Journal of Consulting and Clinical Psychology
2006, Vol. 74, No. 6, 1180–1191

0022-006X/06/$12.00 DOI: 10.1037/0022-006X.74.6.1180

Couple and Individual Adjustment for 2 Years Following a Randomized Clinical Trial Comparing Traditional Versus Integrative Behavioral Couple Therapy

Andrew Christensen
University of California, Los Angeles

David C. Atkins
Fuller Graduate School of Psychology

Jean Yi
University of Washington

Donald H. Baucom
University of North Carolina, Chapel Hill

William H. George
University of Washington

Follow-up data across 2 years were obtained on 130 of 134 couples who were originally part of a randomized clinical trial comparing traditional versus integrative behavioral couple therapy (TBCT vs. IBCT; A. Christensen et al., 2004). Both treatments produced similar levels of clinically significant improvement at 2 years posttreatment (69% of IBCT couples and 60% of TBCT couples). Both treatments showed a "hockey-stick" pattern of change in which satisfaction dropped immediately after treatment termination but then increased for most of follow-up. The break point when couples reversed courses and gained in satisfaction occurred sooner for IBCT than TBCT couples, and those couples who stayed together generally fared better in IBCT than in TBCT. Finally, there was evidence of greater stability during follow-up in IBCT than in TBCT couples. There was little change in individual functioning over follow-up, but when change occurred it was strongly related to change in marital satisfaction. Given that this sample was selected for its significant and chronic distress, the data are encouraging about the long-term impact of behavioral couple therapy.

Keywords: couple therapy, marital therapy, clinical trial, follow-up data

Behavioral couple therapy (BCT) is both widely practiced and extensively researched. A recent national survey of practicing clinicians in the American Association for Marriage and Family Therapy found that 31% indicated that their primary treatment modality was either behavioral or cognitive behavioral (Northey, 2002). Many practitioners, particularly psychologists, who are not members of this association but who nevertheless conduct couple therapy also practice BCT. In addition to BCT's clinical use, more research studies have examined BCT than any other type of couple therapy (Christensen & Heavey, 1999).

Two recent meta-analytic reviews of investigations comparing BCT with control groups have concluded that BCT is more effective than no treatment for distressed couples. On the basis of 17 published investigations of randomized clinical trials, Baucom, Hahlweg, and Kuschel (2003) estimated the mean effect size as $d = 0.82$. On the basis of a larger pool of 30 randomized clinical trials, both published investigations as well as unpublished dissertations, Shadish and Baldwin (2005) estimated a mean effect size of $d = 0.59$. These recent reviews are consistent with earlier reviews that documented the effectiveness of BCT (Baucom, Shoham, Mueser, Daiuto, & Stickle, 1998; Christensen & Heavey, 1999). Shadish and Baldwin concluded that BCT "might be viewed as a mature therapy, with relatively fewer questions left to research compared with these newer therapies" (p. 11).

Two important questions that do remain largely unanswered about BCT concern its long-term effectiveness and its relative effectiveness compared with other treatments. Although many clinical trials have been conducted on BCT, relatively few have included long-term follow-up assessments. Only two studies have examined couple functioning more than a year after treatment termination, and both indicated significant deterioration from pretreatment levels. In a study of BCT by Jacobson, Schmaling, and

Andrew Christensen, Department of Psychology, University of California, Los Angeles (UCLA); David C. Atkins, Travis Research Institute, Fuller Graduate School of Psychology; Jean Yi and William H. George, Department of Psychology, University of Washington; Donald H. Baucom, Department of Psychology, University of North Carolina, Chapel Hill.

This research was supported by National Institute of Mental Health Grants MH56223, awarded to Andrew Christensen at UCLA, and MH56165, awarded to Neil S. Jacobson at the University of Washington, for a two-site clinical trial of couple therapy. After Jacobson's death in 1999, William H. George served as primary investigator at the University of Washington. We are grateful for the enormous contributions that Neil S. Jacobson made to this research.

Correspondence concerning this article should be addressed to Andrew Christensen, Department of Psychology, UCLA, Box 156304, Los Angeles, CA 90095-1563. E-mail: Christensen@psych.ucla.edu

Holtzworth-Munroe (1987), 56% of couples were unchanged or had deteriorated from their pretreatment status at a 2-year follow-up. In a 4-year follow-up of BCT by Snyder, Wills, and Grady-Fletcher (1991), 58% of couples were unchanged or had deteriorated from their pretreatment status. Thus, Christensen and Heavey (1999) concluded that many BCT couples "may relapse some 1 to 4 years posttreatment" (p. 169).

Because the short-term efficacy of BCT has been demonstrated, recently investigators have been less interested in comparisons of BCT with no treatment and "more interested in . . . comparative or incremental treatment effectiveness" (Shadish & Baldwin, 2005, p. 11). Also, the apparent problems with the long-term effectiveness of BCT have encouraged the development of alternative treatments that might prove not only more effective in the short run, but also more durable in the long term.

In the current study, we examined the comparative efficacy of a traditional form of BCT, called here *traditional behavioral couple therapy* (TBCT; Jacobson & Margolin, 1979), with a new form of treatment called *integrative behavioral couple therapy* (IBCT; Jacobson & Christensen, 1998). TBCT creates positive changes in couples through direct instruction as well as through training in communication and problem solving. In contrast to the change focus in TBCT, IBCT focuses on increasing emotional acceptance between partners. IBCT assumes that relationship problems lie not just in the egregious actions of each partner, but also in their emotional reactivity to those actions. Therefore, IBCT attempts to alter the emotional context between partners and achieve greater intimacy between partners as well as make concrete changes in target problems.

Two small studies have documented the efficacy of IBCT by comparing it with a wait-list control group (Wimberly, 1998) and with TBCT (Jacobson, Christensen, Prince, Cordova, & Eldridge, 2000). However, the first major clinical trial of this approach was reported by Christensen et al. (2004), who described the initial treatment results from 134 significantly and chronically distressed married couples at two sites who were randomly assigned to either TBCT or IBCT. Both treatments achieved improvements in marital satisfaction that were statistically as well as clinically significant. Although the two different treatment groups were not significantly different at the end of treatment, the trajectory of their change in marital satisfaction was significantly different. Couples in IBCT made gradual improvements in satisfaction throughout the course of treatment whereas couples in TBCT made more rapid gains early in treatment and then, in contrast to IBCT, plateaued later in treatment. In the current study, we examined the outcome of these couples through 2 years following treatment.

As noted earlier, researchers who have conducted couple therapy studies have generally neglected to examine long-term follow-up. However, even those who have examined it are limited by an assessment or two during follow-up periods. Thus, in no previous study have researchers examined the trajectory of change over long-term follow-up. Therefore, we do not know whether deterioration is gradual following the end of therapy, whether there is a sudden deterioration when therapy ends, or whether some other pattern occurs. Short- and long-term follow-up assessments of couple therapy have also been limited by a focus on marital outcomes to the exclusion of individual outcomes. Another problem with follow-up assessments, noted by Baucom et al. (1998), is that "few investigators track or report whether clients have obtained additional treatment during the follow-up period." Snyder et al. (1991) was a welcome exception to these last two limitations.

Our objectives in the current study were to overcome some of the limitations of past research on the long-term outcome of BCT by investigating over the 2 years after therapy ended: (a) the trajectory of marital satisfaction, (b) the change over time in other couple behaviors (e.g., marital stability, affective and problem-solving communication), (c) the effect of treatment condition and other covariates on the previously mentioned outcomes, (d) the association of individual functioning and marital satisfaction over time, (e) the clinical significance of change in marital satisfaction, and (f) the impact of additional therapy during follow-up.

Method

Participants

Through media announcements and clinic referrals, 134 chronically and seriously distressed couples were recruited for a marital therapy program offered in two sites: Los Angeles (71 couples) and Seattle (63 couples). Average age of the spouses was in the early 40s, and couples had been married an average of 10 years. Most had children; most of the participants were Caucasian, but more than 20% of both husbands and wives were of another ethnicity. All procedures were approved by the institutional review boards of the respective universities; couples signed an informed consent form at their intake session that covered their participation in treatment and in assessments through the 2-year follow-up assessment. See Christensen et al. (2004) for more detail on the participants.

