American Psychological Association

Journal of Consulting and Clinical Psychology

ISSN: 0022-006X

2020, Vol. 88, No. 7, 583–596
http://dx.doi.org/10.1037/ccp0000514

Meta-Analysis of Couple Therapy: Effects Across Outcomes, Designs, Timeframes, and Other Moderators

McKenzie K. Roddy, Lucia M. Walsh, Karen Rothman, S. Gabe Hatch, and Brian D. Doss
University of Miami

Objectives: This study updated existing meta-analyses of couple therapy that typically do not include multiple treatment modalities, various research designs, long-term outcomes, or recent studies. ***Eligibility Criteria:*** Studies published in English that reported relationship satisfaction or other outcomes of couple therapy were included; over 70% of studies have not been included in previous meta-analyses. ***Methods of Synthesis:*** Using random effects models across 58 studies representing 40 unique samples and 2,092 couples, effect sizes were summarized within measure domains as mean gains for treatment groups and waitlist groups as well as between-groups comparisons. ***Results:*** Couple therapy has a large effect on relationship satisfaction (pre to post within-group Hedges $\bar{g} = 1.12$, CI [0.92, 1.31], $p < .001$) and couples assigned to waitlists do not significantly improve (pre-to-post within-group satisfaction Hedges $\bar{g} = 0.12$, CI [−0.04, 0.29], $p > .05$). Additionally, couple therapy has significant impacts on key domains including self-reported and observed communication, emotional intimacy, and partner behaviors. Moderation analyses of pre-to-post gains in relationship satisfaction for treatment groups were generally nonsignificant; however, greater baseline distress was associated with larger gains. ***Conclusions:*** Couple therapy has large effects on key relationship domains and gains are generally maintained over short- and long-term follow-up with minimal impact of tested moderators. Limitations include sample of exclusively opposite sex couples and inability to fully model dependencies within studies. The relationship between mean gain effect sizes and between-groups comparisons is discussed with implications for future research.

What is the public health significance of this article?
Couple therapy positively impacts multiple domains of relationship functioning (e.g., satisfaction, communication) during treatment, and gains remain evident at short-term and long-term follow-up. Couples assigned to waitlist control groups generally do not improve. Couple therapy's effects on relationship satisfaction are consistent across individual, couple, and study characteristics.

Keywords: meta-analysis, couple therapy, waitlist, relationship distress, moderation

Supplemental materials: http://dx.doi.org/10.1037/ccp0000514.supp

Relationship distress is highly prevalent among couples in the United States, with as many as one third of couples endorsing dissatisfaction with their current partner (Whisman, Beach, & Snyder, 2008). Lower levels of relationship satisfaction are associated with higher likelihood of divorce (Lavner & Bradbury, 2010) and poorer physical and mental health outcomes (Robles, Slatcher, Trombello, & McGinn, 2014; Slatcher & Selcuk, 2017). While previous meta-analyses have demonstrated that couple therapy is an effective treatment for relationship distress, there are a number of questions that remain unanswered.

Unfortunately, our current knowledge of the effectiveness of couple therapy is constrained by a number of important limitations of the existing literature. First, existing meta-analyses are not comprehensive and only include a subset of the broader field. Indeed, behavioral marital therapy (BMT) is the most frequently studied couple therapy modality and one of the few to be consistently featured in meta-analytic reviews (Shadish & Baldwin, 2005). While some researchers have begun to explore the effects of more recently developed treatment modalities such as cognitive–behavioral couple therapy (CBCT; Fischer, Baucom, & Cohen, 2016) and emotionally focused couple therapy (EFCT; Rathgeber, Bürkner, Schiller, & Holling, 2019; Wiebe & Johnson, 2016), existing meta-analyses fail to com-

McKenzie K. Roddy, Lucia M. Walsh, Karen Rothman, S. Gabe Hatch, and Brian D. Doss, Department of Psychology, University of Miami.

There are no funding or conflicts of interest to report. McKenzie K. Roddy completed this work at the University of Miami but has since moved to Baylor College of Medicine. This study was registered with Open Science Framework: https://osf.io/j8y6h/?view_only=a7d38de0197f426984c24c7649bd9f27.

Correspondence concerning this article should be addressed to McKenzie K. Roddy, who is now at Menninger Department of Psychiatry and Behavioral Sciences, Baylor College of Medicine, 1977 Butler Boulevard, Houston, TX 77030. E-mail: mkr19@miami.edu

bine across approaches to get an overall estimate of the effectiveness of couple therapy.

Furthermore, the most common reference for an overall summary of effectiveness of couple therapy (Shadish & Baldwin, 2003) is a review—not a meta-analysis—summarizing previous meta-analyses, resulting in the commonly cited mean effect size of 0.84. However, this average does not take into account the fact that individual studies were included in more than one meta-analysis nor does it consider that the various meta-analyses had different numbers of comparisons using different sample sizes—which should be weighted accordingly. To address these limitations, the current study purposely casts a wide net to encompass couple therapy from a variety of theoretical modalities, treatment lengths, and settings.

A second important limitation of the existing literature is that previous meta-analyses have focused exclusively on comparisons to a no-treatment or waiting list control treatment (Dunn & Schwebel, 1995; Hahlweg & Markman, 1988; Shadish & Baldwin, 2005). Although this study design provides more confidence in the magnitude of the intervention effects, many studies of couple therapy conducted in the past 20 years have omitted a control group given results from a meta-analysis of couple therapy control groups indicating that those couples do not significantly improve over time (Baucom, Hahlweg, & Kuschel, 2003). Instead, modern studies of couple therapy rely on within-group changes over time—either in a single group (e.g., Doss et al., 2012) or with couples randomized to two or more active interventions (e.g., Christensen et al., 2004).

A third limitation of the existing meta-analyses on couple therapy is that they do not typically include an estimate of the effect of couple therapy over follow-up. Therefore, the current study includes estimates of both between-groups changes (compared with a no-treatment or waitlist control) and within-group changes. Fourth, previous meta-analyses have only included efficacy studies of couple therapy, omitting effectiveness studies of couple therapy conducted in real-world settings. This limitation likely serves to overestimate and upwardly bias effect sizes yielded from previous meta-analyses, given that effectiveness studies generally yield smaller effect sizes compared to efficacy studies (Doss et al., 2012).

A fifth limitation of previous meta-analyses of couple therapy is that they often omit assessment of moderation effects (e.g., Hahlweg & Markman, 1988; Plattor, 1990) or lack necessary variability to test moderators of interest (e.g., Wood, Crane, Schaalje, & Law, 2005). The field has, however, also seen successful implementation of moderation effects. Shadish and Baldwin (2005) tested moderation effects of clinical representativeness and found that BMT outcomes do not statistically differ between nonclinical and clinical samples. Additionally, Dunn and Schwebel (1995) tested for moderation by treatment orientation comparing 21 treatment groups that fell within the following domains: BMT, cognitive–behavioral marital therapy (CBMT), insight-oriented marital therapy (IOMT), or waitlist control. Results indicated that all treatment groups, relative to control couples, improved significantly from pre-to-post treatment in positive and negative relationship behaviors (weighted mean ES $d = 0.54$ to $d = 0.87$), with no significant between-groups differences. However, there were significant between-groups differences in relationship quality from pre-to-post evaluation such that IOMT was more effective in improving relationship quality than BMT and CBMT.

Finally, it been over a decade since a more comprehensive meta-analysis of multiple therapy modalities has been published (Shadish & Baldwin, 2005; Wood et al., 2005). Although narrowly focused meta-analyses of emotionally focused couple therapy (EFCT) have been conducted in the past few years (Rathgeber et al., 2019; Wiebe & Johnson, 2016), these meta-analyses have not included recent developments in third-wave behavior therapy such as integrative behavioral couple therapy (IBCT; Christensen et al., 2004) or enhanced cognitive-behavioral couple therapy (ECBT; Epstein, Baucom, & Baucom, 2002).

Current Study

To build on previous meta-analyses, update the existing literature, and account for the limitations of previous work, we conducted a meta-analysis of couple therapy in adults. The current meta-analysis includes the use of follow-up data, a large variety of theoretical orientations, and a diversity of study designs (single group designs, nonrandomized or quasi-experimental trials, randomized controlled trials, as well as studies without waitlist comparison groups). Additionally, most prior meta-analyses have collapsed across measurement types (e.g., self-report measures and observationally coded tasks); here we keep these distinct. This meta-analysis aimed to summarize (Aim 1) mean gains for treatment groups during the intervention period, (Aim 2) mean gains for waitlist groups during the intervention period, (Aim 3) between-groups comparisons during the intervention period, and (Aim 4) mean gains for treatment groups from pretreatment to follow-up. Second, (Aim 5) this meta-analysis examined moderators of effects of relationship satisfaction. The full protocol code book is available on Open Science Framework.

Method

This study was registered with Open Science Framework (OSF; Foster & Deardorff, 2017): https://osf.io/j8y6h/?view_only=a7d38de0197f426984c24c7649bd9f27.

Search Procedures

We conducted an exhaustive search for articles on couple therapy, including both published and unpublished literature using *PsycINFO, Web of Knowledge,* and *Dissertation & Theses Global* databases. We used a combination of keywords to identify relevant studies including but not limited to: "couple therapy," "marital therapy," "couple treatment," "marriage counseling"; see Supplementary Table 1 for full list of search terms. Wildcard terms were used to allow for results to indicate all possible variations of a word (e.g., couple* to include couple, couples, couple's). In addition to the database searches, we identified additional appropriate articles form past reviews and meta-analyses and searched the reference lists of relevant studies (Dunn & Schwebel, 1995; Shadish & Baldwin, 2003; Shadish & Baldwin, 2005; Wood et al., 2005). All citations were exported and saved into an EndNote ×8 database and analyzed for duplicate references.

