td>NR</td>
<td>89</td>
<td>Australia</td>
<td>Clinical EFT</td>
<td>Overweight and obese adults</td>
<td>SA-45</td>
<td>Group</td>
</tr>
</tbody>
</table>

BDI II = Beck Depression Inventory II; DASS = DASS Depression Anxiety Stress Scales; DB = diaphragmatic breathing; EFT = emotional freedom techniques; EMDR = eye movement desensitization and reprocessing; GDS-K = Geriatric Depression Scale in Korea; HADS = Hospital Anxiety and Depression Scale; NT = no treatment; POMS = profile of mood states; SA-45 = symptom assessment–45; SHE = sleep hygiene education; SI = supportive interview; TAU = treatment as usual; WL = wait list; NR = not reported.
posttest, <90 days, and 90 days or more (Tables 4–6). We calculated weighted mean effect size for each stratum. For the multivariate model, we took the one comparison per study, the 90-day assessment in an EFT group if the study extended 90 days or longer. The outcome variable was depression scale scores. For the studies included in the meta-analysis multivariate model, we calculated as the outcome the standardized mean difference in 90-day follow-up versus pre-intervention scores.

Tables 2 and 3 illustrate the effect sizes pretest and posttest for all studies. Tables 4–6 illustrate time point–specific measurements. The letter $d$ represents Cohen’s $d$, while SE ($d'$) represents the standard errors in each respective Cohen’s $d$ measurement. Finally, $\rho(d)$ represents the probability of observing this or more larger effect size by chance. As can be seen from Tables 2 and 3, the majority of studies demonstrated a statistically significant change in depression compared to pretest. The weighted effect size for outcome studies is 0.7046. The pretest–posttest effect size for RCTs is 1.8498. Overall, the pre–post effect size is 1.3087.

The same pattern holds for measurements taken at less than 90 days (Table 5) where five of the eight studies reported a statistically significant change. Table 6 reports measurements at 90 days and greater. All 90-day or less versus pretest effect sizes were reported as being at least 0.29, with effects ranging from 0.29 to 3.78. For less than 90 days follow-up versus pretest comparisons, the outcome studies have an effect size of 0.8102. The RCT’s effect size is 1.3832. The weighted effect size for these eight studies is 1.2119. Similarly, the effect size for 90 days follow-up or greater versus pretest is 0.5253 and 1.73 for outcome and RCT studies, respectively.

We also assessed heterogeneity using a Q statistic to determine which type of model should be used for computing estimates of effect sizes 90 days or more after treatment. We found a Q of 22.39 ($P = .0004$), which indicates significant heterogeneity among samples within this time period. We also calculated the between-study variance ($\tau^2$) as 0.4505. Based on this calculation, we used a random effects model to estimate the effect size due to EFT as $d = 1.11$ ($P = .0007$). This effect size is very large and indicates a consistent depression score difference from baseline to

### Table 2. Pretest and Posttest Percent Change and $P$ Values for Outcome Studies

<table>
<thead>
<tr>
<th>Number</th>
<th>Study</th>
<th>Analytical Sample Size</th>
<th>Country</th>
<th>Depression Measurement (Instrument)</th>
<th>Pre Versus Post % Change (%)</th>
<th>$P$ Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Baker and Hoffman (2015)</td>
<td>41</td>
<td>UK</td>
<td>POMS</td>
<td>−16</td>
<td>.006</td>
</tr>
<tr>
<td>2</td>
<td>Church and Brooks (2013)</td>
<td>39</td>
<td>US</td>
<td>SA-45</td>
<td>−26</td>
<td>.024</td>
</tr>
<tr>
<td>3</td>
<td>Church and Brooks (2010)</td>
<td>216</td>
<td>US</td>
<td>SA-45</td>
<td>−60</td>
<td>.001</td>
</tr>
<tr>
<td>4</td>
<td>Church et al. (2009)</td>
<td>7</td>
<td>US</td>
<td>SA-45</td>
<td>−34</td>
<td>.001</td>
</tr>
<tr>
<td>5</td>
<td>Rowe (2005)</td>
<td>102</td>
<td>US</td>
<td>SA-45</td>
<td>−60</td>
<td>.001</td>
</tr>
<tr>
<td>6</td>
<td>Stapleton et al. (2014)</td>
<td>6</td>
<td>Australia</td>
<td>BDI II</td>
<td>−23</td>
<td>.1</td>
</tr>
<tr>
<td>7</td>
<td>Stewart et al.</td>
<td>39</td>
<td>UK</td>
<td>HADS</td>
<td>−48</td>
<td>.001</td>
</tr>
<tr>
<td>8</td>
<td>Church (2010)</td>
<td>11</td>
<td>US</td>
<td>SA-45</td>
<td>−87</td>
<td>.005</td>
</tr>
</tbody>
</table>

