Prediction of Response to Treatment in a Randomized Clinical Trial of Marital Therapy

David C. Atkins
Fuller Graduate School of Psychology

Sara B. Berns and William H. George
University of Washington

Brian D. Doss
Texas A&M University

Krista Gattis and Andrew Christensen
University of California, Los Angeles

This study investigated demographic, intrapersonal, and interpersonal predictors of treatment response in a randomized clinical trial of 134 distressed married couples, which examined traditional (N. S. Jacobson & G. Margolin, 1979) and integrative (N. S. Jacobson & A. Christensen, 1996) behavioral couple therapy. Results based on hierarchical linear modeling revealed that interpersonal variables were the strongest predictors, but their effects were largely limited to predicting initial marital dissatisfaction; greater individual mental health was also associated with less distress initially. Couples who were married longer demonstrated stronger treatment gains, and exploratory analyses suggested that sexually dissatisfied couples showed slower initial, but overall more consistent, gains in the integrative versus the traditional approach. Findings are considered in light of the previous literature on predicting response to marital therapy.

Keywords: marital therapy, treatment response, couple functioning

Dozens of randomized clinical trials attest to the efficacy of couple therapy for improving relationship satisfaction (see Baucom, Shoham, Mueser, Daiuto, & Stickle, 1998; Christensen & Heavey, 1999, for recent reviews). This research demonstrates that a number of distinct treatment interventions produce greater change in marital satisfaction than exhibited in control groups. Despite these promising findings, a substantial number of couples—in some cases almost 50% (Jacobson & Addis, 1993)—are not responsive to treatment. Currently, we know little about the factors that distinguish between those couples who respond to treatment and those who do not. As a result, we have little information to guide treatment revisions, make them more powerful, or apply them more efficiently with couples likely to benefit

from them. The purpose of the present investigation was to address this limitation in the literature by examining pretreatment measures that predict which couples benefit from behaviorally oriented couple therapy. The focus is upon measures that are cost-effective and can be administered easily by a practicing clinician.

A number of studies have examined predictors of outcome—typically, marital satisfaction—in different types of couple therapy, with the majority of this research focused on behavioral approaches. However, these studies are plagued with methodological problems and inconsistent findings that make interpretation and synthesis difficult. Predictors from previous research can be grouped into the following three types of variables: (a) demographic variables (e.g., age and years married), (b) interpersonal variables (e.g., communication, intimacy, and commitment), and (c) intrapersonal variables (e.g., personality and psychopathology). We briefly review the previous research with respect to these classes of predictors.[1]

Regarding demographic variables, four studies have found evidence of more favorable treatment outcome for younger couples (Baucom & Aiken, 1984; Bennun, 1985; Hahlweg, Schindler, Revenstorf, & Brangelmann, 1984; O'Leary & Turkewitz, 1981), but three other studies failed to replicate these findings (Crowe, 1978; Jacobson, Follette, & Pagel, 1986; Mendonca, Lumley, & Hunt, 1982). Two studies have revealed conflicting relationships between education and success in treatment, and one study found that less educated couples responded more favorably (Crowe, 1978); however, another study found no relationship between education and treatment success (Mendonca et al., 1982). Other

---

[1] Each of the studies reviewed in the following section used behavioral treatments with one exception: Snyder, Mangrum, and Wills (1993) compared behavioral marital therapy with insight-oriented marital therapy.

demographic variables found to be unrelated to marital therapy outcome have included occupational status, years married, number of previous marriages, and number of children (Crowe, 1978; Mendonca et al., 1982).

The research on interpersonal variables has also yielded contradictory and counterintuitive findings. Cain (1997) found that negative communication predicted poorer outcomes, whereas Snyder, Mangrum, and Wills (1993) found that pretreatment communication variables were unrelated to treatment outcome. Regarding intimacy, two studies found that couples reporting conflicts over sex were less likely to improve in therapy (Cain, 1997; Hahlweg et al., 1984), and one study found that differences in desired levels of intimacy was negatively related to improvement (Jacobson et al., 1986). In a rare show of consistency, three studies of commitment (Beach & Broderick, 1983; Crowe, 1978; Hahlweg et al., 1984) found that the lower the commitment prior to therapy—measured as steps taken toward divorce—the more likely the couple was to either separate or fail to show improvements in satisfaction.