Articles were included if they: (a) were written in English; (b) examined treatment outcomes of couple therapy seeking to ame-

liorate relationship distress; (c) reported data on both members of the dyad in a distressed relationship; and (d) reported data on at least one domain including relationship satisfaction, emotional intimacy, cognitions about the relationship, communication, violence (both physical and psychological), relationship behaviors (both self and partner), and relationship status (e.g., divorce/separation as an outcome). Studies were excluded if: (a) the treatment was focused on an individual's problem rather than a relationship problem (e.g., depression); (b) the treatment(s) were majority self-help or relationship education/prevention, per the author's description of the intervention or selection of nondistressed couples; (c) the treatment was delivered in individual or group format; (d) the sample was not adults (i.e., <18 years); and/or (e) the sample was smaller than 10 couples (20 individuals). Studies that met all other inclusion criteria but did not include necessary statistical information to compute effect sizes were initially included and study authors were contacted for this information. Studies that compared group delivery or individual versus couple delivery were included and coded for the couple delivery, but not group or individual delivery. We included studies published in any year. Additionally, there were no exclusion criteria regarding randomization; therefore, randomized controlled trials, quasi-experimental, and single group design studies were included provided they met all other inclusion requirements.

The search process initially identified 6,418 articles. We removed 1,340 duplicates and 4,961 studies for relevance in the first screen of titles and abstracts. Studies were removed for several reasons, including only treating one member of the dyad, therapeutic focus other than relationship satisfaction/adjustment, nonexperimental article, or lack of reported outcomes. We conducted a detailed review of the remaining 117 articles.

To ensure consistent inclusion and exclusion of screened studies, the first author and two trained undergraduate research assistants coded the remaining 117 articles as "included" or "excluded" as well as the specific reason they were excluded (i.e., treatment, design, or client factors) by reviewing the full text. To examine interrater reliability, kappas were calculated between the two students. Interrater reliability ranged from 0.77 to 0.88, above the "acceptable" threshold of 0.70 (McHugh, 2012). After reliability was met, the two students divided the remaining articles in half and coded them independently, and the first author reviewed all their notes and coding to ensure appropriate inclusion/exclusion and avoid coder drift. Fifty-seven articles were excluded, leaving 60 articles to review and code. Two studies that met inclusion criteria were not included (Denton, Burleson, Clark, Rodriguez, & Hobbs, 2000; O'Leary & Turkewitz, 1981) because they did not provide enough information to calculate effect sizes and efforts to obtain data from the authors were not successful. If multiple publications reported on the same sample, all articles were included and coded simultaneously to retrieve complete information on the study. The final sample consisted of 58 articles representing 40 unique samples; the full data set is available from the first author (see Figure 1).

Study and Effect Size Coding

We coded included articles to calculate effect sizes and characterize their design. Study coding was conducted by four doctoral student coders with experience in statistics and/or meta-analyses. Codes for study design were continuous (average baseline satisfaction [converted to z-scores using published norms for analyses as studies used different measures of satisfaction], year of publication, average age of participants, percentage of race/ethnicity of participants, mean number of sessions, total number of couples, percent male, and percent married) as well as categorical (e.g., randomization, treatment orientation, peer-reviewed publication status, and clinical representativeness). Clinical representativeness codes were based on the Clinical Representativeness coding system (Shadish, Matt, Navarro, & Phillips, 2000) and summed to a total score. Coders classified the measure type (e.g., self-report, behavioral tasks) as well as the domain of the measure (e.g., relationship satisfaction, communication, emotional intimacy). We initially included the following measure domains based on an informal review of the couple therapy literature: satisfaction, emotional intimacy, cognitions about the relationship, communication, violence (both physical and psychological), relationship behaviors (both self and partner), and relationship status (e.g., divorce/separation as an outcome). We excluded violence, self-relationship behaviors, and relationship status from analyses because too few studies included them as outcomes. The full codebook is available on OSF.

Coders computed mean gain effect sizes and standard errors using the Practical Meta-Analysis Effect Size Calculator (Wilson, 2001) from pre- posttreatment, pre to short-term follow-up (= < 6 months), and pre to long-term follow-up (>6 months to 2 years). Additionally, coders computed between-groups mean gains of treatment versus control during the intervention period for study designs that included a control group. All effect sizes were calculated such that a positive effect represents the desired treatment effect (i.e., both an increase in satisfaction as well as a decrease in conflict would result in a positive effect size). When data were presented separately by gender, effect sizes were calculated separately for husbands and wives then combined to a couple level for analyses. Effect sizes were windsorized within each measure domain to ±3 standard deviations from the mean.

Coders simultaneously coded 30% of articles to establish interrater reliability. Disagreements between coders were discussed with all four coders and brought to the senior author for a final decision in the uncommon event coders could not reach consensus within the group. Reliability for categorical codes (e.g., measure domain) was calculated with Cohen's kappa and reliability for continuous codes (e.g., number of sessions) was calculated with intraclass correlation coefficients (ICC). Reliability ranged from .81 to 1, above the .70 cutoff typically considered acceptable (McHugh, 2012). Once coders met reliability, articles were coded by pairs of coders and disagreements discussed with all coders. All effect sizes were calculated as Cohen's d and adjusted using Hedges' g to attain an unbiased estimate (Hedges & Olkin, 1985).

Analyses

All analyses were run using the *metafor* package in R v3.6.2 (Viechtbauer, 2010); code available on OSF. We decided a priori to use a random-effect model, allowing the true effect size to vary from study to study, due to the heterogeneity in methods, client sample, and treatment modalities included in this meta-analysis (Pigott, 2012). We calculated the overall effect size of each measure domain for mean gains during intervention for treatment

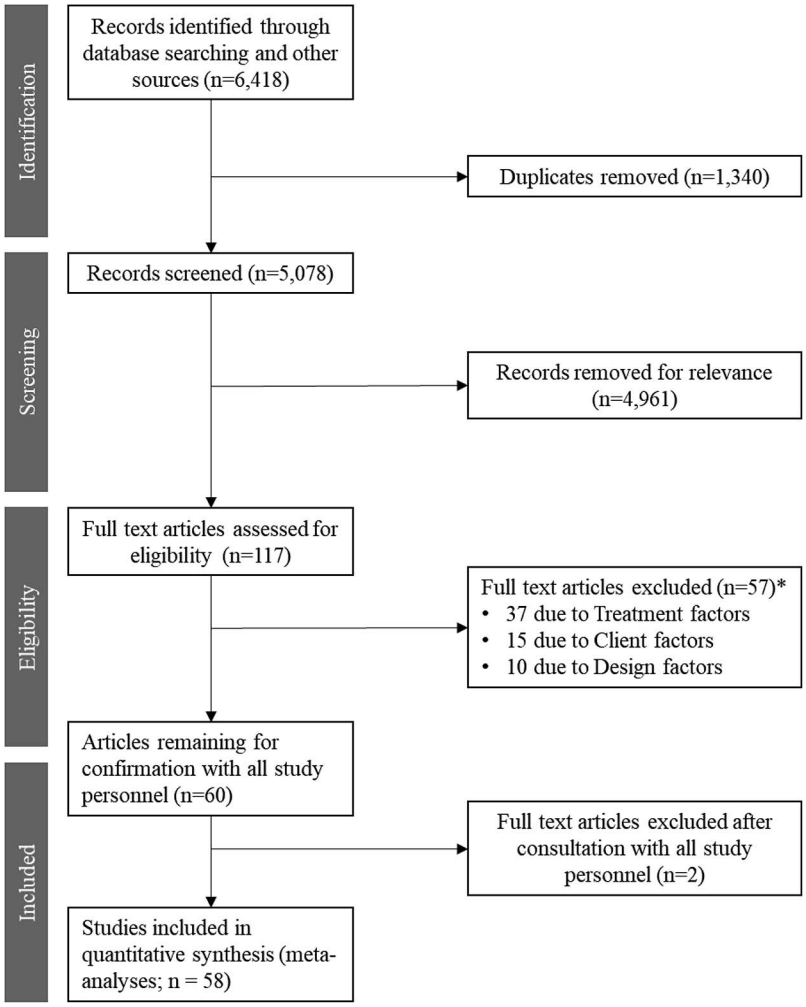

Figure 1. Study selection flow diagram. * Articles could be excluded for multiple reasons therefore numbers do not sum.

groups, mean gains during waiting period, between-groups comparisons during intervention, and mean gains pre to follow-up, with each effect size weighted by the inverse of that independent sample's variance. Moderation analyses were conducted using mixed effects models. Continuous moderators were centered prior to analysis and categorical moderators were dummy coded with the "zero group" as the intercept.

We explored the possibility of evidence of publication bias in this meta-analysis using funnel plots (Greenhouse & Inyengar, 2009). Interpreting funnel plot symmetry can be subjective (Sutton, 2009), so we also calculated Egger's linear regression test using the "regtest" function of the *metafor* R package. We calculated a fail-safe *N*, a calculation that estimates the number of studies needed for the *p* value to become insignificant; a test is considered robust if $N > 10 + 5k$ when k is the original number of studies (Rosenberg, 2005). In subgroup meta-analyses where there were less than 10 effect sizes, we did not create funnel plots or use Egger's regression test, as the power of the regression test can be too low to distinguish chance from true asymmetry (Higgins & Green, 2011).

