

Accounting for Therapist Variability in Couple Therapy Outcomes: What Really Matters?

Jesse Owen

Department of Educational and Counseling Psychology, Counseling, and College Student Personnel, College of Education, University of Louisville, Louisville, Kentucky, USA

Barry Duncan

Heart and Soul of Change Project, Jensen Beach, Florida, USA

Robert Jeff Reese

Counseling Psychology, University of Kentucky, Lexington, Kentucky, USA

Morten Anker

Vestfold Counseling Office of Family Affairs, Tønsberg, Norway

Jacqueline Sparks

Department of Human Development and Family Studies, University of Rhode Island, Kingston, Rhode Island, USA

This study examined whether therapist gender, professional discipline, experience conducting couple therapy, and average second-session alliance score would account for the variance in outcomes attributed to the therapist. The authors investigated therapist variability in couple therapy with 158 couples randomly assigned to and treated by 18 therapists in a naturalistic setting. Consistent with previous studies in individual therapy, in this study therapists accounted for 8.0% of the variance in client outcomes and 10% of the variance in client alliance scores. Therapist average alliance score and experience conducting couple therapy were salient predictors of client outcomes attributed to therapist. In contrast, therapist gender and discipline did not significantly account for the variance in client outcomes attributed to therapists. Tests of incremental validity demonstrated that therapist average alliance score and therapist experience uniquely accounted for the variance in outcomes attributed to the therapist. Emphasis on improving therapist alliance quality and specificity of therapist experience in couple therapy are discussed.

Research confirms what everybody knows: therapists differ in their ability to help clients (Crits-Christoph et al., 1991; Kraus, Castonguay, Boswell, Nordberg, & Hayes, 2011). Across

Address correspondence to Jesse Owen, University of Louisville, Counseling Psychology, College of Education, Louisville, KY 40203, USA. E-mail: jesse.owen@louisville.edu

efficacy and effectiveness studies of individual therapy, therapists account for 5% to 10% of the variance in client outcomes (e.g., Kim, Wampold, & Bolt, 2006; Lutz, Leon, Martinovich, Lyons, & Stiles, 2007; Wampold & Brown, 2005). As Wampold (2005) summarized, “. . . the variance of outcomes due to therapists (8% to 9%) is larger than the variability among treatments (0%–1%), the alliance (5%), and the superiority of an EST to a placebo treatment (0%–4%)” (p. 204). What the psychotherapist brings to the table, therefore, is an important area of investigation. It is surprising that there are few studies that examine therapist variability in couple therapy outcomes (Blow, Sprenkle, & Davis, 2007; Sparks & Duncan, 2010).

Attempts to discern what accounts for therapist variability have resulted in limited success and inconsistent findings. Demographic variables such as gender, race/ethnicity, and professional discipline have not generally shown a relation with therapy outcomes in individual therapy (Beutler et al., 2004; Blatt, Sanislow, Zuroff, & Pilkonis, 1996; Huppert et al., 2001; Maramba & Nagayama Hall, 2002; Okiishi et al., 2006; Wampold & Brown, 2005). Investigations of therapist training and experience have also yielded mixed results (Beutler et al., 2004). Some researchers have concluded that therapist experience is weakly and unreliably related to better therapy outcomes (e.g., Christensen & Jacobson, 1994; Nyman, Nafziger, & Smith, 2010). For example, in a naturalistic study with 581 therapists and more than 6,000 clients, Wampold and Brown (2005) found that therapist experience conducting therapy was not related to outcome.

On the other hand, a meta-analysis by Crits-Christoph and colleagues (1991) found that therapist experience accounted for a significant proportion of the variability among therapists (approximately 25% of the therapist variability). In an interesting analysis, Leon, Martinovich, Lutz, and Lyons (2005) explored whether therapists learn from the experience of treating a particular client with a particular problem and consequently are able to more effectively treat a similar client at a later point in time. Their results indicated that second clients in the matched pairs fared better than the first after an interval of 15–75 days, suggesting that therapists learn, at least in the short run, from experience.

The research regarding the association between specific therapist characteristics and couple outcomes is scant but similar to the findings for individual therapy. Regarding therapist demographics, Bischoff and Sprenkle (1993) did not find evidence that therapist age, gender, and ethnicity affected retention in their study of marriage and family therapy dropouts. Regarding experience, Raytek, McCrady, Epstein, and Hirsch (1999) found a significant positive association between therapist experience and observer ratings of the alliance and completion of treatment for a partner’s substance use, but not for overall outcome.

A criticism often leveled at research investigating therapist experience is that it is not operationally defined according to specific treatment model or client population (Beutler et al., 2004) and that a more sophisticated look at therapist experience may yield more positive findings. In support, some have found that therapist experience may be related to therapy outcomes when the type of treatment and client presentation is considered (e.g., Blatt et al., 1996; Leon et al., 2005; Luborsky, McLellan, Diguier, Woody, & Seligman, 1997). For example, Kraus and colleagues (2011) found that therapists’ competencies can be domain specific; some therapists were on average better than others at treating certain conditions. Consequently, specificity in the definition of experience may be important in understanding the effect of experience on client outcomes.