Measures

With the exception of the Marital Activities Questionnaire (MAQ) and therapy information sheet (discussed below), all the measures used during follow-up were also used to assess marital and individual functioning during the treatment phase of the study. More detail on the measures can be found in Christensen et al. (2004).

Relationship satisfaction. We used the Dyadic Adjustment Scale (DAS; Spanier, 1976), a well validated measure consisting of 32 Likert items, to assess marital satisfaction.[1] In the current sample, measures of internal consistency were .89 and .87 for husbands and wives, respectively.

Relationship stability. The Marital Status Inventory (MSI; Weiss & Cerreto, 1980) consists of 14 true–false items that measure movement toward separation or divorce, ranging from items that almost all couples would endorse ("I have occasionally thought of divorce or wished that we were separated, usually after an argument or other incident") to items that indicate an imminent or completed divorce ("I have filed for divorce or we are divorced"). Couples completed the measure for the time period since their last assessment. In the current sample, internal consistency was .80 for both husbands and wives.

Communication. Two subscales of the Marital Satisfaction Inventory—Revised (MSI–R; Snyder, 1997) are designed to assess problem-solving communication (PSC) and affective communication (AFC). PSC consists of 19 true–false items that assess lack of skills in discussing

[1] The Global Distress Scale (GDS) from the Marital Satisfaction Inventory—Revised (Snyder, 1997) was also completed by couples. Christensen et al. (2004) found that the DAS was more sensitive to change, likely because of certain historical items on the GDS that once endorsed would always be endorsed (e.g., "My partner and I have never come close to ending our relationship."). The GDS was also highly correlated with the DAS in the present sample ($r = -.79$). For these reasons, the present analyses focus only on the DAS to assess relationship satisfaction.

differences and sensitive issues. AFC consists of 13 true–false items that assess lack of skills in providing support, affection, and disclosure of feelings. Internal consistency for the PSC was .71 and .77 for husbands and wives, respectively; for AFC it was .76 and .77.

Individual functioning. We used the Mental Health Index (MHI) from the Compass Outpatient Treatment Assessment System (Sperry, Brill, Howard, & Grissom, 1996) to assess individual spousal functioning. The MHI is composed of three subscales: Subjective Well-Being, Current Symptoms, and Current Life Functioning. We also examined Current Symptoms (CS) independently to specifically examine change in psychopathology. In the present sample, internal consistency for the MHI was .86 and .88 for men and women, respectively; for CS, it was .93 and .94.

Activities promoted by therapy. The MAQ (Christensen, 1999) consists of 10 items designed to measure how often clients continue to do the activities they presumably learned in therapy. Five items refer specifically to activities that were the focus of TBCT (e.g., "Used specific communication skills when talking with spouse, such as 'I statements'"), and five refer specifically to activities that were the focus of IBCT (e.g., "Empathize or sympathize with the emotions that spouse experiences during disagreements or arguments"). For each item, the respondent rates on a 9-point scale the frequency with which he or she does the item and the frequency with which his or her spouse does the item. For some items that refer to joint activities, such as "Set aside specific times to communicate with each other about the relationship or about problems we face," respondents rated the frequency with which they jointly did the activity. We calculated summary scores for TBCT behaviors and IBCT behaviors. Internal consistency for TBCT behaviors was .83 for both men and women; for IBCT behaviors it was .77 for men and .75 for women.

Additional treatment. The therapy information sheet was designed to assess additional treatment, such as psychotherapy.

Procedures

Treatments. During the treatment phase of this study (described in Christensen et al., 2004), couples were randomly assigned to receive a maximum of 26 sessions of TBCT (66 couples) or IBCT (68 couples) stratified within two levels of initial distress: moderately (66) versus severely (68) distressed couples. TBCT was guided by the classic monograph by Jacobson and Margolin (1979); couples were also given readings from the communication guide by Gottman, Notarius, Gonso, and Markman (1977). IBCT was guided by the treatment manual by Jacobson and Christensen (1998); couples were also given a self-help book about IBCT by Christensen and Jacobson (2000).

After the conclusion of the treatment program, couples were prohibited from further contact with their therapist for 2 years. Of course, they were free to seek other therapeutic services, but this contact was monitored with the therapy information sheet.

Follow-up assessments. To understand the statistical analyses presented below, it is critical to understand the assessment schedule used in this study. The original study plan called for regular assessments at specific time periods measured from the pretreatment assessment. Given that the course of treatment was approximately 6 months (i.e., up to 26 sessions, typically scheduled weekly), follow-up assessments would begin approximately 6 months following the presumed end of treatment. Thus, the first follow-up assessment was scheduled at 1 year following the pretreatment assessment, and subsequent follow-up assessments occurred in 6-month increments to 2.5 years following the pretreatment assessment.

This schedule would lead to equally spaced time intervals for assessments; however, an important decision to increase the external validity of the study sharply affected how assessments were timed posttreatment. To enhance the external validity of the study, the frequency of sessions was left open to the couple and therapist with the only caveats being that a maximum of 26 sessions was allowed and all sessions had to be completed within a year of the pretreatment assessment. As a result, couples took varying lengths of time to finish therapy ($Mdn = 36$ weeks) but completed a termination assessment immediately after the final session, no matter when therapy ended. Thus, the time from the final session of therapy and termination assessment to the initial follow-up assessment varied quite dramatically across couples, which in turn affected the timing of other follow-up assessments relative to the last session of therapy and termination assessment. The average times from termination to each of the follow-up assessments were 17.3, 43.3, 69.9, and 96.6 weeks for the 6-, 12-, 18-, and 24-month assessments, respectively. The *SD* of the time since final session was approximately 9 weeks at each assessment. Therefore, even though follow-up assessments were scheduled according to a calendar defined by the couples' pretreatment assessment, because therapy ended at widely differing points, the follow-up assessments varied widely in the length of time since the termination of therapy and the termination assessment. See Figure 1 for a flow chart of assessments through treatment and follow-up periods.

Data Analyses

Hierarchical linear modeling (HLM; Raudenbush & Bryk, 2002) was our primary analytic tool as the research questions focused on change post-therapy and timing of assessments was highly unbalanced as noted above. HLM is ideally suited for modeling longitudinal data on couples as it flexibly handles correlated data (e.g., repeated measures and spouses) and does not require balanced data with respect to the timing of assessment or the total number of assessments. In addition, time-varying covariates are easily included. We anticipated that change in individual functioning would covary across time with marital satisfaction; thus, we used time-varying covariates for these models. We used a three-level model, reflecting that repeated measures were nested within individuals who were nested within couples (see Atkins, 2005, for a discussion of three-level models for couples). Typically, a model with only a representation of time and a therapy predictor (TBCT = 0, IBCT = 1) was fit first; next, a series of covariates that were important during therapy were tested, including gender, initial distress (moderate vs. severe), site (Seattle vs. Los Angeles), and whether or not the couple was part of the pilot procedures. We completed all analyses using R Version 2.1.1 (R Development Core Team, 2005) and made extensive use of the nlme package (Pinheiro, Bates, DebRoy, & Sarkar, 2005) of functions for mixed-effects modeling (i.e., HLM).

Results

Trajectory of Change

Means and standard deviations for each of our outcome variables at each of our follow-up time periods are presented in Table 1. The first step in modeling longitudinal data is to create an accurate model of the change over time. Trajectories of marital satisfaction using the DAS were plotted both individually and grouped by important covariates (e.g., therapy, initial distress, gender). These spaghetti plots (Diggle, Heagerty, Liang, & Zeger, 2002) revealed a notably nonlinear pattern of change during the 2 years following therapy. Specifically, there appeared to be an initial, rapid period of deterioration in marital satisfaction that stabilized into a slow period of increasing marital satisfaction later in follow-up. In addition, this pattern appeared to be somewhat different between the two therapies in that the initial period of deterioration was shorter for IBCT couples than TBCT couples.