Handling Dependency

Most independent samples included multiple measures, and therefore yielded multiple effect sizes. For all analyses, we chose to use the Borenstein, Hedges, Higgins, and Rothstein (2009; BHHR) method of aggregation, as it is the univariate method found to be least biased and most precise in large simulation studies of such methods (Del Re, 2015; Scammacca, Roberts, & Stuebing, 2014; Wei & Higgins, 2013). Specifically, we aggregated effect sizes so there would only be one effect size per measure domain per independent sample in each analysis. We aggregated all outcome measures using the *MAd* package in R, which averages all within-study effect sizes and variances, considering the correlations among the within-study outcome measures consistent with BHHR procedures (Cooper, Hedges, & Valentine, 2009). We kept the default correlation for between effect sizes at $r = .70$ (Wampold et al., 1997) when we aggregated all outcome measures, due to the high correlation between outcomes in couple therapy (e.g., Christensen et al., 2004).

Results

Final Study Sample for Meta-Analytic Coding

The 40 independent samples (from 58 articles; noted with * in the reference list) ranged in publication date from 1977 to 2017. More than 70% of included studies had not been included in any previous meta-analyses of couple therapy; 13 of these articles were included in previous meta-analyses (noted with † in reference list). Analyses yielded 517 effect sizes due to multiple outcomes reported per independent sample, representing 76 independent treatment groups totaling 4,184 clients (2,092 couples). Due to small sample sizes within measure domains, only within-group effect sizes (not between-groups effect sizes) over follow-up were calculated.

The client sample was on average 38-years-old (10% of studies failed to report clients' age). All couples were in different-sex relationships. Of the 21 studies (52.5%) that reported racial information, most couples were Caucasian (87.18%) with fewer African American/Black clients (5.96%), Asian (2.65%), Native American (0.03%), or Other (0.94%). Of the one quarter of studies that reported ethnicity, a small portion of clients identified as Hispanic (2.74%). Three quarters of studies reported relationship status at baseline; of those, 91.10% of couples were married. About 40% of studies reported lengths of treatment and/or waitlist; for those studies, the average amount of time between pre- and posttreatment was 12 weeks. Most studies (80%) reported a measure of relationship satisfaction that has established norms. Using community norms (Funk & Rogge, 2007), baseline levels of relationship satisfaction were converted into z-scores; the mean z-score for the sample was −1.26. This z-score corresponds to a raw score of 86.8 on the 32-item Dyadic Adjustment Scale (Spanier, 1976), which is below an established cutoff for relationship distress (97.5; Funk & Rogge, 2007).

We identified 15 negative pre-to-post treatment effects of the 269 effect sizes that were calculated, representing 13 treatment groups from 10 different studies. The majority of the negative outcomes identified ($n = 12$) were from measures of observed communication. Of effect sizes that were at least small in magnitude (Hedges' $g > .20$; $n = 7$), the original authors reported that one was statistically significant, four were nonsignificant, and two were not tested.

After aggregating effect sizes within measure domains across time points, we screened for outliers. One outlier (>3 *SD* from mean) was identified and, upon further investigation, it was determined the study had a very small sample standard deviation which resulted in a very large effect size for relationship satisfaction (Bennun, 1985). The funnel plot went from significantly asymmetrical to nonsignificant when removed. We therefore dropped this datapoint from all analyses as it was having an undue influence on results (Lipsey & Wilson, 2001). Results including this datapoint are available from the authors.

Aim 1: Mean Gains for Treatment Groups

Full results for within-treatment group pre-to-post mean gains are presented in Table 1. All measure domains consist of self-report measures except for observed communication. Couples in active couple therapy conditions reported large, positive changes in key target areas such as global relationship satisfaction (Hedge's $\bar{g} = 1.12$; Figure 2), emotional intimacy (Hedge's $\bar{g} = 1.48$), relationship cognitions (Hedge's $\bar{g} = 0.81$), self-reported communication (Hedge's $\bar{g} = 1.62$), and partner behaviors (Hedge's $\bar{g} = 1.15$). The effect size for observed communication measures was more modest (Hedges' $\bar{g} = 0.40$). Although some measure domains other than relationship satisfaction had funnel plots that were significantly asymmetrical per Egger's regression test, the fail-safe *N*s were all sufficiently large (all >700; Rosenberg, 2005).

Aim 2: Mean Gains for Waitlist Control Groups

During a waiting period, couples did not experience significant change in global relationship satisfaction (Hedge's $\bar{g} = 0.12$),

Table 1
Random Effects Models at Posttreatment of Within-Group Effect Sizes

Outcome	k	Hedge's $\bar{g}$	95% CI		Failsafe *N*	Regression test for funnel plot asymmetry		Q_{within}	I^2(%)
			Lower	Upper		Mixed-effects meta-regression model	Weighted regression with multiplicative dispersion		
Within-treatment groups									
Relationship satisfaction	59	1.12***	0.92	1.31	55,164	z = .78	$t(57) = 1.81$	1275.62***	96.13
Emotional intimacy	17	1.48***	0.89	2.07	10,806	z = .64	$t(15) = 1.28$	875.03***	98.94
Relationship cognitions	10	0.81***	0.43	1.19	701	z = 3.68***	$t(8) = 3.25$*	100.36***	93.85
Self-reported communication	15	1.62***	1.10	2.13	8,383	z = 1.80	$t(13) = 2.85$*	630.27***	98.37
Observed communication	23	0.40***	0.27	0.54	899	z = .87	$t(21) = 1.58$	75.18***	70.88
Partner behaviors	21	1.15***	0.89	1.42	10,241	z = 5.77***	$t(19) = 5.86$***	411.53***	96.21
Within-waitlist groups									
Relationship satisfaction	14	0.12	−0.04	0.29	104	z = −1.88	$t(12) = -3.78$**	140.25***	81.87
Emotional intimacy	5	0.09	−0.27	0.45	8			46.27***	92.48
Observed communication	6	−0.11	−0.41	0.20	0			20.64***	73.37
Partner behaviors	6	0.36***	0.28	0.46	158			5.91	15.01

Note. k = unique treatment groups. Within-waitlist relationship cognitions and self-reported communication had fewer than five unique treatment groups and were omitted from analyses pre to post. Funnel plot asymmetry was not tested for analyses with fewer than 10 studies due to power issues.
* $p < .05$. *** $p < .001$.

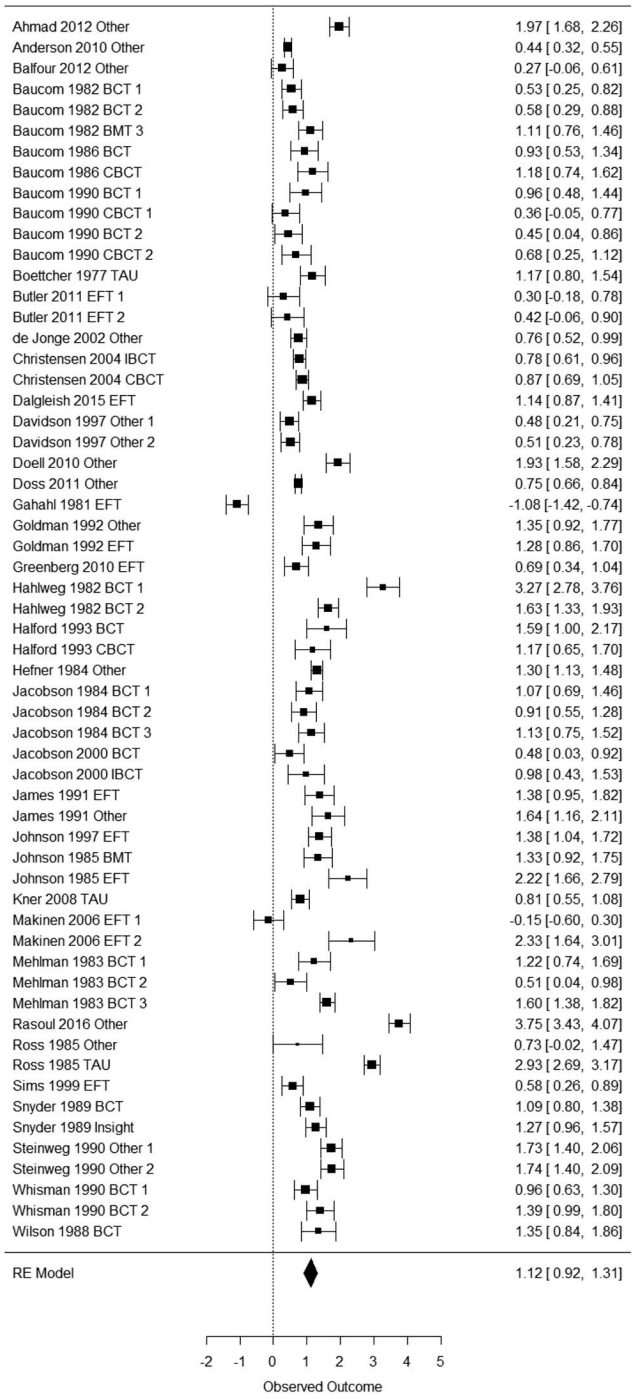

Figure 2. Forest plot of within treatment mean gains in relationship satisfaction.

emotional intimacy (Hedge's $\bar{g} = 0.09$), or observed communication (Hedge's $\bar{g} = -0.11$). On average, couples reported small but significant improvements in relationship behaviors during waiting periods (Hedge's $\bar{g} = 0.36$; Table 1). There were too few studies reporting measures of relationship cognitions and self-reported communication to reliably analyze pre-to-post changes within waitlist groups; see Supplementary Table 2 for a descriptive summary of these data.

