SYSTEMATIC REVIEW

Massage therapy for the treatment of depression: a systematic review

H. F. Coelho, K. Boddy, E. Ernst

SUMMARY

Background: People with depressive disorders or subsyndromal symptoms of depression (SSD) often use complementary and alternative therapies, including massage therapy (MT). This systematic review evaluates the evidence, from randomised clinical trials (RCTs), for the effectiveness of multiple sessions of classical European (Swedish) MT for the treatment of depression. Methods: Eligible RCTs were identified via eight electronic databases and manual searches of references. Two reviewers independently selected trials, assessed trial quality and extracted data. Results: Four RCTs met our inclusion criteria. Three of these RCTs compared MT with relaxation therapies, but provided insufficient data and analyses to contribute meaningfully to the evaluation of MT for depression. The fourth included RCT used MT as a control condition to evaluate a depression-specific acupuncture treatment. This trial provided limited evidence that, in the early stages of treatment, MT is less effective than acupuncture for treating depression, a treatment which itself is not accepted for this condition. Conclusions: Despite previous research suggesting that MT may be an effective treatment for depression, there is currently a lack of evidence to support this assertion from RCTs that have selected participants for depression or SSD.

Introduction

A large number of people with clinically significant depression do not receive formal treatment. In Europe, only 36.5% (95% CI: 32.5–40.5) of people with a 12-month mood disorder contact formal healthcare services, and 15.1% (95% CI: 10.6–19.7) of these do not receive any treatment (1). Subsyndromal depressive symptoms (SSD) are also common, are associated with reduced functioning and quality of life, and can occur as prodromal and residual (interepisodic) features of major depression (2,3). Many people use a complementary or alternative therapy (CAT) to treat their symptoms of depression. Surveys indicate that 53.6% of those with ‘severe depression’ (4) and 22.4% of those with major depression (5) have used CAT to treat their symptoms in the preceding 12 months. Massage therapy (MT) is a popular CAT used by those with depression. Amongst those with ‘severe depression’, 2.1% use MT to treat their symptoms (4). The use of MT declined in the early 20th century (at the time of the pharmaceutical revolution), but its popularity has been regained, with consumers spending increasing amounts on this treatment (6).

The bio-physiological rationale for using MT for depressed mood is twofold. First, electroencephalogram activation changes to a symmetrical or left frontal pattern with MT. Such a pattern is associated with happy affect. Second, MT stimulates facial expressions and increases vagal activity, subsequently reducing depressed affect (7).

There are two systematic reviews of all complementary and alternative medicine (CAM) for depression which have included MT. However, one of these reviews did not critically evaluate the included studies (8) and the other only identified one trial of MT that had selected participants for depressive symptoms (9). A meta-analysis evaluating the effectiveness of MT for all indications reported largest effects for the reduction of depression and anxiety. However, data from general medical populations who were not specifically selected for having depression were included in the meta-analysis (10). Because of the limitations of these previous reviews (8–10) it is still not clear whether MT is effective for treating those with depressive disorders or SSD. Thus, there is a...
need for a systematic and critical review of the evidence for MT as a treatment for depression that focuses on randomised clinical trials (RCTs) using depressed populations. The objective of the current study was to address this need. Our review concentrates on multiple sessions of classical European (Swedish) massage, as a sole or adjunctive treatment, for depressive disorders or SSD.

Materials and methods

Search
The first author searched PUBMED, EMBASE, SCOPUS, ISI Web of Science, CENTRAL, PsychINFO, CINAHL and AMED from inception until June 2006, using a predefined search strategy (Table 1). No language restrictions were imposed. The reference lists of all articles retrieved in full were also searched.

Study selection
Abstracts of all articles identified from electronic searches were read. Articles clearly not meeting inclusion criteria were filtered out. All other articles were obtained in full (Figure 1). HFC and KB independently considered articles written in English. EE and an independent reviewer considered articles written in other languages. Inclusion of a trial was subject to both reviewers reaching agreement that the inclusion criteria were met.