Studies of intrapersonal predictors of outcome are sparse, and the results are inconsistent. For example, Jacobson et al. (1986) found that couples in which one spouse exhibited depression at pretreatment were more likely to respond positively to marital therapy, whereas Snyder and colleagues (1993) found that depression at pretreatment was negatively related to marital satisfaction at posttreatment. Other intrapersonal variables, such as personality, have yet to be studied as predictors of improvement in satisfaction despite evidence that neuroticism is among the strongest and most consistent predictors of marital instability and dissatisfaction cross-sectionally (Karney & Bradbury, 1995) though not longitudinally (Karney & Bradbury, 1997).

A number of methodological limitations in previous studies further complicate conclusions about which variables affect outcome in marital therapy. Many studies have used large numbers of predictor variables with relatively small samples, which leads to low power and the possibility of chance findings. In addition, almost all prediction studies have assessed couples only twice during acute treatment (at pretreatment and posttreatment); however, two points of data can only describe the amount of change during therapy and reveal nothing about how this change occurred. In addition to examining spouses at only two time points, most previous studies averaged husband and wife outcome scores together, rather than examining them individually. Averaging in this manner reduces the sample size and loses crucial information regarding the differential impact of predictor variables on each individual spouse.

The current study examined the ability of pretreatment variables to explain changes in marital satisfaction over time in a randomized clinical trial comparing two types of behaviorally oriented couple therapy (Christensen et al., 2004). Methodological and data analytic techniques were implemented to overcome many of the limitations of previous studies. First, with 134 couples, this study is the largest randomized clinical trial of marital therapy. The larger sample increases statistical power, greatly improving the ability to detect true effects. Second, whereas most studies have examined change over time with only two time points, couples in the current study were assessed at four separate time points during acute treatment, providing a clearer portrait of how couples change over time. Thus, predictors of initial status can be distinguished from predictors of change.

We examined three sets of predictors corresponding to the classes of variables reviewed in the previous literature: (a) demographic variables, including age, education, income, years married, and whether the couple has children; (b) intrapersonal variables, including overall mental health and the personality characteristic of neuroticism; and (c) interpersonal variables, including communication, commitment, and intimacy. These three types of variables have rarely been examined in the same investigation, and one of the goals of the present study was to compare these sets of variables by estimating their unique effects on the trajectory of satisfaction through treatment. Because no previous studies have examined the trajectory of change in a treatment sample, there is little basis for making predictions of the differential effects of these three classes of variables on initial distress, linear change, or curvilinear change.

The clinical trial that provides the data for the present study compared two behavioral treatment conditions: traditional behavioral couple therapy (TBCT; Jacobson & Margolin, 1979) and integrative behavioral couple therapy (IBCT; Jacobson & Christensen, 1996). The goal of TBCT is to improve the exchange of positively reinforcing behaviors and to improve couples' skills in communication and problem solving. In contrast, the primary goal of IBCT is to foster emotional understanding and acceptance between partners, although strategies from TBCT are also implemented in IBCT. In the current prediction study, our initial focus is on treatment response, regardless of therapy, because (a) there was no significant difference between the two treatments in their final outcome at treatment termination (Christensen et al., 2004), (b) we were interested in generic predictors of response to behaviorally oriented couple therapy rather than treatment specific predictors, and (c) we wanted maximal statistical power for detecting changes over time. However, although the termination outcomes were not significantly different, the two treatments did follow significantly different trajectories of change in relationship satisfaction. Furthermore, gender and level of initial distress also generated different trajectories of change (Christensen et al., 2004). Therefore, we conducted exploratory analyses of differential predictors of change on the basis of therapy condition, gender, and severity of initial distress.