Initial attempts to fit this nonlinear trajectory using polynomials (i.e., linear, quadratic, cubic) were unsatisfactory, as seen in plots of the Level 1 residuals against time. As a result, we fit a piecewise linear model that allowed for two separate phases of linear change on either side of a break point. In the statistics literature, this model would be called a linear spline model with a single knot (Snijders

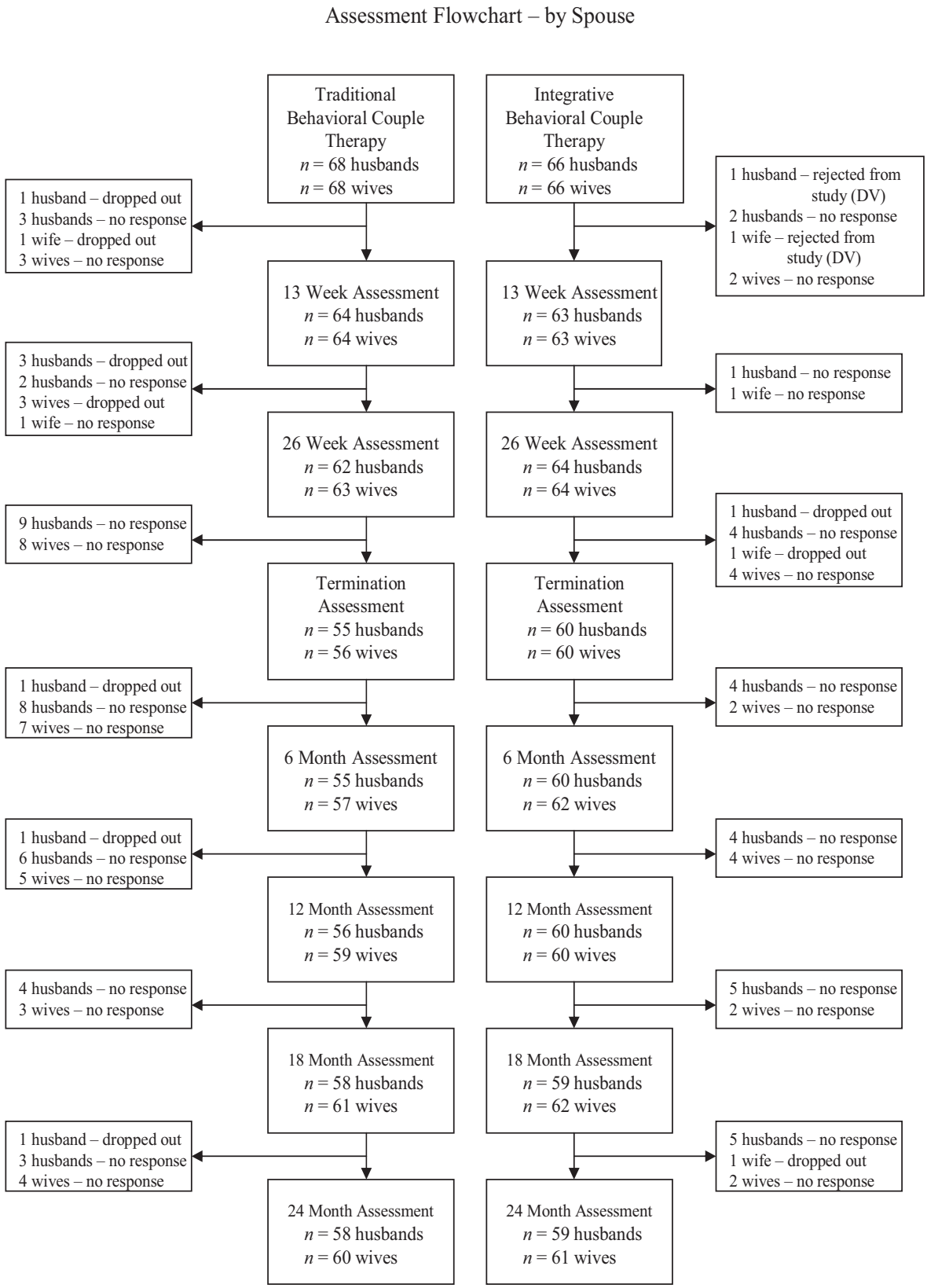

Figure 1. Assessment flowchart from randomization to treatment condition through 2-year follow-up assessment. DV = domestic violence.

Table 1
Means and Standard Deviations for Major Outcome Variables

Measure	Final session		6 months		12 months		18 months		24 months	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Dyadic Adjustment Scale										
TBCT	95.27	18.77	90.19	19.20	92.95	19.11	90.58	20.06	92.87	19.49
IBCT	97.75	18.82	94.84	18.94	97.05	18.61	96.45	17.39	100.45	13.82
Marital Status Inventory										
TBCT	—	—	3.08	3.53	3.20	3.87	3.69	4.00	3.63	4.06
IBCT	—	—	2.76	3.68	3.07	4.17	3.38	4.22	3.31	4.48
Current symptoms										
TBCT	—	—	38.26	6.54	37.99	5.93	38.40	6.00	37.48	5.93
IBCT	—	—	38.04	5.66	38.03	5.93	38.31	5.90	37.38	5.63
Mental Health Index										
TBCT	—	—	63.07	9.05	63.11	8.23	63.22	8.76	64.39	8.56
IBCT	—	—	63.30	8.37	63.61	8.54	62.88	8.56	64.66	8.46
Affective communication										
TBCT	—	—	—	—	60.44	8.53	—	—	60.33	7.74
IBCT	—	—	—	—	58.85	9.28	—	—	57.33	7.78
Problem-solving communication										
TBCT	—	—	—	—	59.33	7.89	—	—	59.59	8.05
IBCT	—	—	—	—	58.01	8.56	—	—	56.27	7.39

Note. Because the length of therapy was allowed to vary but follow-up assessments were determined by length of time since the pretreatment assessment, the actual timing of assessments varies widely around the follow-up time points listed above. Dashes indicate that data were not collected on this measure at this particular time point. TBCT = traditional behavioral couple therapy; IBCT = integrative behavioral couple therapy.

& Bosker, 1999, Section 12.2), or more descriptively, a so-called hockey-stick model. To determine the break point of the hockey-stick model, we tested the fit of successive models across a range of break points suggested by the spaghetti plots for each treatment separately, with final break points at 14 weeks and 22 weeks posttherapy for IBCT and TBCT, respectively. We compared the hockey-stick model with separate break points for each treatment versus models with a polynomial representation of time and a hockey-stick model with a common break point using the Bayesian information criterion (BIC; Raftery, 1995). BIC assesses the relative fit of a given model against other models and balances the complexity of the model (i.e., number of predictors) with the sample size. Raftery (1995) suggested heuristic guides for selecting one model over another according to differences in BIC values: less than 2 is "weak," 2–6 is "positive," 6–10 is "strong," and over 10 is "very strong." The hockey-stick model with separate break points for each treatment was superior to linear, quadratic, and cubic models of time as well as the common break point model with BIC differences of 35, 28, 68, and 7, respectively.

Because no previous couple therapy study has shown a piecewise model of change during follow-up, we also examined the hockey-stick model of change on an individual basis. HLM fits subject-specific trajectories to each participant's data. Thus, for each individual in the analysis it is possible to examine the slopes prior to and after the break point. Eighty-four percent of participants (103/123) who received TBCT had negative slopes prior to the break point at 22 weeks and positive slopes following the break point. Similarly, 89% of participants (112/126) who received IBCT had negative slopes prior to the break point at 14 weeks and positive slopes following the break point. In examining the data at a group level and individual level, we found support for the hockey-stick model of change in the present data.

