Psychotherapy
2016, Vol. 53, No. 4, 377–387

0033-3204/16/$12.00 http://dx.doi.org/10.1037/pst0000098

The Effectiveness of Couple Therapy: Clinical Outcomes in a Naturalistic United Kingdom Setting

David Hewison, Polly Casey, and Naomi Mwamba
Tavistock Relationships, London, United Kingdom

Couple therapy outcomes tend to be judged by randomized controlled trial evidence, which comes primarily from the United States. United Kingdom and European outcome studies have tended to be naturalistic and there is a debate as to whether "laboratory" (RCT) studies are useful benchmarks for the outcomes of "clinic" (naturalistic) studies, not least because the therapies tested in the RCTs are hardly used in these settings. The current paper surveys the naturalistic studies in the literature and presents results from a U.K. setting of 877 individually and relationally distressed participants who completed at least 2 sessions of psychodynamic couple therapy and completed self-report measures assessing psychological well-being (CORE-OM) and relationship quality (Golombok Rust Inventory of Marital State, GRIMS). A clinical vignette is given that demonstrates the psychodynamic approach used. Analysis of the measure data conducted using hierarchical linear modeling showed an overall significant decrease in individual psychological distress for both male and female clients at the end of therapy, with a large effect size of $d = -1.04$. There was also a significant improvement in relationship satisfaction for both male and female clients, with a medium effect size of $d = -0.58$. These findings suggest that psychodynamic couple therapy is an effective treatment for couples experiencing individual and relational distress, with effect sizes similar in strength to those reported in RCTs. It argues that naturalistic effectiveness studies should be given a stronger role in assessments of which therapies work.

Keywords: couple therapy, effectiveness studies, clinical outcomes, naturalistic, psychodynamic

Couple therapy outcomes tend to be judged by the existing evidence base that comes from randomized controlled trials (RCTs). However, such evidence is missing in the United Kingdom and is rare elsewhere, found predominantly in the United States. In the U.K., at least, this absence of RCT evidence is partly because couple therapy has been delivered primarily through voluntary organizations that have their roots in pastoral care, marriage guidance, casework, and family and community interventions. These organizations have not had the backing of university psychology departments, and have not developed in the context of measurement of subjects, with salaried research staff. In addition, the strong systemic and psychoanalytic influences on these U.K. organizations have not been particularly sympathetic to an apparently impartial assessment of outcomes, preferring to rely on a general agreed sense of the success of the therapy between the clients and the therapists or counselors. Although there has been some slight change in this situation with the development of Improving Access to Psychological Therapies services in the National Health Service (NHS), which require session-by-session measurement of targeted mental health outcomes, and the place for Couple Therapy for Depression (Hewison, Clulow, & Drake, 2014) within it, systematic outcome measurement is a relatively new technology for couple therapy in the U.K. Additionally, RCTs are expensive to administer, suffer from problems of recruitment, can lead to equivocal outcomes, and are generally felt to be a hard-to-justify drain on such not-for-profit and charity organizations' finances and resources (Petch, Lee, Huntingdon, & Murray, 2014). As a result, in a recent review of couple and family therapy outcome evidence, Stratton and colleagues (Stratton et al., 2015) concluded that the vast amount of evidence about the outcomes of couple and family therapy is not in a format that would be picked up by meta-analyses of RCTs, and that although the evidence is patchy in methodology and reporting, it nonetheless suggests that couple and family therapies are effective. Not insignificantly, papers on couple therapy only amounted to 6% of publications in the family area ($n = 9$ out of 225 papers).

Given that there is an absence of U.K.-based RCT evidence for couple therapy, what can couples seeking therapy here expect as a reasonable outcome? Lundblad and Hansson (2006) point out that "all couples therapies that have been reasonably well tested have been empirically proven to be effective" (p 137); this is in line with the evidence for individual therapies (Hill & Lambert, 2004; Wampold, 2001). So the question is then: how effective *should* they be? RCTs tell us that something around 60% to 70% of couples receiving treatment should get some improvement in relationship quality, with between 35% to 50% of clinically relationally distressed couples moving to nondistressed levels. Baucom and colleagues (Baucom, Hahlweg, & Kuschel, 2003) felt that the RCT research evidence for the efficacy of behavioral couple therapy was so well proven that any new studies could simply use it as a benchmark, without the need for a waiting list control group, as untreated couples make no improvements in relationship quality

David Hewison, Department of Clinical, Training & Research Faculty, Tavistock Relationships, London, United Kingdom; Polly Casey and Naomi Mwamba, Research Department, Tavistock Relationships.

Correspondence concerning this article should be addressed to David Hewison, Tavistock Relationships, 70 Warren Street, London W1T 5PB, United Kingdom. E-mail: dhewison@tavistockrelationships.ac.uk

and communication and in fact decline slightly on average, with a negative effect size of $d = -0.06$. This benchmark would reduce the burden on clinics of an RCT that compares an active treatment against a waiting list group and has the ethical advantage of enabling couples to enter directly into therapy without an artificial delay in treatment. Their meta-analysis found an effect size of $d = 0.82$ for impacts on relationship distress and communication skills in behavioral couple therapy across 17 studies, which dropped slightly to $d = 0.71$ in comparison with the waiting list control arms. Shadish and Baldwin (2005), on the other hand estimated a lower effect size for behavioral couple therapy of $d = 0.585$ in RCTs, once the effects of a few early studies with small samples sizes and large effect sizes, and publication bias were controlled for. Even this was significantly higher than the effect size of $d = 0.28$ across all measures found in Hahlweg and Klann's, 1997 study of German and Austrian couple therapists in ordinary clinical practice (Hahlweg & Klann, 1997; Hahlweg, Markman, Thurmaier, Engl, & Eckert, 1998), replicated some years later by Klan and colleagues with similar results (Klann, Hahlweg, Baucom, & Kroeger, 2011). This suggests a difference between the outcomes that can be expected in a controlled trial and those that can be expected in a naturalistic clinical setting, casting doubts on the utility of Baucom and colleagues' benchmark for the latter, and underlining Shadish and Baldwin's concerns about the real-life clinical utility of randomized trial results.

As is widely known, for RCTs to have strong internal validity they need to control as many variables as possible in order to make their results more reliable. There is usually an unambiguous prescription of the therapy being offered (managed through manualization and fidelity checks), a defined and generally noncomorbid diagnosis, and extensive support to the therapists in the study through supervision, as well as an identical group of patients used as a control group. A survey by Wright and colleagues (Wright, Sabourin, Mondor, McDuff, & Mamodhoussen, 2007) showed that 62% of efficacy studies excluded cohabiting couples in favor of married ones who may be at less risk of relationship breakdown, especially in the early years of the relationship (Hewitt & Baxter, 2014). In naturalistic clinical settings, however, couples coming for treatment may be a mix of cohabiting and married, and therapists tend to work according to their original trainings in idiosyncratic ways (and not all therapists in any one clinic may have been trained in the same orientation). Couples coming for treatment may well not have a formal diagnosis, are likely to be suffering from a range of concurrent problems, emotional, physical, and social, and there is likely to be less supervisory assistance than there would be in a formal trial. Because of these differences, Leichsenring (2004) points out the need to distinguish evidence from RCTs and from naturalistic studies as two different but equal forms of knowledge, on the basis that they are testing separate interventions—a *laboratory* therapy that tests efficacy and a *field* therapy that tests effectiveness (also known as *clinic* therapy (Weisz, Donenberg, Han, & Weiss, 1995)). Wright and colleagues echo this, pointing out that such efficacy studies are not very representative of ordinary clinical practice (Wright et al., 2007). Shean (2015) points out the "circularity" involved in the methodology of efficacy studies, summarizing the objections to their influence on psychotherapy research and going as far as to suggest it is harmful rather than helpful for people seeking therapy. Carey and Stiles (2016) even suggest that the strict causal logic of RCTs means they cannot in any way be considered as a "gold standard" for psychological psychotherapy studies: psychotherapy is hugely more complex to investigate than pharmacotherapy.