Method

Participants

One hundred thirty-four seriously and stably distressed married couples were recruited for a therapy program in Los Angeles (71 couples) and Seattle (63 couples). To determine eligibility for the study, we had couples participate in a three-step screening procedure. These three steps included a telephone screen (Step 1), a mail-home questionnaire screen (Step 2), and an in-lab assessment (Step 3). To be included in the study, couples had to be legally married and living together, request couple therapy, and meet criteria on the three consecutive screening assessments for serious and stable marital distress. Neither partner could be in another form of therapy during the course of the study, and if either partner were on psychotropic medication, that partner had to be stabilized on the medication prior to participation in the study. All spouses were given a diagnostic psychiatric interview; individuals who met criteria for current Axis I disorders of substance abuse or dependence, schizophrenia, bipolar disorder or current Axis II disorders of borderline, schizotypal, or antisocial personality disorder were excluded from the study. In addition, potential male batterers were excluded from participation. A flowchart of participant eligibility and final randomization to treatment is shown in Figure 1. Further details of our
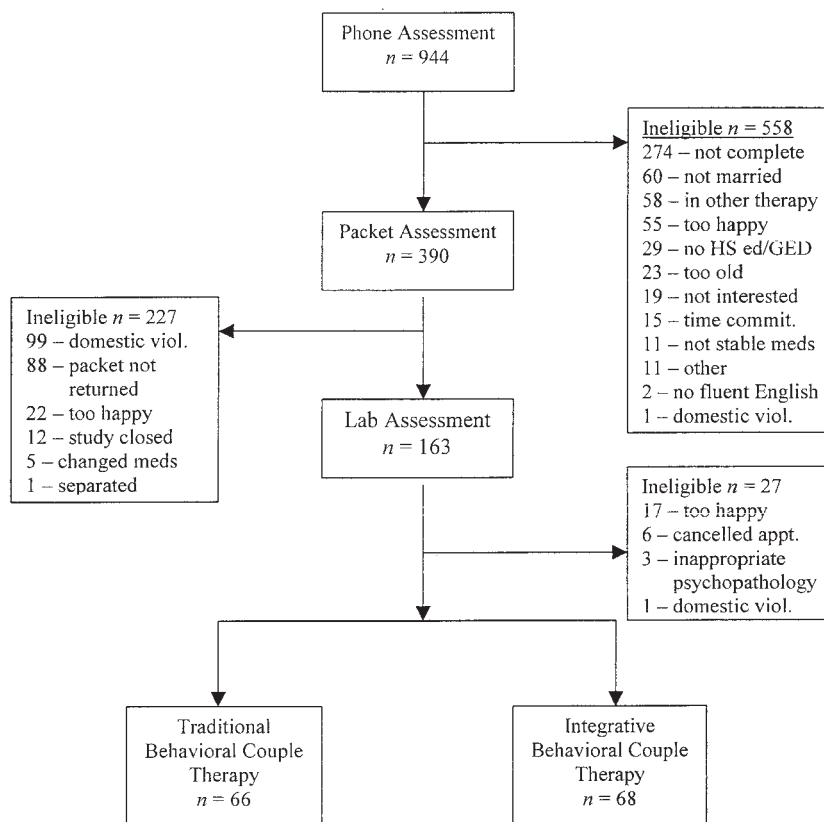
Phone Assessment
n = 944

Ineligible *n* = 558
274 – not complete
60 – not married
58 – in other therapy
55 – too happy
29 – no HS ed/GED
23 – too old
19 – not interested
15 – time commit.
11 – not stable meds
11 – other
2 – no fluent English
1 – domestic viol.

Packet Assessment
n = 390

Ineligible *n* = 227
99 – domestic viol.
88 – packet not
returned
22 – too happy
12 – study closed
5 – changed meds
1 – separated

Lab Assessment
n = 163

Ineligible *n* = 27
17 – too happy
6 – cancelled appt.
3 – inappropriate
psychopathology
1 – domestic viol.

Traditional
Behavioral Couple
Therapy
n = 66

Integrative
Behavioral Couple
Therapy
n = 68

Figure 1. Flowchart of participant eligibility. Commit. = commitment; viol. = violence.

screening measures and inclusion–exclusion criteria are presented in Christensen et al. (2004).

On average, couples had been married 10 years, had at least one child, were in their early 40s, and had a college education. The majority of the participants were Caucasian (husbands = 79.1%, wives = 76.1%). Also represented in the sample were African Americans (husbands = 6.7%, wives = 8.2%), Asian or Pacific Islanders (husbands = 6.0%, wives = 4.5%), Latino/Latinas (husbands = 5.2%, wives = 5.2%) and Native American or Alaskan Natives (husbands = 0.7%).