The fitted regression lines from the hockey-stick model with separate break points are shown in Figure 2. As seen in the figure, the HLM analysis shows TBCT couples ending therapy with a DAS score of 94.4 and IBCT couples nonsignificantly greater at 97.1: $B = 2.7$, $SE = 2.9$, $t(806) = 0.92$, $p = .36$, CI = −3.1, 8.5. Couples in both treatment conditions showed a steep, initial decline in marital satisfaction, losing .27 DAS points per week: $SE = 0.06$, $t(806) = 4.6$, $p < .0001$, CI = −0.39, −0.16. Both treatments had a significant, sharp change in slope: $B = 0.30$, $SE = 0.07$, $t(806) = 4.29$, $p < .0001$, CI = 0.17, 0.43, following the break points at 14 and 22 weeks for IBCT and TBCT, respectively. Thus, on average, IBCT couples lost approximately 3.8 DAS points in the first 14 weeks posttherapy and gained back 2.7 DAS points over the remainder of follow-up; similarly, TBCT couples on average lost approximately 5.9 DAS points during the first 22 weeks posttherapy and regained approximately 2.5 DAS points over the remainder of follow-up. The greatest difference between the treatments occurred at Week 22 and following. A post hoc contrast of the two treatments at 22 weeks revealed a trend for IBCT couples being more satisfied than TBCT couples: $B = 5.0$, $SE = 2.9$, $t(123) = 1.73$, $p = .09$, CI = −0.7, 10.8.

Missing Data

Comparisons of the descriptive data on the DAS in Table 1 with the HLM results just presented reveal several discrepancies. The descriptive data show higher DAS values for both therapies at each time point and also a greater difference between the two treatments. These discrepancies are due both to the presence of missing data and their pattern across the two treatments. There were 1,058 data points included in the HLM analyses of the DAS out of 1,340 total possible (i.e., five time points on two spouses on 134 couples); thus, 21.1% of the possible data are missing. The majority of the missing data on the DAS was due to couples separating and divorcing, at which point they would no longer provide marital satisfaction data (statistical analyses of divorce are presented be-

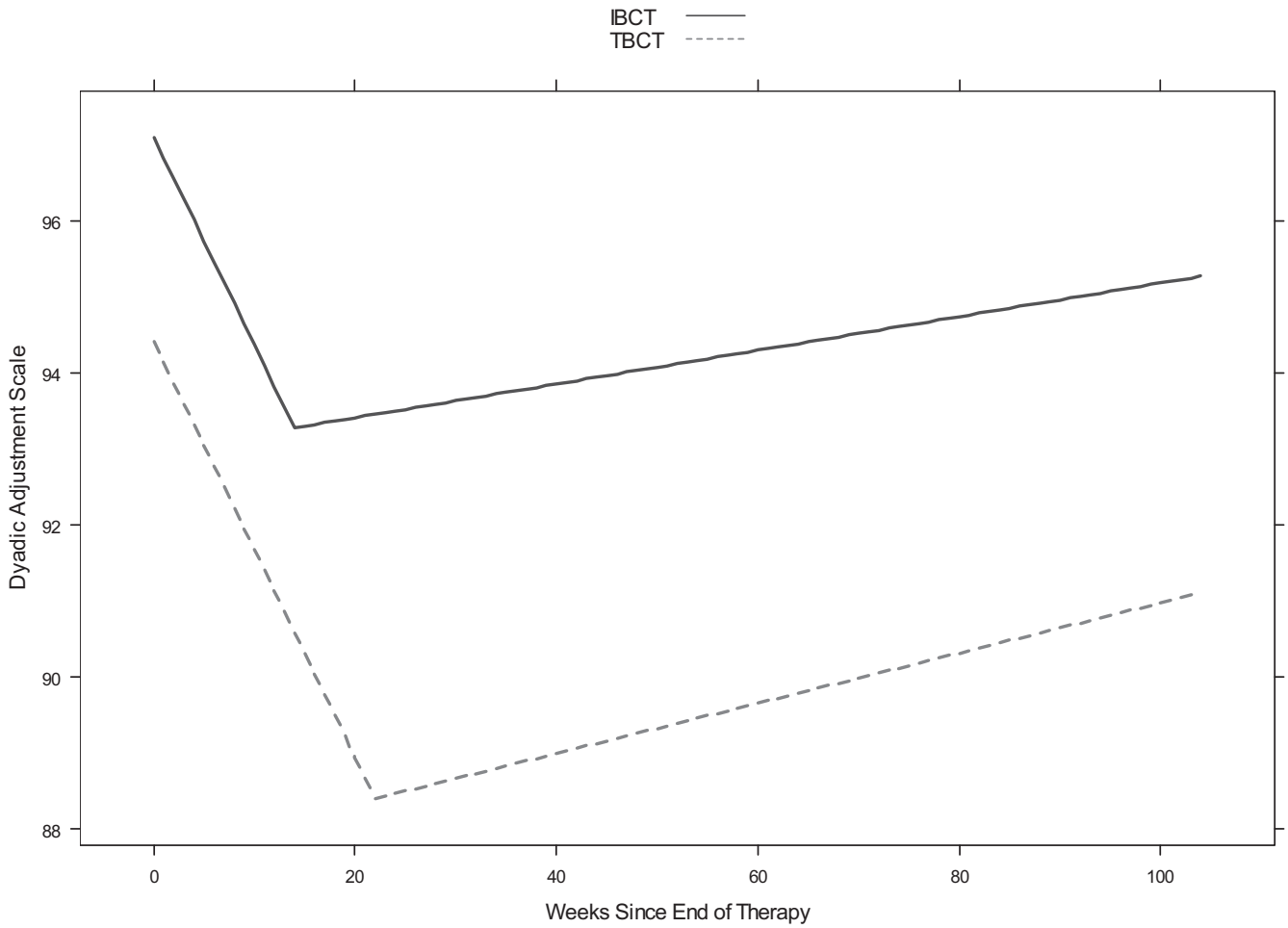

Figure 2. Predicted regression lines of the Dyadic Adjustment Scale (DAS) for the hockey-stick model. IBCT = integrative behavioral couple therapy; TBCT = traditional behavioral couple therapy.

low), thereby elevating the means of the descriptive data at subsequent time points. However, the vast majority of individuals and couples are still being tracked by the study; no data were available on only 4 of the 134 original couples.

HLM is robust to missing data as long as the data can be considered *missing at random*, which means that data can be missing dependent on other variables included in the analysis (see Atkins, 2005, for a discussion of missing data assumptions). If the missing data are dependent on variables not included in the analysis, then HLM may provide biased estimates. To evaluate the effects of divorce and missing data attributable to divorce, we included a divorce dummy variable in the analyses. Not surprisingly, couples who went on to divorce reported significantly lower marital satisfaction regardless of treatment condition: $B = -10.3$, $SE = 5.0$, $t(121) = -2.1$, $p = .04$, CI $= -20.2, -0.4$; in addition, there was some support for a Treatment $\times$ Divorce interaction in predicting marital satisfaction: $B = -12.3$, $SE = 6.8$, $t(121) = -1.8$, $p = .07$, CI $= -25.7, 1.2$. Including the Divorce $\times$ Treatment interaction is similar to fitting separate analyses by those who divorced and those who did not. For those who divorced, TBCT couples had a final session DAS score of 86, whereas IBCT couples had a final session DAS score of 79, on the basis of the HLM analysis. TBCT couples who stayed together had an estimated final session DAS score of 96, whereas IBCT couples had an estimated DAS score of 102. Although the significance of the interaction term is just above the conventional .05 cutoff, the Divorce $\times$ Therapy interaction is retained in the following models as without this term the data would questionably meet the missing at random assumption.[2]

There was no evidence of an effect of divorce on the trajectories of change for either treatment, influenced in part by the paucity of data over time for couples who divorced. After we included the divorce predictor, post hoc comparisons revealed that IBCT couples who stayed together reported higher levels of marital satisfaction than TBCT couples at Week 22 and following: $B = 7.2$, $SE = 2.8$, $t(99) = 2.6$, $p = .01$, CI $= 1.6, 12.9$.

[2] Because it is impossible to definitively know what caused the missing data, often called the *missingness mechanism* (Schafer & Graham, 2002), there are no statistical tests for missing data assumptions.