A relatively new and promising area of research addressing therapist variability is through client ratings of the therapeutic alliance. Baldwin and Imel (2013) meta-analyzed therapist variability in

client alliance scores and found that therapists accounted for approximately 9% of the variance in client-rated alliance across 15 samples ranging from college counseling centers, in-patient units, to couples interventions. The client's view of the alliance—the emotional bond and the agreement between the client and therapist about the goals and tasks of therapy (Bordin, 1979)—is not only a robust predictor of therapy outcomes, including couple therapy (Anker, Owen, Duncan, & Sparks, 2010; Friedlander, Escudero, Heatherington, & Diamond, 2011; Horvath, Del Re, Flückiger, & Symonds, 2011), but also may offer an avenue to understand therapist differences. Marcus, Kashy, and Baldwin (2009) noted:

... high levels of consensus in client ratings of their therapist indicate that clients of the same therapist tend to agree about the traits or characteristics of their therapist, suggesting that there is something about the therapist's manner or behavior that evokes similar response from all of his or her clients. (p. 538)

In the process of examining what differentiates more (or less) effective therapists, their average ability to form alliances can address an important question: Are therapists who are generally better at facilitating positive outcomes more likely to be rated as forming stronger alliances across their clients? Baldwin, Wampold, and Imel (2007), the first researchers to address this question, found only modest therapist variability (2%) compared with other studies, but reported that therapist average alliance quality accounted for 97% of that variability. In a study of couple relationship education, Owen, Rhoades, Stanley, and Markman (2011) found that leaders accounted for a larger proportion of the variance in couples' alliance (29.6%) and that leaders' average alliance quality accounted for nearly all of the variance (99%) in couples' communication and relationship adjustment outcomes attributed to the leaders.

The present study is the first, to our knowledge, to investigate therapist variability in couple therapy. First, we expected that therapists would account for a small to medium proportion of the variance in therapy outcomes and working alliance (Hypothesis 1). Next, we tested whether therapist experience conducting couple therapy, average alliance score, discipline, and gender would account for the variability between therapists in therapy outcomes. We did not expect gender and discipline to be related to outcome, and given the mixed evidence regarding experience, we did not make a formal prediction. We hypothesized that therapist average alliance quality would account for the majority of the variance with less proportion of the variance being attributable to experience conducting couple therapy, discipline, and gender (Hypothesis 2).

METHOD

Participants

The data from this study came from Anker et al. (2010), and a full description of the participant recruitment can be found there. For the purpose of this study, we used a subsample of clients who attended a minimum of three sessions ($n = 316$ clients, 158 couples, and 18 therapists). Three sessions was the minimum number of sessions for clients to have a second-session alliance score, a common time to assess early alliance, and to have a therapy outcome score at a future session. Couples were White, Euro-Scandinavian, heterosexual, and had a mean age of 38.54 years ($SD = 8.47$). Most participants (70%) were employed full-time; and 28.9% had completed lower

secondary school, 33.9% had completed upper secondary school, and 37.2% had completed university or college. Couples had been together an average of 11.8 years ($SD = 8.7$ years, range = 1–39 years). Couples self-referred with a broad range of typical relationship problems, including communication difficulties, loss of feeling for partner, jealousy/infidelity, conflict, and coping with a partner's physical or psychological problem. Diagnosis is not required, nor is it a routine convention in this setting. However, initial distress levels based on the Outcome Rating Scale (ORS; Miller, Duncan, Brown, Sparks, & Claud, 2003) score at pretreatment was 19.01 ($SD = 7.59$), indicative of a clinical population and similar to distress levels of other clinical sites (Duncan, 2011; Miller & Duncan, 2004). In addition, the mean marital satisfaction score on the Locke Wallace Marital Adjustment Test (LW; Locke & Wallace, 1959) was 72.62 ($SD = 24.11$), indicative of a dissatisfied relationship, well under the traditional cutoff score of 100. The average number of sessions for this subsample was 5.38 ($SD = 2.37$).

Therapists

The couples were seen by 18 therapists who worked at one of two family counseling agencies in Norway. There were 11 female therapists and 7 male therapists (all were White, Euro-Scandinavian), 9 of whom were licensed psychologists, 8 were licensed social workers, and 1 a licensed psychiatric nurse. All therapists professed an eclectic orientation, using a variety of approaches—solution-focused, narrative, cognitive-behavioral, humanistic, and systemic—similar to those typically practiced in Norway family counseling agencies. The average age of the therapists was 44 years ($SD = 12.6$ years, range = 26–61 years). The mean years of experience with couple therapy was 7.28 years ($SD = 7.14$ years, range = 0–19 years). The average number of couples treated by each therapist was 8.77 (range = 3 to 18 couples).

Measures

The Outcome Rating Scale (ORS)

Psychological functioning and distress was assessed at the beginning of every session using the ORS (Miller & Duncan, 2004), but analyses were derived from pre- and posttreatment. The ORS is a brief four-item visual analog self-report instrument designed to measure client progress repeatedly throughout the course of therapy. The ORS assesses key areas of client functioning, including personal well-being, interpersonal functioning (i.e., family, close relationships), social (work, school), and overall well-being. The scores are totaled, ranging from 0 to 40, with lower scores reflecting more distress. The reliability and validity of the scores for this measure has been supported in previous studies of individual and couple therapy (e.g., Anker et al., 2010; Campbell & Hemsley, 2009; Miller et al., 2003; Reese, Nosworthy, & Rowlands, 2009). Internal consistency estimates of the ORS have ranged from .84 to .93 (Miller et al., 2003; Reese et al., 2009). In the present sample, the internal consistency of ORS scores was .91. Concurrent validity of ORS scores has been demonstrated in independent studies as adequate through correlates with the Outcome Questionnaire 45.2 (Duncan, 2011; Lambert et al., 1996; $r = .74$, Campbell & Hemsley, 2009; $r = .59$, Miller et al., 2003).

Using formulas developed by Jacobson and Truax (1991), clinical and normative data for the ORS were analyzed by Miller and Duncan (2004) to provide cutoff scores for the reliable change

index and clinically significant change. Using a sample of 34,790 participants, clients who change in a positive or negative (deterioration) direction by at least five points are regarded as having made reliable change. This degree of change exceeds measurement error based on the reliability of the ORS and is one of the two criteria posited by Jacobson and Truax (1991) as indicative of clinically meaningful change. The second criterion requires movement from a score typical of a clinical population to one typical of a functional population. The cutoff on the ORS for marking the point at which a person's score is more likely to come from a dysfunctional population than from a nondysfunctional population is 25 (Miller et al., 2003).