Inclusion criteria were: (i) the study was a clinical trial, described as randomised; (ii) participants were selected for having a depressive disorder or SSD. Trials were excluded if participants with bipolar disorder or cyclothymic disorder were included; (iii) the intervention resembled classical European (Swedish) MT, delivered as a series of sessions, and therefore excluding single-session trials or analyses. Trials investigating self-massage techniques and MT involving the use of mechanical devices were excluded. The primary reason for excluding mechanical devices was the fact that skin-to-skin contact, which plays a major role in classical MT, would be either reduced or absent. RCTs of MT administered as part of a multi-component therapy (including trials using aromatherapy oils) were also excluded, unless the trial design enabled the evaluation of the MT component as an adjunctive treatment [see (iv) below]. Trials providing background music, which is often utilised in MT and considered to be part of the therapeutic session, were included; (iv) the study design enabled the evaluation of the sole or additive benefit of MT by using at least one appropriate control condition. Acceptable controls for sole MT included comparison treatments not involving massage, attention controls, and no treatment or waiting list groups. When MT was part of a multi-component therapy, control conditions must have been identical therapy minus the MT component; (v) there was at least one measure of the depressive disorder or SSD, or a clinical aspect or construct of such a disorder, and the measurement was made before treatment and at least once after treatment. Physiological outcomes were excluded.

Data extraction
Two reviewers (HFC and KB) independently extracted data from all included studies using data extraction forms specifically designed for the study. The data extracted corresponds with those given in Tables 2 and 3. Disagreements between reviewers were resolved through discussion.

Quality assessment
The Jadad et al. scale (11), modified to allocate one point for single blinding of the outcome assessor rather than for double blinding (maximum score = 4), was used to evaluate the methodological quality of each included RCT. To supplement the Jadad scale, Table 2 provides additional methodological details for each trial, such as those described in Verhagen et al. (12).

Data presentation and interpretation
Table 3 provides general information and the main findings of all included trials. Only results of between-group comparisons, using appropriate statistical techniques to evaluate the effectiveness of multiple-session MT, are included. If a study did not provide between-group comparisons, but presented sufficient data, we analysed the data using
software from the Cochrane Collaboration (Rev-Man 4.2).

Results

Thirty-three articles were retrieved in full for a detailed evaluation (Figure 1). Of these, 30 were excluded. Reasons for exclusion were: the study was a trial but was not randomised (13,14), participants were not all selected for depressive symptoms (13–37), participants with bipolar disorder or cyclothymia were included within the depressed sample (38), the intervention was not MT (36), the intervention was an excluded type of MT (14,28), the intervention was a single session of MT and physiological outcome only: n = 1 (1), and not randomised: n = 1 (1), and outcome not depression relevant: n = 1 (6), and not MT: n = 1 (9), and excluded type of MT: n = 1 (10), and excluded type of MT: n = 1 (6).

The remaining three RCTs were included in this review. One further RCT was identified from manual searches of reference lists. All four of the included RCTs were reported in English. Jadad scores ranged from 1 to 3 (Table 2). The number of MT sessions ranged from 5 to 32 sessions, with the length of the sessions ranging from 20 to 30 min. None of the included trials reported using background music as part of the MT.

MT vs. relaxation therapy

Three of the included trials compared MT with relaxation therapy (43–45). Participants were depressed adolescents (43), postnatally depressed women (44) and women with depression during pregnancy (45). The latter trial included a second control group who received usual prenatal care (Table 3).

