

THE EFFICACY OF EMOTIONALLY FOCUSED COUPLES THERAPY AND BEHAVIORAL COUPLES THERAPY: A META-ANALYSIS

Maren Rathgeber, Paul-Christian Bürkner, Eva-Maria Schiller, and Heinz Holling
University of Muenster

Behavioral couple therapy (BCT) and emotionally focused couples therapy (EFCT) are well-established treatments to reduce couple distress. This meta-analysis summarizes the current state of knowledge on the efficacy of these two therapy methods by focusing on randomized controlled trials only. A literature search revealed 33 suitable primary studies (2,730 participants in total), all of them measuring relationship satisfaction. Robust-variance random-effects meta-analysis revealed medium effect sizes at post-test (overall: $g = 0.60$; BCT: $g = 0.53$; EFCT: $g = 0.73$) and small effect sizes of 6 months after treatment (overall: $g = 0.44$; BCT: $g = 0.35$; EFCT: $g = 0.66$), but these gains were not maintained after 12 months (BCT only: $g = 0.06$). Between the two therapy methods, no significant effect size differences could be found. Results have to be interpreted with caution due to potential publication bias.

COUPLE DISTRESS AND COUPLE THERAPY

Marital dissatisfaction and couple distress are widespread problems in many families, resulting not only in increasing divorce rates in most developing nations (as defined by the Organization for Economic Cooperation and Development, 2017), but also in psychosocial problems and psychological disorders in couples and their children (Grych & Fincham, 1990; Lebow, Chambers, Christensen, & Johnson, 2012; Uebelacker & Whisman, 2006; Whisman & Bruce, 1999).

One of the main issues leading to couple distress are automatized conflict patterns between partners. Often, each partner has a specific conflict behavior with a specific underlying position; for example, one partner is regularly the one who criticizes while the other partner frequently reacts with withdrawal, establishing a vicious circle (Steinglass, 1978). Effective therapies that can alleviate couple distress are needed. While couple therapy approaches abound (Gurman, Lebow, & Snyder, 2015), two theory-based, manualized approaches stand out as having the longest history of successful rigorous evaluations as to their effectiveness: The *behavioral couple therapy* (BCT) and the *emotionally focused couple therapy* (EFCT). Both aim to solve the conflict patterns distressed couples often display (Jacobson & Margolin, 1979; Johnson & Greenberg, 1985). For both, BCT (Jacobson & Margolin, 1979) and EFCT (Johnson & Greenberg, 1985) several studies have confirmed their efficacy (e.g. EFCT: Ahmadi, Rasouli, Alaf, & Zadi, 2014; Dalton, Greenman, Classen, & Johnson,

Maren Rathgeber received her Master degree in psychology from the Department of Psychology at the University of Muenster, Germany. She is in training for cognitive-behavioral-therapy at the University of Muenster, Germany; Paul-Christian Bürkner received his Master degrees in psychology and mathematics from the University of Muenster, Germany and received his PhD from the Faculty of Psychology at the University of Münster. He is research assistant at the Department of Statistics and Methods at the University of Muenster, Germany; Dr. Eva-Maria Schiller received her PhD from the Faculty of Psychology at the University of Vienna, Austria. She is the managing director of the Counselling Lab and postdoctoral researcher and lecturer at the Department of Psychology, University of Muenster, Germany; Prof. Dr. Heinz Holling is full professor and the head of the Department of Statistics and Methods at the University of Muenster, Germany.

Two anonymous reviewers gave very thoughtful and detailed feedback that helped us a lot in improving the manuscript. We acknowledge Dr. Celeste Brennecke for proofreading.

Address correspondence to Paul-Christian Bürkner, Department of Psychology and Sport Sciences, University of Muenster, Fliednerstr. 21, D-48149, Münster, Germany; E-mail: paul.buerkner@uni-muenster.de

2013; Denton, Wittenborn, & Golden, 2012; Sims, 1999; BCT: Baucom, 1982; Beach & Daniel O'Leary, 1992; Fals-Stewart, Yates, & Klostermann, 2005; Kavitha, Rangan, & Nirmalan, 2014; Lam, Fals-Stewart, & Kelley, 2009). In addition, prior meta-analytic evidence documented medium to large effects on relationship satisfaction for BCT (Byrne, Carr, & Clark, 2004; Hahlweg & Markman, 1988; Powers, Vedel, & Emmelkamp, 2008; Shadish & Baldwin, 2005; Wood, Crane, Schaalje, & Law, 2005) and large effects for EFCT (Byrne et al., 2004; Johnson, 1999). However, previous meta-analyses either investigated only one of the two approaches or also included primary studies without a rigorous randomization, thus not allowing to systematically compare the efficacy of BCT and EFCT. In addition, they often neglected to consider different types of biases (e.g., publication bias, attrition bias, allocation bias, performance bias), and they focused only on a small group of potential moderators. For instance, the moderating role of the type of control group or comorbidity has been neglected. Finally, existing meta-analyses included only studies published until 2005 for EFCT and 2008 for BCT. Those characteristics of previous meta-analyses indicate that an extended and updated meta-analysis is required. Consequently, the aim of the present meta-analysis was to address all aforementioned aspects: (a) We examined and compared the effects of BCT and EFCT on relationship satisfaction. Furthermore, (b) we selected only randomized controlled trials. According to recommendations by the Cochrane Collaboration, randomization is needed to avoid systematic differences between participants in different experimental groups due to both known and unknown confounding factors. Empirical evidence suggests that, on average, nonrandomized studies produce effect estimates that indicate more extreme benefits of the effects of health care than randomized trials (Higgins & Green, 2011). Thus, potential differences in efficacy between BCT and EFCT are yet to be explored on the basis of stricter inclusion criteria. In addition, (c) we considered a broader range of potential biases. Following the Cochrane Collaboration's tool for assessing risk of bias (Higgins & Green, 2011) the risk of biases should be assessed separately. Due to the individual assessment, each bias's impact can be analyzed in a more differentiated way. Beyond that, (d) we included 32 potentially moderating variables, because a highly relevant question in clinical research is which treatment works for whom under which conditions (Wood et al., 2005). Moreover, (e) we included studies published until 2015 and thus provide an update to the state of knowledge about the two primary couple therapy approaches.

BEHAVIORAL COUPLE THERAPY AND EMOTIONALLY FOCUSED COUPLES THERAPY: BASIC PRINCIPLES AND EMPIRICAL EVIDENCE

Both, BCT and EFCT aim to ameliorate interaction patterns between partners. The goal of BCT is to increase positive, relationship-supporting interactions between partners using behavioral techniques (Byrne et al., 2004; Jacobson & Margolin, 1979). In the therapy, partners are asked to identify and collect relationship-reinforcing events and behaviors; this step is part of the behavior exchange (Byrne et al., 2004). To facilitate the occurrence of these interactions, BCT involves three central components: negotiation, problem-solving and communication skills training (Christensen et al., 2004; Jacobson, Schmalings, & Holtzworth-Munroe, 1987). Studies examining effects of these components suggest that problem-solving and communication-skills training explain most of the treatment's efficacy (Shadish & Baldwin, 2005).

In prior meta-analyses the outcome variable was labeled *couple distress* or *relationship satisfaction*, which are two poles of the same construct. In the following, we use the expression *relationship satisfaction* to describe the empirical evidence.

Hahlweg and Markman first published a meta-analysis differentiating between couple therapy and family therapy. They found that married couples receiving BCT showed significantly higher improvement in relationship satisfaction than did waiting-list controls (Cohen's $d = 0.95$)¹ These therapy gains persisted for at least 1 year after treatment (Hahlweg & Markman, 1988). Shadish and Baldwin analyzed 30 randomized studies evaluating the efficacy of BCT compared to no treatment conditions, and they reported an effect size of $d = 0.59$ (Shadish & Baldwin, 2005). The authors tried to explain the difference between the large effect size found by Hahlweg and Markman (1988) and the moderate effect size reported in their own study (Shadish & Baldwin, 2005) by arguing that in the previous meta-analysis by Hahlweg and Markman, unpublished material was excluded, leading to a publication bias. Additionally, according to Shadish and Baldwin, BCT did

not show significant superiority compared to other couple treatment approaches (Shadish & Baldwin, 2005); this finding was also replicated by Wood and colleagues (Wood et al., 2005). In 2008, Powers and colleagues published a meta-analysis about the efficacy of BCT concerning relationship satisfaction as well as substance-related outcomes. This analysis was limited to randomized trials, but EFCT was not examined. They stated that BCT fared better than any other control treatment for relationship outcomes ($d = 0.64$), but was not superior concerning substance-related outcomes such as abstinence at post-test. At follow-up assessment, BCT outperformed the control treatments concerning relationship satisfaction as well as substance-related outcomes (Powers et al., 2008).