In addition to these criticisms, Philips (2009) points out that different therapies have different therapeutic aims and suggests that judging them all by the same outcome is an error. In his study of therapy provided at a large addiction clinic in Stockholm, he found that family therapy worked toward improving family relationships, group therapy worked toward improving relationships with others, psychodynamic therapy worked toward increasing insight and reflective functioning as well as improving functioning, and CBT focused on behavioral change and increasing motivation. The management of presenting symptoms—the outcome that is most likely to be measured in RCTs—was only the third most frequent goal of therapy (p. 334). The same goes for what patients/clients want too; these are not always the removal of clearly defined symptoms but can include such things as a general increase in functioning, a better sense of meaning in life, and improved relationships with others (which may mean an increase in the tolerance of things that are not optimal, rather than their removal (Dirmaier, Harfst, Koch, & Schulz, 2006)). Hill and Lambert (2004) suggest that multiple measures across multiple domains are, in any case, needed to properly identify differences between therapy outcomes in RCTs. These are rare in the literature. Effectiveness research is therefore needed to help clinicians understand more about their real-life practice of couple therapy as distinct from that of the laboratory (Halford, Pepping, & Petch, 2016).

There are a small number of prospective naturalistic studies in the literature from the United States, from Canada, from Scandinavia, and from Germany and Austria with sample sizes that range from $N = 36–410$, and Effect Sizes that range from $d = 0.17–1.55$ (see Table 4). Varieties of integrative and mixed-model couple therapy have been studied in Norway by Anker and colleagues (Anker, Duncan, & Sparks, 2009), in Germany by Hahlweg and Klann (1997) with a repetition by Klann et al. in 2011, and in Sweden by Lundblad and Hansson (2006). This later study was able to show that treatment effects were maintained at the 2-year follow-up period. Systems-based couple therapy has been studied in Finland (Kuhlman, Tolvanen, & Seikkula, 2013a, 2013b; Seikkula, Aaltonen, Kalla, Saarinen, & Tolvanen, 2013) and by Reese and colleagues in the United States (Reese, Toland, Slone, & Norsworthy, 2010). Doss and colleagues looked at behavioral couple therapy with military veterans in the United States (Doss et al., 2012) and in Canada, Knobloch-Fedders and colleagues (Knobloch-Fedders, Pinsof, & Haase, 2015) used Integrative Problem-Centered Metaframeworks (Pinsof, Breunlin, Chambers, Solomon, & Russell, 2015) in their study. Psychodynamic couple therapy with a small sample has been studied by Balfour and colleagues in the U.K. (Balfour & Lanman, 2012).

Other naturalistic couple therapy studies not comparable with those reported above include Petch and colleagues' large Australian nonprospective, cross-sectional study of couple therapy outcomes against recollected relationship distress at intake (Petch et al., 2014). It is not stated what model of couple therapy was used in the study. A further study, again difficult to compare directly, was that by Ward and McCollum (2005) of systemic treatment effectiveness in an American university marriage and family therapy training clinic. This study did not report couple therapy out-

comes separately from those of individual and family therapy, and used the therapist report of outcomes achieved against presenting problems.

The present largescale, naturalistic, retrospective study is a contribution toward understanding the effectiveness of couple therapy. It explores *field* data collected routinely using standardized outcome measurements to try and understand client-reported changes from pre- and posttherapy with respect to global psychological distress and couple relationship quality.

The Context of the Current Study

The results reported here are from Tavistock Relationships (formerly the Tavistock Centre for Couple Relationships (TCCR)). Tavistock Relationships is a U.K. voluntary sector clinical training, practice, and research organization that specializes in the understanding of the adult couple relationship primarily from a psychoanalytic perspective. It is a not-for-profit charity and has never been part of the U.K. NHS despite having historically strong links with the Tavistock and Portman NHS Foundation Trust (the former Tavistock Clinic). Its standard couple therapy is conducted along psychodynamic lines and is not manualized, relying instead on a 3- or 4-year in-house part-time Masters-level clinical training in psychodynamic/psychoanalytic couple therapy and frequent case consultation and supervision to ensure quality and sufficient consistency. There is no specific number of sessions offered, with therapy available on an open-ended basis until such time as the couple, usually in consultation with their therapist, feels they have had enough. Although couples coming to the service have to pay a fee, this is set on a sliding scale from very low to substantial based on the ability to pay. Most couples self-refer, having found Tavistock Relationships by recommendation or via its website; others are referred by GPs, individual therapists, lawyers, or child and adolescent mental health services involved with the couples' children. No couple is excluded from therapy because they cannot afford it. The clinical service is based in London, with the majority of couples coming to it from the capital, and with others traveling from across the country and abroad. Couples come without a formal diagnosis and usually present with a range of difficulties as would be expected from any sample of an adult population of distressed couples: these include relationship conflict, affairs, the impact of first children or of adoption, the empty nest syndrome as children leave home, sexual difficulties, physical illness including life-threatening conditions, depression and other mental health difficulties, economic challenges, retirement, and problems in divorce and separation processes. Tavistock Relationships works across ethnicities and sexualities.

Method

Design

The study is a nonrandomized, single group, largescale retrospective clinical study. It constitutes all clients who scored as both relationally and individually distressed on standardized measures who attended Tavistock Relationships' couple therapy service for two or more therapy sessions between 2008 and 31st August, 2014, who had given us baseline and at least one other set of measures, and who were no longer in therapy at this point (N = 877). Couples complete self-rating forms at assessment, 6 weeks, 12 weeks, 24 weeks, 36 weeks, and 48 weeks if still in treatment, and at end of sessions whenever this falls even if it is prior to the 6-week assessment point. Demographic data are gathered at clients' initial assessment.

Measures

Clinical Outcomes in Routine Evaluation-Outcome Monitoring. The Clinical Outcomes in Routine Evaluation-Outcome Monitoring (CORE-OM) form is a 34-item tool used to measure global psychological distress (Evans et al., 2000). The self-report measure has a variety of positively and negatively worded questions that fall into four subscales (Subjective well-being, Problems, Functioning, and Risk) and an overall mean score, which measure how the client has been feeling over the last week. This report uses only the overall mean score. Clients respond to each question on a 5-point Likert scale ranging from 0 "Not at all" to 4 "Most or all the time." Scores can range from 0 to 40, with higher scores representing greater distress (Leach et al., 2006). Overall mean scores can be categorized in the following way to indicate levels of severity with regard to psychological distress: 0–5 "Healthy," 6–9 "Low level," 10–14 "Mild," 15–19 "Moderate," 20–24 "Moderately Severe," and 25–40 "Severe." The CORE can be used at different intervals to track progress or decline, as it has good test–retest reliability and sensitivity to change (Evans et al., 2002); however, it is not designed to be a diagnostic tool. In addition, the authors recommend that a mean score of 10 be used as indication of clinical levels of distress (Connell et al., 2007). That is, clients scoring below 10 are considered "nonclinical cases" and clients scoring 10 or above are considered "clinical cases" experiencing levels of distress for which one might seek professional support. Reliable change is indicated by a move of 5 or more points, and clinically reliable change also involves a move across the threshold into "nonclinical cases" (Jacobson & Truax, 1991). Internal consistency for the CORE (all items) in this sample was excellent (Cronbach's alpha = .93).