Session Rating Scale (SRS)

The client's perspective of the alliance with the therapist was measured using the SRS (Duncan et al., 2003). The SRS was administered at the end of every session, but the analyses were derived from the second session. The SRS is also a four-item visual analog scale. Specifically the clients rated their therapist on the following items: relationship with the therapist ("I felt heard, understood, and respected."), goals and topics ("We worked on or talked about what I wanted to work on or talk about"), the approach used in therapy (The therapist's approach is a good fit for me"), and the overall rating of the session ("Overall, today's session was right for me"). Clients' scores for the four items were totaled for a score ranging from 0 to 40, with higher scores indicating a stronger alliance.

The SRS was developed from two key sources: (a) Bordin's (1979) classic delineation of the components of the alliance; and (b) Hatcher and Barends (1996) factor analysis of three popular alliance measures where in addition to the general factor measured by all alliance scales (i.e., strength of the alliance), two other factors were predictive: *confident collaboration* and *the expression of negative feelings*. Confident collaboration describes the level of confidence that the client has that therapy and the therapist will be helpful. Although overlapping with question three on the SRS (the fit of the therapist's approach), the fourth item of the SRS directly addresses this factor. The other factor predictive beyond the general strength of the alliance is the client's freedom to voice negative feelings and reactions to the therapist, which is again captured in the nature of the items. The reliability and validity for SRS scores has been demonstrated in previous studies. For example, the reliability estimates have ranged from .88 to .90 for Cronbach's alphas and .54 to .74 for test-retest correlations (e.g., Duncan et al., 2003; Reese et al., 2009). In addition, the validity for SRS scores has been supported through correlations with the Helping Alliance Questionnaire-II, $r = .48$ (Duncan et al., 2003) and the Working Alliance Inventory, $r = .63$ (Campbell & Hemsley, 2009). The SRS has predicted therapy outcomes in similar magnitude with other alliance measures (Anker et al., 2010; Reese et al., 2009). In the present sample, the internal consistency of SRS scores was .89.

The SRS was developed to encourage clinicians to routinely assess and discuss the alliance with clients. The ORS and SRS have a database of more than 400,000 administrations and have been used in three randomized controlled trials comparing feedback and Treatment as Usual (TAU), two of which examined couple therapy¹ (Anker, Duncan, & Sparks, 2009; Reese et al., 2009; Reese, Toland, Slone, & Norsworthy, 2010).

¹The ORS and SRS are free for individual clinician use and can be downloaded at www.heartandsoulofchange.com.

Couple Distress Measures

We used two measures to assess the level of adjustment/distress for couples. Couples were administered these measures at pretherapy only. For our purposes, we tested whether therapists' caseloads differed significantly on these measures to help ensure similarities across therapists' cases.

Locke Wallace Marital Adjustment Test (LW; Locke & Wallace, 1959)

The LW is a commonly used self-report measure to assess marital functioning. The LW is considered a reliable and valid measure of marital satisfaction and still relevant to clinical practice and research (Freeston & Plechaty, 1997). It is highly correlated with the often-used Dyadic Adjustment Scale ($r = .93$; Spanier, 1976). The LW cutoff score of 100, which differentiates satisfied couples from dissatisfied couples, is widely accepted (Christensen et al., 2004; Freeston & Plechaty, 1997). In the present study, the coefficient alpha for the LW was .75. The LW was administered at pretreatment and 6-month follow-up (but not at posttreatment). Accordingly, we used the pretreatment scores to test whether therapists differed in their clients' initial relationship distress.

Goal to Better the Relationship

We used a one-item question to assess whether clients' initial therapy goal was to better their relationship, which was rated dichotomously (yes/no). Clients completed this item on their intake forms and their responses were not available to their partner. Similar to the LW, we used this item to assess whether therapists differed in the clients' desire to work on the relationship at the beginning of treatment.

Procedure

This was a naturalistic study conducted in community-based outpatient centers in Norway. This study was approved by the university's institutional review board. Clients and therapists gave informed consent to participate in a study regarding improving therapeutic practices. Clients were then assigned randomly and weekly to available therapist intake slots. Therapists could exchange one case for another if they felt uncomfortable with the couple's clinical presentation as depicted on the intake paperwork or if they had had previous nonclinical contact with the couple. Such an exchange happened 20 times over the course of the study, primarily because of previous nontherapy contact with the couple.

Therapists worked with couples, using the Partners for Change Outcome Management System (Duncan, 2010, 2012; Miller, Duncan, Sorell, & Brown, 2005) where therapists had access to alliance (SRS) and outcome (ORS) feedback from each spouse every session (Anker et al., 2009). All therapists attended two days of training (8 hr total) before the study and three 3-hr follow-up trainings during the investigation. Therapists were instructed to follow the general protocol outlined in the scoring and administration manual for the ORS and SRS (Miller & Duncan, 2004) as well as the transparent, collaborative process of monitoring outcome and the alliance with clients described in these authors' other publications (e.g., Duncan et al., 2004). Clients were

administered the ORS by the therapists at the beginning of every session. Therapists were also instructed in the use of the SRS (Duncan et al., 2003) to detect potential breaches in the alliance. Toward the end of every session, the SRS was administered to the client and scored allowing therapists to openly discuss any concerns and how the services may better fit client expectations.

Data Analysis

Given the interdependencies in the data and to estimate the variance in therapy outcomes and working alliance attributed to clients, couples, and therapists, we conducted multilevel regression models (i.e., partners nested within couples who were nested within therapists). Specifically, we used multilevel modeling with Bayesian estimation. Bayesian statistical models differ in philosophy and estimation from frequentist (or conventional) statistical models that are typically based on null hypothesis significance testing (see Hamaker & Klugkist, 2011, for a discussion). The key advantages of Bayesian models include (a) the results are not based on normality assumptions or asymptotic results and (b) there is less biased estimation for unbalanced and small sample sizes (Hamaker & Klugkist, 2011). These strengths are important here because our sample size at the therapist level is relatively small and there is an unbalanced number of clients treated by each therapist.² In addition, in contrast with other estimation approaches, in Bayesian models there are credible intervals (vs. confidence intervals). According to Yuan and MacKinnon (2009), “a 95% credible interval means that there is a 95% chance that the credible interval contains the true value of the parameter on the basis of the observed data” (p. 304). Confidence intervals describe a range of scores if the same study was repeated multiple times with samples from the target population (note that the replication process is seldom done in practice; Yuan & MacKinnon, 2009).