Aim 3: Between-Group Effects

The 14 studies that randomized couples to treatment or waitlist groups were coded for between-groups effect sizes comparing the active treatment group(s) to the waitlist control group in addition to inclusion in the analyses above (see Supplementary Tables 3 and 4 for description of measures coded from each group and study). Compared with a waitlist control group, couple therapy had a large, significant impact on global relationship satisfaction (Hedge's $\bar{g} = 0.91$) and a medium, significant impact on self-reported communication (Hedge's $\bar{g} = 0.76$), observed communication (Hedge's $\bar{g} = 0.57$), and partner behaviors (Hedge's $\bar{g} = 0.60$). The between-groups effects on emotional intimacy (Hedge's $\bar{g} = 0.39$) and relationship cognitions (Hedge's $\bar{g} = 0.35$) were small but also statistically significant (see Table 2).

Aim 4: Mean Gains for Treatment Groups Over Follow-Up

Over short-term follow-up ($< = 6$ months), participants continued to report large gains from their pretreatment levels of global relationship satisfaction (Hedge's $\bar{g} = 0.89$), emotional intimacy (Hedge's $\bar{g} = 0.91$), self-reported communication (Hedge's $\bar{g} = 1.97$), and partner behaviors (Hedge's $\bar{g} = 1.31$). In contrast, initial improvements in observed communication observed at the end of treatment were no longer significant over short-term follow-up (see Table 3). There were too few studies reporting measures of relationship cognitions to analyze over short-term follow-up; see Supplementary Table 2 for summary of data not included in meta-analyses.

Only five samples in this meta-analysis reported follow-ups between 6 months and 2 years in length, resulting in 10 unique treatment groups. There was a large significant effect size from pretreatment to long-term follow-up for relationship satisfaction (Hedge's $\bar{g} = 0.91$; Table 3). All other measure domains had too few studies to complete meta-analytic analyses; however, a descriptive summary can be found in Supplementary Table 2.

Aim 5: Moderation Analyses

As there was significant heterogeneity in the effect sizes of relationship satisfaction for treatment groups mean gain pre-to-post, $Q(58) = 1275.62$, $p < .001$, we conducted moderation analyses on these results. We decided to only examine moderation of prepost gains in relationship satisfaction for treatment groups because it was the most commonly included outcome across studies, describes a global measure of relationship functioning rather than a specific domain (e.g., communication or emotional intimacy), and limited the number of atheoretical tests to be conducted.

Study characteristics. Peer-reviewed publication status, clinical representativeness, number of couples per study, randomization, and year of publication were not significant moderators of changes in relationship satisfaction.

Participant characteristics. Percent of the sample that was married and average age of the sample were not moderators of

Table 2
Random Effects Models at Posttreatment Between Treatment and Waitlist Groups

Outcome	k	Hedge's $\bar{g}$	95% CI Lower	95% CI Upper	Failsafe *N*	Regression test for funnel plot asymmetry: Mixed-effects meta-regression model	Regression test for funnel plot asymmetry: Weighted regression with multiplicative dispersion	Q_{within}	I^2(%)
Relationship satisfaction	28	0.91***	0.75	1.06	4,469	z = .27	*t*(26) = .76	84.15***	69.63
Emotional intimacy	11	0.39***	0.22	0.56	258	z = 2.69**	*t*(9) = 2.56*	33.63***	71.34
Relationship cognitions	6	0.35*	0.07	0.63	28			12.44*	59.62
Self-reported communication	6	0.76***	0.51	1.01	346			22.84***	78.09
Observed communication	15	0.57***	0.32	0.81	300	z = 1.24	*t*(13) = 1.84	43.80***	66.62
Partner behaviors	16	0.60***	0.44	0.77	1,257	z = 1.40	*t*(14) = 1.01	58.09***	76.52

Note. k = unique treatment groups. Funnel plot asymmetry was not tested for analyses with fewer than 10 treatment groups.
* $p < .05$. ** $p < .01$. *** $p < .001$.

effects on prepost mean gains in relationship satisfaction for treatment groups. However, couples' level of baseline relationship satisfaction moderated prepost changes in relationship satisfaction such that samples with more distressed couples at baseline reported greater gains during treatment (moderator b = −0.59, p = .014; Table 5).

Treatment orientation. As BCT has been tested over the greatest period of time and in the most studies, we chose to use this group as the comparison for moderation analyses of treatment orientation (i.e., BCT was coded as zero in analyses). There were no significant differences between BCT and any other treatment modality tested (CBCT, IBCT, EFCT, treatment as usual, or other; Table 5). Additionally, when we included baseline satisfaction as a control variable in second analyses, there continued to be no significant differences by treatment orientation.

Discussion

This study adds to the literature on couple therapy in several ways. First, calculating within-treatment, within-waitlist, and between treatment and waitlist effect sizes of the same studies provided a more comprehensive picture of changes during couple therapy. Specifically, it allowed us to summarize studies with any treatment design including quasi-experimental, single group designs, and trials without waitlist control groups—all of which all had previously been omitted from the meta-analytic literature. Indeed, less than 30% of the studies included in this meta-analysis have been included in previous meta-analyses. Second, we included several notable recent trials of couple therapy that had previously been absent from previous meta-analyses; these trials represent the state-of-the-art of couple therapy, including emotionally focused couple therapy (EFCT) and the main outcome studies of IBCT. Third, we presented results on both short-term and long-term follow-up from couple therapy, allowing us to draw conclusions about maintenance of effects. Finally, we analyzed moderators of improvements in relationship satisfaction during treatment, replicating some previous findings as well as testing moderators that had been impossible to test due to limitations in previous meta-analyses.

Effects of Couple Therapy

Previous meta-analyses of couple therapy have generally reported medium to large prepost between-groups effect sizes for global relationship satisfaction (0.585, Shadish & Baldwin, 2005; 0.43 to 1.36, Wood et al., 2005; 0.71 to 1.37, Dunn & Schwebel,

Table 3
Random Effects Models Over Follow-Up Within Treatment Groups

Outcome	k	Hedge's $\bar{g}$	95% CI Lower	95% CI Upper	Failsafe *N*	Regression test for funnel plot asymmetry: Mixed-effects meta-regression model	Regression test for funnel plot asymmetry: Weighted regression with multiplicative dispersion	Q_{within}	I^2(%)
Short-term follow-up (<=6 months)									
Relationship satisfaction	36	0.89***	0.71	1.08	12,998	z = 2.75**	*t*(34) = 3.24**	386.60***	94.79
Emotional intimacy	10	0.91***	0.45	1.38	2,180	z = −1.32	*t*(8) = −2.28	508.77***	96.92
Self-reported communication	6	1.97***	1.09	2.66	1,850			224.85***	97.85
Observed communication	5	0.37	−0.09	0.84	30			20.49***	85.81
Partner behaviors	17	1.31***	1.12	1.49	7735	z = 2.94**	*t*(15) = 1.84	90.38***	84.70
Long-term follow-up (>6 months to 2 years)									
Relationship satisfaction	10	0.91***	0.40	1.41	1363	z = 1.28	*t*(8) = 1.21	172.43***	97.49

Note. k = unique treatment groups. Within-treatment relationship cognition as well as all domains within-waitlist had fewer than five treatment groups at short-term follow-up as well as all domains other than satisfaction at long-term follow-up and were omitted from analyses. Funnel plot asymmetry was not tested for analyses with fewer than 10 treatment groups.
* $p < .05$. ** $p < .01$. *** $p < .001$.

1995). This study found a between-groups effect size of Hedge's $\bar{g}$ of 0.91 for relationship satisfaction, or an increase of 18 points on the DAS-32 compared with a waitlist control. The inclusion of newer studies with larger effect sizes likely influenced this shift; additionally, analyses were weighted by sample size and therefore newer, larger studies had a greater influence. These results are an encouraging update to the literature on couple therapy—participants are, on average, experiencing large improvements in satisfaction during treatment compared to waitlist control groups.

Not only does couple therapy have a positive impact on global relationship satisfaction, but our results reveal that numerous specific domains of relationship functioning also improved—including emotional intimacy, relationship cognitions, self-reported communication, observed communication, and partner behaviors. Previous meta-analysis have either collapsed across measures to the study level (e.g., Dunn & Schwebel, 1995; Shadish & Baldwin, 2005) or selected one measure or domain, omitting other measures (e.g., Wood et al., 2005). In contrast, the approach used here allowed us to capture change across a broader range of relationship functioning. However, some of these constructs are highly related, and the broad impacts suggested by our analyses may be due in part to a "glop" problem—shared method variance (see Supplementary Table 5 for correlations). A previous meta-analysis of premarital education found differences in communication outcomes when measured with self-report versus observed measures (Fawcett, Hawkins, Blanchard, & Carroll, 2010). Future research should include more objective measures of couple functioning.