Golombok-Rust Inventory of Marital State. The Golombok-Rust Inventory of Marital State (GRIMS) is a 28-item self-report measure developed in the U.K. that assesses relationship quality in couples, and is filled out individually by both partners (Rust, Bennun, Crowe, & Golombok, 1986). Individuals indicate agreement to each statement using a 4-point Likert scale ranging from 0 "Strongly disagree" to 3 "Strongly Agree." Total scores can range from 0–84, with higher scores indicating poorer relationship quality. Norm and criterion referencing were used to produce a transformed score, based on which relationship quality can be categorized in the following way: 0–16 "undefined," 17–21 "Very Good," 22–25 "Good," 26–29 "Above Average," 30–33 "Average," 34–37 "Poor," 38–41 "Bad," 42–46 "Severe Problems," and 47–84 "Very Severe Problems." The reliability of the measure has been shown to be 0.91 for men and 0.87 for women (Rust & Golombok, 2007). The GRIMS has been shown to discriminate between couples who are about to separate and those who are not. Internal consistency for the GRIMS in this sample was good (Cronbach's alpha = .88). There are no published guidelines on how to judge clinical and reliable change, but our calculations based on the standardization values given in Rust and Golombok

(2007) suggest a functional/dysfunction threshold of 29 for men and 31 for women which equates to the "Above-average/average" band. We estimate that the amount of movement needed for reliable change is 11 points.

Participants and Procedure

The sample described in this paper ($N = 877$) comprised 508 female (57.9%) and 369 male (42.1%) clients who were both individually distressed (CORE scores within the clinical range) and relationally distressed (those falling within the "bad" to "very severe" range with respect to GRIMS scores) according to their pretherapy questionnaire responses. All clients in the sample had attended two or more sessions as a couple, and had finished their therapy. The majority of female and male clients fell into the 26–35 (39.1% and 30.8%, respectively) or 36–45 (38.3% and 42.3%, respectively) age-groups. Most clients identified as White (White British/Irish/other White background: 77.0%), 7.5% identified as Asian (which in the U.K. context refers to people whose heritage is the Indian subcontinent), 4.9% identified as being from a Mixed background, 6.4% identified as Black, 0.8% identified as Chinese, and 3.4% selected the "Other ethnic background" option. 76.0% were in full or part time employment. Most clients identified as heterosexual (93.9%); of the 3.7% that identified as gay, just over half (55.2%) were lesbian clients. All clients were attending therapy with their partner. 58.4% were married or in a civil partnership, 28.7% were cohabiting, and 7.3% were with noncohabiting partners. Only a small proportion (5.7%) indicated that they were separated or divorced at intake. The majority of clients had been in a relationship for between 1 and 5 years (29.3%), or 6–10 years (30.9%). Around half of the clients (51.0%) reported having children under the age of 18; for those with children from their current relationship, the mean number of children was 1.8 (*SD* .77). A small percentage (9.5%) had children from previous relationships, with a mean number of children of 1.5 (*SD* .69). Descriptive statistics for clients' psychological distress (CORE) and relationship satisfaction (GRIMS) according to gender and ethnicity are presented in Table 1.

The Therapists

Clinicians were the usual staff in our clinic: a mix of Masters-level couple therapy trainees and qualified couple therapists who had graduated from our training. We use only our own trainees and graduates in our clinic. Some clinicians had previous therapy qualifications as child or individual adult psychotherapists, or as social workers or psychologists, but most had a variety of other nonclinical backgrounds and life experience such as journalism, law, banking, HR, or academia, as is common in the U.K. All clinicians were graduates. The majority were in their 40s and 50s (with a mean of 50 and a range of 26–71 years old). 45.20% of cases in our cohort were seen by qualified couple therapists and 54.80% by couple therapists in training—because the cohort has been gathered over time, a therapist may have seen one case when a trainee and another when qualified. The level of couple therapy experience ranged from 0–24 years. The therapist cohort changes in its make-up slightly each year as students graduate from the training and others enter, but in general more than 60% are White women (as is common in psychodynamically oriented settings in the U.K. and similar to Northey's survey of marital and family therapists in the United States (Northey, 2002).

Table 1
Pre- and Post-Group Means and Standard Deviations for All Measures, According to Client Gender and Ethnicity ($N = 877$)

	CORE				GRIMS			
	Pretherapy		Posttherapy		Pretherapy		Posttherapy	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
All clients	16.35	4.50	11.10	5.93	47.62	6.97	41.13	12.32
Male	15.99	4.37	11.01	5.84	46.26	6.28	41.29	11.67
Female	16.61	4.58	11.15	5.98	48.61	7.28	44.22	12.59
White	16.33	4.42	10.73	5.67	47.57	6.94	42.80	12.80
Non-White	16.38	4.68	11.79	6.19	48.17	7.18	42.96	11.38

Note. CORE = Clinical Outcomes in Routine Evaluation; GRIMS = Golombok-Rust Inventory of Marital State.

Treatment Description

The couple therapy in Tavistock Relationships' standard clinical service is psychodynamically oriented (Fisher, 1999; Ruszczynski, 1993b; Scharff & Savege Scharff, 2014), focusing on insight and emotional connection and expression within the context of a therapeutic relationship that recognizes the impact of the unconscious on thinking, feeling, and behavior. It does not set a formal agenda for sessions, nor does it teach communication skills or use homework and other exercises to change or reinforce behaviors. Couples come for an initial assessment meeting in order to understand more about their difficulties and to confirm that couple therapy at Tavistock Relationships is a suitable treatment and there are no undue risks to themselves, their children, or others in them entering therapy. They are then allocated to any in-training or qualified couple therapist with a matching vacancy. The couple therapists are trained to listen to the couple in a way that attends to unconscious dynamics between the couple, based on each partner's individual life experience and family-of-origin influences, and to intervene through interpretations that make these unconscious structures and influences known. In the course of doing this, communication, behavior, and emotional expression are all variously subjects of the therapy in order to understand the couple relationship and to ensure that the couple understand the underlying meaning of their interaction. The therapist's own internal thoughts and feelings (countertransference) are also used as potential evidence as to what is going on under the surface with the couple. There is an assumption that the psychodynamics of the couple relationship work in an open systems way, and are therefore amenable to influence and change. The overall technique is to make the *relationship*, rather than the two individual partners, the object of the therapy. Therapists will move their attention between these three in a fluid but deliberate way, depending on the needs of the individual partners and the therapeutic task at any one point. Sessions are not audio or video recorded and supervision is done on the basis of therapists' detailed write-ups of the session. The essential elements of the model are itemized in italics in the following clinical vignette, as a way of bringing them to life.

The Example of Adrian and Carol[1]

The presenting problem is not seen automatically as the "real" problem

A heterosexual White couple in their early 30s, Adrian and Carol, presented on the edge of breaking up complaining of bitter arguments that appeared to have their roots in the lack of a sexual life between them. Both scored as cases for individual and relationship distress. It appeared that Adrian wanted sex, but was always rejected by Carol who complained that he was not approaching her in the right way or at the right time. She scolded and scorned him angrily whenever this happened, and he felt beaten and guilty. They alternately attacked and cold-shouldered each other. One night, drunk, Carol made all the running and "seduced" Adrian and they had what each of them agreed was satisfying sex. The arguments between them got worse, however, and Adrian no longer made any sexual overtures at all, which suggested that the problem was not simply a sexual one.