RESULTS

We initially conducted three multilevel models to determine whether therapists differed in their clients' distress (ORS-pre), relationship distress (LW-pre), and desire to work on enhancing the relationship (yes/no). In these models, the ORS-pre, LW-pre, and enhancing the relationship were entered as the dependent variables, respectively, at Level 1 (individual client) and there were no predictor variables at Level 2 (couples) or Level 3 (therapists). To determine therapist effects, we divided the proportion of variance in therapy outcomes attributed to therapists by the total variance in the model. Specifically, we used the following equation for therapist intraclass correlation (ICC):

$$ICC_{therapist} = \frac{\tau_{therapist}}{\tau_{therapist} + \tau_{couple} + \sigma_{client}^2}$$

The therapist ICC for ORS-pre, LW-pre, and enhancing the relationship (yes/no) were all less than 1% ($ps > .05$). This suggests that therapists' caseloads did not significantly differ in terms of

²Bayesian models seek to reduce the uncertainty in a model by combining prior knowledge with current data [i.e., $P(\text{theory} | \text{data})$]. The use of prior knowledge is commonly described as the prior distribution that can be represent previous findings in the literature or in cases when there is less previous knowledge to inform the model, the prior distribution is more uncertain or wider (Dienes, 2011). We used diffuse priors for the present study.

TABLE 1
Associations Among Therapist Level Variables (N = 18)

	1	2	3	4
1. SRS-2nd	—			
2. Gender	-.57	—		
3. Experience	-.24	.28	—	
4. Discipline	.13	.12	-.47	—

Note. For SRS-2nd, clients' scores were aggregated. Gender was coded 1 = men, 0 = women. Discipline was coded 1 = psychologist, 0 = social worker/psychiatric nurse.

*p < .05.

their clients' initial distress and desire to work on the relationship, which provides some support for the random assignment to therapists. Table 1 shows the associations among the therapist-level variables (note that these associations account for the interdependences in the data).

First, we tested whether therapists would account for a meaningful proportion of the variance in their clients' second-session SRS scores. We conducted a baseline model, where client second-session SRS score was the outcome variable and there were no predictor variables. Results revealed that the client level variance was 12.74 (95% CI [10.23, 16.09], $p < .05$), couple level variance was 6.01 (95% CI [3.16, 9.85], $p < .05$), and therapist level variance was 2.08 (95% CI [0.38, 7.05], $p < .05$). The ICC for therapists was .10, or, in other words, therapists accounted for 10% of the variance in their clients' alliance scores.

Second, we tested whether therapists would account for a significant proportion of the variance in their clients' therapy outcomes. To do so, we conducted a baseline model for client ORS-post. In Model 1, client ORS posttherapy scores were the dependent variables, and we controlled for client ORS pretherapy scores (grand mean centered at Level 1), but there were no other predictor variables at Level 2 or 3. Using the aforementioned formula, the therapist ICC was .08 for the ORS-post (see Table 2, Model 1). That is, therapists accounted for 8.0% of the variance in clients' therapy outcomes (supporting Hypothesis 1). For a descriptive illustration, Figure 1 shows the

TABLE 2
Summary of Random Effects for Bayesian Multilevel Models

	Model 1 ORS-post	Model 2 + exp only	Model 3 + SRS 2nd only	Model 4 + SRS 2nd and exp
Client variance	27.35*	27.37*	27.33*	27.22*
Couple variance	32.54*	32.22*	31.85*	30.86*
Therapist variance	5.11*	3.71*	2.60*	1.21*
Client level	42%	43%	44%	46%
Couple level	50%	51%	52%	52%
Therapist level	8%	6%	4%	2%

Note. For all models, we controlled for ORS pretherapy scores. Exp = experience, ORS = Outcome Rating Scale, SRS = Session Rating Scale.

*p < .05.