The three studies (43–45) used multiple outcome measures at various time points, resulting in numerous analyses: for each outcome, both within- and between-group analyses were conducted to assess change over the first and the last session (single-session effects), and to assess change over the series of sessions (multiple-session effects). However, for the outcomes relevant to this review, all three studies omitted to report the between-group, pair-wise analyses that would have enabled evaluation of multiple-session MT. Also, all three of the studies comparing MT with relaxation therapies did not provide standard deviations alongside the mean values given in tables (43–45), precluding reanalysis of the data.
MT vs. acupuncture

In the fourth RCT, MT was used as a control condition when investigating the effectiveness of a depression-specific acupuncture treatment (SPEC) for depressed pregnant women (46). A non-specific acupuncture control condition (NSPEC) was also used. Analyses conducted for an 8-week acute treatment phase found that SPEC resulted in a greater proportion of participants responding to treatment than MT (see Table 3 for criteria used to define treatment response; Fisher exact $p = 0.03$, number to treat $(NTT)^2$ effect size $= 2.7$, 95% CI: 1.5–16.0). Because of the fact that MT and NSPEC were intended as control treatments, similar pair-wise analysis were not conducted between these two groups. Using the data provided in the trial report (46), our analysis demonstrated no significant difference between MT and NSPEC in proportion of participants responding to treatment following the acute treatment phase (OR $= 0.51$, 95% CI: 0.14–1.92). The authors of the trial (46) analysed Beck Depression Inventory (BDI) data, obtained weekly during the acute treatment phase, by fitting individual regression lines to each participant’s data (to represent individual change in BDI scores), and then comparing rates of change between groups using $t$-tests. These $t$-tests demonstrated a greater rate of improvement with SPEC than MT during the first month of acute treatment ($t = 1.72$, $p = 0.047$ and $d = 0.70$), but not during the second month ($p = 0.083$, $d = 0.54$). Similar $t$-tests were not conducted between MT and NSPEC, but insufficient data were available to enable reanalyses.

Following the acute treatment phase, Manber et al. (46) offered continuation treatment, to treatment responders, until 8 weeks after labour. Participants who did not receive continuation treatment also participated in the continuation phase assessments and were included in subsequent analyses. Using baseline and postcontinuation-phase data, repeated measures ANOVA demonstrated no significant group by time interaction effects on the BDI or the Hamilton Rating Scale for Depression. Likewise, pair-wise analyses conducted for this review revealed no significant difference between MT and SPEC, or between MT and NSPEC, in the odds of achieving full remission (OR $= 0.33$, 95% CI: 0.05–2.10; OR $= 2.00$, 95% CI: 0.49–8.24 respectively; see Table 3 for criteria used to define full remission).

Discussion

This systematic review cannot generate conclusive evidence for the effectiveness of MT as a treatment for depression. We identified only four relevant
Table 3 Randomised clinical trials of massage therapy for the treatment of depression