Covariates of the DAS

We then tested the impact of a series of covariates that were shown to be important during treatment (Christensen et al., 2004), including gender, initial distress, site, and whether the couple was seen during the pilot phase. The only covariate that revealed a significant effect on marital satisfaction was initial distress. It was not surprising that those couples who were severely distressed at the start of therapy reported lower marital satisfaction at the end of therapy: $B = –16.0$, $SE = 2.2$, $t(120) = –7.2$, $p < .0001$, CI = –20.4, –11.6. After accounting for the variance attributable to initial distress, the Therapy × Divorce interaction was more robust: $B = –12.7$, $SE = 5.7$, $t(120) = –2.2$, $p = .03$, CI = –24.0, –1.4.

An additional effect of initial distress was revealed through the spaghetti plots mentioned earlier. There appeared to be notable differences in the variability of the trajectories—specifically, the intercepts—across the combinations of therapy and initial distress: IBCT couples who were moderately distressed initially evidenced less variability in their trajectories relative to the other groups (i.e., TBCT couples regardless of initial distress and IBCT couples who were severely distressed). We examined the variability statistically in two ways. First, we ran separate HLM analyses for each combination of therapy and initial distress and compared the variances for the random effect of the intercept using an F ratio. These analyses revealed that there was significantly less variability in the intercepts of the moderately distressed IBCT couples compared with other groups (p values were .02 or lower; ratio of intercept variances were 2.1 or higher). Second, we also examined the variability of the Level 1 residuals in the HLM analysis, including initial distress. We included a complex variance function that fit a separate Level 1 variance term for each combination of therapy and initial distress, which provided a significantly better fit of the data relative to a model with a common Level 1 error term, $\chi^2(3, N = 125) = 36.1$, $p < .0001$. The pattern of Level 1 variances was similar to what was found with the intercept variances: IBCT moderately distressed couples had 31% to 43% less variability in the scatter of their actual DAS points around the fitted regression lines as compared with other groups. Thus, it appears that IBCT moderately distressed couples had more consistent change as a group relative to the greater variability seen in the other groups.

MSI

Because the MSI assesses steps taken toward separation and divorce, its distribution is strongly skewed; at the end of therapy, many individuals reported no or few steps taken toward separation and divorce whereas a smaller percentage of individuals endorsed greater levels of marital instability. Outcome data such as these are not well fit by methods that assume normally distributed residuals. Thus, we modeled the MSI using a Poisson hierarchical generalized linear model (HGLM; Raudenbush & Bryk, 2002), which assumes that the Level 1 residuals are distributed as a Poisson random variable whereas all other random effects are assumed to be normally distributed.

The initial model with the MSI included therapy, time in weeks (as there was no evidence of nonlinear effects), and a dummy variable indicating whether or not the couple divorced at some point during follow-up. The Poisson HGLM estimated that TBCT couples had taken 1.3 steps toward divorce at the end of therapy: $B = 0.29$, $SE = 0.13$, $t(673) = 2.2$, $p = .03$, CI = 0.03, 0.56; whereas IBCT couples had taken 0.8 steps toward divorce, significantly fewer than TBCT couples: $B = −0.41$, $SE = 0.17$, $t(124) = −2.4$, $p = .02$, CI = −0.76, −0.07.[3] There was a small but significant increase in steps taken toward divorce for all couples over the follow-up period: $B = 0.003$, $SE = 0.0006$, $t(673) = 4.2$, $p < .0001$, CI = 0.001, 0.004. At 2 years posttherapy, all couples are predicted to have taken 1.3 more steps toward divorce relative to the end of treatment. Finally, as we would expect, couples who went on to divorce reported 6.2 more steps taken toward divorce at the end of therapy relative to those who did not go on to divorce: $B = 1.8$, $SE = 0.2$, $t(124) = 8.8$, $p < .0001$, CI = 1.4, 2.2.

In examining the covariates, only initial distress revealed any significant association in the HGLM analysis of the MSI. The model containing time, therapy, initial distress, and divorce dummy variable showed a three-way interaction between therapy, initial distress, and time: $B = −0.009$, $SE = 0.003$, $t(670) = −3.1$, $p = .002$, CI = −0.014, −0.003. Plotting the regression lines from this model revealed that moderately distressed IBCT couples reported 0.5 steps taken toward divorce, and moderately distressed TBCT reported 1.5 at the end of therapy; the steps taken toward divorce were relatively constant across the 2-year follow-up for both groups. Conversely, IBCT couples who were severely distressed reported approximately 3 steps taken toward divorce at the end of therapy with little change during follow-up, and TBCT couples who were severely distressed reported just over 2 steps taken toward divorce at the end of therapy; however, this increased notably during follow-up. Thus, the three-way interaction was driven by the increasing steps toward divorce for the TBCT severely distressed group compared with little to no change over time in the other groups.

MHI and CS

Both the MHI and CS were highly skewed, though in opposite directions. The majority of study participants reported few symptoms of psychopathology; however, a smaller proportion of the sample reported elevated levels of psychopathology. Similarly, the majority of study participants reported relatively high levels of overall mental health with a tail of the distribution stretching into lower levels of the MHI. We used the Box-Cox procedure to estimate optimal transformations for normalizing CS and MHI. CS was largely normalized by raising it to the –3 power, whereas MHI was normalized by raising it to the + 3 power. Quantile-quantile plots revealed that the transformations led to satisfactory normal distributions.[4] Both variables showed slight, smooth nonlinearity and were modeled using linear and quadratic terms for time.

[3] Poisson mixed-effects models use a log-link to transform the relationship of the predictors with the outcome. The reported regression coefficients are on the log scale, but numbers reported in the text have been back-transformed to the original metric of the MSI.

[4] Because of the odd transformations, the regression coefficients were either fantastically small or large and uninterpretable. Thus, we present only t statistics and p values; interpretations of the findings were based on plots of the fitted regression lines, back-transformed to the original metric of the variables.

We initially examined how CS and MHI changed over the course of follow-up and found that there was little change on average. However, our hypotheses about individual functioning concern primarily how these variables change with alterations in marital satisfaction. To assess the covariation over time between individual functioning and marital satisfaction, we treated the DAS as a time-varying covariate. Following the recommendations of Diggle et al. (2002) and Singer and Willett (2003), we transformed each individual's series of DAS scores into two components. DAS at posttherapy was included as a time-invariant predictor, and the deviations of subsequent DAS points from posttherapy were treated as a time-varying predictor. These two components reflect the overall level of marital satisfaction and the change in marital satisfaction throughout the posttherapy follow-up period.

Using these two components of the DAS as predictors, there was reliable evidence for the covariation of marital satisfaction with individual functioning, using both CS and MHI. Posttherapy DAS, $t(107) = 4.2$, $p = .0001$, and DAS deviations, $t(551) = 5.5$, $p < .0001$, showed highly reliable associations with CS, reflecting an association between marital satisfaction and psychopathology at the end of treatment and also during the follow-up period. However, the overall magnitude of the effects was not huge. One *SD* changes in either component of the DAS were associated with approximately 1.5 point changes in CS. After accounting for the association with the DAS, there was no other change in CS. Thus, there was a reliable but small association between marital satisfaction and reported psychopathology: As marital satisfaction increased, psychopathology decreased.

In a very similar fashion, a model examining the covariation across time of the DAS and MHI revealed strong effects for the posttherapy DAS, $t(107) = 6.0$, $p < .0001$, and for the DAS deviations, $t(546) = 7.3$, $p < .0001$. Similar to CS, one *SD* changes in either posttherapy DAS or DAS deviations were associated with 2.5–3.0 points of change in the MHI.

AFC and PSC

Because the MSI-R, which includes AFC and PSC, was given only twice during follow-up, we modeled time as a simple dummy variable, where 0 represented the 12-month assessment and 1 represented the 24-month assessment. An initial model with AFC and PSC included therapy, time, and divorce as covariates. The HLM analyses revealed that TBCT couples at the 12-month assessment were right at the distress cutoff for the AFC of 60 on average: $B = 60.1$, $SE = 0.9$, $t(107) = 64.8$, $p < .0001$, CI = 58.2, 61.8. IBCT couples were nonsignificantly less distressed in their AFC: $B = -2.1$, $SE = 1.2$, $t(107) = -1.7$, $p = .09$, CI = −4.6, 0.3. There was no evidence that AFC changed over time for couples: $B = 0.1$, $SE = 0.5$, $t(107) = 0.3$, $p = .77$, CI = −0.8, 1.0. As we might expect, those couples who went on to divorce reported significantly more problems with AFC: $B = 6.7$, $SE = 2.3$, $t(107) = 2.9$, $p = .005$, CI = 2.1, 11.3.