In addition to calculating a summary effect size for treatment groups, this study also summarized the effect of waitlist control groups. In contrast to other outcomes frequently targeted by psychotherapy (e.g., depression), couples were generally not able to improve or recover on their own, highlighting the need for efficacious treatments for relationship distress. Specifically, couples waiting for treatment did not experience significant gains in relationship satisfaction, emotional intimacy, or observed communication. We were unable to calculate summary effect sizes for self-reported communication or relationship cognitions due to limited sample sizes ($k < 5$). These results replicate a previous study that found no effect of waitlist control groups when averaged across measure domains (Baucom et al., 2003). Therefore, future study designs seeking to understand overall effects may not need to include a waitlist control group, and can instead provide treatment to all couples and benchmark results against the within-treatment effect sizes calculated here. However, studies analyzing mechanisms of effects and other questions would continue to benefit from inclusion of a control group. In contrast, partner behaviors significantly improved at a small effect size for waitlist groups, possibly due to increased awareness or attention on the relationship as a result of initiating the help-seeking process or through participation in the research study.

As several more recent trials of couple therapy have not included a waitlist control group, calculating between-groups effect sizes was impossible for these trials. How then should we understand the true effect of these newer interventions (relative to a control group)? The between-groups effect sizes reported here can be roughly interpreted as the difference between the within-treatment effect sizes and the within-waitlist effect sizes (see Figure 3). Specifically, for relationship satisfaction, observed communication, and partner behaviors, the difference of the within-treatment and within-waitlist effect sizes is roughly equal to the between-groups effect sizes for those domains, suggesting that even though studies omitted a waitlist control, the overall effectiveness of these newer treatments was generally the same as that of previous types of couple therapy (that included a control group in their randomized controlled trial). In contrast, applying the same logic for the emotional intimacy effect sizes suggest the within-treatment effect size for the newer studies is much larger than what would be expected based on previous studies that included a between-groups effect size. It could be that many of the more recent methodologies (e.g., EFCT, IBCT) which place a heavier

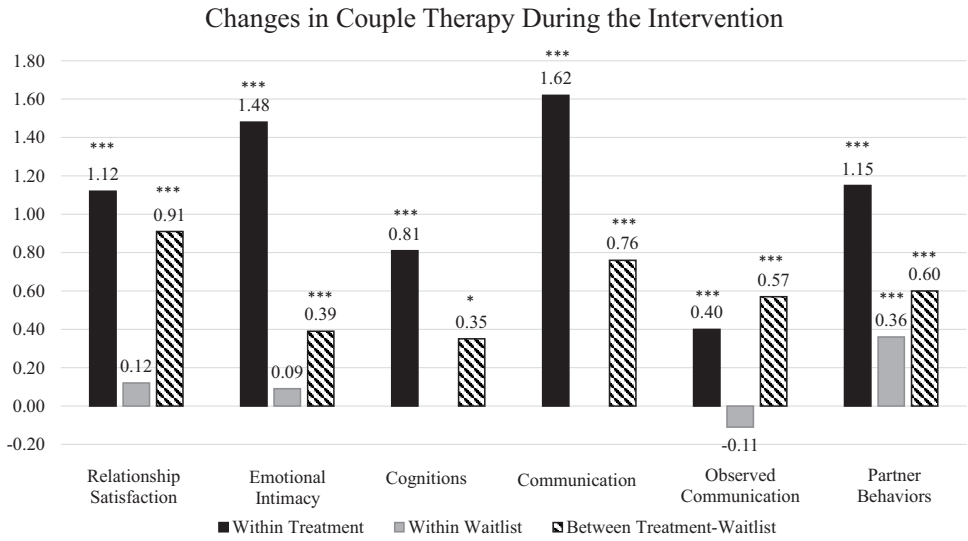

Figure 3. The sample sizes for within-waitlist relationship cognitions and self-reported communication were too small for meta-analytics ($k < 5$); therefore, they are omitted here. Summary effect sizes for these domains are available in Supplementary Table 3. * $p < .05$. *** $p < .001$.

focus on emotional intimacy (e.g., Johnson, Hunsley, Greenberg, & Schindler, 1999; Roddy, Nowlan, Doss, & Christensen, 2016), and which also have not included a waitlist control group, yield larger effect sizes in this domain. Unfortunately, such comparisons are impossible for cognitions about the relationship and self-reported communication because the number of waitlist groups that reported these measures were too few for analysis.

Finally, this is the first meta-analysis of couple therapy to summarize effect sizes over follow-up. At short-term follow-up, effect sizes for relationship satisfaction and emotional intimacy were slightly smaller than at posttreatment, while effect sizes for self-reported communication and partner behaviors were slightly larger than at posttreatment. At long-term follow-up, couples continued to report significant, large improvements in relationship satisfaction from baseline that are roughly equivalent in magnitude to the short-term follow-up estimate. However, these results should be interpreted with caution due to small sample sizes and limited number of examined domains at long-term follow-up.

Moderation of Changes in Satisfaction

The results of the moderation analyses offer both an encouraging and discouraging look at the field of couple therapy. Baseline satisfaction moderated results such that more distressed samples resulted in larger gains, replicating work in couple therapy (Baucom, Atkins, Simpson, & Christensen, 2009) and relationship education (Hawkins & Erickson, 2015). The nonsignificant moderation by randomization, clinical representativeness, and number of couples per study suggests that couple therapy is robust to these design elements. Further, the lack of moderation by publication status suggests the "file drawer" problem may not be a significant issue for this field. Further, tests of publication bias were nonsignificant for pre-to-post relationship satisfaction within treatment groups, within waitlist groups, and between treatment and waitlist groups. Finally, the lack of moderation by percent married and age of the participants suggests couple therapy is helpful across participant characteristics (at least for the ranges of these variables included in these studies). Unfortunately, year did not moderate results, suggesting that couple therapy on average has not improved in its effectiveness as new advances or approaches have been introduced.

Furthermore, across all studies, moderation by treatment orientation demonstrated no significant differences between any therapy modality and BCT. Although individual studies comparing treatment orientations in randomized trials have shown superiority of BCT (e.g., Hahlweg, Revenstorf, & Schindler, 1982), inferiority of BCT (e.g., Johnson & Greenberg, 1985), or equivalence of BCT and newer approaches (e.g., Christensen et al., 2004) at the end of treatment, our results suggest there are no differences on average. The lack of significant differences by modality is consistent with the "common principles" conceptualization of couple therapy (Benson, McGinn, & Christensen, 2012), which emphasizes the similarities rather than differences across empirically based couple therapies. However, it is important to note that one third of the treatment groups represented here were BCT, one third were categorized as "other," and the remaining third were divided between the remaining orientations (see Table 4), which limited power to detect differences. Although the nonsignificant differ-

Table 4
Descriptive Statistics of Moderators for Treatment Groups

Moderators	k	Percent or Mean	*SD*	Range
		Study characteristics		
Randomization	40			
Single group or no randomization	16	40%		
Quasirandomized or randomized	24	60%		
Publication	40			
Unpublished	7	17.5%		
Published	33	82.5%		
Clinical representativeness	40	1.7	1.54	0 to 5
Number of couples per study	40	52.3	55.97	12 to 250
Year	40	1996.78	11.97	1977 to 2016
		Participant characteristics		
Baseline satisfaction (z-score)	32	−1.26	0.36	−2.53 to −0.40
Married	30	94.92	11.95	49.4 to 100
Age	36	37.51	5.16	27 to 48.15
		Treatment group characteristics		
Treatment orientation	63			
Behavioral	21	33.3%		
Cognitive behavioral	5	7.9%		
Integrative behavior	2	3.2%		
Emotionally focused	9	14.3%		
Treatment as usual	5	7.9%		
Other	21	33.3%		

Note. k = number of studies for study and participant characteristics and number of unique treatment groups for treatment group characteristics.

Table 5
Random Effects Moderation Models Within Treatment Groups for Relationship Satisfaction Pre to Post

Moderators	k	Hedge's $\bar{g}$	95% CI Lower	95% CI Upper	$Q_{moderators}$
Study characteristics					
Randomization	59				0.04
Single group (intercept)		1.08***	0.69	1.47	
Randomized		−0.02	−0.55	0.51	
Publication	59				0.35
Unpublished (intercept)		0.98***	0.48	1.48	
Published		0.16	−0.38	0.71	
Clinical representativeness	59				0.57
Unrepresentative (intercept)		1.21***	0.90	1.53	
Clinically representative (continuous)		−0.06	−0.23	0.10	
Number of couples per study	59				1.60
Average number of couples (intercept)		1.10***	0.90	1.29	
Number of couples (centered)		0.00	−0.01	0.00	
Year	59				0.07
Average year (intercept)		1.11***	0.91	1.31	
Year (centered)		0.00	−0.02	0.02	
Participant characteristics					
Baseline satisfaction	49				6.07**
Average satisfaction z-score (intercept)		1.04***	0.89	1.20	
Satisfaction (centered)		−0.59*	−1.06	−0.12	
Married	49				0.05
Average percent married (intercept)		1.14***	0.91	1.38	
Percent married (centered)		0.00	−0.02	0.03	
Age	54				0.49
Average age (intercept)		1.12***	0.92	1.32	
Age (centered)		−0.01	−0.06	0.03	
Treatment group characteristics					
Treatment orientation	59				0.95
Behavioral (intercept)		1.12***	0.77	1.47	
Cognitive behavioral		−0.29	−1.07	0.50	
Integrative behavior		−0.20	−1.37	0.97	
Emotionally focused		0.07	−0.56	0.71	
Treatment as usual		−0.03	−0.89	0.84	
Other		0.06	−0.44	0.56	

Note. k = unique treatment groups.
* $p < .05$. ** $p < .01$. *** $p < .001$.

ences between orientations held when baseline satisfaction was added to the model, there could be other important moderators that were not tested here that could potentially reveal differences by treatment orientation.