Whereas BCT concentrates on eliciting positive interactions through helping partners identify relationship-reinforcing events and ameliorate their communication to facilitate the occurrence of these events (Byrne et al., 2004; Christensen et al., 2004), EFCT pursue another objective. Based on attachment theory (Byrne et al., 2004; Johnson, Hunsley, Greenberg, & Schindler, 2006), EFCT proposes that temporal insecure attachment (e.g., fear of abandonment) and negative emotions can derive from disputes between partners (Johnson et al., 2006). In this context, EFCT distinguishes between primary and secondary emotions. Primary emotions include, for example, fear of abandonment and need for secure attachment; secondary emotions include, for example, anger and aggression (Byrne et al., 2004). EFCT is based on the assumption that distressed couples often communicate secondary emotions, for example via attacking, nagging, and withdrawing, instead of sharing primary emotions (Byrne et al., 2004). Thus, EFCT aims to ameliorate interaction patterns through helping partners to identify and to accept previously unconscious primary emotions related to the conflict patterns, reframe the conflict patterns in light of the newly conscious primary emotions and facilitate partners' identification with these primary emotions (Greenberg & Johnson, 1988). Through this process, EFCT focuses on the rebuilding of attachment bonds (Byrne et al., 2004). This aim is promoted through a process consisting of nine steps subdivided into the following three phases: (a) assessment and de-escalation of the conflict pattern, (b) encouraging, supporting, and validating of secure attachment experiences between partners, and (c) consolidation of the new secure attachment base into everyday life (Byrne et al., 2004). Process studies indicate that the main elements of EFCT consist of intense emotional experiences in therapy sessions and the creation of new interaction patterns between partners, such as communicating attachment needs and responding to the partner's needs (Lebow et al., 2012). In EFCT two versions exist, the Johnson version and the Greenberg version. Whereas the Johnson version focuses on attachment processes, the Greenberg version focuses on affect regulation (APA, 2007). Both versions of EFCT have been shown by randomized trials to be effective (Dalton et al., 2013; Greenberg, Warwar, & Malcolm, 2008).

The only meta-analysis exclusively examining the efficacy of EFCT on relationship satisfaction reported an average effect size of $d = 1.31$, computed of seven partly randomized studies (Johnson, 1999). Results of EFCT seemed to be stable at the 2-year follow-up time point (Cloutier, Manion, Walker, & Johnson, 2002). In 2004, Byrne and colleagues published a meta-analysis examining the efficacy of BCT and EFCT on quality of couples' relationships compared to waiting-list controls. The study revealed large effect sizes for both treatments ($d_{BCT} = 0.95$, $d_{EFCT} = 1.27$, respectively), and the therapy gains persisted at follow-up periods ranging from 6 months to 4 years. On average, couples receiving BCT were more distressed after treatment than couples receiving EFCT (Byrne et al., 2004). The therapy gains associated with BCT were maintained for up to 4 years compared to a waiting-list control group. EFCT follow-up testing showed that significant improvements in relationship satisfaction lasted for up to 2 years (Byrne et al., 2004). Finally, it should be emphasized that in contrast to BCT, the meta-analyses examining EFCT have never been limited to randomized controlled trials.

OBJECTIVES

This meta-analysis addresses the following research questions and hypotheses.

1. What effect do BCT and EFCT have on relationship satisfaction?

The present analysis includes both versions of EFCT (Greenberg's and Johnson's version) and the traditional version of BCT following Jacobson and Margolin. We assume that combined effect sizes of BCT and EFCT will reveal moderate effect sizes comparable to

those found by Shadish and Baldwin (2005) and Powers et al. (2008), who both included larger and randomized samples. Due to previous findings, considerable long-term effects are also expected (Byrne et al., 2004). Still, long-term effects should be smaller than short-term effects (Davis, Lebow, & Sprenkle, 2012).

2. Do BCT and EFCT differ in their effects on relationship satisfaction?
Based on previous results, the efficacy of EFCT is expected to be somewhat greater than the effect size of BCT when both are compared to waiting-list controls (Johnson, 1999). However, differences are not assumed to be large (Davis et al., 2012).
3. Which variables moderate the effect of couple therapy on relationship satisfaction?
Lower levels of initial relationship satisfaction are expected to be positively associated with the efficacy of couple therapy (McCrary, Epstein, Cook, Jensen, & Hildebrandt, 2009). Here, it is also hypothesized that longer duration of the relationship will be associated with larger effects (Atkins et al., 2005). As stated by Lambert and Hawkins (2001), the implementation of supervision should elicit better results. Unpublished results should decrease the overall effect sizes and, consequently, the combination of published and unpublished data should display lower effect sizes (Shadish & Baldwin, 2005). Finally, results of several studies have suggested that using active treatments as a control intervention result in lower effect sizes than when no treatment (such as a waitlist control) is used as the control intervention (e.g. Powers et al., 2008; Wood et al., 2005). Therefore, in this meta-analysis, 32 potential moderators – including some that have never been analyzed before – were included to give indications for clinical practice. All hypotheses concerning moderating variables are displayed in Appendix A1 (accessible via Online Supplemental Files).
4. Which biases are found to be present in the data?
Randomization bias, allocation bias, performance bias, attrition bias, and other bias were examined separately. An overview of all biases can be found in Appendix A2 (accessible via Online Supplemental Files).

METHOD

Inclusion Criteria

We included studies in this meta-analysis if the following criteria were met: (a) Married or unmarried adult couples (aged 18 years or older) were involved in the treatment. (b) Couple therapy was defined as a psychological intervention involving the presence of both partners, in regular sessions led by a therapist with the aim of augmenting relationship satisfaction. (c) At least one intervention met BCT or EFCT criteria (Jacobson & Margolin, 1979; Johnson, 1999). (d) EFCT or BCT efficacy was evaluated through randomized clinical trials where couple therapy was compared to non-active controls, drug therapy, individual psychotherapy, psychoeducation, or counseling. (e) Studies measured relationship satisfaction by at least one measure. (f) Studies provided sufficient data to calculate effect sizes. (g) Studies assessed outcome measures concerning relationship satisfaction at least at baseline and at post-test. (h) Both versions of EFCT and the traditional version of BCT were included. (i) Study versions were available in English, German, French, or Spanish.

Search Strategy

Database searches were performed on MEDLINE, PsycINFO, PsycARTICLES, PSYINDEX; Literature and audiovisual media were found with PSYINDEX Tests, the Cochrane Central Register of Controlled Trials, PubMed, Springerlink, BASE and ProQuest Dissertations & Theses A&I. We used a similar search strategy as Byrne et al. (2004), except we further restricted the search term to controlled and randomized trials, and the concepts *attachment theory* and *traditional behavioral couple therapy* were added. Following recommendations of the Cochrane Handbook, the Boolean search term was subdivided into the *PICO criteria*, summarizing the inclusion criteria: Participants, interventions, controls, and outcomes (Higgins & Green, 2011). The search concepts forming the Boolean search term and the complete search term can be found in Appendix A3. The study was conceptualized according to the recommendations of the Meta-Analysis Reporting

Standards (MARS; APA, 2008), but it is also in general accordance with the PRISMA guidelines (Moher, Liberati, Tetzlaff, & Altman, 2009; the PRISMA checklist is provided in the Appendix). Advanced search features were applied in the Cochrane Central Register of Controlled Trials in order to specify the results (see Appendix A2).

Concerning manual searches in bibliographies, the publication lists of authors who published many studies or who were important in this field of research (Christensen et al., 2004) were searched separately. To complete this analysis via backward search, we scanned a recently published review (Lebow et al., 2012), several previous meta-analyses (Barbato & D'Avanzo, 2008; Byrne et al., 2004; Dunn & Schwebel, 1995; Hahlweg & Markman, 1988; Johnson, 1999; Powers et al., 2008; Shadish & Baldwin, 2005; Wood et al., 2005) and the reference lists of each matching study (including those published before 2002). Only studies that met all the inclusion criteria outlined above were included. An explicit review protocol was not created.