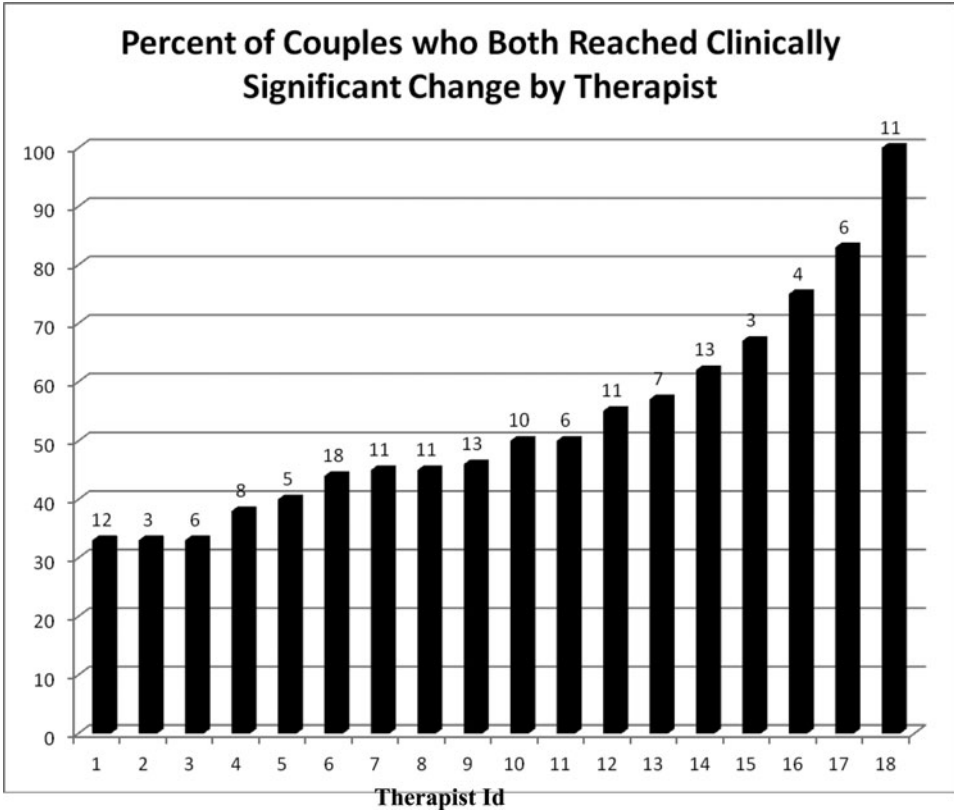


FIGURE 1 Percentage of couples who reached reliable and clinical significant change by therapist. The number above each therapist bar represents the number of couples treated by the therapist.

percentage of couples who reached reliable and clinically significant change for each of the therapists in the study.

Next, we tested whether therapist factors—gender, discipline, and experience conducting couple therapy, and average second-session alliance score—would account for the variability in outcomes attributed to therapists. To do so, we conducted a series of models where client ORS-post score was the dependent variable and therapist variables were entered at level 3 as the predictor variables in a stepwise manner. We also controlled for client ORS-pre scores (grand mean centered) at Level 1.

The results demonstrated that therapists’ gender did not significantly account for variance in the model ($b = -0.63, 95\% \text{ CI} [-3.96, 2.93], p > .05$). The variance at the therapist level in client outcomes when therapist gender was included in the model was 5.53 (slightly up from 5.11 in the baseline model). The therapist ICC when therapist gender was included in the model was .08, consistent with the baseline model. Accordingly, we removed this variable from our subsequently models. Next, we tested whether therapists’ discipline would account for a significant proportion of the variance of clients’ outcomes attributed to therapists. Similar to the previous model, therapists’ discipline did not significantly account for any of the variance in client outcomes

($b = -0.48$, 95% CI $[-3.73, 3.08]$, $p > .05$). The variance at the therapist level in client outcomes was 5.63, again slightly higher than the baseline model. The increase in variance when a variable is added to the model is typically referred to as negative variance, which simply implies that the variable is not adding to greater understanding in the model (Snijders & Bosker, 1999). Given this, we eliminated it from our final model.

In contrast, when therapist experience was added to the model, the results demonstrated that therapist experience accounted for 2% of the initial 8% of variance in client outcomes that were attributed to therapists (see Table 2, Model 2). In other words, therapists' experience accounted for 25% of the variance in client outcomes that were associated with therapists. We then added therapist second-session SRS score to Model 3, but removed therapist experience from the model. The results demonstrated that therapist second-session SRS score accounted for 4% of the initial 8% of variance in client outcomes that were attributed to therapists (see Table 2, Model 3). Last, we added therapist experience and therapist second-session SRS score simultaneously to the model. The results showed that in combination these two variables explained 6% of the initial 8% of therapist effects in client outcomes (see Table 2, Model 4). The fixed effects for therapist experience was statistically significant (median $b = 0.25$, 95% CI $[0.06, 0.42]$, $p = .01$) as was for therapist second-session SRS score ($b = 1.21$, 95% CI $[0.51, 1.96]$, $p = .002$). Simply put, 25% of therapist effects in the present study were accounted for by therapist experience and 50% was accounted for therapist second-session SRS score.³ It is interesting that these variables did not notably overlap as the association between them was $r = -.24$, $p = .38$.

DISCUSSION

The present study investigated whether therapists differ in client outcomes in couple therapy as well as what therapist factors may account for these differences. Although the treatment setting was naturalistic where therapists were eclectic in their theoretical orientation and diverse in their experience conducting couple therapy, the random assignment of couples seemed to create a level playing field for comparison as evidenced by a lack of significant differences among therapists in client pretherapy functioning and initial goals. Accordingly, the variability between therapists likely reflects differences in therapeutic effectiveness. It is not surprising that we found therapists differed in their ability to secure positive outcomes, accounting for 8.0% of the variance, an estimate consistent with prior studies in the individual therapy literature (e.g., Crits-Christoph et al., 1991; Lutz et al., 2007; Okiishi et al., 2006; Wampold & Brown, 2005). The magnitude of this therapist effect may seem modest in absolute terms; however, it is a medium-sized effect and consistent with many other predictors in the psychotherapy process and outcome literature (Wampold, 2001). Therapists also accounted for 10% of the variance in client alliance scores (a medium to large-sized effect), which is consistent with the Baldwin and Imel (2013)

³When using a stepwise procedure to determine the proportion of the variance explained by each variable, it is important to consider which order the variables were entered. However, the order of therapist experience or therapist second-session SRS score did not affect the results.

meta-analysis estimate of 9.0%. Our results suggest that therapists have a significant role in the client's experience of both the process and outcome of couple therapy.

The differences among therapists in average alliance quality as well as therapist experience conducting couple therapy were meaningful contributors to outcome. Specifically, therapist average alliance quality accounted for 50%; whereas, therapist experience accounted for 25% of the variability in therapy outcomes attributed to therapists, suggesting that there may be at least two pathways for therapists to become more effective in conducting couple therapy. Therapists who were more effective in facilitating successful outcomes were also rated, on average, as forming stronger alliances, a finding that parallels previous studies for individual therapy (Baldwin et al., 2007) and relationship education (Owen et al., 2011). Therapist average alliance quality may be but one indication of the facilitative interpersonal skills required that allow the identification and processing of key interpersonal messages in therapy (Anderson, Ogles, Patterson, Lambert, & Vermeersch, 2009). In addition, therapists who were more experienced conducting couple therapy achieved, on average, better outcomes. It is interesting that therapist experience and therapist average alliance quality were weakly and nonsignificantly associated. Accordingly, our findings provide no clear evidence that experienced therapists have, on average, better alliance scores. One way of interpreting these findings is that therapist experience and therapist average alliance quality are important, yet notably distinct in effect on outcomes.