The therapist's countertransference is a key source of information about the couple

The therapist was a man about 10 years older than them and he found himself alternately feeling really close and comfortable with the couple, sympathetic to their difficulties, and then bored, distant, and lifeless. He noted these feelings inside himself as the couple spoke about their interactions, and he realized that his countertransference was a kind of parallel to the problem that the couple brought: closeness and liveliness was followed by distance and deadliness.

The couple's histories are attended to as information about their interaction and the initial fit between the couple is identified

Carol was the older of two daughters, and her 18-month younger sister had been born with a congenital disability which had taken up a lot of her mother's attention. Carol's father had worked long hours to support the family, and Carol's sister died from her disability when Carol was 10; her parents split-up 6 months later. Carol's mother became very depressed and Carol spent her teenage and early adult years looking after her. Her father met another woman quickly after the break-up and moved to another part of the country where they soon had a family of their own. Adrian was the second child of 6. He had an older sister, followed by two brothers and another sister. There were less than 2 years between each of the children. He said that his parents were very close and that the family was lively and loud, with much talking and arguing and connecting to each other. Carol had been welcomed and made a part of the extended family with ease. She enjoyed the rapid-fire connections and conversations that went on—so different to her own family which was silent and disconnected in comparison. For his part, Adrian said the he had been attracted to Carol's stillness and self-reliance, as it was a welcome relief from the flurry of his own family. The therapist could see the way the couple fitted together: each got the thing that had been missing from their families of origin. He could also see how getting it could become difficult and the very thing that attracted each to the other could be the thing that pushed them apart: Adrian's easy excitement could feel impinging and irritating to Carol, and Carol's habitual quietness could feel disengaged and rejecting to Adrian.

Surface progress can mask deeper difficulties which then repeat

The therapist had talked to the couple, making explicit links to their families of origin, and it made sense to them of the ways in which they interacted. They made progress in that they were able to take the other's personality more into consideration, to recognize when they were becoming "too much themselves" and to remain, generally, more connected to each other. When they now argued, however, things were much worse, and they appeared to distance themselves from the therapist. The therapist realized that these changes were directly related to the couple feeling closer again, and he recognized the parallel with his countertransference feeling. He noted that while the couple moved from closeness to passionate fury, his feelings moved from closeness to deadness, not the liveliness of rage. He felt he was picking something more up—a projection of disallowed feelings or parts of the internal self of the couple.

The couple's problems are formulated as an unconscious dynamic

The psychoanalyst Henry Ezriel (1956) suggested a psychodynamic formulation that says that a particular kind of internal relationship is required in order to avoid another kind of relationship that is linked in the mind to an unknown catastrophe. This "required-avoided-catastrophe" template of the internal states of individuals is applied at Tavistock Relationships to the dynamics in couple relationships. Adrian and Carol were repeating over and over again the move away from closeness and back to angry rejection, despite their stated wishes, suggesting that rejection was a preferable relationship to closeness between them. What then was the unconscious catastrophe that needed to be avoided?

The repetition of the unconscious dynamic is understood by reference to the presenting problem, the couple's histories, the current state of the therapy (including the transference), and the therapist's countertransference. Interpretation is used to make this conscious and so enable change

The therapist linked his countertransference to elements of the couple's individual histories, to their presenting problem, and to their reactions to the success of the therapy in making them closer. He interpreted that, as a couple, they were caught by two contrasting and contradictory needs. The first need was to connect to each other, to give each other the things that they each had not received during their growing-up. This was what had brought them together and that kept them returning to each other after the horrible rows—the passion for connection was there in the way they fought. However, the second need was opposed to this and was the need to not have a child. He noted that the deepening of the arguments had happened when they began to talk about the future possibility of becoming parents, and that they had unconsciously divided-up the responsibility for not having sex (which could produce children) between them: the rule seemed to be that the relationship should not be sexual. One of them would carry the sexual feeling for the couple and the other would successfully reject it. This was shown by their response to Carol's breaking the rule and becoming sexually active: they switched from Carol rejecting sex to Adrian rejecting it, keeping the overall couple dynamic the same. The therapist went on to say that Carol's experience was that childbirth brought disaster in the form of a

[1] This case is a composite based on Ruszczynski (1993a). It has been reviewed by a number of Tavistock Relationships' couple therapists who agree it is representative of the kinds of case seen in the clinic and the way of working with them.

damaged sister who took up all her mother's time and attention and deprived Carol of both her mother and her father. Indeed, her father disappeared into work and then into another family with new children that took him away. Her mother disappeared into depression and Carol had to stay enmeshed with her as an adult rather than separate into a family of her own. For Adrian, the danger posed by having a child was that of a repetition of being crowded-out by yet another sibling who took the parents' attention away from him even more. He would lose Carol if they had a child. He appeared to welcome the excitement and bustle of a large family but had chosen as his partner someone much more disconnected and isolated. This seemed to be for two reasons: first, that he would get her sole attention; and second, that she could carry all his denied depressed feelings as she was so familiar with them from her own background. As a result, their relationship—difficult and painful as it was—served a specific function for each of them.

The couple are enabled to see their relationship as a 3rd entity and to use it as a resource for them individually and together

As a result of the intervention, the couple were more able to be in touch with their mixed feelings about relating to another person and to hold them in mind in a more mature way—their emotions could now be experienced and thought about, not just acted on (Klein, 1946). They developed a "couple state of mind" (Morgan, 2004) rather than just an individual perspective and began to see their relationship as a thing that was separate from them individually and which needed work and investment. When they were upset with each other, they could make the effort to reconnect because the relationship was more important than their individual feeling. They were now able to tolerate further work on the movement between liveliness and deadliness between them in different aspects of the relationship and by the end of the therapy they were each more individually psychologically well and better satisfied as a couple. They were no longer scoring above the clinical cut-off in either domain.

Data Analysis

The sample's data were analyzed with hierarchical linear modeling (HLM; Raudenbush & Bryk, 2002), also called multilevel modeling (Snijders & Bosker, 1999) using Stata version 13.1 (StataCorp LP, 2013). HLM allows researchers to study the trajectory of individual change over time, and has several advantages over more traditional statistical techniques. First, HLM is well suited to the study of couple data because it accounts for the dependency between data from individuals within a couple. Second, HLM does not assume that individuals have been assessed at identical and equally spaced time points. This is rarely the case in naturalistic settings. The present paper uses data from questionnaires completed at clients' assessment (baseline) and at their last session or—if data from this is missing—the last time at which they completed measures, the length between which varies for each individual. Third, HLM does not require that individuals provide data at each interval. Unlike more traditional statistical methods such as repeated measures analysis of variance, in which an individual's entire dataset is discounted from the analysis if they have one data point missing, HLM accounts for missing data provided that the data is "Missing at Random."

Because we are only looking at clients' change over two time points in the present paper (baseline and the last measurement point), we characterized individual change using a linear model, which indicates whether an individual's score on a certain measure has increased or decreased over time in a consistent fashion. We used a three-level model with repeated measures, individuals, and the couple unit as the three levels. Individuals' data were fitted with an intercept, indicating their initial score on a given measure, and a slope, indicating their rate of change over time on that measure. Data were represented at the individual level by random effects for the intercept, and by random effects for both the intercept and the slope at the couple level as recommended by Atkins (2005).