Remaining Questions

In addition to expanding our understanding of the outcomes of couple therapy, this meta-analysis also highlights a number of gaps in our knowledge. First, many studies failed to include information about the demographic composition of the sample. Future research on couple therapy should seek to include more diverse samples, including same-sex and other diverse relationships, and fully report on participant demographics. Second, most studies included only married couples. Recent research in the field of couple therapy has shifted away from marriage as an inclusion criteria and instead focused on committed relationships—including cohabiting couples, coparents, and nonmarried partners—to reflect the fact that other types of relationship unions are becoming more common. Third, some regression tests revealed the possibility of publication bias within certain domains of relationship functioning other than relationship satisfaction. Thus, the suspected publication bias may be due to selective reporting of measures or missing studies from this work. We believe that the impact of bias on these results is not trivial but the overall finding remains valid (Borenstein, Hedges, Higgins, & Rothstein, 2011). Additionally, more information is needed on the long-term effects of couple therapy; indeed, only two studies with follow-ups longer than 2 years were identified (5 years, Christensen, Atkins, Baucom, & Yi, 2010; 4 years, Snyder, Wills, & Grady-Fletcher, 1991). If more studies were able to collect and report follow-ups along these time ranges, future meta-analyses could speak to effects of couple therapy more than 2 years following treatment. Finally, the high I^2 values in results here suggest there is a large amount of unexplained variability in the effect sizes. Therefore, although the moderators

examined in the present study do not explain variation in effect sizes, there are likely other moderators (many likely not reported in published studies) that affect change during treatment.

Limitations

Methodology for meta-analyses has focused on summarizing between-groups effect sizes; therefore, the analyses here for within-group mean gain effect sizes were informed by these methodologies and adapted as needed following previously published guidelines (Cooper et al., 2009). However, continued work in the field of methodology is critical if future research designs continue to omit a waitlist control group. Additionally, there was variability in whether studies presented data at the individual level (e.g., separately for husbands and wives), at the couple level averaged across members of the couple, or at the couple level as a sum of the two partners' scores. Further, most moderators were coded at the study level (e.g., randomization, participant characteristics) versus the treatment group level (e.g., theoretical orientation) and therefore there is a degree of dependency we were unable to model; more work is needed in the area of nested data and meta-analyses. Although multilevel meta-analyses (Van den Noortgate, López-López, Marín-Martínez, & Sánchez-Meca, 2015) have sought to find new methodological ways to include multiple dependent measures rather than aggregating measures, as was done in this study, these techniques are difficult to implement due to the specificity of information needed—information that is rarely available in published articles. Third, some analyses were omitted due to small sample sizes when effect sizes were subdivided into measure domains. While this approach potentially provides a more accurate estimate for specific outcome domains, we were not able to represent the full scope of the field. Finally, the included samples were entirely same-sex couples and, based on the minority of studies that reported demographic information, a mostly White, upper-middle class sample; thus, results may not generalize to more diverse populations.

Conclusions

Overall, this meta-analysis is an optimistic update to the field of couple therapy; treatments developed for distressed couples are producing large-size changes across multiple, key domains of relationship functioning. Moreover, these effects were generally maintained for as much as 2 years after the end of treatment. However, continued work is needed in this field to improve the transparency of reporting and expanding the study of couple therapy in more diverse samples.

References

* Study included in this meta-analysis. † Previous meta-analyses or study included in previous meta-analysis.

*Ahmad, S. (2012). *Enhancing the quality of South Asian marriages through systemic-constructivist couple therapy (SCCT): Investigating the role of couple identity in marital satisfaction* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (Order No. NR92761)

*Ahmad, S., & Reid, D. W. (2016). Enhancing the relationship adjustment of South Asian Canadian couples using a systemic-constructivist approach to couple therapy. *Journal of Marital and Family Therapy, 42,* 615–629. http://dx.doi.org/10.1111/jmft.12161

*Anderson, S. R., & Johnson, L. N. (2010). A dyadic analysis of the between- and within-system alliances on distress. *Family Process, 49,* 220–235. http://dx.doi.org/10.1111/j.1545-5300.2010.01319.x

*Anker, M. G., Duncan, B. L., & Sparks, J. A. (2009). Using client feedback to improve couple therapy outcomes: A randomized clinical trial in a naturalistic setting. *Journal of Consulting and Clinical Psychology, 77,* 693–704. http://dx.doi.org/10.1037/a0016062

*Anker, M. G., Owen, J., Duncan, B. L., & Sparks, J. A. (2010). The alliance in couple therapy: Partner influence, early change, and alliance patterns in a naturalistic sample. *Journal of Consulting and Clinical Psychology, 78,* 635–645. http://dx.doi.org/10.1037/a0020051

*Balfour, A., & Lanman, M. (2012). An evaluation of time-limited psychodynamic psychotherapy for couples: A pilot study. *Psychology and Psychotherapy, 85,* 292–309. http://dx.doi.org/10.1111/j.2044-8341.2011.02030.x

Baucom, B. R., Atkins, D. C., Simpson, L. E., & Christensen, A. (2009). Prediction of response to treatment in a randomized clinical trial of couple therapy: A 2-year follow-up. *Journal of Consulting and Clinical Psychology, 77,* 160–173. http://dx.doi.org/10.1037/a0014405

*†Baucom, D. H. (1982). A comparison of behavioral contracting and problem-solving/communications training in behavioral marital therapy. *Behavior Therapy, 13,* 162–174. http://dx.doi.org/10.1016/S0005-7894(82)80060-9

Baucom, D. H., Hahlweg, K., & Kuschel, A. (2003). Are waiting-list control groups needed in future marital therapy outcome research? *Behavior Therapy, 34,* 179–188. http://dx.doi.org/10.1016/S0005-7894(03)80012-6

*†Baucom, D. H., & Lester, G. W. (1986). The usefulness of cognitive restructuring as an adjunct to behavioral marital therapy. *Behavior Therapy, 17,* 385–403. http://dx.doi.org/10.1016/S0005-7894(86)80070-3

*†Baucom, D. H., Sayers, S. L., & Sher, T. G. (1990). Supplementing behavioral marital therapy with cognitive restructuring and emotional expressiveness training: An outcome investigation. *Journal of Consulting and Clinical Psychology, 58,* 636–645. http://dx.doi.org/10.1037/0022-006X.58.5.636

*Baucom, K. J. W., Baucom, B. R., & Christensen, A. (2015). Changes in dyadic communication during and after integrative and traditional behavioral couple therapy. *Behaviour Research and Therapy, 65,* 18–28. http://dx.doi.org/10.1016/j.brat.2014.12.004

*Baucom, K. J. W., Sevier, M., Eldridge, K. A., Doss, B. D., & Christensen, A. (2011). Observed communication in couples two years after integrative and traditional behavioral couple therapy: Outcome and link with five-year follow-up. *Journal of Consulting and Clinical Psychology, 79,* 565–576. http://dx.doi.org/10.1037/a0025121

*Bennun, I. (1985). Prediction and responsiveness in behavioural marital therapy. *Behavioural Psychotherapy, 13,* 186–201. http://dx.doi.org/10.1017/S0141347300011058

Benson, L. A., McGinn, M. M., & Christensen, A. (2012). Common principles of couple therapy. *Behavior Therapy, 43,* 25–35. http://dx.doi.org/10.1016/j.beth.2010.12.009

*Berns, S. B. (2002). *Predicting response to treatment in a randomized clinical trial of marital therapy* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (Order No. 3062919)

*Boettcher, R. E. (1977). Interspousal empathy, marital satisfaction, and marriage counseling. *Journal of Social Service Research, 1,* 105–113. http://dx.doi.org/10.1300/J079v01n01_08

Borenstein, M., Hedges, L. V., Higgins, J., & Rothstein, H. R. (2009). Subgroup analyses. In M. Borenstein, L. V. Hedges, J. P. T. Higgins, & H. R. Rothstein (Eds.), *Introduction to meta-analysis* (pp. 149–186). West Sussex, UK: Wiley Ltd.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2011). *Introduction to meta-analysis*. West Sussex, UK: Wiley.

*Butler, M. H., Harper, J. M., & Mitchell, C. B. (2011). A comparison of attachment outcomes in enactment-based versus therapist-centered therapy process modalities in couple therapy. *Family Process, 50,* 203–220. http://dx.doi.org/10.1111/j.1545-5300.2011.01355.x

*Chambers, J. (2008). *The durability of support focused marital therapy* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3318555)

Christensen, A., Atkins, D. C., Baucom, B., & Yi, J. (2010). Marital status and satisfaction five years following a randomized clinical trial comparing traditional versus integrative behavioral couple therapy. *Journal of Consulting and Clinical Psychology, 78,* 225–235. http://dx.doi.org/10.1037/a0018132

*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,* 176–191. http://dx.doi.org/10.1037/0022-006X.72.2.176

*Christensen, A., Atkins, D. C., Yi, J., Baucom, D. H., & George, W. H. (2006). Couple and individual adjustment for 2 years following a randomized clinical trial comparing traditional versus integrative behavioral couple therapy. *Journal of Consulting and Clinical Psychology, 74,* 1180–1191. http://dx.doi.org/10.1037/0022-006X.74.6.1180

Cooper, H., Hedges, L. V., & Valentine, J. C. (Eds.). (2009). *The handbook of research synthesis and meta-analysis*. New York, NY: Russell Sage Foundation.