Coding Procedure

A coding scheme was developed by MR and PCB in order to extract important information from the studies. All studies were initially coded by MR, who applied the final version of the coding scheme after being trained by PCB and HH in the coding of meta-analyses. The coding scheme consisted of the following categories: study information, study quality, sample characteristics, treatment conditions, methodological features, outcome information and assessment of results. Central outcome information and study quality were verified by PCB, and disagreements were resolved by consensus. In case of missing or unclear data, authors of the primary studies were contacted and asked for clarification. An overview of all variables coded in this meta-analysis can be found in Appendix A4. A complete description of the content of all variables and coding options can be found in Appendix A5. Study quality was assessed according to the Cochrane Collaboration's tool for assessing risk of bias (Higgins & Green, 2011), which includes assessments for selection bias (randomization bias and allocation bias), performance bias, detection bias, attrition bias, reporting bias, and other bias. These are coded as *high risk*, *low risk*, and *unclear risk*. In this meta-analysis, only randomization bias, allocation bias, performance bias, attrition bias, and other bias could potentially differ between studies and were therefore relevant.

Statistical Procedure

Data were selected and inserted in an Excel spreadsheet (see Appendix A5) and were then analyzed using the *System for Statistical Computation and Graphics R*, version 3.3.2 (R Core Team, 2016). Hedges' *g* overall effect sizes were computed by applying the R package *metafor* (Viechtbauer, 2010). Hedges' *g* is essentially the same as Cohen's *d*, but it corrects for the small positive bias that Cohen's *d* has in small samples. Hence, the standard interpretation of a small ($d = 0.2$), medium ($d = 0.5$) and large ($d = 0.8$) effect applies to Hedges's *g* as well.

Some studies reported more than one effect size because of several measures per experimental group or because of several experimental/control groups (in the latter case, all experimental groups were compared to all control groups). As a result, the number of effect sizes (k_s) was larger than the number of included studies (k). To respect this dependence, random effects robust variance meta-analyses were conducted via the R package *robumeta*, which uses robust variance estimation (Hedges, Tipton, & Johnson, 2010). This technique serves as a robust alternative to multilevel meta-analyses. All statistical analyses were computed at a 5% significance level and one-tailed when directional hypotheses about results were available. Potential outliers were detected via visual inspection of the forest plots.

Mean values, standard deviations and number of observations at baseline and post-test were coded for each study in order to compute effect sizes. When this information was missing, baseline/post-change scores were coded instead. When 6- or 12-month follow-up data were available, it was included in the analysis. We decided in favor of these follow-up assessments because they were provided in most of the studies. Before computing effect sizes, mean scores of all studies were pooled in the same direction, with higher scores indicating more relationship satisfaction. Detailed information about the combination of subgroup scores can be found in Appendix A6.

Effect sizes were computed in two different ways. First, standardized mean differences in each time point were computed separately by comparing the experimental and control groups at post-

test, 6-month follow-up, and 12-month follow-up, respectively. Second, controlled standardized mean differences were computed for three comparisons of different time points: (a) between baseline and post-test (post), (b) between baseline and 6-month follow-up (FU6), and (c) between baseline and 12-month follow-up (FU12). See Appendix A7 for the mathematical details. This meta-analysis focuses on the second type of effect sizes, because it controls both for baseline values and for differences between experimental and control group. In two cases, controlled mean differences were computed on the basis of baseline/postchange scores [14, 16]. One single study provided only baseline/post-test effect sizes [1].

Moderator analysis was conducted by means of meta-regression of controlled effect sizes (Higgins & Green, 2011). Each continuous and categorical variable was analyzed separately due to reasons of statistical power. An analysis of all time points (post, FU6, FU12) was carried out when sufficient data was available. A one-tailed analysis was conducted when a hypothesis about the variable's direction of influence was accessible. For all exploratory analyses of potential moderators, two-tailed tests were carried out. To assess the degree of heterogeneity, the I^2 statistic as well as the between sample standard deviation τ was computed. If the I^2 statistic reveals a high amount of heterogeneity, studies do not share a common effect size. In this case, it is probable that moderators explain the difference between studies or a random effects model has to be applied.

Risk of publication bias was graphically investigated by means of funnel plots (Egger, Davey Smith, Schneider, & Minder, 1997). Publication bias may be present if the studies are asymmetrically distributed around the averaged effect size. In this case, the overall effect is likely overestimated. To test statistically for publication bias, the *trim-and-fill* method (Duval & Tweedie, 2000) as well as *Egger's test* (Egger et al., 1997) were applied.

Studies included in the current meta-analysis are cited by their Study-ID rather than by authors and publication year for reasons of brevity (see Appendix A8 and A10 for more details).

RESULTS

Study Selection

Figure 1 illustrates the study selection process. The electronic search strategy was applied in the following databases: ProQuest Dissertations & Theses A&I, Cochrane Central Register of Controlled Trials and PubMed, a combined search in MEDLINE, PsycINFO, and PsycARTICLES, PSYNDEX (Literature and Audiovisual Media with PSYNDEX Tests). After duplicates were removed, 15 of 5847 articles met the inclusion criteria [2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, 17, 18]. In the Springerlink database, a modified Boolean search term was applied, revealing 1256 studies. A separate search was conducted for several concepts in Springerlink (see Appendix A3). Of 635 studies, none met the inclusion criteria. Once relevant articles were found, the reference lists were scanned to identify more studies that fit the inclusion criteria. No additional corresponding studies were identified. In 31 cases, potentially matching studies were not accessible in English, German, French, or Spanish, despite many efforts to get a version in one of these languages. In some cases no contact information was available or the authors did not respond. In Appendix A9 all inaccessible studies can be found. In sum, the keyword search revealed 13 studies that have

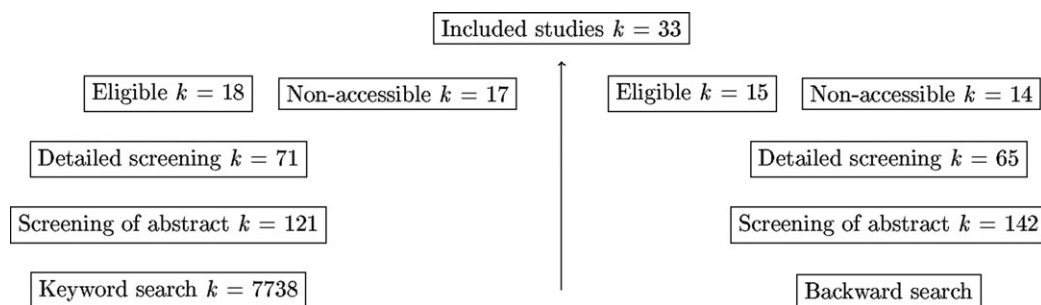


Figure 1. Keyword search and backward search, k , number of studies.

never been evaluated in a meta-analysis [1, 3, 4, 7, 8, 10, 11–17]. Within the manual search in bibliographies, 15 relevant studies were identified [19–33]. The publication lists of important researchers in this field did not reveal additional articles. Three studies were found coincidentally while searching for other studies [1, 14, 16]; one study was excluded from further analyses, because of an unrealistically large effect size (Ahmadi, Zarei, & Fallahchai, 2014; $d = 5.58$) that we believe resulted from a reporting error. A list of all studies included in this meta-analysis can be found in Appendix A10. The date of latest search was 16.6.2015.

Study Characteristics

The following information provides an overview of the conceptualizations of the studies included in this meta-analysis. Further information can be found in Appendix A12. Studies differed concerning their inclusion criteria. In 19 studies (57%), couples were primarily seeking help because of couple distress [1–3, 10, 12–14, 16, 19, 20, 24, 26–33]. In 14 studies (42%), participants were predominantly involved in couple treatment due to a psychological disorder [4–9, 11, 15, 17, 18, 21–23, 25]. Altogether, 2730 individuals were included in the 33 treatment studies. Of these, 278 individuals (10%) received EFCT [1–4, 12–14, 16, 29–32] and 1168 individuals (43%) were treated with BCT [5–11, 15, 17–28, 33]. Participants were predominantly Caucasian, with an average age of 36.61 years ($SD = 6.39$) ranging from 28.58 to 55.5 years. The average length of relationships was 9.74 years ($SD = 4.62$); individuals had an average of 1.67 children ($SD = 0.43$). In the majority of studies (76%), participants showed moderately distressed ranges. In only two studies (6%), individuals in the experimental groups were severely distressed at baseline [18, 23]. In only four studies (12%), participants were found to be within the range of mild couple distress [2, 15, 25, 32].