There was a trend that women reported less problems with AFC than men: $B = -1.5$, $SE = 0.8$, $t(105) = -1.8$, $p = .07$, CI = −3.2, 0.1. In addition, there was a significant interaction between time and initial severity: $B = -2.7$, $SE = 0.9$, $t(177) = -2.9$, $p = .004$, CI = −4.5, −0.9. When we examined the predicted means for the interaction, we found that moderately distressed couples reported fewer problems with AFC during follow-up ($M = 56.3$ across both assessments), with little change between the 1-year and 2-year assessments. Conversely, couples who were initially severely distressed reported worse communication at the 1-year assessment ($M = 63.4$) that continued to improve up to the 2-year assessment ($M = 61.1$). There was no evidence that therapy interacted with either initial distress or assessment period: $B = -2.3$, $SE = 1.6$, $t(105) = -1.4$, $p = .15$, CI = −5.5, 0.9.

The effects for PSC were virtually identical to those for AFC. The initial model with therapy, assessment phase, and divorce revealed that TBCT couples had a predicted mean close to the distress cutoff for PSC at the 1-year assessment: $B = 59.45$, $SE = 0.9$, $t(107) = 64.0$, $p < .0001$, CI = 57.6, 61.3. Results revealed a trend that IBCT couples were lower than TBCT couples across both assessment points: $B = -2.3$, $SE = 1.2$, $t(107) = -1.9$, $p = .07$, CI = −4.8, 0.2. There was no evidence for change over time for any couples, and couples who later divorced reported significantly more distress in PSC: $B = 4.8$, $SE = 2.3$, $t(107) = 2.1$, $p = .04$, CI = 0.3, 9.4. In considering the covariates, there were no apparent gender differences in couples' reported difficulties with PSC, though there was again a significant interaction between initial distress and assessment period: $B = -1.8$, $SE = 0.9$, $t(178) = -2.0$, $p = .05$, CI = −3.5, 0. The pattern of the predicted means was very similar to what was seen with AFC. Moderately distressed couples reported fewer problems and little change in PSC during follow-up ($M = 56.6$ over both assessment periods). Severely distressed couples reported greater difficulties with PSC at the 1-year assessment ($M = 61.0$) that reduced somewhat at the 2-year assessment ($M = 59.2$). After accounting for initial distress, the therapy differences were somewhat more pronounced: $B = -3.5$, $SE = 1.7$, $t(105) = -2.1$, $p = .04$, CI = −7.0, −0.1.

In sum, for both measures of communication, there was no general evidence of change over follow-up. Severely distressed couples reported more problems with communication than moderately distressed couples, but unlike moderately distressed couples, they showed improvement in communication from the first to the second follow-up point. Not surprisingly, those couples who later divorced evidenced poorer communication.

MAQ

The MAQ consisted of two subscales that measured marital behaviors that would be promoted by TBCT and IBCT. At 2 years posttherapy, couples who had received IBCT reported using IBCT behaviors between rarely and sometimes on the MAQ-IBCT ($M = 4.1$, $SD = 1.1$); couples who had received TBCT reported levels just below those of the IBCT couples ($M = 3.7$, $SD = 1.0$). For TBCT behaviors (MAQ-TBCT), IBCT couples reported an average of 3.0 behaviors ($SD = 1.1$), whereas TBCT couples reported an average of 2.8 behaviors ($SD = 1.2$).

HLM analyses confirmed these descriptive results: IBCT couples reported significantly more IBCT behaviors than TBCT couples: $B = 0.4$, $SE = 0.2$, $t(101) = 2.1$, $p = .04$, CI = 0.02, 0.68. Conversely, there was not a statistically reliable difference between the two treatment conditions with TBCT behaviors: $B = 0.2$, $SE = 0.2$, $t(101) = 1.3$, $p = .20$, CI = −0.1, 0.6. In looking at the covariates, there were no effects for couples' initial level of distress or whether the couple was in the pilot phase of the project. However, for both sets of behaviors, there was a significant three-way interaction between therapy condition, gender, and site: For

Table 2
Clinically Significant Outcome on the Dyadic Adjustment Scale for Couples in Traditional (TBCT) and Integrative (IBCT) Behavioral Couple Therapy

	Termination						2-year follow-up					
	IBCT		TBCT		Total		IBCT		TBCT		Total	
Outcome	Count	% within therapy	Count	% within therapy	Count	% within therapy	Count	% within therapy	Count	% within therapy	Count	% within therapy
Deteriorated	5	7.8	12	18.2	17	13.1	14	21.9	15	25.0	29	23.4
Unchanged	14	21.9	14	21.2	28	21.5	6	9.4	9	15.0	15	12.1
Improved	12	18.8	11	16.7	23	17.7	14	21.9	14	23.3	28	22.6
Recovered	33	51.6	29	43.9	62	47.7	30	46.9	22	36.7	52	41.9
Total	64	100.0	66	100.0	130	100.0	64	100.0	60	100.0	124	100.0

IBCT behaviors, $B = 1.4$, $SE = 0.5$, $t(94) = 2.8$, $p = .01$, CI = 0.4, 2.4; for TBCT behaviors, $B = 1.3$, $SE = 0.6$, $t(94) = 2.3$, $p = .02$, CI = 0.2, 2.4. The predicted means showed that UCLA husbands were driving this effect. UCLA husbands who received IBCT reported more IBCT and more TBCT behaviors than any other group, and UCLA husbands who had received TBCT reported fewer IBCT behaviors and fewer TBCT behaviors than any other group.

Separation and Divorce

Of the 134 couples who began the study, relationship status at 2 years was known on all but 4 (3 couples in TBCT and 1 in IBCT). Of the couples with known relationship status, 10 out of 65 couples (or 15%) in TBCT had divorced by the 2-year follow-up assessment (which is 2.5 years into the study). Fourteen out of 65 (or 22%) couples who received IBCT had divorced by the 2-year assessment. There were no statistically significant differences in either the rate, $\chi^2(1, N = 130) = 0.28$, $p = .60$, or timing of the divorces between the two treatments.

Clinical Significance

Clinical significance data for both treatments using the method of Jacobson and Truax (1991) are presented in Table 2. Results were based on couple averages of the DAS, and couples who divorced were counted as deteriorated to be consistent with the reports during therapy (Christensen et al., 2004). For these estimates we used the pretherapy and 2-year assessment DAS to calculate clinical significance.[5] At 2 years posttherapy, just under half of IBCT couples were recovered and just over a third of TBCT couples were classified as recovered. Moreover, approximately two thirds of IBCT couples were either improved or recovered as compared with 60% of TBCT couples. There was no evidence for therapy differences in classification at the 2-year assessment, $\chi^2(3, N = 124) = 1.74$, $p = .63$.

We also computed clinical significance statistics focusing on the change from posttherapy to 2-year follow-up assessment. Our primary interest here was whether those couples who had improved (i.e., improved or recovered classifications) at the end of therapy would maintain those gains across follow-up, and conversely, whether couples who had not improved (i.e., deteriorated or unchanged classifications) at the end of therapy would improve following therapy. Maintenance during follow-up was designated by any classification other than deteriorated for those who had improved previously. Seventy-four percent of IBCT couples maintained their gains during follow-up (i.e., 29/39 couples), and 69.7% of TBCT couples maintained their gains (i.e., 23/33 couples). Of those couples who did not improve during therapy (and also were still together at the 2-year follow-up assessment), 55.6% of IBCT couples improved during follow-up (i.e., 5/9 couples) and 21.4% of TBCT couples who had not improved during therapy did so in the follow-up period (i.e., 3/14 couples).