BDI II = Beck Depression Inventory II; DASS = DASS Depression Anxiety Stress Scales; DB = diaphragmatic breathing; EFT = emotional freedom techniques; EMDR = eye movement desensitization and reprocessing; GDS-K = Geriatric Depression Scale in Korea; HADS = Hospital Anxiety and Depression Scale; NT = no treatment; POMS = profile of mood states; SA-45 = symptom assessment–45; SHE = sleep hygiene education; SI = supportive interview; TAU = treatment as usual; WL = wait list; NR = not reported.

### Table 3. Pretest and Posttest Percent Change in Symptoms and $P$ Values After EFT in RCTs

<table>
<thead>
<tr>
<th>Number</th>
<th>Study</th>
<th>Analytical Sample Size</th>
<th>Country</th>
<th>Depression Measurement (Instrument)</th>
<th>Pre versus Post % Change (%)</th>
<th>$P$ Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Brattberg</td>
<td>330</td>
<td>Sweden</td>
<td>HADS</td>
<td>−29</td>
<td>.02</td>
</tr>
<tr>
<td>2</td>
<td>Church et al.</td>
<td>49</td>
<td>US</td>
<td>SA-45</td>
<td>−58</td>
<td>.001</td>
</tr>
<tr>
<td>3</td>
<td>Church et al.</td>
<td>18</td>
<td>US</td>
<td>SA-45</td>
<td>−47</td>
<td>.001</td>
</tr>
<tr>
<td>4</td>
<td>Church et al.</td>
<td>28</td>
<td>US</td>
<td>SA-45</td>
<td>−49</td>
<td>.001</td>
</tr>
<tr>
<td>5</td>
<td>Stapleton et al.</td>
<td>96</td>
<td>Australia</td>
<td>SA-45</td>
<td>−23</td>
<td>.01</td>
</tr>
<tr>
<td>6</td>
<td>Karatzias et al.</td>
<td>23</td>
<td>UK</td>
<td>HADS</td>
<td>−28</td>
<td>.001</td>
</tr>
<tr>
<td>7</td>
<td>Lee et al.</td>
<td>10</td>
<td>South Korea</td>
<td>GDS-K</td>
<td>−60</td>
<td>.005</td>
</tr>
<tr>
<td>8</td>
<td>Church et al.</td>
<td>9</td>
<td>Philippines</td>
<td>BDI</td>
<td>−74</td>
<td>&lt;.05</td>
</tr>
<tr>
<td>9</td>
<td>Geronilla et al.</td>
<td>58</td>
<td>US</td>
<td>SA-45</td>
<td>−48</td>
<td>.001</td>
</tr>
<tr>
<td>10</td>
<td>Church and Nelms</td>
<td>16</td>
<td>US</td>
<td>SA-45</td>
<td>−44</td>
<td>.001</td>
</tr>
<tr>
<td>11</td>
<td>Church et al.</td>
<td>16</td>
<td>US</td>
<td>SA-45</td>
<td>−38</td>
<td>.001</td>
</tr>
<tr>
<td>12</td>
<td>Stapleton et al.</td>
<td>45</td>
<td>Australia</td>
<td>SA-45</td>
<td>−21</td>
<td>.001</td>
</tr>
</tbody>
</table>
after EFT treatment, and that the effects of EFT treatment are lasting, up to and through 90 days after the intervention. The weighted mean of reductions in depressive symptoms was measured and found to be 41%.

Six RCTs used an active treatment (not wait list) as the comparison group. Two provided treatment as usual (TAU), one provided supportive interviews (SI), one provided dia-phragmatic breathing (DB), one provided eye movement desensitization and reprocessing (EDMR) therapy, and one provided sleep hygiene education (SHE). Table 7 illustrates the statistical differences between EFT and DB, SI, EDMR, and SHE. We did not include treatment as usual (TAU) as an active control group. EFT were more efficacious than DB and SI in the posttest measurements (P = .06 versus DB; P < .001 versus SI). EFT was more efficacious than SHE at the ninth week assessment (P = .036). We found no statistically significant difference between EFT and EDMR at posttest. Most studies reported that no adverse reactions were noted, and that participant distress decreased rapidly across all symptom domains simultaneously.