Design and Procedure

On the basis of their status on screening measures of relationship satisfaction, couples were categorized as either moderately or severely distressed. Within these two categories, couples were randomly assigned to receive up to 26 sessions of TBCT or IBCT at no cost. The therapist and the couple decided the total number of sessions on the basis of the couple's needs up to a maximum of 26 sessions to be completed within 1 year. Interventions in TBCT included behavioral exchange and communication and problem-solving training (Jacobson & Margolin, 1979). IBCT interventions included empathic joining, unified detachment, and tolerance building, as well as the strategies of TBCT (Jacobson & Christensen, 1996).

Couples were assessed at four time points: at the pretreatment assessment, at 13 weeks into treatment, at 26 weeks into treatment, and after the final treatment session. The first three assessments were conducted at the university laboratory. However, the final assessment occurred immediately after the couple's final session; couples completed the questionnaires independently and mailed them to the project immediately. As a result of these procedures, the first three time points are similar across couples, but the final time point varies. All study procedures were approved by the

University of California, Los Angeles, and University of Washington Institutional Review Boards. Informal consent was obtained during the initial telephone and packet assessments; written informed consent was obtained at the in-person pretreatment assessment.

Criterion Measure: Marital Satisfaction

We used the Dyadic Adjustment Scale (DAS; Spanier, 1976) as our criterion variable in the present study. In the parent study, the Global Distress Scale of the Marital Satisfaction Inventory—Revised (MSI–R; Snyder, 1997) was also used as a measure of marital satisfaction; however, previous research by Whisman and Jacobson (1992), as well as our own outcome research (Christensen et al., 2004), showed that the DAS is a more sensitive measure of change than the Global Distress Scale, so it was designated as our measure of satisfaction.

The DAS is a widely used, 32-item self-report measure of marital satisfaction. Items reflect amount of disagreement in the relationship and levels of affection, cohesion, and satisfaction. Scores can range from 0 to 151, with higher scores indicating greater marital satisfaction. The DAS has been shown to have good internal consistency (alphas typically in the low to mid .90s; Spanier, 1989), test–retest reliability, and discriminant validity (Sharpley & Cross, 1982; Spanier, 1976, 1989). In the current sample, internal consistencies of .89 and .87 were found for husbands and wives, respectively.

Predictor Measures: Demographics Questionnaire

Participants were asked to provide a variety of demographic information about themselves. The variables of interest in the current study were age, years of education, monthly pretax income, years married,

presence of children (yes/no), and wife employment outside the home (yes/no).

Predictors: Intrapersonal Variables

Personality: NEO Five-Factor Inventory (NEO-FFI; Costa & McCrae, 1989). The 60-item NEO-FFI is a well-validated short form of the NEO Personality Inventory (NEO-PI; Costa & McCrae, 1985). The NEO-PI is the most widely used measure of the five-factor model of personality and provides scores on five personality dimensions: Neuroticism, Extraversion, Openness to Experience, Agreeableness, and Conscientiousness. Correlations between the NEO-PI and the NEO-FFI factors range from .75 for Conscientiousness to .89 for Neuroticism (Costa & McCrae, 1989). In the current study, only the Neuroticism subscale was examined, and internal consistencies of .88 and .85 were found for men and women, respectively.

Mental health: Compass Outpatient Treatment Assessment System (COMPASS; Sperry, Brill, Howard, & Grissom, 1996). The COMPASS includes three self-report scales related to patient functioning: Subjective Well-Being, Current Symptoms, and Current Life Functioning. The Mental Health Index (MHI) is the overall summary measure of the COMPASS, taken as the mean of the well-being, symptom distress, and life-functioning scales. This index is reported as a T score and is scored so that a higher score represents greater mental health. A score of 60 or less has been shown to be representative of an outpatient population; 35.7% of the present sample scored in this range. In the present sample, the MHI had an internal consistency of .86 and .88 for men and women, respectively.