Informed by previous research pointing to the importance of therapist characteristics in psychotherapy outcomes, we looked into the possibility of including the therapist as a fourth level in the model, clustering couples within therapists. We found, however, that this did not improve the fit of the model. It involved reducing the sample size to $n = 490$ (55% of the sample) because nearly half the clients predated the customer relationship database in which therapists' names were available. The data were relatively sparse—cluster sizes were small with the majority of these therapists covering less than five cases (71%) and with around 20% of therapists covering just one case there was a high proportion of singletons. When group (cluster) sizes are very small, studies have shown that parameter estimates may be distorted (fixed and random effects—e.g., Bell, Morgan, Kromrey, & Ferron (2010); Theall et al. (2011)). Akaike information criteria (AIC) values of −2308.83 with the therapist level included and 2308.88 without the therapist level strongly suggested that the model fit was not improved with this fourth level. We also found that the therapist level only accounted for 0.008% of the variance in CORE scores. We concluded that, with the data available to us for this sample, couples belonging to a particular therapist do not appear to be reacting to treatment in a more similar way than they are to couples being seen by a different therapist—that there was no apparent therapist effect.

For both psychological distress and relationship satisfaction, we analyzed the data to investigate first whether couples had reported overall change between pre- and posttherapy, and second, to examine whether there was a difference in the amount of change according to client gender or ethnicity. Because of the very low number of clients belong to each of the nonwhite ethnic groups, in order to investigate the effect of ethnicity on reported change over therapy we started by collapsing all of the nonwhite ethnic groups in to one single group. In effect, this first look at the impact of ethnicity was a comparison of white ($n = 613$) and nonwhite clients ($n = 183$). Eighty-one clients did not give information on their ethnicity. If there was a substantial difference between the two groups with respect to the amount of change in either of the two dependent variables we would investigate this further by refining the nonwhite category. The statistics relevant to interactions between gender or ethnicity and change during therapy are noted in Table 2, but we describe only statistically significant interactions in the text below.

Because the model of couple therapy at Tavistock Relationships does not have a prescribed length, couples attended varying numbers of sessions ($M = 23.3$, $SD = 23.5$, range = 2–150). The possibility that more sessions of couple therapy is associated with more improvement (Ward & McCollum, 2005), and so could skew the results, was accounted for by including in each model the

Table 2
Estimated Model Parameters for All Measures, Showing Effect of Client Gender and Ethnicity on the Intercept, Effect of Time (Slope), and the Effect of Client Gender and Ethnicity on Slope (N = 877)

	Intercept				Slope				Gender × Time Interaction			
Measures	*B*	*SE B*	z	95% CI	*B*	*SE B*	z	95% CI	B	*SE B*	z	95% CI
Gender												
CORE	−.68	.32	−2.15*	[−1.29, −.06]	−4.93	.31	−15.71***	[−5.55, −4.32]	.60	.53	1.14	[−.43, 1.63]
GRIMS	−2.44	.49	−5.01***	[−3.40, −1.49]	−4.28	.66	−6.47***	[−5.58, −2.98]	.03	.89	.03	[−1.72, 1.77]
Ethnicity												
CORE	.07	.39	.18	[−.69, .83]	−5.10	.37	−13.71***	[−5.77, −4.33]	.35	.69	.51	[−1.01, 1.71]
GRIMS	.63	.62	1.03	[−.58, 1.84]	−4.64	.78	−5.93***	[−6.17, −3.11]	−.60	1.36	−.44	[−3.26, 2.07]

Note. CORE = Clinical Outcomes in Routine Evaluation; GRIMS = Golombok-Rust Inventory of Marital State.
* $p < .05$. *** $p < .001$.

number of sessions centered around the grand mean. Effect sizes were calculated by multiplying the square root of the sample size by the standard error of the constant.

Data Completeness/Missing Data

As is often the case in naturalistic settings, it has not been possible to obtain end-of-therapy data from every client. Despite an organizational commitment to rigorous outcome monitoring, situations in which clients terminate therapy unexpectedly, do not return, or simply do not wish to complete outcome monitoring forms are unavoidable and no couple is denied therapy just because they do not comply with ongoing measures. Tavistock Relationships, however, routinely obtains end-of-therapy data from between 40 and 50% of clients, a comparatively large proportion of clients for clinics delivering similar services (Bewick, Trusler, Mullin, Grant, & Mothersole, 2006). In the present sample, end-of-therapy data were available from 48.41% ($n = 425$) of clients, with in-therapy data available for the remainder ($n = 452$). As the progress of Tavistock Relationships' clients is monitored at regular intervals throughout therapy, where end-of-therapy data were missing but other postbaseline data were available we used the last point at which a questionnaire was completed for our follow-up data. Although this can only provide approximation of clients' self-assessments following their final session of therapy, it allows us more information for statistical modeling. Follow-up CORE and GRIMS data (i.e., end-of-therapy data or otherwise) were available for 48.2% ($n = 423$) and 47.0% ($n = 412$) of our final population, respectively, meaning that 49.08% had at least either a follow up CORE or GRIMS form, and 47.0% ($n = 412$) had both. Clients with only baseline data in one of the measures and postbaseline data in the other were nevertheless included in the analysis as afforded by HLM. This step was taken in order to avoid any potential bias associated with including only clients who are compliant with outcome monitoring, or only "completers." Missing-data analyses were conducted to explore whether there was a perceptible pattern or mechanism, such as levels of distress/dissatisfaction at intake, behind the missing data. The analyses indicated that data were Missing At Random (Little & Rubin, 2002; Rubin, 1976), and logistical regression showed that there was no influence of baseline scores on the likelihood of follow-up measures being completed. Therefore, we proceeded with the HLM analysis using maximum likelihood estimation. Maximum likelihood estimation does not discard nor impute missing data, but uses the available data to estimate parameters of interest for the entire sample (Baraldi & Enders, 2010). As a result, we were able to make use of a sample of $n = 877$ of clients who had given us at least baseline and one other timepoint data for at least one of the measures and we, in effect, did an intention to treat analysis. We felt that to do otherwise would be to distort the reality of data collection in naturalistic settings and populations.

Results

Psychological Distress

Overall, clients reported a significant decrease in individual psychological distress over the course of therapy ($B = -4.99$, $SE = .31$, $z = -16.25$, $p < .001$). This corresponds to an effect size of $d = -1.04$, considered "large" according to Cohen's criteria (Cohen, 1988). Next we examined whether there was a difference in the amount of pre- to posttherapy change reported by male and female clients. Despite the significant impact of client gender on the intercept, such that female clients reported significantly greater psychological distress than male clients prior to starting therapy (B = −.68, $SE = .32$, z = −2.15, $p < .05$), there was no significant effect of client gender on the slope. In other words, both female and male clients reported a similar improvement in terms of psychological distress. There was no impact of client ethnicity on either baseline distress or the amount of change in distress reported posttherapy. Reliable and clinically significant change is reported in Table 3.