Table 4. Posttest Effect Sizes for Both Outcome Studies and RCTs

<table>
<thead>
<tr>
<th>Study Number</th>
<th>Study Type</th>
<th>d</th>
<th>SE(d)</th>
<th>p(d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Baker and Hoffman (2015)</td>
<td>Outcome</td>
<td>0.19</td>
<td>0.29</td>
</tr>
<tr>
<td>2</td>
<td>Church and Brooks (2013)</td>
<td>Outcome</td>
<td>0.26</td>
<td>0.23</td>
</tr>
<tr>
<td>3</td>
<td>Church and Brooks (2010)</td>
<td>Outcome</td>
<td>0.75</td>
<td>0.1</td>
</tr>
<tr>
<td>4</td>
<td>Church et al. (2009)</td>
<td>Outcome</td>
<td>1.1</td>
<td>0.57</td>
</tr>
<tr>
<td>5</td>
<td>Rowe (2005)</td>
<td>Outcome</td>
<td>\</td>
<td>\</td>
</tr>
<tr>
<td>6</td>
<td>Stapleton et al. (2014)</td>
<td>Outcome</td>
<td>0.04</td>
<td>0.58</td>
</tr>
<tr>
<td>7</td>
<td>Stewart et al. (2019)</td>
<td>Outcome</td>
<td>0.84</td>
<td>0.31</td>
</tr>
<tr>
<td>8</td>
<td>Church (2010)</td>
<td>Outcome</td>
<td>2.64</td>
<td>0.57</td>
</tr>
<tr>
<td>9</td>
<td>Brattberg (2019)</td>
<td>RCT</td>
<td>0.62</td>
<td>0.26</td>
</tr>
<tr>
<td>10</td>
<td>Church et al. (2019)</td>
<td>RCT</td>
<td>8.02</td>
<td>0.78</td>
</tr>
<tr>
<td>11</td>
<td>Church et al. (2014)</td>
<td>RCT</td>
<td>3.11</td>
<td>0.68</td>
</tr>
<tr>
<td>12</td>
<td>Church et al. (2012)</td>
<td>RCT</td>
<td>1.12</td>
<td>0.29</td>
</tr>
<tr>
<td>13</td>
<td>Stapleton et al. (2011)</td>
<td>RCT</td>
<td>0.28</td>
<td>0.22</td>
</tr>
<tr>
<td>14</td>
<td>Karatzias et al. (2011)</td>
<td>RCT</td>
<td>0.69</td>
<td>0.39</td>
</tr>
<tr>
<td>15</td>
<td>Lee et al. (2015)</td>
<td>RCT</td>
<td>1.41</td>
<td>0.41</td>
</tr>
<tr>
<td>16</td>
<td>Church et al. (2012)</td>
<td>RCT</td>
<td>7.57</td>
<td>1.29</td>
</tr>
<tr>
<td>17</td>
<td>Geronilla et al. (2014)</td>
<td>RCT</td>
<td>1.93</td>
<td>0.3</td>
</tr>
<tr>
<td>18</td>
<td>Church and Nelms (2017)</td>
<td>RCT</td>
<td>0.88</td>
<td>0.37</td>
</tr>
<tr>
<td>19</td>
<td>Church et al. (2015)</td>
<td>RCT</td>
<td>0.9</td>
<td>0.36</td>
</tr>
<tr>
<td>20</td>
<td>Stapleton et al. (2013)</td>
<td>RCT</td>
<td>0.37</td>
<td>0.23</td>
</tr>
</tbody>
</table>

Weighted effect size—outcome studies = 0.7046.
Weighted effect size—RCTs = 1.8498.
Weighted effect size—all studies = 1.3087.