Methodological Characteristics

The studies varied in methodological characteristics. Further information is documented in Appendix A11. On an average, individuals participated in 10.88 sessions ($SD = 4.07$). Mostly, studies included a waiting-list control group [1–3, 14, 16, 19–21, 24, 26–33]. Twelve studies included an individual psychotherapy control condition [5–9, 11, 15, 17, 18, 21–23]. Regarding the study quality, almost all of the studies were coded at low risk for randomization bias [exception: 26], at high risk for allocation concealment [exception: 4, 10, 26], at high risk for performance bias [no exception] and at low risk for other biases [no exception]. With the exception of attrition bias (high risk: 3, 4, 10, 18, 19, 21, 23), no study quality measure was sufficiently heterogeneous across studies to be statistically analyzed further. The majority of the studies included in this meta-analysis were published; only three of the 33 studies were unpublished dissertations [2, 12, 29]. Furthermore, most of the investigations were published in the USA and in Canada; only two studies were conducted in Europe [17, 26] and only four studies were conducted in other places (Iran and in India) [1, 10, 14, 16].

Measures

This section focuses on the dyadic adjustment scale (DAS; Spanier, 1976), because it is the most applied measurement tool. DAS was used in 18 of 33 studies [2, 3, 5–8, 11, 12, 15, 18, 20, 21, 23, 25, 28–31]. All the other measures were applied in a maximum of four studies. The DAS defines relationship satisfaction as dyadic adjustment, whereby dyadic adjustment is considered any item intuitively related to relationship satisfaction that is normally distributed and different between married couples and divorced couples (Heyman, Sayers, & Bellack, 1994; Spanier, 1976). The DAS shows good reliability (Cronbach's $\alpha = .96$) as well as convergent validity (Heyman et al., 1994). The DAS differentiates between mildly distressed couples (DAS score 96–107), moderately distressed couples (DAS score 80–95.9), and severely distressed couples (DAS score <80, Wood et al., 2005). Apart from DAS, all the other measures included in this meta-analysis directly measure couple relationship satisfaction or couple intimacy, emotional support or longing for union with another.

Overall Effect

Effects from baseline to post-test were significantly larger in couple therapy conditions than in control conditions ($g_{\text{post}} = 0.60$, 95%-CI = [0.45, 0.74], $p < .001$, $\tau = 0.34$, $I^2 = 80\%$), revealing a

medium effect size with substantial between sample heterogeneity. This finding is based on 56 study samples from the 33 studies. Figure 2 represents a forest plot of the study-level effect sizes from baseline to post-test, including the 95% confidence interval for each study. Because all treatment groups were compared to all control groups, several studies are displayed in two or more lines to display the results of all comparisons.

Calculations of effect sizes from baseline to 6-month follow-up included 14 studies (42%; see A.12), providing 24 study samples. The average effect size of couple therapy versus control from baseline to 6-month follow-up was $g_{FU6} = 0.44$ (95%-CI = [0.26, 0.61], $p < .001$, $\tau = 0.25$, $I^2 = 74\%$), indicating a small to medium effect (Cohen, 1988). Eight studies (24%) provided 12 months follow-ups (see Appendix A12) contributing 18 study samples, all of them analyzing BCT. From baseline to 12-month follow-up, data revealed an averaged effect size of $g_{FU12} = 0.06$ (95%-CI = [-0.08, 0.19], $p = .35$, $\tau = 0.01$, $I^2 = 6\%$), indicating no effect 1 year after couple therapy. The forest plots illustrating the results from baseline to the follow-up assessments are shown

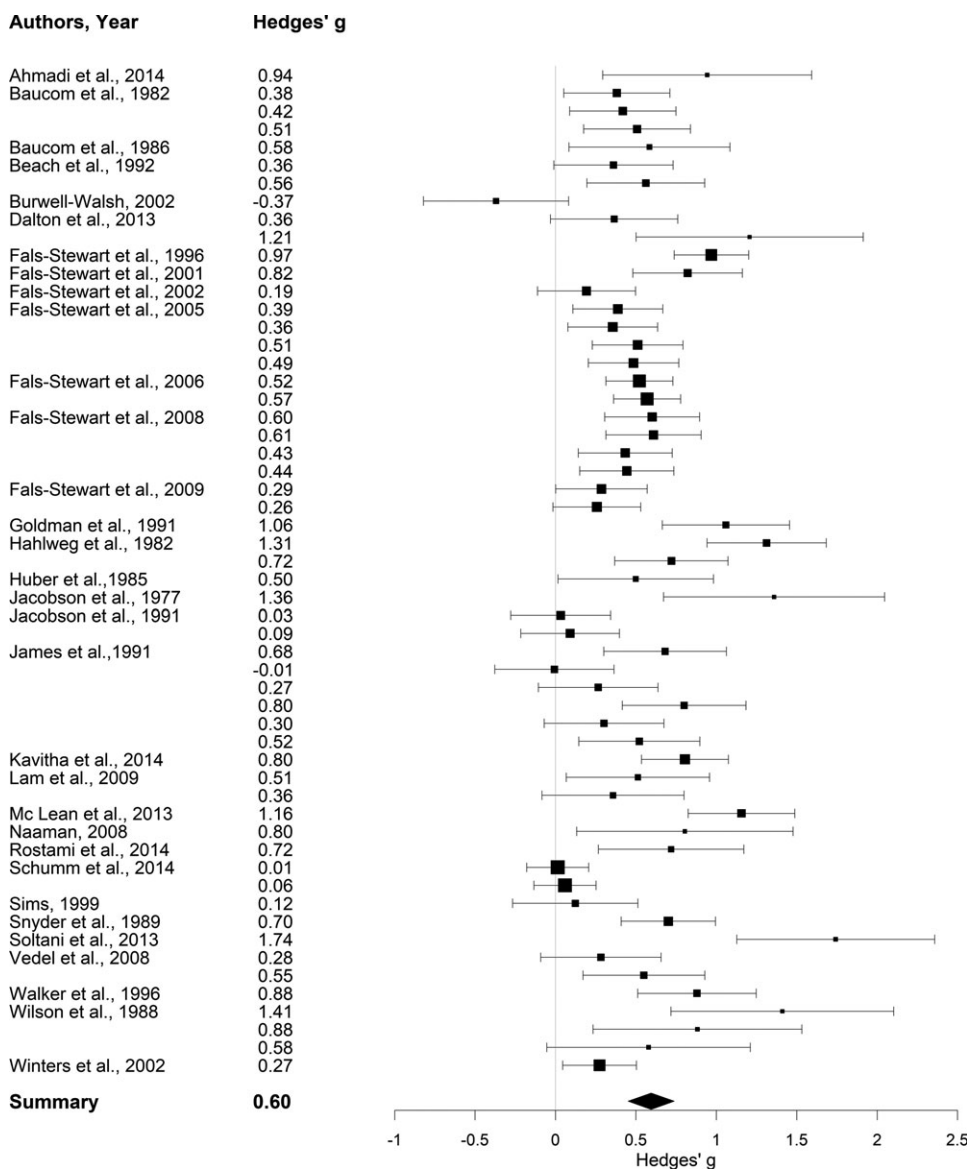


Figure 2. Forest plot of effect sizes and 95%-CIs from baseline to post-test.

in Appendix A13.1. Overall, a relatively large amount of heterogeneity could be attributed to the variation between samples as shown by the I^2 and τ estimates above. Results are summarized in Table 1.

In addition to controlled effect sizes, effect sizes were computed individually for the different assessment times (post-test, 6-month follow-up, and 12-month follow-up) to investigate the difference between the treatment group and the control group at each time point. Results of cross-sectional effects can be found in Appendix A13.2.

Comparison of BCT and EFCT

The effect sizes for BCT studies were averaged, which resulted in a mean effect size of $g_{\text{post}} = 0.53$ ($k = 21, k_s = 39, 95\%-\text{CI} = [0.39, 0.67], p < .001, \tau = 0.29$) from baseline to post-test. From baseline to 6-month follow-up, an average effect size of $g_{\text{FU6}} = 0.35$ was found ($k = 9, k_s = 19, 95\%-\text{CI} = [0.17, 0.53], p < .01, \tau = 0.20$). From baseline to 12-month follow-up, data revealed an average effect size of $g_{\text{FU12}} = 0.06$ ($k = 8, k_s = 18, 95\%-\text{CI} = [-0.08, 0.19], p = .35, \tau = 0.09$). Computing the effect sizes for EFCT produced an average effect size of $g_{\text{post}} = 0.73$ ($k = 12, k_s = 17, 95\%-\text{CI} = [0.38, 1.08], p < .001, \tau = 0.48$) and $g_{\text{FU6}} = 0.66$ ($k = 5, k_s = 5, 95\%-\text{CI} = [0.15, 1.17], p < .05, \tau = 0.34$), referring to the comparisons of baseline and post-test and 6-month follow-up data, respectively. The results suggest that there were no significant differential effects of couple therapy for BCT and EFCT, neither from baseline to post-test ($g_{\text{post (BCT vs. EFCT)}} = -0.18, k = 33, k_s = 56, 95\%-\text{CI} = [-0.53, 0.17], p = .29$), nor from baseline to 6-month follow-up assessment ($g_{\text{FU6 (BCT vs. EFCT)}} = -0.31, k = 14, k_s = 24, 95\%-\text{CI} = [-0.74, 0.13], p = .15$). However, it has to be noted that EFCT showed somewhat higher effect sizes, but the current evidence is not strong enough to clearly favor one over the other. Because no EFCT study provided data at 12-month follow-up, a comparison for 12-month follow-up was not possible.