Relationship Satisfaction

Overall, clients also reported a significant increase in relationship satisfaction between their initial assessment prior to therapy and their final assessment point ($B = -4.19$, $SE = .65$, $z = -6.41$, $p < .001$). This reflects a "medium" overall effect size of $d = -0.58$. Next, we investigated whether there was a difference in the progress of male and female clients with regard to relationship satisfaction. Again, although female clients reported significantly lower levels of satisfaction with their relationship prior to starting therapy as indicated by the significant effect of client gender on the intercept (B = −2.44, $SE = .49$, z = −5.01, $p < .001$), there was no impact of client gender on the slope. That is,

Table 3
Proportion of Clients Demonstrating Movement Into the Functional Range, Reliable Change, Clinically Significant Change, and Deterioration

	CORE (%) (N = 432)				GRIMS (%) (N = 412)			
	Moving into functional distribution	Reliable change	Clinically significant change	Deterioration[a]	Moving into functional distribution	Reliable change	Clinically significant change	Deterioration[b]
All clients	46.3	49.9	36.6	3.6	12.6	26.0	12.1	7.8
Male	44.6	47.1	33.8	1.9	11.7	22.1	11.7	5.8
Female	47.4	51.5	38.4	4.5	13.2	28.3	12.4	8.9

Note. CORE = Clinical Outcomes in Routine Evaluation; GRIMS = Golombok-Rust Inventory of Marital State.
[a] An increase of 5 points or more. [b] An increase of 11 points or more.

both male and female clients reported similar improvements in relationship satisfaction over the course of therapy. As with clients' individual distress, there was no impact of client ethnicity on either baseline relationship satisfaction or the amount of change in relationship satisfaction reported posttherapy. Reliable and clinically significant change is reported in Table 3.

Discussion

This is the largest naturalistic prospective study of the outcomes of couple therapy for adults experiencing both relationship and individual distress in the literature to date as far as we know. As such, there can be some confidence that its results are not statistical anomalies stemming from low numbers. It shows that psychodynamic couple therapy is effective in reducing comorbid relationship distress and psychological distress from intake to end of therapy in a nonmanualized, open-ended, clinical service as measured by validated instruments. It raises two broad questions: first, how its results compare with the other field effectiveness studies; and second, what it says about the applicability of RCT benchmarks to couple therapy in ordinary practice. Table 4 summarizes the effectiveness studies in the literature to date and indicates the type of therapy, the number of participants, and the effect sizes of outcomes on relationship distress and psychological distress.

It is clear that this large naturalistic study of psychodynamic couple therapy at Tavistock Relationships has one of the largest effect sizes for change in relationship distress and for change in individual psychological distress reported in the effectiveness literature to date. It is striking that, like Lundblad and Hansson's, 2006 study of integrative couple therapy, the impact on relationship distress is the same as that estimated for RCTs by Shadish and Baldwin in 2005, even though RCTs are generally agreed to produce higher results because of the ways in which they select their participants (Shadish, Ragsdale, Glaser, & Montgomery, 1995).

It suggests that, as in many of the other effectiveness studies, gender has an influence on distress at presentation in heterosexual

Table 4
Couple Therapy Effectiveness Studies and Their Results

Study	Type of couple therapy	(N)	Relationship distress ES (d)	Psychological distress ES (d)
Anker et al. (2009)	Solution-focused, narrative, cognitive–behavioral, humanistic, systemic with feedback (*f*) and without (*w*)	410	.52*f* .21*w*	1.05*f* .44*w*
Balfour & Lanman (2012)	Psychodynamic	36	*ns*	.64
Doss et al. (2012)	Problem solving & communication; goal-focused behavioral/ cognitive/emotion-focused	354	.43	
Hahlweg & Klann (1997)	Integrative, systems-communication, psychodynamic, gestalt, behavioral, Rogerian	252	.37	.46
Klann et al. (2011)	Integrative, systems-communication, psychodynamic, gestalt, behavioral, Rogerian	230	.52	.72
Knobloch-Fedders et al. (2015)	Integrative problem-centered metaframeworks	96	.39	.80
Kuhlman et al. (2013a)	Systemic	51	.17	1.55[a]
Lundblad & Hansson (2006)	Integrative (combining systemic, psychodynamic, cognitive, & solution-focused)	262	.58	.72
Reese et al. (2010)	Family systems approach with feedback (*f*) and without (*w*)	92		.94*f*.38*w*
Current study	Psychodynamic	877	.58	1.04
Weighted effect size across all studies			N = 2,532 d = .51*f* d = .46*w* d = .48 mean	N = 2,306 d = .88*f* d = .75*w* d = .82 mean

[a] Only one partner with a diagnosis of depression was given this measure (SCL-90) n = 25.

couples—with women tending to be more distressed than men—but that it does not affect how much change is experienced, suggesting that both women and men can benefit from it. Although results by ethnicity were not reported on in many of the effectiveness studies, our analysis did not show a difference by ethnicity, suggesting that the better results for African Americans found by Doss et al. (2012) in their veterans' study might be due to something particular about their intervention or setting, though it cannot be definitive about this. It certainly indicates that couple therapy is not just for White/Caucasian couples, but it leaves open the continuing question as to how to identify who is likely to benefit most from couple therapy (Baucom, Atkins, Rowe, Doss, & Christensen, 2015).

It underlines the interactions found by other researchers between relationship distress and individual psychological distress and physiological illness, including onset, course, and likelihood of remission (Baucom, Belus, Adelman, Fischer, & Paprocki, 2014; Whisman, 2007; Whisman & Baucom, 2012). What makes individual psychological distress a particularly important measure of the impact of couple therapy is that it not only captures the impact of distressed relationships but it is also applicable to couples who are planning on separating as well as those working to stay together, as Sparks (2015) notes. Couples who are planning or working toward separation will not be aiming at making the relationship better. As a result, we would echo Owen and colleagues' suggestion that individual measures may be better than those that capture relationship outcomes as indicators of the success of couple therapy (Owen, Duncan, Anker, & Sparks, 2012). Certainly in this current study the CORE-OM measure of individual psychological functioning showed the most change, suggesting an area for future research as to why this is the case.

Results for reliable and clinically significant change given in Table 3 are similar to those reported by Halford et al. (2016) in their comparison of efficacy and effectiveness studies, though not all studies give reliable change separately from clinically significant change, or report on both relationship distress and individual distress, making a precise comparison difficult. It is clear that psychodynamic couple therapy, like other forms of couple therapy, clearly brings about individual and relationship change for distressed partners.

The current study is also of particular interest in the context of debates about the benefits and disadvantages of improving therapy outcomes via session-by-session monitoring (de Jong, van Sluis, Nugter, Heiser, & Spinhoven, 2012; Miller, Hubble, Chow, & Seidel, 2015), as our results, without such monitoring, were very similar to those reported in Anker et al.'s Norwegian study (Anker et al., 2009) for their feedback arm, though it leaves open the question as to what added benefits or disadvantages there might be to using such a system in Tavistock Relationships' clinical service.

This study has limitations that need to be mentioned. As a prepost study, it has not tracked the process of change, nor has it been able to conduct posttherapy follow-up to see if the changes reported have been maintained. The client group were mixed with comorbidity, and with at least some willingness to pay for therapy (even if that amount is nominal). The therapists were mixed in experience. The clinical model was psychodynamic, so the results may not transfer to nonpsychodynamic models (though research on the difference between therapeutic outcomes across models suggests that it may well do). There was no control group and no randomization. Fidelity adherence was not strictly controlled, though there was frequent supervision. Measures were self-report and not observational. All this, of course, reduces internal validity but, as Halford and colleagues point out (Halford et al., 2016), clinicians and the couples they see are more interested in external validity, as this represents the world in which they live and engage in therapy.

This large study suggests that psychodynamic couple therapy is at least as effective an intervention for individual and relationship distress as any other couple therapy that has been tested for effectiveness, that its results are similar to those found in RCTs, and that it is likely to be efficacious if tested in an RCT. It gives a clear positive answer to Lebow and colleagues' laboratory-centric question as to whether psychodynamic therapy could possibly be effective given that it does not "share a common ground" in the strategies and techniques of those therapies that have been tested successfully in RCT efficacy studies (Emotionally focused therapy; Behavioral Couple Therapy; and Integrative Behavioral Couple Therapy; Lebow, Chambers, Christensen, & Johnson, 2012, p. 159).