Two tentative implications emerge from these findings. The first implication concerns therapist professional development and clinical experience. Novice therapists wishing to treat couples should likely focus their experience specifically with couples. We can speculate that couple therapy experience may enhance therapist confidence and skills in working with couples, who often present with highly challenging and often conflictual in session interactions—an often daunting situation for many therapists. Consequently, new therapists may want to seek consultation or supervision with an experienced couple therapist. Couple therapy involves a skill set that is different from that of individual therapy, and therapists need to be mindful that experience in and knowledge of couple therapy can prevent negative outcomes. To ensure that they are learning from their experience, therapists may want to monitor their effectiveness using available process and outcome measures. Ongoing measuring of outcomes allows therapists to plot their cumulative career development (Orlinsky & Rønnestad, 2005), which enables a trial and error application of new learning and a continual process of professional reflection. The combination of supervision and monitoring outcomes perhaps provide the best possibility that any experience gained will promote therapist development and better couple outcomes.

Our study, although preliminary, suggests that gaining specific experience as a couple therapist is a worthy avenue to improve outcomes, but enhancing alliance abilities may provide a clear pathway to better results (see Duncan, 2010). There are many ways to understand alliance skills (Ackerman & Hilsenroth, 2003; Owen, Reese, Quirk, & Rodolfa, 2013) as well as many systems to improve relational abilities, from Safran & Muran's (2000) ways of therapeutically addressing alliance ruptures, to specific models that are attentive to relational aspects, such as motivational interviewing (Miller & Rollnick, 2002). Moreover, there is evidence suggesting that therapists can improve their ability to form strong alliances (Crits-Christoph et al., 2006). Consistently, therapists wishing to improve outcomes over the course of their careers might consider formally tracking their ability to form strong alliances to allow a more proactive effort to develop their alliance skills (Duncan, 2010). Systematically tracking couples outcome and alliance with available instruments would enable proactive efforts by psychotherapists to continue to improve without guesswork,

leaving their growth to chance, or waiting for the platitudes about experience being the best teacher to manifest. Routine alliance assessment allows the therapist to identify problems before they exert a negative effect on outcome as well as the opportunity to tailor services to client preferences and expectations. Many scholars have recently recommended continuous measurement of the alliance as a way of directly improving client care (e.g., Castonguay, Constantino, & Holtforth, 2006; Crits-Cristoph et al., 2006). In addition, the Division 29 Task Force on Empirically Supported Relationships (Ackerman et al., 2001) has endorsed the routine assessment of the alliance. Learning to tailor the alliance to the vast array of couples that a typical therapist treats may enhance one's average alliance abilities.

There are several limitations to this study. The use of only one outcome measure in the pre-/posttreatment analysis as well as one alliance measure limits the conclusions that may be drawn. We do not know whether more extensive alliance and progress assessments would have given different results or whether other measures from clinician or observer perspectives would alter our findings. However, our findings are consistent with other studies examining therapist variability. This study was intentionally designed to more closely replicate what happens in typical clinical practice to demonstrate the feasibility of doing routine assessment and was consequently limited in terms of the instruments chosen as well as the frequency of administration. Moreover, this study was conducted in Norway at two agencies; thus, we do not know whether these results will generalize to other couple therapists.

In addition, we do not know whether clients were influenced by demand characteristics (Orne, 1962). That is, clients could have inflated their scores because the measures were completed in the therapist's presence and clients knew therapists would likely discuss their meaning. This is more likely with the SRS (the alliance scale) than the ORS. Some clients do hide things from their therapist, but they are more likely to withhold a negative reaction to the therapist or session than to hide or misrepresent their level of distress (Farber, 2003). This limitation is tempered, however, by a recent study (Reese et al., 2013) that found therapist presence did not lead to increased SRS scores.

Another concern is that therapists were trained to discuss alliance ratings and therefore were encouraged to intervene when alliance problems existed which could have resulted in inflated alliance scores. Alliance scores, however, tend to be positively skewed and clients tend to score alliance measure high regardless of whether they are in the presence of the therapist or not (e.g., Reese et al., 2013). Nonetheless, further research is needed that directly addresses these issues and our findings should be viewed with these important considerations in mind. Last, our study was conducted in Norway with a racially and ethnically homogenous sample of clients and therapists, thus the degree to which our findings will replicate in other settings, countries, or populations should be explored.

In conclusion, our study highlights two promising factors that account for therapist effectiveness in couple therapy: the effect of the alliance and specific therapist experience with couple therapy. Thus, we suggest that therapists balance their focus on the technical aspects of couple therapy with gaining specific experience with couples and continually honing alliance skills to accommodate more clients across their careers. Perhaps the best way to accomplish both is to (a) systematically monitor the alliance and outcome and (b) take a more proactive approach to improving alliance ratings across clients and ensuring that experiences with couples translate to better outcomes.