*Cordova, J. V. (1996). *Acceptance versus change interventions in behavioral couples therapy: Impact on client communication process in the therapy sessions* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (Order No. 9637925)

*Cordova, J. V., Jacobson, N. S., & Christensen, A. (1998). Acceptance versus change interventions in behavioral couple therapy: Impact on couples' in-session communication. *Journal of Marital and Family Therapy, 24,* 437–455. http://dx.doi.org/10.1111/j.1752-0606.1998.tb01099.x

*Dalgleish, T. L., Johnson, S. M., Burgess Moser, M., Lafontaine, M. F., Wiebe, S. A., & Tasca, G. A. (2015). Predicting change in marital satisfaction throughout emotionally focused couple therapy. *Journal of Marital and Family Therapy, 41,* 276–291. http://dx.doi.org/10.1111/jmft.12077

*Dalgleish, T. L., Johnson, S. M., Burgess Moser, M., Wiebe, S. A., & Tasca, G. A. (2015). Predicting key change events in emotionally focused couple therapy. *Journal of Marital and Family Therapy, 41,* 260–275. http://dx.doi.org/10.1111/jmft.12101

*Davidson, G. N. S., & Horvath, A. O. (1997). Three sessions of brief couples therapy: A clinical trial. *Journal of Family Psychology, 11,* 422–435. http://dx.doi.org/10.1037/0893-3200.11.4.422-435

*de Jonge, M. M. (2002). *Support -focused marital therapy: Assessing change in the wait -list control group and predictors of outcome* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (Order No. NR92761)

Del Re, A. C. (2015). A practical tutorial on conducting meta-analysis in R. *The Quantitative Methods for Psychology, 11,* 37–50.

Denton, W. H., Burleson, B. R., Clark, T. E., Rodriguez, C. P., & Hobbs, B. V. (2000). A randomized trial of emotion-focused therapy for couples in a training clinic. *Journal of Marital and Family Therapy, 26,* 65–78. http://dx.doi.org/10.1111/j.1752-0606.2000.tb00277.x

*Doell, F. K. (2010). *Enhancing couples' communication through systemic-constructivist couple therapy: The relationship between marital listening and relationship quality* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (Order No. NR64890)

*Doss, B. D., Mitchell, A., Georgia, E. J., Biesen, J. N., & Rowe, L. S. (2015). Improvements in closeness, communication, and psychological distress mediate effects of couple therapy for veterans. *Journal of Consulting and Clinical Psychology, 83,* 405–415. http://dx.doi.org/10.1037/a0038541

*Doss, B. D., Rowe, L. S., Carhart, K., Madsen, J. W., & Georgia, E. J. (2011). Sudden gains in treatment-as-usual couple therapy for military veterans. *Behavior Therapy, 42,* 509–520. http://dx.doi.org/10.1016/j.beth.2010.12.001

*Doss, B. D., Rowe, L. S., Morrison, K. R., Libet, J., Birchler, G. R., Madsen, J. W., & McQuaid, J. R. (2012). Couple therapy for military veterans: Overall effectiveness and predictors of response. *Behavior Therapy, 43,* 216–227. http://dx.doi.org/10.1016/j.beth.2011.06.006

*Doss, B. D., Thum, Y. M., Sevier, M., Atkins, D. C., & Christensen, A. (2005). Improving relationships: Mechanisms of change in couple therapy. *Journal of Consulting and Clinical Psychology, 73,* 624–633. http://dx.doi.org/10.1037/0022-006X.73.4.624

†Dunn, R. L., & Schwebel, A. I. (1995). Meta-analytic review of marital therapy outcome research. *Journal of Family Psychology, 9,* 58–68.

*Eldridge, K. A. (2000). *Demand-withdraw communication during marital conflict: Relationship satisfaction and gender role considerations* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (Order No. 9979032)

Epstein, N., Baucom, D. H., & Baucom, D. H. (2002). *Enhanced cognitive-behavioral therapy for couples: A contextual approach* (Vol. 1). Washington, DC: American Psychological Association.

Fawcett, E. B., Hawkins, A. J., Blanchard, V. L., & Carroll, J. S. (2010). Do premarital education programs really work? A meta-analytic study. *Family Relations, 59,* 232–239. http://dx.doi.org/10.1111/j.1741-3729.2010.00598.x

Fischer, M. S., Baucom, D. H., & Cohen, M. J. (2016). Cognitive-behavioral couple therapies: Review of the evidence for the treatment of relationship distress, psychopathology, and chronic health conditions. *Family Process, 55,* 423–442. http://dx.doi.org/10.1111/famp.12227

Foster, E. D., & Deardorff, A. (2017). Open science framework (OSF). *Journal of the Medical Library Association, 105,* 203–206. http://dx.doi.org/10.5195/jmla.2017.88

Funk, J. L., & Rogge, R. D. (2007). Testing the ruler with item response theory: Increasing precision of measurement for relationship satisfaction with the Couples Satisfaction Index. *Journal of Family Psychology, 21,* 572–583. http://dx.doi.org/10.1037/0893-3200.21.4.572

*Ganahl, G. F. (1981). *Effects of client, treatment, and therapist variables on the outcome of structured marital enrichment* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (Order No. 8209778)

*Goldman, A., & Greenberg, L. (1992). Comparison of integrated systemic and emotionally focused approaches to couples therapy. *Journal of Consulting and Clinical Psychology, 60,* 962–969. http://dx.doi.org/10.1037/0022-006X.60.6.962

*Greenberg, L., Warwar, S., & Malcolm, W. (2010). Emotion-focused couples therapy and the facilitation of forgiveness. *Journal of Marital and Family Therapy, 36,* 28–42. http://dx.doi.org/10.1111/j.1752-0606.2009.00185.x

Greenhouse, J. B., & Inyengar, S. (2009). Sensitivity analysis and diagnostics. In H. Cooper, L. V. Hegdes, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 417–433). New York, NY: Russell Sage Foundation.

†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,* 440–447. http://dx.doi.org/10.1037/0022-006X.56.3.440

*†Hahlweg, K., Revenstorf, D., & Schindler, L. (1982). Treatment of marital distress: Comparing formats and modalities. *Advances in Behaviour Research and Therapy, 4,* 57–74. http://dx.doi.org/10.1016/0146-6402(82)90005-4

[*]Halford, W. K., Sanders, M. R., & Behrens, B. C. (1993). A comparison of the generalization of behavioral marital therapy and enhanced behavioral marital therapy. *Journal of Consulting and Clinical Psychology, 61,* 51–60. http://dx.doi.org/10.1037/0022-006X.61.1.51

Hawkins, A. J., & Erickson, S. E. (2015). Is couple and relationship education effective for lower income participants? A meta-analytic study. *Journal of Family Psychology, 29,* 59–68. http://dx.doi.org/10.1037/fam0000045

Hedges, L. V., & Olkin, L. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.

[*]Hefner, C. W., & Prochaska, J. O. (1984). Concurrent vs. conjoint marital therapy. *Social Work, 29,* 287–291.

Higgins, J. P. T., & Green, S. (2011). Recommendations on testing for funnel plot asymmetry. *Cochrane Handbook for Systematic Reviews of Interventions (Version 5)*.

[*†]Jacobson, N. S. (1984). A component analysis of behavioral marital therapy: The relative effectiveness of behavior exchange and communication/problem-solving training. *Journal of Consulting and Clinical Psychology, 52,* 295–305. http://dx.doi.org/10.1037/0022-006X.52.2.295

[*†]Jacobson, N. S., Christensen, A., Prince, S. E., Cordova, J., & Eldridge, K. (2000). Integrative behavioral couple therapy: An acceptance-based, promising new treatment for couple discord. *Journal of Consulting and Clinical Psychology, 68,* 351–355. http://dx.doi.org/10.1037/0022-006X.68.2.351

[*†]Jacobson, N. S., Follette, V. M., Follette, W. C., Holtzworth-Munroe, A., Katt, J. L., & Schmaling, K. B. (1985). A component analysis of behavioral marital therapy: 1-year follow-up. *Behaviour Research and Therapy, 23,* 549–555. http://dx.doi.org/10.1016/0005-7967(85)90102-0

[*]James, P. S. (1991). Effects of a communication training component added to an emotionally focused couples therapy. *Journal of Marital and Family Therapy, 17,* 263–275. http://dx.doi.org/10.1111/j.1752-0606.1991.tb00894.x

[*†]Johnson, S. M., & Greenberg, L. S. (1985). Differential effects of experiential and problem-solving interventions in resolving marital conflict. *Journal of Consulting and Clinical Psychology, 53,* 175–184. http://dx.doi.org/10.1037/0022-006X.53.2.175

[†]Johnson, S. M., Hunsley, J., Greenberg, L., & Schindler, D. (1999). Emotionally focused couples therapy: Status and challenges. *Clinical Psychology: Science and Practice, 6,* 67–79. http://dx.doi.org/10.1093/clipsy.6.1.67

[*]Johnson, S. M., & Talitman, E. (1997). Predictors of success in emotionally focused marital therapy. *Journal of Marital and Family Therapy, 23,* 135–152. http://dx.doi.org/10.1111/j.1752-0606.1997.tb00239.x

[*]Knerr, M. R. (2008). *Differentiation and power in couples therapy* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (Order No. 3327336)

Lavner, J. A., & Bradbury, T. N. (2010). Patterns of change in marital satisfaction over the newlywed years. *Journal of Marriage and Family, 72,* 1171–1187. http://dx.doi.org/10.1111/j.1741-3737.2010.00757.x

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis* (Vol. 49). Thousand Oaks, CA: Sage.