Structured Clinical Interviews for DSM–IV for Axis I and II (SCID; First, Spitzer, Gibbon, & Williams, 1994; Spitzer, Williams, Gibbon, & First, 1994). Graduate students were trained to conduct SCID interviews, and each participant was interviewed prior to the start of treatment. All SCID interviews were audio- or videotaped, and approximately 15% of these tapes were independently rated by an interviewer at the alternative site. Raters reached 88% overall agreement on the presence or absence of a current diagnosis ($\kappa = .66$, $p < .01$) and 90% overall agreement on the presence or absence of a past diagnosis ($\kappa = .75$, $p < .01$). Fifty-six percent of individuals met criteria for a lifetime psychiatric diagnosis, whereas 15.6% of individuals met criteria for a current psychiatric diagnosis. A dichotomous rating of presence versus absence of a current *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.; *DSM–IV*; American Psychiatric Association, 1994) diagnosis was examined as an indicator of mental health.[2]

Family History of Distress Scale (FAM) from the MSI–R (Snyder, 1997). This scale from the MSI–R assesses the disruption of relationships within the respondent's family of origin. This scale is based on the premise that either unresolved conflicts evolving from the family of origin or an absence of adequate parental models may contribute significantly to distress in the current relationship. Higher scores reflect greater distress in the family of origin. As FAM scores increase, respondents are more likely to describe an unhappy childhood, disruption in the parents' marriage, and lack of affection among family members. The scale demonstrated an internal consistency for men and women of .76 and .79, respectively.

Predictors: Interpersonal Measures

Communication Patterns Questionnaire (CPQ; Christensen & Sullaway, 1984). The CPQ is a 35-item questionnaire that assesses the occurrence of problematic interaction and communication patterns in close relationships, rating the likelihood of specific interaction patterns occurring before, during, and after an argument. Heavey, Christensen, and Zumbotel (1996) constructed a 7-item Constructive Communication subscale (CPQ-CC) from the CPQ that contains three constructive items (mutual discussion, mutual expression, and mutual negotiation) and four reverse-scored de-

structive items (mutual blame, mutual threat, husband verbal aggression, and wife verbal aggression). Heavey et al. demonstrated high levels of internal consistency in this subscale ($\alpha = .81$–.84) as well as high levels of consistency with objective ratings of problem-solving behavior from videotaped conflict interactions. Higher scores on the CPQ-CC indicate higher levels of positive communication. Internal consistencies in the present sample for men and women were .69 and .67.

Intimacy: Closeness and Independence Inventory (CII; Heavey & Christensen, 1991). This measure assesses spouse's desired level of closeness in the relationship. Spouses rate themselves and their partners using six items, each on 9-point scales that reflect greater levels of closeness versus independence in the relationship. Sample items include, "Would you like to have more sharing of feelings with your partner or more respect for privacy in your relationship?" and "Would you like to spend more or less time talking with your partner about his or her thoughts and feelings?" Higher scores reflect a higher desire for closeness in the relationship. Internal consistencies for men and women were found to be .76 and .75, respectively.

Affective Communication Scale (AFC) from the MSI–R (Snyder, 1997). The items of the AFC scale evaluate the respondent's dissatisfaction with the amount of affection and understanding expressed by his or her partner. Higher scores reflect greater dissatisfaction with the amount of affection and understanding expressed by the respondent's partner. As scores on the AFC scale increase, respondents are less likely to describe their partners as loving, tender, trusting, or confiding and more likely to describe their partners as unsympathetic, insensitive, overly critical, or withdrawn. The AFC subscale demonstrated strong internal consistency with alphas of .76 and .77 for men and women, respectively.

Sexual Dissatisfaction Scale (SEX) from the MSI-R (Snyder, 1997). Items on the SEX scale reflect the level of discontent with the frequency and quality of intercourse and other sexual activity. The scale includes an assessment of general dissatisfaction with the sexual relationship, partner's lack of interest, and inadequate affection during sexual exchanges. Higher scores indicate greater dissatisfaction with the sexual activity in the relationship. The SEX subscale demonstrated strong internal consistency with an alpha of .84 for both men and women.