Moderators of Treatment Effect

Effect sizes may vary due to the kind of treatment provided in the control groups. In the present meta-analysis, most of the studies included a waiting-list control group or compared couple therapy to individual psychotherapy. From baseline to post-test, treatment groups compared to nontreatment groups revealed an effect size of $g = 0.66$. When comparing couple therapy with individual-based treatments, an effect size of $g = 0.41$ was found. As predicted, baseline to post-test effect sizes between couple treatment and control using waiting list controls were significantly

Table 1
Summary of Meta-Analytic Effect Sizes

Time	Therapy	Hedges' g	k	k_s	95%-CI	p	τ	I^2
Post	Overall	0.60	33	56	[0.45, 0.74]	<.001	0.34	80%
	BCT	0.53	21	39	[0.39, 0.67]	<.001	0.29	70%
	EFCT	0.73	12	17	[0.38, 1.08]	<.001	0.48	81%
6-month follow-up	Overall	0.44	14	24	[0.26, 0.61]	<.001	0.25	74%
	BCT	0.35	9	19	[0.17, 0.53]	.002	0.20	64%
	EFCT	0.66	5	5	[0.15, 1.17]	.024	0.34	69%
12-month follow-up	Overall	0.06	8	18	[-0.08, 0.19]	.35	0.01	6%
	BCT	0.06	8	18	[-0.08, 0.19]	.35	0.01	6%
	EFCT	—	0	0	—	—	—	—

Notes All effect-sizes were corrected for pretreatment values. BCT, behavioral couple therapy; EFCT, emotionally focused couple therapy; k , number of studies; k_s , number of samples; p , p -value; τ , between study standard deviation; I^2 , percentage based measure of heterogeneity between studies.

larger than effect sizes between treatment and individual treatment ($g = 0.26$, 95%-CI = $[-0.02, 0.54]$, $p < .05$, one-tailed, $k = 29$, $k_s = 47$). Due to the small sample size at follow-up assessments, no further statistical analysis was conducted.

Other variables were examined to detect additional sources of heterogeneity among studies. The following variables were found not to have a significant moderating role on the data: Publication year, attrition bias, age, gender, number of children, years of education, comorbidity, medication, inclusion criteria, distress level DAS, inclusion criteria for duration of relationship, number of sessions, number of participated sessions, population criteria, spacing of sessions, simultaneous interventions, therapists' expertise, supervision, focus of therapy, treatment fidelity, manualization, implementation and differences between outcome measures. Too few data was available to analyze the following variables: Length of therapy sessions, ethnicity, percentage of separated couples, control group designs (psychoeducation and counseling) as well as influence of simultaneous intervention. However, some variables showed a significant association with improvement of relationship satisfaction. One of these was the baseline score on the DAS: When participants reported more initial distress, they showed more improvement from baseline to post-test compared to less distressed persons ($g = -0.002$ per DAS point, $k = 32$, $k_s = 55$, 95%-CI = $[-0.004, -0.000]$, $p < .05$). From baseline to 12-month follow-up, one further variable was associated with the outcome: participants having longer relationships benefited more from the treatment ($g = 0.02$, $k = 8$, $k_s = 18$, 95%-CI = $[-0.00, 0.03]$, $p < .05$, one-tailed). Two other significant results were found: First, unpublished dissertations were related to lower effect sizes compared with published studies from baseline to post-test ($g = -0.51$, $k = 33$, $k_s = 56$, 95%-CI = $[-1.05, 0.02]$, $p < .05$, one-tailed). Second, studies conducted in USA and Canada provided significantly lower effect sizes compared to studies from Iran and India ($g = 0.47$, $k = 33$, $k_s = 56$, 95%-CI = $[0.05, 0.89]$, $p < .05$). These two findings have to be interpreted with caution because of small sample sizes of the dissertation/Iran and India subgroups. Also, it has to be noted that the difference between countries was primarily driven by one Iranian study [6], which reported the highest effect size in the present meta-analysis with $g = 1.74$. Appendix A13.3 provides a complete overview of the moderator analyses.

Publication Bias

The effect sizes of the studies analyzed in this meta-analysis showed an asymmetrical distribution from baseline to post-test. Several studies seemed to be missing in the lower left part of the funnel plot, indicating a lack of studies with small sample sizes and small or negative effect sizes, potentially due to publication bias. Additionally, several studies with very small sample sizes contributed to the largest effect sizes found in this meta-analysis. Results of the trim and fill method revealed that from baseline to post-test, an estimated 15 studies are missing from the left side of the distribution ($SE = 4.88$); Egger's test was significant as well ($z = 3.68$, $p < .001$). Figure 3 illustrates this result. After correcting for the apparent funnel plot asymmetry in the pre- to post-test estimates, the overall effect was visibly smaller ($g = 0.38$, 95%-CI = $[0.30, 0.49]$) than before correction ($g = 0.60$, 95%-CI = $[0.45, 0.74]$).

From baseline to 6-month follow-up, only one study was estimated to be missing on the left side of the distribution ($SE = 3.14$; Egger's test: $z = 1.19$, $p = .234$). From baseline to 12-month follow-up, trim-and-fill did not reveal any missing studies ($SE = 2.50$), and Egger's test was non-significant ($z = -0.48$, $p = .633$). Funnel plots illustrating the results from baseline to the follow-up assessments can be found in Appendix A13.4.

DISCUSSION

Major Findings

The primary aim of this meta-analysis was to investigate the overall effect of BCT and EFCT on relationship satisfaction. Despite stricter inclusion criteria than in previous meta-analyses and a different method of computing effect sizes, both BCT and EFCT showed moderate, positive effects on relationship satisfaction from the beginning of the therapy to the end. From the beginning of couple therapy until 6 months after the treatment, both therapy approaches still showed a small, positive effect on relationship satisfaction. No significant effect of BCT (in the absence of EFCT

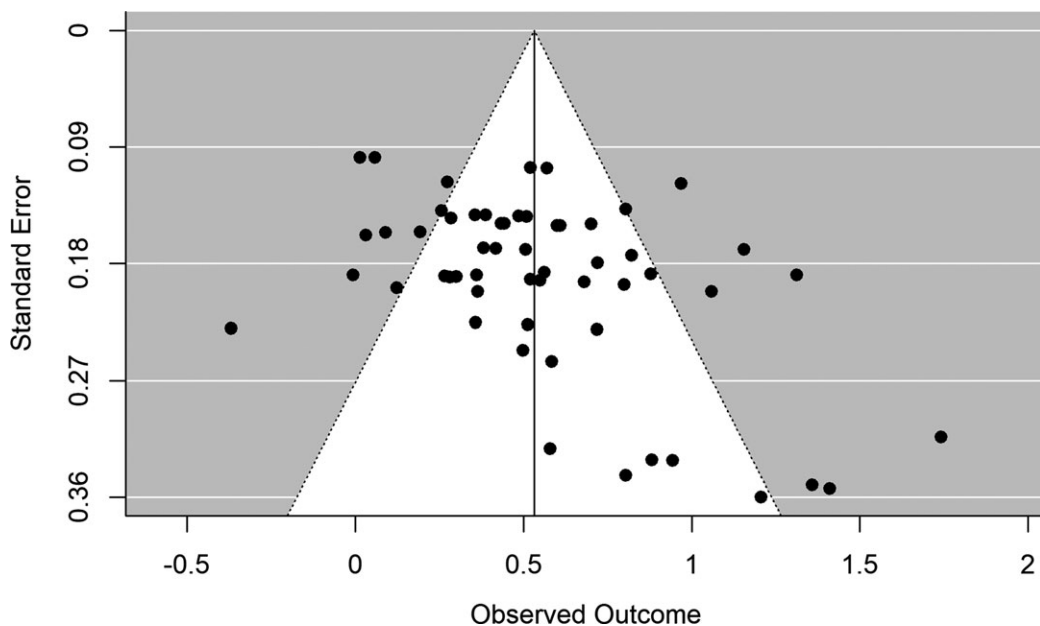


Figure 3. Funnel Plot of outcomes from baseline to post-test.

studies) was found 12 months after treatment. Additionally, substantial publication bias was found in the data for the post-test effect sizes. When results were corrected for publication bias, effects were smaller, but still significant and moderate in size. Furthermore, as expected, studies with waiting-list control groups revealed larger effect sizes than studies implementing active control groups.