More broadly, this survey of effectiveness studies suggests that couple therapy around the world is being delivered effectively with a range of therapeutic strategies and techniques—most of which are not those which have been used in RCTs—to a range of different couples, cohabiting and married, with different presenting problems, measured in similar but not identical ways. As such, it strongly supports the view that results from field couple therapy naturalistic effectiveness studies need to have greater recognition and weight in discussions about which couple therapies work, and that the debate must not be left to unrepresentative laboratory efficacy studies alone.

References

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

Atkins, D. C. (2005). Using multilevel models to analyze couple and family treatment data: Basic and advanced issues. *Journal of Family Psychology, 19,* 98–110. http://dx.doi.org/10.1037/0893-3200.19.1.98

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

Baraldi, A. N., & Enders, C. K. (2010). An introduction to modern missing data analyses. *Journal of School Psychology, 48,* 5–37. http://dx.doi.org/10.1016/j.jsp.2009.10.001

Baucom, B. R., Atkins, D. C., Rowe, L. S., Doss, B. D., & Christensen, A. (2015). Prediction of treatment response at 5-year follow-up in a randomized clinical trial of behaviorally based couple therapies. *Journal of Consulting and Clinical Psychology, 83,* 103–114. http://dx.doi.org/10.1037/a0038005

Baucom, D. H., Belus, J. M., Adelman, C. B., Fischer, M. S., & Paprocki, C. (2014). Couple-based interventions for psychopathology: A renewed direction for the field. *Family Process, 53,* 445–461. http://dx.doi.org/10.1111/famp.12075

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

Bell, B. A., Morgan, G. B., Kromrey, J. D., & Ferron, J. M. (2010). The impact of small cluster size on multilevel models: A Monte Carlo examination of two-level models with binary and continuous predictors. *JSM Proceedings, Survey Research Methods Section,* 4057–4067. Retrieved from https://www.amstat.org/sections/srms/proceedings/y2010/Files/308112_60089.pdf

Bewick, B. M., Trusler, K., Mullin, T., Grant, S., & Mothersole, G. (2006). Routine outcome measurement completion rates of the CORE-OM in primary care psychological therapies and counselling. *Counselling and Psychotherapy Research, 6,* 33–40. http://dx.doi.org/10.1080/14733140600581432

Carey, T. A., & Stiles, W. B. (2016). Some problems with randomized controlled trials and some viable alternatives. *Clinical Psychology and Psychotherapy, 23,* 87–95. http://dx.doi.org/10.1002/cpp.1942

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York, NY: Erlbaum.

Connell, J., Barkham, M., Stiles, W. B., Twigg, E., Singleton, N., Evans, O., & Miles, J. N. (2007). Distribution of CORE-OM scores in a general population, clinical cut-off points and comparison with the CIS-R. *The British Journal of Psychiatry, 190,* 69–74. http://dx.doi.org/10.1192/bjp.bp.105.017657

de Jong, K., van Sluis, P., Nugter, M. A., Heiser, W. J., & Spinhoven, P. (2012). Understanding the differential impact of outcome monitoring: Therapist variables that moderate feedback effects in a randomized clinical trial. *Psychotherapy Research, 22,* 464–474. http://dx.doi.org/10.1080/10503307.2012.673023

Dirmaier, J., Harfst, T., Koch, U., & Schulz, H. (2006). Therapy goals in inpatient psychotherapy: Differences between diagnostic groups and psychotherapeutic orientations. *Clinical Psychology and Psychotherapy, 13,* 34–46. http://dx.doi.org/10.1002/cpp.470

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

Evans, C., Connell, J., Barkham, M., Margison, F., McGrath, G., Mellor-Clark, J., & Audin, K. (2002). Towards a standardised brief outcome measure: Psychometric properties and utility of the CORE-OM. *The British Journal of Psychiatry, 180,* 51–60. http://dx.doi.org/10.1192/bjp.180.1.51

Evans, C., Mellor-Clark, J., Margison, F., Barkham, M., Audin, K., Connell, J., & McGrath, G. (2000). CORE: Clinical outcomes in routine evaluation. *Journal of Mental Health, 9,* 247–255. http://dx.doi.org/10.1080/713680250

Ezriel, H. (1956). Experimentation within the psycho-analytic session. *British Journal for the Philosophy of Science, 7,* 29–48. http://dx.doi.org/10.1093/bjps/VII.25.29

Fisher, J. (1999). *The uninvited guest. Emerging from narcissism towards marriage*. London, United Kingdom: Karnac.

Hahlweg, K., & Klann, N. (1997). The effectiveness of marital counseling in Germany: A contribution to health services research. *Journal of Family Psychology, 11,* 410–421. http://dx.doi.org/10.1037/0893-3200.11.4.410-421

Hahlweg, K., Markman, H., Thurmaier, F., Engl, J., & Eckert, V. (1998). Prevention of marital distress: Results of a German prospective longitudinal study. *Journal of Family Psychology, 12,* 543–556. http://dx.doi.org/10.1037/0893-3200.12.4.543

Halford, W. K., Pepping, C. A., & Petch, J. (2016). The gap between couple therapy research efficacy and practice effectiveness. *Journal of Marital and Family Therapy, 42,* 32–44. http://dx.doi.org/10.1111/jmft.12120

Hewison, D., Clulow, C., & Drake, H. (2014). *Couple therapy for depression: A clinician's guide to integrative practice*. Oxford: Oxford University Press. http://dx.doi.org/10.1093/med:psych/9780199674145.001.0001

Hewitt, B., & Baxter, J. (2014). Relationship dissolution in Australia. In D. Arunachalam & G. Heard (Eds.), *Family formation in 21st century Australia* (pp. 77–99). New York, NY: Springer.

Hill, C. E., & Lambert, M. J. (2004). Methodological issues in studying psychotherapy processes and outcomes. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (5th ed., pp. 84–135). New York, NY: Wiley.

Jacobson, N. S., & Truax, P. (1991). Clinical-significance - A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59,* 12–19.

Klann, N., Hahlweg, K., Baucom, D. H., & Kroeger, C. (2011). The effectiveness of couple therapy in Germany: A replication study. *Journal of Marital and Family Therapy, 37,* 200–208. http://dx.doi.org/10.1111/j.1752-0606.2009.00164.x

Klein, M. (1946). Notes on some schizoid mechanisms. In M. Klein (Ed.), *Love, guilt and reparation* (pp. 1–24). London, United Kingdom: Hogarth Press.