Table 5. Effect Sizes Measured at Less Than 90 Days

<table>
<thead>
<tr>
<th>Study Number</th>
<th>Study Type</th>
<th>d</th>
<th>SE(d)</th>
<th>p(d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Baker and Hoffman (2015)</td>
<td>Outcome</td>
<td>0.29</td>
<td>0.3</td>
</tr>
<tr>
<td>6</td>
<td>Stapleton et al. (2014)</td>
<td>Outcome</td>
<td>0.79</td>
<td>0.6</td>
</tr>
<tr>
<td>7</td>
<td>Stewart et al. (2019)</td>
<td>Outcome</td>
<td>2.3</td>
<td>0.87</td>
</tr>
<tr>
<td>8</td>
<td>Church (2010)</td>
<td>Outcome</td>
<td>1.1</td>
<td>0.46</td>
</tr>
<tr>
<td>11</td>
<td>Church et al. (2014)</td>
<td>RCT</td>
<td>3.78</td>
<td>0.76</td>
</tr>
<tr>
<td>15</td>
<td>Lee et al. (2015)</td>
<td>RCT</td>
<td>0.94</td>
<td>0.38</td>
</tr>
<tr>
<td>18</td>
<td>Church and Nelms (2017)</td>
<td>RCT</td>
<td>1.1</td>
<td>0.38</td>
</tr>
<tr>
<td>19</td>
<td>Church et al. (2015)</td>
<td>RCT</td>
<td>0.53</td>
<td>0.35</td>
</tr>
</tbody>
</table>

Weighted effect size—outcome studies = 0.8102.
Weighted effect size—RCTs = 1.3832.
Weighted effect size—all studies = 1.2119.
The effect size found for EFT in this meta-analysis was 0.22. The effect size found for EFT in this meta-analysis exceeded the treatment effect typically found in psychopharmacology and psychotherapy trials. Further, we discovered that the impact of EFT, especially in interventions performed over multiple sessions, endures up to and surpasses 90 days, even up to 6 months post-intervention.

The findings of two studies that were excluded since they were reanalyses of work on depression in general and among population subgroups. We found that EFT was efficacious in mitigating depression in a variety of demographic groups. A meta-analysis of psychopharmacology for major depression found Cohen’s $d = 0.20$, while high-quality trials of psychotherapy found an effect size of $d = 0.22$. The effect size found for EFT in this meta-analysis exceeded the treatment effect typically found in psychopharmacology and psychotherapy trials. Further, we discovered that the impact of EFT, especially in interventions performed over multiple sessions, endures up to and surpasses 90 days, even up to 6 months post-intervention.

Table 6. Effect Sizes for 90 Days and Greater

<table>
<thead>
<tr>
<th>Study Number</th>
<th>Study Type</th>
<th>Type</th>
<th>d</th>
<th>SE(d)</th>
<th>p(d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Church and Brooks (2013)</td>
<td>Outcome</td>
<td>0.32</td>
<td>0.25</td>
<td>0.2</td>
</tr>
<tr>
<td>4</td>
<td>Church et al. (2009)</td>
<td>Outcome</td>
<td>1.24</td>
<td>0.58</td>
<td>0.04</td>
</tr>
<tr>
<td>8</td>
<td>Church (2010)</td>
<td>Outcome</td>
<td>0.75</td>
<td>0.44</td>
<td>0.09</td>
</tr>
<tr>
<td>11</td>
<td>Church et al. (2014)</td>
<td>RCT</td>
<td>4.32</td>
<td>0.83</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>19</td>
<td>Church et al. (2015)</td>
<td>RCT</td>
<td>0.87</td>
<td>0.36</td>
<td>0.02</td>
</tr>
<tr>
<td>20</td>
<td>Stapleton et al. (2013)</td>
<td>RCT</td>
<td>0.87</td>
<td>0.23</td>
<td>&lt;0.01</td>
</tr>
</tbody>
</table>

Weighted effect size—outcome studies = 0.5253.
Weighted effect size—RCTs = 1.73.
Weighted effect size—all studies = 1.1093.

DISCUSSION

This study evaluated the impact that EFT has on depression in general and among population subgroups. We found that EFT was efficacious in mitigating depression in a variety of demographic groups. A meta-analysis of psychopharmacology for major depression found Cohen’s $d = 0.20$, while high-quality trials of psychotherapy found an effect size of $d = 0.22$. The effect size found for EFT in this meta-analysis exceeded the treatment effect typically found in psychopharmacology and psychotherapy trials. Further, we discovered that the impact of EFT, especially in interventions performed over multiple sessions, endures up to and surpasses 90 days, even up to 6 months post-intervention.