Interpretation of the Results

Several results found in this meta-analysis need further interpretation. All the effect sizes found in this meta-analysis were lower than most of the effect sizes reported before (Byrne et al., 2004; Hahlweg & Markman, 1988; Johnson, 1999). As expected, this finding could be related to the relatively strict inclusion criteria compared to other meta-analyses; meta-analyses using similar inclusion criteria also reported similar effect sizes. Therefore, the first hypothesis was partially confirmed. Consistent with predictions, therapy gains seemed to diminish over time. This finding has to be considered with caution due to the small number of studies providing a 12 months follow-up assessment. Also, all studies assessing 12-month gains compared BCT to individual psychotherapy or psychoeducation including counseling. Still, it has to be noted that large drop in relationship satisfaction over time contradicted the first hypothesis, which states that couple therapy evokes sustainable rates of improvement in relationship satisfaction compared to other treatments.

With regard to the second hypothesis, efficacy differences between BCT and EFCT were noteworthy in size ($g = 0.17$ at post-test and $g = 0.31$ at 6-month follow-up both favoring EFCT), but did not reach significance. A comparison at 12-month follow-up was not possible, because none of the EFCT studies provided 12-month follow-up data. To summarize, there is not (yet) enough evidence to favor on treatment over the other and future research will need to clarify this point. It is, however, clinically important to point out that both couple therapy approaches were clearly superior both to no treatment as well as to individual therapy, which may be the result of relevant conceptual overlap between processes in BCT and EFCT.

Regarding the analysis of moderating variables, we found, congruent with the hypotheses about moderating variables, that effect sizes comparing couple therapy to a waiting-list control were significantly larger than those comparing couple therapy to individual psychotherapy. Another moderating variable assessed was relationship duration. Participants with longer

relationships showed descriptively more improvement at follow-up assessment, supporting the hypothesis that relationship duration positively influences the efficacy of couple therapy. Each year of relationship increases the efficacy of couples therapy by an estimated effect size of $g = 0.02$, which appears clinically relevant (despite not being statistically significant), as relationships often span several decades. Still, relationship duration was not associated with relationship satisfaction at post-test. In the literature, the impact of this variable is ambiguous. Some studies have found an association with the success of couple therapy (Atkins et al., 2005; Kirkpatrick & Davis, 1994), while others did not (Kirkpatrick & Davis, 1994; Levinger, Senn, & Jorgensen, 1970). Furthermore, moderator analysis revealed higher efficacy of couple therapy in Iran and India as compared to the USA and Canada at post-test. We think that this might be an artifact of the low or unknown reliability of two out of the four outcome measures used in the studies conducted in Iran and India. Thus, this result may rather point to methodological differences than to differences in the efficacy itself. More generally, we want to point out that the results of all moderator analyses should not be overstated. Specifically, non-significant results should not be over-interpreted in terms of evidence for no effect. Rather, these results may just be the result of low statistical power, variance restrictions, or confounding variables. Due to reasons of statistical power, we did not analyze moderators for BCT and EFCT separately.

With regard to potential biases in the primary studies, we want to point out that risk of allocation concealment and performance bias were consistently high. Thus, future studies should aim for reducing these biases as far as possible, even though it is not always possible to avoid them completely in psychological therapy. Due to insufficient variation across studies, we could only analyze attrition bias further, for which we could not find a relevant moderating effect on the outcomes of couple therapy.

Lastly, when compared to other meta-analysis, lower effect sizes could also be the result of the *type* of effect sizes emphasized in the present meta-analysis, since we controlled both for differences between experimental and control groups and for pretreatment values. As far as reported, all previous meta-analyses computed effect sizes not controlling for at least one of these two aspects (Barbato & D'Avanzo, 2008; Byrne et al., 2004; Dunn & Schwebel, 1995; Hahlweg & Markman, 1988; Johnson, 1999; Powers et al., 2008; Shadish & Baldwin, 2005; Wood et al., 2005). Combined with our stricter inclusion criteria, this points to the conclusion that efficacy of couple therapy may have been overestimated in the existing literature. We believe our results to be more accurate in this regard.

Limitations and Generalizability of Results

At least four factors limit or complicate the meta-analytic examination of couple therapy. First, the full texts of several studies that were likely to match the inclusion criteria were not accessible despite a detailed search and contacting the authors. Second, in some studies, necessary information to compute effect sizes was missing and efforts to obtain information by contacting authors remained unsuccessful. Third, only studies published in English, German, French or Spanish could be included in the current meta-analysis. Several studies in other languages examined the efficacy of BCT or EFCT and provided an abstract in English, but the full texts could not be translated. Fourth, due to lacking information in the primary studies, we were not able to disentangle systematically whether therapists included certain techniques belonging to another therapy approach or not (e.g., if a randomized trial had a treatment condition that included communication skills, it could be coded as BCT, even if it were cognitive behavioral couple therapy, or psychoeducational couple therapy, or even EFCT). Thus, a more rigorous method is needed for determining if studies are BCT or EFCT.

Furthermore, due to the relatively small sample size of this meta-analysis (33 studies) and small subgroups for many categorical moderators, nonsignificant results do not provide satisfactory evidence for no moderating effects. Instead, such findings may just be the result of low statistical power. Also, we did not control for potential confounding of moderators, as every moderator was analyzed separately in order to not reduce statistical power even further. For these combined reasons, we strongly caution against over-interpretation of the moderator analyses. Moreover, some moderators such as length of therapy sessions, participant ethnicity, or couple relationship

status rates (e.g., separation, divorce), could not be analyzed at all, as too few studies reported the related information. We hope that more studies will report this information in the future.

We decided to only include studies comparing (BCT or EFCT) couple therapy groups to individual therapy groups or nonactive control groups. Comparisons of multiple couple therapy groups we not included in order to ensure a more homogenous and better comparable sample of studies, and also in order not to broaden the present meta-analysis too much. Note that we did not find studies comparing BCT with EFCT, directly.

As to the generalizability of the meta-analysis, there is considerable evidence that several studies with small sample sizes and with small effect sizes were not published. Integrating these studies would eventually lead to more accurate effect sizes and to more information about potential moderators explaining the high level of heterogeneity between studies.

From a practical perspective, long-term gains of couple therapy are of great importance. On the basis of the identified data, it was not possible to identify if EFCT treatment could maintain short-term improvements of relationship satisfaction to longer time periods. A previous meta-analysis reported maintenance of BCT and EFCT gains, but it included nonrandomized studies and non-controlled effect sizes (Byrne et al., 2004). Therefore, there is still a lack of evidence of long-term gains due to EFCT treatment.

Furthermore, this meta-analysis mainly displays results including white, heterosexual North American participants. Also, the majority of studies were based on moderately distressed couples; mildly and severely distressed participants were underrepresented. Almost all samples consisted of highly educated individuals (see Appendix A11). Therefore, multiple unrepresentative sample characteristics limit the generalization of results.

In addition, the most applied outcome measure DAS was introduced in 1976 and revalidated in 1994 (Heyman et al., 1994). It combines the assessment of relationship behaviors (mainly relationship agreement) and relationship satisfaction. Nonetheless, the DAS focuses more on consensus than on satisfaction (Graham, Liu, & Jeziorski, 2006; Heyman et al., 1994). Therefore, it should be noted that most of the results rely on relatively old outcome measures which address partners' agreement (e.g. agreement on child education or agreement on leisure activities) more than on relationship satisfaction.

Finally, most of the studies did not assess individual satisfaction supplementary to relationship satisfaction. Only three studies focused also on personal happiness. All empirically tested couple therapy approaches seem to take a relationship-supporting, antiseperation point of view (Lebow et al., 2012). The only exception are cases of intimate partner violence (Lebow et al., 2012). This attitude may lead to an underrepresentation of individual goals during the therapy, thus possibly preventing deeper individual satisfaction for the sake of maintaining the relationship.