Knobloch-Fedders, L. M., Pinsof, W. M., & Haase, C. M. (2015). Treatment response in couple therapy: Relationship adjustment and individual functioning change processes. *Journal of Family Psychology, 29,* 657–666. http://dx.doi.org/10.1037/fam0000131

Kuhlman, I., Tolvanen, A., & Seikkula, J. (2013a). Couple therapy for depression within a naturalistic setting in Finland: Factors related to change of the patient and the spouse. *Contemporary Family Therapy: An International Journal, 35,* 656–672. http://dx.doi.org/10.1007/s10591-013-9246-6

Kuhlman, I., Tolvanen, A., & Seikkula, J. (2013b). The therapeutic alliance and couple therapy for depression: Predicting therapy progress and outcome from assessments of the alliance by the patients, the spouse, and the therapists. *Contemporary Family Therapy: An International Journal, 35,* 1–13. http://dx.doi.org/10.1007/s10591-012-9215-5

Leach, C., Lucock, M., Barkham, M., Stiles, W. B., Noble, R., & Iveson, S. (2006). Transforming between Beck Depression Inventory and CORE-OM scores in routine clinical practice. *British Journal of Clinical Psychology, 45,* 153–166. http://dx.doi.org/10.1348/014466505X35335

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,* 145–168. http://dx.doi.org/10.1111/j.1752-0606.2011.00249.x

Leichsenring, F. (2004). Randomized controlled versus naturalistic studies: A new research agenda. *Bulletin of the Menninger Clinic, 68,* 137–151. http://dx.doi.org/10.1521/bumc.68.2.137.35952

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd ed.). Hoboken, NJ: Wiley. http://dx.doi.org/10.1002/9781119013563

Lundblad, A.-M., & Hansson, K. (2006). Couples therapy: Effectiveness of treatment and long term follow up. *Journal of Family Therapy, 28,* 136–152. http://dx.doi.org/10.1111/j.1467-6427.2006.00343.x

Miller, S. D., Hubble, M. A., Chow, D., & Seidel, J. (2015). Beyond measures and monitoring: Realizing the potential of feedback-informed treatment. *Psychotherapy, 52,* 449–457. http://dx.doi.org/10.1037/pst0000031

Morgan, M. (2004). On being able to be a couple: The importance of a "creative couple" in psychic life. In F. Grier (Ed.), *Oedipus & the couple* (pp. 9–30). London, United Kingdom: Karnac.

Northey, W. F. J., Jr. (2002). Characteristics and clinical practices of marriage and family therapists: A national survey. *Journal of Marital and Family Therapy, 28,* 487–494. http://dx.doi.org/10.1111/j.1752-0606.2002.tb00373.x

Owen, J., Duncan, B., Anker, M., & Sparks, J. (2012). Initial relationship goal and couple therapy outcomes at post and six-month follow-up. *Journal of Family Psychology, 26,* 179–186. http://dx.doi.org/10.1037/a0026998

Petch, J., Lee, J., Huntingdon, B., & Murray, J. (2014). Couple counselling outcomes in an Australian not for profit: Evidence for the effectiveness of couple counselling conducted within routine practice. *Australian and New Zealand Journal of Family Therapy, 35,* 445–461. http://dx.doi.org/10.1002/anzf.1074

Philips, B. (2009). Comparing apples and oranges; how do patient characteristics and treatment goals vary between different forms of psychotherapy? *Psychology and Psychotherapy: Theory, Practice, Research, 82,* 323–336. http://dx.doi.org/10.1348/147608309X431491

Pinsof, W. M., Breunlin, D. C., Chambers, A. L., Solomon, A. H., & Russell, W. P. (2015). Integrative, multi-systemic and empirically informed couple therapy: The IPCM perspective. In A. Gurman, J. Lebow, & D. K. Snyder (Eds.), *Clinical handbook of couple therapy* (5th ed., pp. 161–191). New York, NY: Guilford Press.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage.

Reese, R. J., Toland, M. D., Slone, N. C., & Norsworthy, L. A. (2010). Effect of client feedback on couple psychotherapy outcomes. *Psychotherapy, 47,* 616–630. http://dx.doi.org/10.1037/a0021182

Rubin, D. B. (1976). Inference and missing data. *Biometrika, 63,* 581–592. http://dx.doi.org/10.1093/biomet/63.3.581

Rust, J., Bennun, I., Crowe, M., & Golombok, S. (1986). The Golombok Rust Inventory of Marital State (GRIMS). *Sexual and Marital Therapy, 1,* 55–60. http://dx.doi.org/10.1080/02674658608407680

Rust, J., & Golombok, S. (2007). *The handbook of the Golombok Rust Inventory of Marital State (GRIMS)*. London, United Kingdom: Pearson Assessment.

Ruszczynski, S. (1993a). Thinking about and working with couples. In S. Ruszczynski (Ed.), *Psychotherapy with couples: Theory and practice at the Tavistock Institute of Marital Studies* (pp. 197–217). London, United Kingdom: Karnac.

Ruszczynski, S. (Ed.), (1993b). *Psychotherapy with couples: Theory and practice at the Tavistock Institute of Marital Studies*. London, United Kingdom: Karnac.

Scharff, D. E., & Savege Scharff, J. (Eds.). (2014). *Psychoanalytic couple therapy: Foundations of theory and practice*. London, United Kingdom: Karnac.

Seikkula, J., Aaltonen, J., Kalla, O., Saarinen, P., & Tolvanen, A. (2013). Couple therapy for depression in a naturalistic setting in Finland: A 2-year randomized trial. *Journal of Family Therapy, 35,* 281–302. http://dx.doi.org/10.1111/j.1467-6427.2012.00592.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., Ragsdale, K., Glaser, R. R., & Montgomery, L. M. (1995). The efficacy and effectiveness of marital and family-therapy - A perspective from metaanalysis. *Journal of Marital and Family Therapy, 21,* 345–360.

Shean, G. D. (2015). Some methodological and epistemic limitations of evidence-based therapies. *Psychoanalytic Psychology, 32,* 500–516. http://dx.doi.org/10.1037/a0035518

Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London, United Kingdom: Sage.

Sparks, J. A. (2015). The Norway couple project: Lessons learned. *Journal of Marital and Family Therapy, 41,* 481–494. http://dx.doi.org/10.1111/jmft.12099

StataCorp, L. P. (2013). *Stata se* (Version 13.1) [Computer software]. College Station, TX: StataCorp LP. Retrieved from www.stata.com

Stratton, P., Silver, E., Nascimento, N., McDonnell, L., Powell, G., & Nowotny, E. (2015). Couple and family therapy outcome research in the previous decade: What does the evidence tell us? *Contemporary Family Therapy: An International Journal, 37,* 1–12. http://dx.doi.org/10.1007/s10591-014-9314-6

Theall, K. P., Scribner, R., Broyles, S., Yu, Q., Chotalia, J., Simonsen, N., & Carlin, B. P. (2011). Impact of small group size on neighbourhood influences in multilevel models. *Journal of Epidemiology and Community Health, 65,* 688–695. http://dx.doi.org/10.1136/jech.2009.097956

Wampold, B. E. (2001). *The great psychotherapy debate: Models, methods and findings*. Hillsdale, NJ: Erlbaum.

Ward, D. B., & McCollum, E. E. (2005). Treatment effectiveness and its correlates in a marriage and family therapy training clinic. *American Journal of Family Therapy, 33,* 207–223. http://dx.doi.org/10.1080/01926180590932960

Weisz, J. R., Donenberg, G. R., Han, S. S., & Weiss, B. (1995). Bridging the gap between laboratory and clinic in child and adolescent psychotherapy. *Journal of Consulting and Clinical Psychology, 63,* 688–701. http://dx.doi.org/10.1037/0022-006X.63.5.688

Whisman, M. A. (2007). Marital distress and *DSM–IV* psychiatric disorders in a population-based national survey. *Journal of Abnormal Psychology, 116,* 638–643. http://dx.doi.org/10.1037/0021-843X.116.3.638

Whisman, M. A., & Baucom, D. H. (2012). Intimate relationships and psychopathology. *Clinical Child and Family Psychology Review, 15,* 4–13. http://dx.doi.org/10.1007/s10567-011-0107-2

Wright, J., Sabourin, S., Mondor, J., McDuff, P., & Mamodhoussen, S. (2007). The clinical representativeness of couple therapy outcome research. *Family Process, 46,* 301–316. http://dx.doi.org/10.1111/j.1545-5300.2007.00213.x

Received June 17, 2016
Revision received September 15, 2016
Accepted September 23, 2016 ■