[*]Makinen, J. A., & Johnson, S. M. (2006). Resolving attachment injuries in couples using emotionally focused therapy: Steps toward forgiveness and reconciliation. *Journal of Consulting and Clinical Psychology, 74,* 1055–1064. http://dx.doi.org/10.1037/0022-006x.74.6.1055

[*]Marschall, D. E. (2002). *The efficacy of support -focused marital therapy in reducing marital distress* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (Order No. 3040413)

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica, 22,* 276–282. http://dx.doi.org/10.11613/BM.2012.031

[*†]Mehlman, S. K., Baucom, D. H., & Anderson, D. (1983). Effectiveness of cotherapists versus single therapists and immediate versus delayed treatment in behavioral marital therapy. *Journal of Consulting and Clinical Psychology, 51,* 258–266. http://dx.doi.org/10.1037/0022-006X.51.2.258

[*]Nowlan, K. M., Georgia, E. J., & Doss, B. D. (2017). Long-term effectiveness of treatment-as-usual couple therapy for military veterans. *Behavior Therapy, 48,* 847–859. http://dx.doi.org/10.1016/j.beth.2017.05.007

[*]O'Leary, K. D., & Arias, I. (1983). The influence of marital therapy on sexual satisfaction. *Journal of Sex & Marital Therapy, 9,* 171–181. http://dx.doi.org/10.1080/00926238308405845

O'Leary, K. D., & Turkewitz, H. (1981). A comparative outcome study of behavioral marital therapy and communication therapy. *Journal of Marital and Family Therapy, 7,* 159–169. http://dx.doi.org/10.1111/j.1752-0606.1981.tb01366.x

Pigott, T. D. (2012). Planning a meta-analysis in a systematic review. In T. D. Pigott (Ed.), *Advances in meta-analysis* (pp. 13–32). New York, NY: Springer.

Plattor, J. R. (1990). *Evaluating the effectiveness of marital therapies: A meta-analytic investigation* (Doctoral dissertation). Available from ProQuest Dissertations & Theses Global database. (UMI No. 2860)

[*]Rasoul, H., Ali, K., Esa, J., & Maryam, M. (2016). Effectiveness of self-regulation couple therapy intervention on marital satisfaction and partner abuse (non-physical). *Mediterranean Journal of Clinical Psychology, 4,* 1–12. http://dx.doi.org/10.6092/2282-1619/2016.4.1195

Rathgeber, M., Bürkner, P. C., Schiller, E. M., & Holling, H. (2019). The efficacy of emotionally focused couples therapy and behavioral couples therapy: A meta-analysis. *Journal of Marital and Family Therapy, 45,* 447–463. http://dx.doi.org/10.1111/jmft.12336

[*†]Revenstorf, D., Schindler, L., & Hahlweg, K. (1983). Behavioral marital therapy applied in a conjoint and a conjoint-group modality: Short- and long-term effectiveness. *Behavior Therapy, 14,* 614–625. http://dx.doi.org/10.1016/S0005-7894(83)80054-9

Robles, T. F., Slatcher, R. B., Trombello, J. M., & McGinn, M. M. (2014). Marital quality and health: A meta-analytic review. *Psychological Bulletin, 140,* 140–187. http://dx.doi.org/10.1037/a0031859

Roddy, M. K., Nowlan, K. M., Doss, B. D., & Christensen, A. (2016). Integrative behavioral couple therapy: Theoretical background, empirical research, and dissemination. *Family Process, 55,* 408–422. http://dx.doi.org/10.1111/famp.12223

Rosenberg, M. S. (2005). The file-drawer problem revisited: A general weighted method for calculating fail-safe numbers in meta-analysis. *Evolution; International Journal of Organic Evolution, 59,* 464–468. http://dx.doi.org/10.1111/j.0014-3820.2005.tb01004.x

[*]Ross, E. R., Baker, S. B., & Guerney, B. G. (1985). Effectiveness of relationship enhancement therapy versus therapist's preferred therapy. *The American Journal of Family Therapy, 13,* 11–21. http://dx.doi.org/10.1080/01926188508250206

Scammacca, N., Roberts, G., & Stuebing, K. K. (2014). Meta-analysis with complex research designs dealing with dependence from multiple measures and multiple group comparisons. *Review of Educational Research, 84,* 328–364. http://dx.doi.org/10.3102/0034654313500826

[*†]Schindler, L., Hahlweg, K., & Revenstorf, D. (1983). Short- and long-term effectiveness of two communication training modalities with distressed couples. *American Journal of Family Therapy, 11,* 54–64. http://dx.doi.org/10.1080/01926188308250135

Shadish, W. R., & Baldwin, S. A. (2003). Meta-analysis of MFT interventions. *Journal of Marital and Family Therapy, 29,* 547–570. http://dx.doi.org/10.1111/j.1752-0606.2003.tb01694.x

[†]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,* 6–14. http://dx.doi.org/10.1037/0022-006X.73.1.6

Shadish, W. R., Matt, G. E., Navarro, A. M., & Phillips, G. (2000). The effects of psychological therapies under clinically representative condi-

tions: A meta-analysis. *Psychological Bulletin, 126,* 512–529. http://dx.doi.org/10.1037/0033-2909.126.4.512

*Sims, A. E. B. (1999). *Working models of attachment: The impact of emotionally focused marital therapy* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (Order No. NQ46545)

Slatcher, R. B., & Selcuk, E. (2017). A social psychological perspective on the links between close relationships and health. *Current Directions in Psychological Science, 26,* 16–21. http://dx.doi.org/10.1177/0963721416667444

*†Snyder, D. K., & Wills, R. M. (1989). Behavioral versus insight-oriented marital therapy: Effects on individual and interspousal functioning. *Journal of Consulting and Clinical Psychology, 57,* 39–46. http://dx.doi.org/10.1037/0022-006X.57.1.39

Snyder, D. K., Wills, R. M., & Grady-Fletcher, A. (1991). Long-term effectiveness of behavioral versus insight-oriented marital therapy: A 4-year follow-up study. *Journal of Consulting and Clinical Psychology, 59,* 138–141. http://dx.doi.org/10.1037/0022-006X.59.1.138

Spanier, G. B. (1976). Measuring dyadic adjustment: New scales for assessing the quality of marriage and similar dyads. *Journal of Marriage and the Family, 38,* 15–28. http://dx.doi.org/10.2307/350547

*Steinweg, C. K. M. (1990). *A comparison of the effectiveness of relationship enhancement and strategic marital therapy using the strength of the therapeutic alliance to predict statistically significant and clinically meaningful outcome* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (Order No. 9116461)

Sutton, A. J. (2009). Publication bias. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 435–454). New York, NY: Russell Sage Foundation.

*Talitman, E. (1995). *Predictors of outcome in emotionally focused marital therapy* (Doctoral thesis, University of Ottawa, Ottawa, Ontario, Canada). Retrieved from https://ruor.uottawa.ca/bitstream/10393/9877/1/NN04886.PDF

Van den Noortgate, W., López-López, J. A., Marín-Martínez, F., & Sánchez-Meca, J. (2015). Meta-analysis of multiple outcomes: A multilevel approach. *Behavior Research Methods, 47,* 1274–1294. http://dx.doi.org/10.3758/s13428-014-0527-2

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36,* 1–48. http://dx.doi.org/10.18637/jss.v036.i03

Wampold, B. E., Mondin, G. W., Moody, M., Stich, F., Benson, K., & Ahn, H. (1997). A meta-analysis of outcome studies comparing bona fide psychotherapies: Empirically, "all must have prizes." *Psychological Bulletin, 122,* 203–215. http://dx.doi.org/10.1037/0033-2909.122.3.203

Wei, Y., & Higgins, J. P. (2013). Estimating within-study covariances in multivariate meta-analysis with multiple outcomes. *Statistics in Medicine, 32,* 1191–1205. http://dx.doi.org/10.1002/sim.5679

*Whisman, M. A. (1990). *The use of booster maintenance sessions in behavioral marital therapy* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (Order No. 9108534)

Whisman, M. A., Beach, S. R., & Snyder, D. K. (2008). Is marital discord taxonic and can taxonic status be assessed reliably? Results from a national, representative sample of married couples. *Journal of Consulting and Clinical Psychology, 76,* 745–755. http://dx.doi.org/10.1037/0022-006X.76.5.745

Wiebe, S. A., & Johnson, S. M. (2016). A review of the research in emotionally focused therapy for couples. *Family Process, 55,* 390–407. http://dx.doi.org/10.1111/famp.12229

Wilson, D. B. (2001). *Practical meta-analysis effect size calculator*. Retrieved from http://www.campbellcollaboration.org/escalc/html/EffectSizeCalculator-SMD1

*†Wilson, G. L., Bornstein, P. H., & Wilson, L. J. (1988). Treatment of relationship dysfunction: An empirical evaluation of group and conjoint behavioral marital therapy. *Journal of Consulting and Clinical Psychology, 56,* 929–931. http://dx.doi.org/10.1037/0022-006X.56.6.929

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. *The American Journal of Family Therapy, 33,* 273–287. http://dx.doi.org/10.1080/01926180590962147

Received October 29, 2019
Revision received April 13, 2020
Accepted April 14, 2020 ■