REFERENCES

- Ackerman, S. J., Benjamin, L. S., Beutler, L. E., Gelso, C. J., Goldfried, M. R., Lambert, M. J., . . . Rainer, J. (2001). Empirically supported therapy relationships: Conclusions and recommendations of the division 29 taskforce. *Psychotherapy, 38*, 495–497.
- Ackerman, S. J., & Hilsenroth, M. J. (2003). A review of therapist characteristics and techniques positively impacting the therapeutic alliance. *Clinical Psychology Review, 23*, 1–33.
- Anderson, T., Ogles, B. M., Patterson, C. L., Lambert, M. J., & Vermeersch, D. A. (2009). Therapist effects: Facilitative interpersonal skills as a predictor of therapist success. *Journal of Clinical Psychology, 65*, 755–768.
- 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.
- Anker, M. G., Owen, J. J., Duncan, B. L., & Sparks, J. A. (2010). The alliance in couple therapy: Gender, partner influence, and the alliance-change relationship in a naturalistic sample. *Journal of Clinical and Consulting Psychology, 78*, 635–645.
- Baldwin, S. A., & Imel, Z. (2013). Therapist effects. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavioral change* (6th ed.). New York, NY: Wiley.
- Baldwin, S. A., Wampold, B. E., & Imel, Z. E. (2007). Untangling the alliance-outcome correlation: Exploring the relative importance of therapist and patient variability in the alliance. *Journal of Consulting and Clinical Psychology, 75*, 842–852.
- Beutler, L. E., Malik, M., Alimohamed, S., Harwood, M. T., Talebi, H., Noble, S., & Wong, E. (2004). Therapist variables. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavioral change* (5th ed., pp. 227–306). New York, NY: Wiley.
- Bischoff, R. J., & Sprenkle, D. H. (1993). Dropping out of marriage and family therapy: A critical review of research. *Family Process, 32*, 353–376.
- Blatt, S. J., Sanislow, C. A., Zuroff, D. C., & Pilkonis, P. A. (1996). Characteristics of effective therapists: Further analyses of data from the National Institute of Mental Health Treatment of Depression Collaborative Research Program. *Journal of Consulting and Clinical Psychology, 64*, 1276–1284.
- Blow, A. J., Sprenkle, D. H., & Davis, S. D. (2007). Is who delivers the treatment more important than the treatment itself? The role of the therapist in common factors. *Journal of Marital and Family Therapy, 33*, 298–317.
- Bordin, E. S. (1979). The generalizability of the psychoanalytic concept of the working alliance. *Psychotherapy: Theory, Research and Practice, 16*, 252–260.
- Campbell, A., & Hemsley, S. (2009). Outcome Rating Scale and Session Rating Scale in psychological practice: Clinical utility of ultra-brief measures. *Clinical Psychologist, 13*, 1–9.
- Castonguay, L. G., Constantino, M. J., & Holtforth, M. G. (2006). The working alliance: Where are we and where should we go? *Psychotherapy, 43*, 271–279.
- Christensen, A., Atkins, D. C., Berns, S., Wheeler, J., Baucom, D., & 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.
- Christensen, A., & Jacobson, N. S. (1994). Who (or what) can do psychotherapy: The status and challenge of nonprofessional therapies. *Psychological Science, 5*, 8–14.
- Crits-Christoph, P., Baranackie, K., Kurcias, J., Beck, A. T., Carroll, K., Perry, K., . . . Zitrin, C. (1991). Meta-analysis of therapist effects in psychotherapy outcome studies. *Psychotherapy Research, 1*, 81–91.
- Crits-Christoph, P., Gibbons, M. B. C., Crits-Christoph, K., Narducci, J., Schamberger, M., & Gallop, R. (2006). Can therapists be trained to improve their alliances? A preliminary study of alliance-fostering psychotherapy. *Psychotherapy Research, 16*, 268–281.
- Dienes, Z. (2011). Bayesian versus orthodox statistics: Which side are you on? *Perspectives on Psychological Science, 6*, 274–290.
- Duncan, B. L., Miller, S. D., Reynolds, L., Sparks, J., Claud, D., Brown, J., & Johnson, L. D. (2003). The session rating scale: Psychometric properties of a “working” alliance scale. *Journal of Brief Therapy, 3*, 3–12.
- Duncan, B. L. (2010). *On becoming a better therapist*. Washington, DC: American Psychological Association.
- Duncan, B. L. (2011). *The Partners for Change Outcome Management System (PCOMS): Administration, scoring, interpreting update for the Outcome and Session Ratings Scale*. Jensen Beach, FL: Author.