<table>
<thead>
<tr>
<th>References</th>
<th>Depression type</th>
<th>Participants</th>
<th>MT details</th>
<th>Comparison details</th>
<th>Timeframe</th>
<th>Relevant measures (and assessment times)</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Field et al. (43)</td>
<td>Depression or dysthymia</td>
<td>Hospitalised adolescents meeting DSM-III-R criteria for depression or dysthymia; $n = 36$</td>
<td>(A) Clothed back MT; administered by psychology students; $n = 26$</td>
<td>(B) Relaxing videotape watched with psychology student; $n = 10$</td>
<td>5 days; 5 sessions (1 × 30 min session per day)</td>
<td>Pre- and postsession at first and last session: (1) Behaviour Observation Scales (2) POMS</td>
<td>Analyses were unclear. Insufficient data provided in tables to enable reanalysis</td>
</tr>
<tr>
<td>Field et al. (44)</td>
<td>Postnatal depression</td>
<td>Adolescent mothers meeting DIS criteria for dysthymia; BDI score &gt; 16; $n = 32$</td>
<td>(A) MT to head, neck, arms, hands, torso, legs, feet; administered by trained therapist; $n = 16$</td>
<td>(B) Relaxation therapy (yoga + PMR); $n = 16$</td>
<td>5 weeks; 10 sessions (2 × 30 min sessions per week)</td>
<td>Pre- and postsession at first and last session: (1) Behaviour Observation Scales (2) POMS</td>
<td>Analyses were unclear. Insufficient data provided in tables to enable reanalysis</td>
</tr>
<tr>
<td>Field et al. (45)</td>
<td>Depression in pregnancy</td>
<td>Pregnant women, criteria for establishing depression not provided; $n = 84$</td>
<td>(A) MT to head, neck, arms, hands, back, legs, feet; administered by significant others (taught by trained therapist); $n = 28$</td>
<td>(B) PMR; $n = 28$</td>
<td>16 weeks; 32 sessions (2 × 20 min sessions per week)</td>
<td>Pre- and postsession at first and last session: (1) POMS At first and last session: (2) CES-D</td>
<td>Significant group × time (over multiple sessions) MANOVA $[F(3,66) = 24.32, p &lt; 0.001]$, reported as favouring massage, but appropriate pair-wise analyses not provided. Insufficient data provided in tables to enable reanalysis</td>
</tr>
<tr>
<td>References</td>
<td>Depression type in pregnancy</td>
<td>Participants</td>
<td>MT details</td>
<td>Comparison details</td>
<td>Timeframe</td>
<td>Relevant measures (and assessment times)</td>
<td>Results</td>
</tr>
<tr>
<td>-----------------</td>
<td>------------------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Manber et al. (46)</td>
<td>Depression in pregnancy</td>
<td>Pregnant women meeting DSM-IV criteria for a current, non-psychotic, major depression episode; HRSD score ≥ 14; n = 61</td>
<td>(A) MT to back, face, head, neck, shoulders, feet; unclear who administered treatment*; n = 20</td>
<td>(B) Depression-specific acupuncture; n = 20</td>
<td>8-week acute treatment phase; 12 sessions (25–30 min sessions); continuation phase: 8–17 additional sessions for acute phase responders</td>
<td>Pre-, mid- and postacute phase: (1) Response to treatment defined by: Failure to meet DSM-IV MDD criteria ≥ 50% reduction in HRSD score HRSD score ≤ 14 Weekly during acute phase: (2) BDI Postcontinuation phase: (3) HRSD (4) BDI (5) Full remission defined by HRSD score ≤ 8</td>
<td>Acute phase results: (1) B favoured over A (Fisher exact p = 0.03, NTT² effect size = 2.7, 95% CI: 1.5–16) Pair-wise contrast between A and C not provided. Analysis conducted for this review found no significant difference between A and C (OR = 0.51, 95% CI: 0.14–1.92) (2) Rate of change greater in B than A in 1st month of treatment (t = 1.72, p = 0.047, d = 0.70), but NSD between A and B in 2nd month (p = 0.083, d = 0.54) Pair-wise contrasts between A and C not provided. Insufficient data provided to enable analysis Continuation phase results: (3) Repeated measures ANOVA demonstrated no significant treatment by time interaction effect [F(2,41) = 2.2, p = 0.12] (4) Repeated measures ANOVA demonstrated no significant treatment by time interaction effect [F(2,41) = 1.6, p = 0.21] (5) Pair-wise contrasts between A and B, and between A and C, not provided. Analyses conducted for this review found no significant difference between A and B (OR = 0.33, 95% CI: 0.05–2.10), or between A and C (OR = 2.00, 95% CI: 0.49–8.24)</td>
</tr>
</tbody>
</table>

*MT details were not included in the trial report but were provided by R. Manber (personal communication). POMS, Profile of Mood States; CES-D, Centre for Epidemiologic Studies Depression Scale; BDI, Beck Depression Inventory; HRSD, Hamilton Rating Scale for Depression.
trials, all of which, through methodological, reporting or design limitations, could not be used to evaluate the effectiveness of MT as a treatment for depression.