Additional Therapy

During the 2 years following therapy, only five couples reported further marital therapy; four of these couples had received IBCT, and one had received TBCT, which is not statistically different using Fisher's exact test: odds ratio = 0.24, 95% CI for odds ratio = 0.01, 2.60. In examining additional therapy besides marital therapy (i.e., individual, family), we found that 33% of individuals who had received IBCT reported receiving some type of therapy in the 2 years following treatment, whereas only 18% of individuals who had received TBCT reported some type of therapy during follow-up. Using a mixed-effects logistic regression, we found an odds ratio of 0.33 comparing TBCT with IBCT: $B = -1.10$, $t(119) = -1.86$, $p = .06$, 95% CI for odds ratio = 0.12, 0.80. In addition, there was a significant gender effect, with an odds ratio of 2.34 comparing women with men: $B = 0.85$, $t(119) = 2.30$, $p = .02$, 95% CI for odds ratio = 1.15, 5.10. Thus, individuals who received TBCT were significantly less likely to be involved in any type of therapy following marital therapy, and women were more likely than men to be involved in therapy.

We also examined whether couples who had received therapy during the follow-up period reported differences in marital satisfaction relative to those who did not report therapy. Because only five couples sought additional couple therapy, we examined the data descriptively.

[5] There are only 124 couples represented in Table 2. In addition to the four couples who have been lost to follow-up, six other couples did not complete the 2-year assessment, and it was impossible to calculate clinical significance statistics for them. A slight difference between the termination data in this table and those reported in Table 2 in Christensen et al. (2004) is because of the misclassification of one couple in the earlier report.

At the end of therapy, couples who received additional therapy during follow-up reported higher marital satisfaction as measured by the DAS ($M = 102.9$, $SD = 14.5$) than couples who did not seek additional marital treatment ($M = 96.3$, $SD = 18.8$). In fact, couples who sought additional couple therapy reported greater levels of satisfaction compared with those who did not at each assessment point of follow-up, with the exception of the 2-year follow-up assessment ($M = 90.5$, $SD = 18.3$ and $M = 97.2$, $SD = 17.1$ for couples who sought couple therapy and did not, respectively). In addition, none of the five couples who sought additional couple therapy separated or divorced during the 2 years following therapy.

Finally, we examined differences in relationship satisfaction trajectories across the follow-up period for individuals who reported additional therapy other than marital therapy (i.e., individual, family, and so forth) versus those who reported no additional therapy. When we included a dichotomous predictor to our hockey-stick model of the DAS, we found highly significant effects of additional nonmarital therapy on early change: $B = 0.32$, $SE = 0.12$, $t(769) = 2.73$, $p = .01$, CI = 0.09, 0.55; and on the therapy-specific slopes for later change: For TBCT, $B = -0.41$, $SE = 0.14$, $t(769) = -2.84$, $p < .01$, CI = −0.69, −0.13; for IBCT, $B = -0.39$, $SE = 0.13$, $t(769) = -2.95$, $p < .01$, CI = −0.65, −0.13. Plots of the predicted regression lines showed that individuals who received nonmarital therapy during follow-up had a linear decline in marital satisfaction, losing approximately 4.5 DAS points over the 2 years following therapy; thus, they did not show the hockey-stick pattern of change. Moreover, the pattern appeared to be similar across the two therapies.

Discussion

The first question any follow-up study must answer is, "What is the final outcome of participants?" The data on clinically significant change using the DAS, our primary outcome measure, indicate that almost two thirds of couples were reliably improved or recovered up to 2 years after the end of treatment. Given this sample of couples was selected for being chronically and significantly distressed (Christensen et al., 2004) and given existing evidence that couples in control groups generally do not change over time (Baucom et al., 2003), the current data are encouraging about the success of BCT.

The only other study to follow couples for 2 years posttreatment (Jacobson et al., 1987) found that across three versions of BCT, 43% of the available sample of 34 couples were reliably improved from intake to 2-year follow-up whereas 56% were unchanged or had deteriorated. A statistical comparison of results from Jacobson et al.'s study with our study indicates that a significantly higher percentage of couples in our study were reliably improved: $\delta = .22$, $SE = .10$, $Z = 2.3$, $p = .02$, CI = 0.03, 0.40. The longest follow-up to date of a randomized trial of couple therapy (Snyder et al., 1991) was conducted 4 years posttreatment and found that 42% of an available sample of 29 couples who received BCT showed improvement whereas the remaining were unchanged or deteriorated. A statistical comparison of results from Snyder et al.'s study with our study indicates that a significantly higher percentage of couples in our study were reliably improved: $\delta = .23$, $SE = 0.10$, $Z = 2.2$, $p = .03$, CI = 0.03, 0.42. Thus, the current data compare quite favorably with these earlier results. A substantial majority of our couples got lasting benefit from behaviorally oriented couple therapy.

In addition to examining final outcome, we examined the trajectory of change in marital satisfaction across follow-up points. This trajectory of change is perhaps most notable for its hockey-stick shape—an initial drop in marital satisfaction immediately following treatment followed by a gradual increase in satisfaction over the course of the 2-year follow-up. There are two features of the study that could account for this shape. First, at the final treatment session, the therapist gave each member of the couple questionnaires and a stamped envelope and asked the couple to independently complete the measures and mail them to the project immediately following the therapy session. Couples may have felt particularly grateful at their final visit with their therapist and may have given an inflated view of their satisfaction. Second, throughout the study but especially during the 2-year follow-up, some of the most distressed couples separated or divorced and thus did not provide measures of marital satisfaction. The removal of the worst cases through divorce may have affected the shape of change. Thus, the hockey-stick shape of the trajectory of change in satisfaction could be somewhat artifactual.

However, participants completed a measure of client satisfaction with services at their final session as well. We included this measure in our HLM analysis of marital satisfaction as a proxy for client gratefulness. Results showed that client satisfaction with services is strongly related to each component of the trajectory during follow-up (i.e., intercept, early slope, and later slope). Those participants who were the most satisfied with services reported greater marital satisfaction at the end of therapy, a steeper drop in satisfaction following therapy, and a more rapid improvement later in the follow-up period. Thus, these couples who were the most satisfied with the services they had received showed the most prominent hockey-stick trajectory. At the same time, client satisfaction with services did not completely account for the piecewise nature of change; regardless of their satisfaction with services, the majority of couples showed a hockey-stick trajectory. Thus, client satisfaction with services moderates but does not completely account for the hockey-stick model.[6]

Moreover, our HLM analyses included a divorce predictor and revealed no evidence for an effect of divorce on the trajectories of change for either treatment, although the lack of data over time from couples who divorced could contribute to this null finding. If the hockey-stick trajectory of change is confirmed by future research, it does provide some optimism about postmarital therapy adjustment. We need not fear that couples will begin a slow, inexorable slide in satisfaction after the end of couple therapy.

In the report of the initial outcome from this study (Christensen et al., 2004), we presented evidence that the two treatments being compared in this study were distinct. Observational ratings of therapy sessions indicated that therapists were adhering to the two different treatment protocols. Furthermore, although the overall outcome at the end of treatment was not significantly different, there was a significantly different trajectory of change during treatment. Couples in TBCT made rapid progress early on in

[6] We thank an anonymous reviewer for suggesting this idea. Details on the analysis are available from David C. Atkins at datkins@fuller.edu

treatment and then flattened out whereas couples in IBCT made slow and steady progress throughout treatment.

In the current follow-up study, we find no dramatic differences in final outcome. Also, on the MAQ, there were few differences between treatment conditions in the therapeutic activities that they reported at 2-year follow-up. Relevant TBCT behaviors, such as setting aside specific times to communicate about a problem, are much more specific and concrete than IBCT behaviors, such as showing empathy for the partner. Therefore, the lack of significant differences between TBCT, which heavily emphasized those activities, and IBCT, which did not, is surprising. However, consistent with our findings, Jacobson et al. (1987) found that most couples treated with a traditional behavioral approach were not doing the activities they had been taught in therapy at 2 years posttreatment.