- Duncan, B. L. (2012). The Partners for Change Outcome Management System (PCOMS): The Heart and Soul of Change Project. *Canadian Psychology, 53*, 93–104.
- Farber, B. A. (2003). Patient self-disclosure: A review of the research. *Journal of Clinical Psychology, 59*, 589–600.
- Freeston, M. H., & Plechaty, M. (1997). Reconsideration of the Locke-Wallace Marital Adjustment Test: Is it still relevant for the 1990s? *Psychological Reports, 81*, 419–434.
- Friedlander, M. L., Escudero, V., Heatherington, L., & Diamond, G. M. (2011). Alliance in couple and family therapy. *Psychotherapy, 48*, 25–33.
- Hamaker, E. L., & Klugkist, I. G. (2011). Bayesian estimation in multilevel modeling. In J. J. Hox & J. K. Roberts (Eds.), *Handbook of advanced multilevel analysis* (pp. 137–161). New York, NY: Taylor & Francis.
- Hatcher, R. L., & Barends, A. W. (1996). Patients' view of the alliance in psychotherapy: Exploratory factor analysis of three alliance measures. *Journal of Consulting and Clinical Psychology, 64*, 1326–1336.
- Hatcher, R. L., Barends, A., Hansell, J., & Gutfreund, M. J. (1995). Patients' and therapists' shared and unique views of the therapeutic alliance: An investigation using confirmatory factor analysis in a nested design. *Journal of Consulting and Clinical Psychology, 63*, 636–643.
- Horvath, A. O., Del Re, A. C., Flückiger, C., & Symonds, D. (2011). Alliance in individual psychotherapy. *Psychotherapy, 48*, 9–16.
- Huppert, J. D., Bufka, L. F., Barlow, D. H., Gorman, J. M., Shear, M. K., & Woods, S. W. (2001). Therapists, therapist variables, and cognitive-behavioral therapy outcome in a multicenter trial for panic disorder. *Journal of Consulting and Clinical Psychology, 69*, 747–755.
- 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.
- Kim, D. M., Wampold, B. E., & Bolt, D. M. (2006). Therapist effects in psychotherapy: A random effects modeling of the National Institute of Mental Health Treatment of Depression Collaborative Research Program data. *Psychotherapy Research, 16*, 161–172.
- Kraus, D. R., Castonguay, L., Boswell, J. F., Nordberg, S. S., & Hayes, J. A. (2011). Therapist effectiveness: Implications for accountability and patient care. *Psychotherapy Research, 21*, 267–276.
- Leon, S. C., Martinovich, Z., Lutz, W., & Lyons, J. S. (2005). The effect of therapist experience on psychotherapy outcomes. *Clinical Psychology and Psychotherapy, 12*, 417–426.
- Locke, H. J., & Wallace, K. M. (1959). Short marital-adjustment and prediction tests: Their reliability and validity. *Marriage and Family Living, 21*, 251–255.
- Luborsky, L., McLellan, A. T., Diguier, L., Woody, G., & Seligman, D. A. (1997). The psychotherapist matters: Comparison of outcomes across twenty-two therapists and seven patient samples. *Clinical Psychology: Science and Practice, 4*, 53–65.
- Lutz, W., Leon, S. C., Martinovich, Z., Lyons, J. S., & Stiles, W. B. (2007). Therapist effects in outpatient psychotherapy: A three-level growth curve approach. *Journal of Counseling Psychology, 54*, 32–39.
- Maramba, G. G., & Nagayama Hall, G. (2002). Meta-analyses of ethnic match as a predictor of dropout, utilization, and level of functioning. *Cultural Diversity and Ethnic Minority Psychology, 8*, 290–297.
- Marcus, D., Kashy, D., & Baldwin, S. (2009). Studying psychotherapy using the one-with-many design: The therapeutic alliance as an exemplar. *Journal of Counseling Psychology, 56*, 537–548.
- Miller, S. D., & Duncan, B. L. (2004). *The Outcome and Session Rating Scales: Administration and scoring manual*. Ft. Lauderdale, FL: Authors.
- Miller, S. D., Duncan, B. L., Brown, J., Sparks, J., & Claud, D. (2003). The Outcome Rating Scale: A preliminary study of the reliability, validity, and feasibility of a brief visual analog measure. *Journal of Brief Therapy, 2*, 91–100.
- Miller, S. L., Duncan, B. L., Sorrell, R., & Brown, G. S. (2005). The partners for change outcome management system. *Journal of Clinical Psychology, 61*, 199–208.
- Miller, W. R., & Rollnick, S. (2002). *Motivation interviewing: Preparing people for change* (2nd ed). New York, NY: Guilford Press.
- Nyman, S. J., Nafziger, M. A., & Smith, T. B. (2010). Client outcomes across counselor training level within a multitiered supervision model. *Journal of Counseling & Development, 88*, 204–209.
- Okiishi, J. C., Lambert, M. J., Eggett, D., Nielsen, L., Dayton, D. D., & Vermeersch, D. (2006). An analysis of therapist treatment effects: Toward providing feedback to individual therapists on their clients' psychotherapy outcome. *Journal of Clinical Psychology, 62*, 1157–1172.

- Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, *17*, 776–783.
- Owen, J., Reese, R. J., Quirk, K., & Rodolfa, E. (2013). Alliance in action: A new look the working alliance. *Psychotherapy Research*, *23*, 67–77.
- Owen, J., Rhodes, G. K., Stanley, S. M., & Markman, H. J. (2011). The role of leaders' working alliance in premarital education. *Journal of Family Psychology*, *1*, 49–57.
- Raytek, H. S., McCrady, B. S., Epstein, E. E., & Hirsch, L. S. (1999). Therapeutic alliance and the retention of couples in conjoint alcoholism treatment. *Addictive Behaviors*, *24*, 317–330.
- Reese, R. J., Gillaspay, J. A., Owen, J. J., Flora, K. L., Cunningham, L. E., Archie, D., & Marsden, T. (2013). The influence of demand characteristics and social desirability on clients' ratings of the therapeutic alliance. *Journal of Clinical Psychology*, *69*, 696–709.
- Reese, R., Norsworthy, L., & Rowlands, S. (2009). Does a continuous feedback model improve psychotherapy outcomes? *Psychotherapy*, *46*, 418–431.
- Reese, R. J., Toland, M. D., Slone, N. C., & Norsworthy, L. A. (2010). Effect of client feedback on couple psychotherapy outcomes. *Psychotherapy: Theory, Research, Practice, Training*, *47*, 616–630.
- Safran, J. D., & Muran, J. C. (2000). Resolving therapeutic alliance ruptures: Diversity and integration. *In Session: Psychotherapy in Practice*, *56*, 233–243.
- Snijders, T., & Bosker, R. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London, England: Sage.
- 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.
- Sparks, J., & Duncan, B. (2010). Common factors in couple and family therapy: Must all have prizes? In B. L. Duncan, S. D. Miller, B. E. Wampold, & M. A. Hubble (Eds.), *The heart and soul of change: Delivering what works* (2nd ed., pp. 357–392). Washington DC: American Psychological Association.
- Wampold, B. E. (2001). *The great psychotherapy debate: Models, methods, and findings*. Hillsdale, NJ: Erlbaum.
- Wampold, B. E. (2005). What should be validated? The psychotherapist. In J. C. Norcross, L. E. Beutler, & R. F. Levant (Eds.), *Evidence-based practices in mental health: Debate and dialogue on the fundamental questions* (pp. 200–208, 236–238). Washington, DC: American Psychological Association.
- Wampold, B. E., & Brown, G. (2005). Estimating therapist variability in outcomes attributable to therapists: A naturalistic study of outcomes in managed care. *Journal of Consulting and Clinical Psychology*, *73*, 914–923.
- Yuan, Y., & MacKinnon, D. P. (2009). Bayesian mediation analysis. *Psychological Methods*, *14*, 301–322.