Three of the studies (43–45) either did not conduct, or did not clearly describe, appropriate between-group analyses to enable evaluation of the multiple-session effects of MT. The fourth trial evaluated acupuncture (MT was intended as a control treatment) (46). The authors state that their study is not a realistic assessment of MT, because the MT sessions were designed to be comparable in length to the SPEC and NSPEC sessions, and were therefore shorter than is typical in clinical practice (46). However, the MT sessions resembled classical European massage (R. Manber, personal communication) and were of a similar length to those used in the other trials (43–45). There is also little information on optimal dosing for MT for the treatment of depression, and the data from all trials should be considered irrespective of the length of treatment sessions. Nevertheless, the results (46) have little bearing on the conclusions that can be drawn about the effectiveness of MT for the treatment of depression because there is currently insufficient evidence to suggest that acupuncture is effective for this condition (47). Comparison between acupuncture and MT only serves as a comparison between two treatments of unknown efficacy.

Our finding that there is currently no good evidence to suggest that MT is an effective treatment for depression differs from the conclusions of two previous reviews, one of which focused on all CAM for depression (8), and the other of which focused on MT for all indications (10). The latter review concluded that ‘reductions in trait anxiety and depression following a course of treatment were among MT’s largest effects’ (10), but included trials where depression was measured as an outcome even if participants were not selected for having depressive symptoms, and did not include any of the four trials identified here. The other review (8) summarised two of the trials included here (43,44) and concluded that ‘MT appears to have short-term benefits’, and that ‘its longer-term effects have not been evaluated’. It is apparent that the reviewers include both single-session and multiple-session effects under ‘short-term benefits’, reserving the phrase ‘longer-term effects’ for the long-term follow-up of participants. The most likely reason for the different conclusions drawn in this previous review (8) and by ourselves is that by focusing on a specific therapy we could devote our review to the critical evaluation of the methodology and analyses of the included trials. In providing a general overview of CAM for depression, the authors (8) succinctly summarised, but did not critically appraise, the available data for each therapy.

Our findings do support, and indeed update, the conclusions of another previous review (9), which identified one of the RCTs included here (43), and concluded that ‘the data are insufficient to judge the value of massage for depression’. Thus, despite the popularity of MT, and the identification of three more trials, including two that have been conducted since the publication of this previous review (9), the conclusions still stand.

It is not entirely unexpected that a critical evaluation of MT trials would identify methodological problems and lead to difficulties in drawing conclusions regarding the effectiveness of MT. This is because it is difficult to provide an adequate control condition for this intervention. However, direct comparisons can be made between MT and standard evidence-based treatments, such as cognitive-behavioural treatment (CBT). However, it is still impossible to ensure that patients and therapists are blind to treatment allocation in such trials. This is undoubtedly an obstacle in MT research, and may account for the general paucity of trials in this area.

Some limitations of our review require discussion. We did not conduct hand searches of the literature, nor did we contact relevant experts or organisations to identify further trials. However, ‘grey literature’ does appear in some of the databases searched. Although we did not exclude non-English language articles, we did not search non-English language databases so we may have unknowingly excluded trials. Finally, our standardised quality assessments reflect the information provided in the trial reports, and may not fairly represent the trials themselves (11). We did supplement the Jadad score with other methodological information, but this too was based upon the reports rather than the trials.

Conclusions

Although a recent review (8) and meta-analysis (10) have indicated that MT may be an effective treatment for depression, this critical evaluation demonstrates a lack of evidence, from RCTs which have selected participants with depressive symptoms, for multiple-session Swedish MT (as either a sole or additive treatment). Of course, lack of evidence does not equate to lack of effectiveness, and Swedish MT delivered by a trained therapist is generally safe (48). Potential indirect risks of MT should also be considered (e.g. hindering access to mainstream services), particularly when patients use MT on its own. Any future trials aiming to estab-
lish whether MT is beneficial for the treatment of depression should draw upon, and seek to improve, the methodological and reporting limitations of the previous research.

Acknowledgements

HFC has a research fellowship from the Pilkington Charitable Trusts, who had no role in this research other than in financing the fellowship. We would like to thank Barbara Wider who, in addition to EE, independently considered non-English language trial reports for inclusion in this review.

References


Paper received June 2007, accepted August 2007