Despite evidence of similarity in final outcome, we find differences in the trajectory of change. Couples in IBCT tended to reverse courses and improve in satisfaction sooner than TBCT couples. Also, our data suggested other important differences between treatments. Although both treatments showed, not surprisingly, a difference in satisfaction scores between those couples who stayed together and those who separated, this difference was greater in IBCT than in TBCT. In general, couples who stayed together fared better in IBCT than in TBCT. Finally, there was less volatility throughout follow-up in IBCT than in TBCT. This difference was apparent in the DAS scores, where IBCT moderately distressed couples showed relatively little volatility, and in the MSI, where the TBCT severely distressed couples showed a steep increase in the number of steps toward divorce over follow-up.

These differences in outcome follow quite naturally from differences in the two treatments. In TBCT, the focus is upon increasing positive behavior, building skills, and incorporating content appropriate to those skills. Attention to major contentious issues between the couple is explicitly delayed until positive interactions can be encouraged and until communication and problem-solving training can be taught and practiced on less contentious issues. If there are difficulties in encouraging positive interactions and/or teaching communication and problem solving, specific attention to major contentious issues may be limited. In contrast, there is no timetable for discussions in IBCT. Couples are encouraged to discuss whatever issues are on their mind. Thus, contentious issues are usually confronted earlier and more thoroughly in IBCT than in TBCT. This difference may lead to a lower level of satisfaction in couples that eventually separate or divorce (they have already faced the worst in therapy) but fewer surprises after treatment is over, perhaps leading to less volatility in IBCT and greater maintenance of treatment gains in IBCT when the couple stays together.

There was little change over follow-up in our two measures of communication, PSC and AFC. However, there were only two measurements of these variables at 1 year and 2 years after treatment. There was also little change in our measures of individual functioning over follow-up: psychological symptoms and the MHI. However, change that did occur was related to change in marital satisfaction.

Very few couples obtained additional couple therapy during the follow-up period. We prohibited them from seeing their project couple therapist for 2 years after treatment and encouraged them to build on what they had learned in couple therapy. A number of participants did pursue other kinds of therapy. The fact that these participants did not do as well as those who did not pursue therapy could reflect poorly on their therapy. More likely, these people pursued individual therapy or child therapy because they were unhappy with their partner and were considering the possibility of separation and divorce.

The most serious weakness of the current study is its complete reliance on self-report measures. However, the primary measure in couple therapy studies has always been relationship satisfaction, which by its very nature must be measured through self report. Also, one of the major limitations of self-report, namely its vulnerability to external influences (demand characteristics, social desirability) is probably minimized in a follow-up study when data are collected long after contact with the therapist and at a distance (mailed questionnaires).

There are a number of strengths of the current study. First, it is the largest follow-up of a randomized controlled study of couple therapy. The number of couples in behavior treatment is approximately three times that of either Jacobson et al. (1987) or Snyder et al. (1991). Second, unlike most previous follow-up studies, which just focused on marital outcomes such as relationship status and satisfaction, the current study also assessed individual functioning, such as psychological symptoms, and broader marital functioning, such as communication. Third, this is one of the few studies to track continued treatment throughout follow-up. Fourth, the study has one of the most ethnically diverse samples and one of the most maritally distressed samples of any available clinical trial. Both factors make generalization of the current findings to clinical work in the real world easier. Fifth, and also noted in the original report, the study provides a rigorous test of the two different treatments because of several methodological features, such as rigorous measures of adherence and competence and the use of experienced therapists receiving intense supervision. Finally, this is the first clinical trial of couple therapy that has assessed couples repeatedly over follow-up and thus been able to examine the trajectory of change in the 2 years following treatment. Thus, we have information not only about functioning at the end point of follow-up, but also how couples traversed to that end point.

Within its limitations of sample, therapists, and measures, the current study suggests that two thirds of couples who were significantly and chronically distressed at pretreatment can be significantly improved at 2 years posttreatment. Although there is a decline in satisfaction immediately posttreatment, there is a gradual increase throughout the remaining follow-up period. Couples in the two behavioral treatments compared in this study are largely similar in outcome, although a number of findings give an edge to IBCT. Clinicians can confidently apply these treatments to even severely distressed couples but perhaps with an effort to taper treatment toward the end of therapy to minimize abrupt drops in satisfaction. Future research should investigate how couples fare over even longer periods than 2 years posttreatment.

References

Atkins, D. C. (2005). Using multilevel models to analyze marital and family treatment data: Basic and advanced issues. *Journal of Family Psychology, 19,* 98–110.

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.

Baucom, D. H., Shoham, V., Mueser, K. T., Daiuto, A. D., & Stickle, T. R. (1998). Empirically supported couple and family interventions for marital distress and adult mental health problems. *Journal of Consulting and Clinical Psychology, 66,* 53–88.

Christensen, A. (1999). *Marital activities questionnaire*. Unpublished questionnaire, University of California, Los Angeles.

Christensen, A., Atkins, D. C., Berns, S., Wheeler, J., Baucom, D. H., & Simpson, L. E. (2004). Traditional versus integrative behavioral couple therapy for significantly and chronically distressed married couples. *Journal of Consulting and Clinical Psychology, 72,* 176–191.

Christensen, A., & Heavey, C. L. (1999). Interventions for couples. In J. T. Spence, J. M. Darley, & D. J. Foss (Eds.), *Annual review of psychology* (pp. 165–190). Palo Alto, CA: Annual Reviews.

Christensen, A., & Jacobson, N. S. (2000). *Reconcilable differences.* New York: Guilford Press.

Diggle, P. J., Heagerty, P., Liang, K. Y., & Zeger, S. L. (2002). *Analysis of longitudinal data* (2nd ed.). New York: Oxford University Press.

Gottman, J. M., Notarius, C., Gonso, J., & Markman, H. (1977). *A couple's guide to communication.* Champaign, IL: Research Press.

Jacobson, N. S., & Christensen, A. (1998). *Acceptance and change in couple therapy: A therapist's guide to transforming relationships.* New York: Norton.

Jacobson, N. S., Christensen, A., Prince, S. E., Cordova, J., & Eldridge, K. (2000). Integrative behavioral couple therapy: An acceptance-based, promising new treatment for couple discord. *Journal of Consulting and Clinical Psychology, 68,* 351–355.

Jacobson, N. S., & Margolin, G. (1979). *Marital therapy: Strategies based on social learning and behavior exchange principles.* New York: Brunner/Mazel.

Jacobson, N. S., Schmaling, K. B., & Holtzworth-Munroe, A. (1987). Component analysis of behavioral marital therapy: 2-year follow-up and prediction of relapse. *Journal of Marital and Family Therapy, 13,* 187–195.

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.

Northey, W. F., Jr. (2002). Characteristics and clinical practices of marriage and family therapists: A national survey. *Journal of Marital and Family Therapy, 28,* 487–494.

Pinheiro, J., Bates, D., DebRoy, S., & Sarkar, D. (2005). *nlme: Linear and nonlinear mixed effects models* (R package Version 3.1–64). Vienna, Austria: R Foundation for Statistical Computing. Available from http://www.r-project.org

Raftery, A. E. (1995). Bayesian model selection in social research (with discussion). In P. V. Marsden (Ed.), *Sociological methodology 1995* (pp. 111–195). Cambridge, MA: Blackwell.

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

R Development Core Team. (2005). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. Available from http://www.r-project.org

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7,* 147–177.

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.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence.* New York: Oxford University Press.

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

Snyder, D. K. (1997). *Marital Satisfaction Inventory—Revised (MSI-R) manual.* Los Angeles, CA: Western Psychological Services.

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

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.

Sperry, L., Brill, P. L., Howard, K. I., & Grissom, G. R. (1996). *Treatment outcomes in psychotherapy and psychiatric interventions.* New York: Brunner/Mazel.

Weiss, R. L., & Cerreto, M. (1980). The Marital Status Inventory: Development of a measure of dissolution potential. *The American Journal of Family Therapy, 8,* 80–86.

Wimberly, J. D. (1998). An outcome study of integrative couples therapy delivered in a group format (Doctoral dissertation, University of Montana, 1997). *Dissertation Abstracts International: Series B. The Physical Sciences and Engineering, 58*(12), 6832B.

Received November 23, 2005
Revision received July 7, 2006
Accepted July 17, 2006 ■