Clinical Implications & Future Research

One of the most important questions in the clinical setting is which therapy works for whom (Wood et al., 2005). The following conclusions can be drawn from this meta-analysis: Both approaches, BCT and EFCT are effective treatments for alleviating relationship distress 6 months after treatment. According to the effect sizes, approx. 73% of the participants in the treatment group showed a higher degree of relationship satisfaction than the average in the control group from baseline to post-test. After 6 months, 66% in the treatment group benefit more than average participants in the control group. After 12 months, participants in both groups showed almost the same level of relationship satisfaction. Eventually, the integration and consolidation of therapy gains in everyday life should play a more important role in the therapy process. Practicing and testing of new interaction patterns could help to maintain achievements. Furthermore, it could be helpful to integrate booster sessions after the end of couple therapy to avoid relapses or to provide immediate sessions in case of an outbreak of a new relationship crisis (Lebow et al., 2012).

In future investigations, the following aspects should be taken into consideration. The present meta-analysis indicates that a publication bias may exist in the data. To provide meaningful and realistic estimates of effects of couple therapy, it is important that journal reviewers and editors do not simply base their evaluation of a study on whether it reports significant findings, but on its overall quality, instead. Furthermore, long-term randomized controlled trials are missing,

especially concerning the efficacy of EFCT. Maintenance of therapy gains is of high importance in the clinical context. Therefore, additional long-term RCTs are needed. Additionally, the current data did not provide satisfying information to generalize the results to different cultural and personal contexts. Consequently, there is a need of more diverse study samples, for example from different places, different cultures and from different educational levels. Another important aspect of future research should be the investigation of comorbid symptoms, which we could not analyze ourselves because of too few primary studies reporting the relevant information. Previous meta-analyses such as a meta-analysis published by Powers et al. (2008) analyzed the impact of couple therapy on psychological disorders. An up-to-date meta-analytic overview of the influence of BCT and EFCT on comorbid symptoms is still required. Moreover, most studies on couple therapy do not assess partners' individual satisfaction. In future research, it would be of high interest to assess individual as well as relationship satisfaction during and after couple therapy.

In light of the limited evidence both in favor of and against the idea that BCT and EFCT show similar effect sizes, it would be interesting to conduct additional dismantling studies to gather more information about differences and overlaps between the two approaches. Although, common factors of psychotherapy such as therapeutic alliance, empathy, and warmth certainly play a role in both approaches (Davis et al., 2012; Lambert & Barley, 2001), it is relevant to uncover which additional factors influence efficacy in couple therapy. It is well possible that future research finds evidence for statistically and clinically relevant differences between BCT and EFCT. At this point, however, it remains unclear which approach is superior and if the treatment of choice might actually vary depending on other variables, for instance, the amount of couple distress. For this purpose, primary studies directly comparing both treatments in a randomized-controlled setting are of great importance.

In the scope of our meta-analysis were only BCT and EFCT, and therefore our focus potentially was too narrow. Excluding all other therapy approaches, promising concepts such as Cognitive Behavioral Couple Therapy (CBCT) have been neglected. Prospectively, meta-analyses should also comprise RCTs on CBCT.

Additionally, further validation of the primary outcome measure DAS is needed. The last validation was conducted in 1994, and it places a strong emphasis on couple's agreement instead of relationship satisfaction. Therefore, further research should be conducted regarding the fit between outcome variable and outcome measure. Finally, the results of this meta-analysis indicate that a considerable amount of participants did not profit from the treatment or did not benefit long enough from BCT or EFCT. Thus, future studies should investigate why some couples do not respond to the treatments and what can be done to remedy that.

CONCLUSION

This meta-analysis adds new evidence on the research of two main approaches in couple therapy including several recently published studies. The present meta-analysis was strictly limited to randomized controlled trials and thus provides an up-to-date overview including only studies that apply the gold standard of experimental design. Results add further evidence to the claim that BCT and EFCT are superior to both no treatment and individual psychotherapy in increasing relationship satisfaction, but there is still a lack of evidence of a gain longer than 6 months.

FUNDING

There was no funding for this study.

NOTE

¹Here and in the remainder of the present paper, in accordance with the conventions of Cohen's *d*, the effect sizes reported can be interpreted as follows: small ($d = 0.2$), medium ($d = 0.5$), and large ($d = 0.8$) effect. .

REFERENCES

- Ahmadi, M., Rasouli, M., Alaf, C. B., & Zadi, F. S. (2014). Effectiveness of emotion-focused couple therapy in marital commitment and emotion regulation. *International Journal of Psychology and Behavioral Research. Special Issue, 1*(1), 54–63.
- Ahmadi, F. S., Zarei, E., & Fallahchai, S. R. (2014). The effectiveness of emotionally-focused couple therapy in resolution of marital conflicts between the couples who visited the consultation centers. *Journal of Educational and Management Studies, 4*(1), 118–123.
- APA. (2007). *Using emotions as a pathway for change*. American Psychological Association. Retrieved November 16, 2015, from <https://www.apa.org/pubs/-videos/4310755.pdf>
- APA. (2008). *MARS Guidelines*: American Psychological Association. Retrieved from <http://www.apastyle.org/manual/related/JARS-MARS.pdf>
- Atkins, D. C., Berns, S. B., George, W. H., Doss, B. D., Gattis, K., & Christensen, A. (2005). Prediction of response to treatment in a randomized clinical trial of marital therapy. *Journal of Consulting and Clinical Psychology, 73*(5), 893–903. <https://doi.org/10.1037/0022-006X.73.5.893>.
- Barbato, A., & D'Avanzo, B. (2008). Efficacy of couple therapy as a treatment for depression: A meta-analysis. *Psychiatric Quarterly, 79*(2), 121–132. <https://doi.org/10.1007/s11126-008-9068-0>.
- Baucom, D. H. (1982). A comparison of behavioral contracting and problem-solving/communications training in behavioral marital therapy. *Behavior Therapy, 13*(2), 162–174. [https://doi.org/10.1016/S0005-7894\(82\)80060-9](https://doi.org/10.1016/S0005-7894(82)80060-9).
- Beach, S. R. H., & Daniel O'Leary, K. (1992). Treating depression in the context of marital discord: Outcome and predictors of response of marital therapy versus cognitive therapy. *Behavior Therapy, 23*(4), 507–528. [https://doi.org/10.1016/S0005-7894\(05\)80219-9](https://doi.org/10.1016/S0005-7894(05)80219-9).
- Byrne, M., Carr, A., & Clark, M. (2004). The efficacy of behavioral couples therapy and emotionally focused therapy for couple distress. *Contemporary Family Therapy, 26*(4), 361–387. <https://doi.org/10.1007/s10591-004-0642-9>.
- Christensen, A., Atkins, D. C., Berns, S., Wheeler, J., Baucom, D. H., & Simpson, L. E. (2004). Traditional versus integrative behavioral couple therapy for significantly and chronically distressed married couples. *Journal of Consulting and Clinical Psychology, 72*(2), 176–191.
- Cloutier, P. F., Manion, I. G., Walker, J. G., & Johnson, S. M. (2002). Emotionally focused interventions for couples with chronically ill children: A 2-year follow-up. *Journal of Marital and Family Therapy, 28*(4), 391–398.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*: L. Erlbaum Associates. Retrieved from <https://books.google.de/books?id=Tl0N2lRAO9oC> [Online; accessed 6-November 2017]
- Dalton, E. J., Greenman, P. S., Classen, C. C., & Johnson, S. M. (2013). Nurturing connections in the aftermath of childhood trauma: A randomized controlled trial of emotionally focused couple therapy for female survivors of childhood abuse. *Couple and Family Psychology: Research and Practice, 2*(3), 209–221. <https://doi.org/10.1037/a0032772>.
- Davis, S. D., Lebow, J. L., & Sprenkle, D. H. (2012). Common factors of change in couple therapy. *Behavior Therapy, 43*(1), 36–48. <https://doi.org/10.1016/j.beth.2011.01.009>.
- Denton, W. H., Wittenborn, A. K., & Golden, R. N. (2012). Augmenting antidepressant medication treatment of depressed women with emotionally focused therapy for couples: A randomized pilot study. *Journal of Marital and Family Therapy, 38*(Suppl 1), 23–38. <https://doi.org/10.1111/j.1752-0606.2012.00291.x>.
- Dunn, R. L., & Schwebel, A. I. (1995). Meta-analytic review of marital therapy outcome research. *Journal of Family Psychology, 9*(1), 58–68. <https://doi.org/10.1037/0893-3200.9.1.58>.
- Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56*(2), 455–463.
- Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ, 315*(7109), 629–634.
- Fals-Stewart, W., Yates, B. T., & Klostermann, K. (2005). Assessing the costs, benefits, cost-benefit ratio, and cost-effectiveness of marital and family treatments: Why we should and how we can. *Journal of Family Psychology, 19*(1), 28–39.
- Graham, J. M., Liu, Y. J., & Jeziorski, J. L. (2006). The dyadic adjustment scale: A reliability generalization meta-analysis. *Journal of Marriage and Family, 68*(3), 701–717. <https://doi.org/10.1111/j.1741-3737.2006.00284.x>.
- Greenberg, L. S., & Johnson, S. M. (1988). *Emotionally focused therapy for couples*. New York: Guilford Press.
- Greenberg, L. J., Warwar, S. H., & Malcolm, W. M. (2008). Differential effects of emotion-focused therapy and psychoeducation in facilitating forgiveness and letting go of emotional injuries. *Journal of Counseling Psychology, 55*(2), 185–196. <https://doi.org/10.1037/0022-0167.55.2.185>.
- Grych, J. H., & Fincham, F. D. (1990). Marital conflict and children's adjustment: A cognitive-contextual framework. *Psychological Bulletin, 108*(2), 267–290.
- Gurman, A.S., Lebow, J.L., & Snyder, D.K. (2015). *Clinical handbook of couple therapy* (5th edn). New York, NY: Guilford Press.