The findings of two studies that were excluded since they were reanalyses of work on depression in general and among population subgroups. We found that EFT was efficacious in mitigating depression in a variety of demographic groups. A meta-analysis of psychopharmacology for major depression found Cohen’s $d = 0.20$, while high-quality trials of psychotherapy found an effect size of $d = 0.22$. The effect size found for EFT in this meta-analysis exceeded the treatment effect typically found in psychopharmacology and psychotherapy trials. Further, we discovered that the impact of EFT, especially in interventions performed over multiple sessions, endures up to and surpasses 90 days, even up to 6 months post-intervention.

Table 7. Significance Tests for Difference Between EFT and Other Active Treatments

<table>
<thead>
<tr>
<th>Control Treatment</th>
<th>Follow-Up Study</th>
<th>Study</th>
<th>P for Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>DB</td>
<td>Pre</td>
<td>Church and Nelms</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td>Post</td>
<td>&amp; 37</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>30-Day</td>
<td></td>
<td>0.2108</td>
</tr>
<tr>
<td>SI</td>
<td>Post</td>
<td>Church et al. (2012)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>EMDR</td>
<td>Post</td>
<td>Karatzias et al. (2011)</td>
<td>0.724</td>
</tr>
<tr>
<td>SHE</td>
<td>Ninth week</td>
<td>Lee et al. (2015)</td>
<td>0.036</td>
</tr>
</tbody>
</table>

DB = diaphragmatic breathing; EMDR = eye movement desensitization and reprocessing; SHE = sleep hygene education; SI = supportive interview.

Thorough the number of studies in which EFT were compared to established treatments in this meta-analysis was small, in all cases EFT were found to have equal or greater efficacy. EFT were also more efficacious than treatment as usual or the passage of time. One of the studies included in this analysis was a dismantling study. It was carefully designed to determine whether or not the acupoint tapping described in EFT was an active ingredient in the treatment protocol. It compared a tapping group to a group that performed an identical protocol but with diaphragmatic breathing (DB) substituted for tapping. While positive effects for the DB group were found, they were not as great as for the tapping group, and they did not persist over time. An earlier meta-analysis found that EFT produced a moderately large treatment effect, but based on the limited number of studies then available to the authors, it was impossible to determine whether treatment effects were due to tapping, or to the non-specific factors found in any therapy (Gilomen and Lee, 2014). The current study allowed us to test this hypothesis, and on the basis of the included study, as well as several other dismantling studies published subsequent to Gilomen and Lee, it appears that tapping is an active ingredient in EFT and not an inert placebo.

As did the extensive review by Shrier et al., we found that there is not a marked difference in effect sizes between outcome studies and RCTs provided that sufficient numbers of both are available for analysis. We therefore suggest that including outcome studies in meta-analyses enriches their utility for clinical decision-making. EFT’s clinical utility and generalizability is supported by several characteristics: The brief treatment time frames found in the studies examined, ranging from 1 to 10 sessions; its efficacy in group format as well as individual counseling sessions; the absence of abractions in participants; the unremarkable dropout rate; and the effects across demographic groups, countries, and treatment settings.

The meta-analysis was strengthened by several factors. One of them was the use of the same assessment, the SA-45, in 13 of the 20 studies included, enabling conclusions to be drawn from a variety of demographic samples. Another was the stated use of The EFT Manual in most of the reports,
indicating consistent application of the therapeutic method. Reported studies were published in a variety of journals, conducted by a variety of investigators, assessed demographically diverse samples including participants in several different countries, and used assessments other than the SA-45 such as the HADS and the BDI. These factors all argue for the generalizability of the results, subject to the limitations outlined below.

**LIMITATIONS**

There are several limitations to this study and virtually any other meta-analysis. Findings from this meta-analysis must be interpreted with caution given limitations of meta-analysis in general and of data collected for this analysis in particular. A critical issue for this meta-analysis, as is true of any systematic review, was deciding which trials or studies to include and which to exclude. As noted, we included all studies published and relevant to our aim(s), independent of their research design, in order to increase the number of studies and participants. All outcome studies including the ones analyzed in this article have the limitation that the improvements they note may be the nonspecific results of any treatment.

As in any review of studies in a given area, it is possible that studies with non-significant results are underreported. Publication bias, the possibility that only studies with significant outcomes have been published and those with null outcomes have not, skews the results of meta-analyses. While we did contact the professional organizations in the field in an attempt to locate unpublished studies, it is possible that some exist unknown to us or to those organizations. It is further important to note that, for some studies, analysis was based on relatively few subjects.

We searched studies in the most important databases for psychology (PsycINFO) and medicine (Medline). Other databases (e.g., CINAHL) were not screened, however, and this may be a limitation to the generalizability of our results, given that other studies might be available in these databases. A high degree of heterogeneity was found; factors that might have contributed to this were the large number of studies, the diversity of populations from clinic clients with major depressive disorder to veterans with PTSD to non-clinical workshop participants, and the varied treatment time frames ranging from 1 to 10 sessions.

Another major limitation generally present in studies, including those of EFT, is a substantial loss to follow-up. Many studies in our analysis experienced up to 50% loss to follow-up after the intervention period. Loss to follow-up occurs for a variety of reasons including: lack of compensation for participation, length of time required to complete follow-up assessments, and the difficulty that participants experience in keeping appointments among life’s other demands. Regardless of the reason, this factor may introduce bias into the measurement of effect or tests of significance. It is impossible to estimate the effect of loss to follow-up on the studies analyzed; therefore, we cannot be certain as to the difference complete follow-up would have made in the treatment or control groups. However, the dropout rate between the start and the end of the EFT intervention period is consistent with the rate of about 20% found in psychotherapy trials.

Our meta-analysis also included relatively few studies comparing EFT to another established treatment. Most studies used a wait list or treatment as usual group as a control. Further studies are needed to gauge the efficacy of EFT relative to established methods. Also, many of the studies used depression as one of several outcomes, including anxiety and PTSD. Most used participant self-report, rather than observer-rated measures. Only two used a categorical diagnosis of major depressive disorder. These limitations make it difficult to generalize the findings of this study to populations that typically seek treatment for severe depression. Instead, only the more conservative conclusion that EFT reduces depressive symptoms in a wide sample of demographic groups can be drawn from this analysis.

Though EFT are generally standardized and manualized throughout the world, the length and number of EFT sessions offered to study participants varies. We calculated the effect of an EFT intervention across studies; however, we were unable to determine if the effect size differed across intervention lengths as study lengths varied greatly across the group of studies. Two approaches could be taken to enhance the ability of researchers to measure the precise effect of EFT over various time points. Firstly, EFT intervention could be further normalized to establish a standard length and time for all interventions. A further normalization of EFT time points would allow for the assessment of EFT effectiveness across virtually any health outcome.

Secondly, daily or weekly monitoring could be performed to assess the change in a health outcome at normalized intervals (e.g., posttest, 3 days, 1 week, 2 weeks, 30 days, 60 days, 90 days, etc.). Such an approach would allow subjects involved in an EFT study to have their progress monitored in any health outcome more frequently, thus providing more data points for study. With more data points available for each participant, even those who drop out could at least present one data point, though they failed to complete the entire intervention.

Furthermore, if many data points are taken across the study period, interpolation, imputing, or other modeling techniques could be employed to estimate the effect at any time point and for any participant that may be missing. Based on existing data, this type of modeling and smoothing is not feasible.

**CONCLUSIONS**

The current study used meta-analytical methods to examine the efficacy of EFT for depression. It identified a large treatment effect across a wide demographic spectrum, with participants maintaining their gains over time. EFT was effective in brief treatment time frames ranging from 1 to 10 sessions, and when delivered in group treatment format as well as individual therapy sessions. It found greater efficacy for EFT than typically found in trials of psychopharmacology and psychotherapy. EFT were more efficacious than usual care and most active controls, and were as effective as EMDR, another evidence-based treatment, in one study. As a safe and
EFT for the Treatment of Depression

EXPLOREx November/December 2016, Vol. 12, No. 6 425

reliable self-help method, EFT demonstrates clinical utility as a low-cost non-drug treatment for depression in a wide variety of settings and demographic groups. Based on the RCTs in our meta-analysis, EFT unequivocally meet the APA Division 12 criteria as an efficacious therapy for depression.

Acknowledgments

The research was supported by donations to the National Institute for Integrative Healthcare (niih.org), which especially thanks Nick Ortner, and Robert and Lynne Hoss, for their contributions.

REFERENCES

40. Anfield B, Papworth, S. Emotional freedom techniques (EFT) for posttraumatic stress disorder: a systematic review and metanalysis. Presented at Tune Your Cells, Los Angeles, California, June 1, 2015. (Submitted for publication).
44. Hartmann N, Boehner M, Groenen F, Kalb R. Telomere length of patients with major depression is shortened but independent from therapy and severity of the disease. Depress Anxiety. 2010;27 (12):1111–1116.