- Hahlweg, K., & Markman, H. J. (1988). Effectiveness of behavioral marital therapy: Empirical status of behavioral techniques in preventing and alleviating marital distress. *Journal of Consulting and Clinical Psychology, 56*(3), 440–447. <https://doi.org/10.1037/0022-006X.56.3.440>.
- Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods, 1*(1), 39–65. <https://doi.org/10.1002/jrsm.5>.
- Heyman, R. E., Sayers, S. L., & Bellack, A. S. (1994). Global marital satisfaction versus marital adjustment: An empirical comparison of three measures. *Journal of Family Psychology, 8*(4), 432–446. <https://doi.org/10.1037/0893-3200.8.4.432>.
- Higgins, J. P. T., & Green, S. (2011). *Cochrane handbook for systematic reviews of interventions: Version 5.1.0 [update March 2011]*: The Cochrane Collaboration, 2011. Retrieved from <http://handbook.cochrane.org/> [Online; accessed 6-November 2017]
- Jacobson, N. S., & Margolin, G. (1979). *Marital therapy: Strategies based on social learning and behavior exchange principles*. Levittown, PA: Brunner/Mazel.
- Jacobson, N. S., Schmalings, K. B., & Holtzworth-Munroe, A. (1987). Component analysis of behavioral marital therapy: 2-year follow-up and prediction of relapse. *Journal of Marital and Family Therapy, 13*(2), 187–195. <https://doi.org/10.1111/j.1752-0606.1987.tb00696.x>.
- Johnson, S. M. (1999). Emotionally focused couples therapy: Status and challenges. *Clinical Psychology: Science and Practice, 6*(1), 67–69. <https://doi.org/10.1093/clipsy/6.1.67>.
- Johnson, S. M., & Greenberg, L. S. (1985). Emotionally focused couples therapy; An outcome study. *Journal of Marital and Family Therapy, 11*(3), 313–317. <https://doi.org/10.1111/j.1752-0606.1985.tb00624.x>.
- Johnson, S. M., Hunsley, J., Greenberg, L., & Schindler, D. (2006). Emotionally focused couples therapy: Status and challenges. *Clinical Psychology: Science and Practice, 6*(1), 67–79. <https://doi.org/10.1093/clipsy.6.1.67>.
- Kavitha, C., Rangan, U., & Nirmalan, P. K. (2014). Quality of life and marital adjustment after cognitive behavioural therapy and behavioural marital therapy in couples with anxiety disorders. *Journal of Clinical and Diagnostic Research: JCDR, 8*(8), WC01.
- Kirkpatrick, L. A., & Davis, K. E. (1994). Attachment style, gender, and relationship stability: A longitudinal analysis. *Journal of Personality and Social Psychology, 66*(3), 502–512. <https://doi.org/10.1037/0022-3514.66.3.502>.
- Lam, W. K. K., Fals-Stewart, W., & Kelley, M. L. (2009). Parent training with behavioral couples therapy for fathers' alcohol abuse: Effects on substance use, parental relationship, parenting, and cps involvement. *Child Maltreatment, 14*(3), 243–254. <https://doi.org/10.1177/1077559509334091>.
- Lambert, M. J., & Barley, D. E. (2001). Research summary on the therapeutic relationship and psychotherapy outcome. *Psychotherapy, 38*(4), 357–361. <https://doi.org/10.1037/0033-3204.38.4.357>.
- Lambert, M. J., & Hawkins, E. J. (2001). Using information about patient progress in supervision: Are outcomes enhanced? *Australian Psychologist, 36*(2), 131–138. <https://doi.org/10.1080/00050060108259645>.
- Lebow, J. L., Chambers, A. L., Christensen, A., & Johnson, S. M. (2012). Research on the treatment of couple distress. *Journal of Marital and Family Therapy, 38*(1), 145–168.
- Levinger, G., Senn, D. J., & Jorgensen, B. W. (1970). Progress toward permanence in courtship: A test of the Kerckhoff-Davis hypotheses. *Sociometry, 33*(4), 427–443. <https://doi.org/10.2307/2786317>
- McCrary, B. S., Epstein, E. E., Cook, S., Jensen, N., & Hildebrandt, T. (2009). A randomized trial of individual and couple behavioral alcohol treatment for women. *Journal of Consulting and Clinical Psychology, 77*(2), 243–256. <https://doi.org/10.1037/a0014686>.
- Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G.; PRISMA Group (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine, 6*(7), e1000097.
- Organization for Economic Cooperation and Development. (2017). *OECD Family Database*. Retrieved from <http://www.oecd.org/els/family/database.htm> [Online; accessed 6-November 2017]
- Powers, M. B., Vedel, E., & Emmelkamp, P. M. G. (2008). Behavioral couples therapy (BCT) for alcohol and drug use disorders: A meta-analysis. *Clinical Psychology Review, 28*(6), 952–962. <https://doi.org/10.1016/j.cpr.2008.02.002>.
- R Core Team. (2016). R: A Language and Environment for Statistical Computing. Retrieved from <https://cran.r-project.org>
- Shadish, W. R., & Baldwin, S. A. (2005). Effects of behavioral marital therapy: A meta-analysis of randomized controlled trials. *Journal of Consulting and Clinical Psychology, 73*(1), 6–14.
- Sims, A. E. B. (1999). Working models of attachment: The impact of emotionally focused marital therapy (Ph.D. Thesis). University of Ottawa.
- Spanier, G. B. (1976). Measuring dyadic adjustment: New scales for assessing the quality of marriage and similar dyads. *Journal of Marriage and Family, 38*(1), 15–28. <https://doi.org/10.2307/350547>.
- Steinglass, P. (1978). The conceptualization of marriage from a systems theory perspective. In T. J. Paolino & B. S. McCrary (Eds.), *Marriage and marital therapy: Psychoanalytic, behavioral, and systems theory perspectives* (pp. 298–365). Levittown, PA: Brunner/Mazel.

- Uebelacker, L. A., & Whisman, M. A. (2006). Moderators of the association between relationship discord and major depression in a national population-based sample. *Journal of Family Psychology, 20*(1), 40–46. <https://doi.org/10.1037/0893-3200.20.1.40>.
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*(3), 1–48. Retrieved from <http://www.jstatsoft.org/v36/i03/> [Online; accessed 6-November 2017]
- Whisman, M. A., & Bruce, M. L. (1999). Marital dissatisfaction and incidence of major depressive episode in a community sample. *Journal of Abnormal Psychology, 108*(4), 674–678. <https://doi.org/10.1037/0021-843X.108.4.674>.
- Wood, N. D., Crane, D. R., Schaalje, G. B., & Law, D. D. (2005). What works for whom: A meta-analytic review of marital and couples therapy in reference to marital distress. *American Journal of Family Therapy, 33*(4), 273–287. <https://doi.org/10.1080/01926180590962147>.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

- Appendix A1.** Hypotheses for moderators.
- Appendix A2.** The Cochrane Collaboration’s tool for assessing risk of bias.
- Appendix A3.** Complete search terms.
- Appendix A4.** Coded variables.
- Appendix A5.** Coding scheme.
- Appendix A6.** Combination subgroup scores.
- Appendix A7.** Computation standardized mean differences.
- Appendix A8.** Equivalent of Study Title and Study-ID.
- Appendix A9.** List of Non-Accessible Studies (full text).
- Appendix A10.** List of all included studies.
- Appendix A11.** Study Characteristics.
- Appendix A12.** Methodological Characteristics.
- Appendix A13.** Calculations.