Commitment/relationship stability: Marital Status Inventory (MSI; Weiss & Cerreto, 1980). The MSI consists of 14 true–false items that measure steps toward separation or divorce, ranging from thoughts (e.g., thinking of separation or divorce after an argument), to tentative steps (e.g., talking to a friend), to actual separation or divorce actions (e.g., moving out). Scores range from 0 to 14 depending on the number of steps the respondent has taken. Research has shown that the MSI is an internally consistent measure (split-half reliability of .86; Crane & Mead, 1980) and can identify couples at risk of divorce (Crane, Newfield, & Armstrong, 1984). In the present sample, the MSI demonstrated strong internal reliability of .80 for both men and women.

Data Analysis

The data in the present study were analyzed with hierarchical linear modeling (HLM; Raudenbush & Bryk, 2002), also referred to as multilevel modeling (Snijders & Bosker, 1999) and mixed-effects modeling (Pinheiro & Bates, 2000). HLM has a number of advantages over traditional statistical methods that make it ideally suited for longitudinal data on couples (see Atkins, 2005, for a discussion of HLM for couples' data). HLM appropriately models the correlations that are due

---

[2] Both current diagnosis and lifetime diagnosis (either current or past) were examined as possible predictors. Neither form of diagnosis demonstrated strong associations with outcome (as described in Results), and they were highly collinear with each other because lifetime diagnosis subsumes current diagnosis. A current diagnosis should affect therapy more than a past one, and thus we chose current diagnosis for our measure of psychopathology in the analyses reported here.

to repeated measures and spouses, providing unbiased estimates of regression parameters and standard errors. Moreover, it does not require complete data on every participant and, thus, flexibly handles missing data (though only if the missing data can be assumed to be missing at random, see Atkins, 2005). Data from all 134 participants were used.

We used a three-level model, including a quadratic component for time, to capture any nonlinearity in couples' change during treatment. A basic model without any predictors is shown in Equation 1:

$$\text{Level 1 (repeated measures): } Y_{tij} = \pi_{0ij} + \pi_{1ij}(Time)_{tij}$$

$$+ \pi_{2ij}(Time^2)_{tij} + e_{tij}$$

$$\text{Level 2 (individuals): } \pi_{0ij} = \beta_{00j} + r_{0ij}$$

$$\pi_{1ij} = \beta_{10j}$$

$$\pi_{2ij} = \beta_{20j} \qquad (1)$$

$$\text{Level 3 (couples): } \beta_{00j} = \gamma_{000} + u_{00j}$$

$$\beta_{10j} = \gamma_{100} + u_{10j}$$

$$\beta_{20j} = \gamma_{200} + u_{20j},$$

where t indexes time within individuals, i indexes individuals within couples, and j indexes couples. A single random intercept at the individual level (r_{0ij}) allows individuals within a couple to have different estimates of initial distress. At Level 3, each change component (i.e., intercept, slope, and quadratic) has a corresponding variance component (u_{00j}, u_{10j}, and u_{20j}) that allows the different couples within the study to have distinct intercepts, slopes, and quadratics. Predictors are added to this basic model either at Level 2 if the predictor is a characteristic of the individual (i.e., gender, age, education, income, MHI, neuroticism, FAM, current *DSM–IV* diagnosis, CPQ, CII, MSI, AFC, or SEX) or at Level 3 if the predictor is a characteristic of the couple (i.e., years married, presence of children) to explain the variability in each of the change components.

Because the previous literature does not lead to precise hypotheses regarding prediction, we used a graduated approach to study covariates of outcome with the present data. First, we considered the three classes of predictors (i.e., demographic, intrapersonal, interpersonal) as blocks and examined their influences on intercept, slope, and quadratic; this analysis demonstrates whether one or more classes of predictors explain more variance in a given change component relative to the other classes. Second, we used an automatic variable selection procedure based on the Bayesian Information Criterion (BIC; Raftery, 1995) to reduce the number of predictors, starting with the model containing all the predictors. Typical stepwise methods of variable selection, in which candidate predictors are entered and removed on the basis of the p values of the regression coefficients, have numerous problems (e.g., ill-fitting models are chosen, nominal p values are often quite different from true p values; see Harrell, 2001). Moreover, stepwise methods maximize the within-sample prediction (i.e., R^2), which can lead to poor generalizability. BIC addresses the same task (i.e., choosing the best predictors); however, in doing so, it balances the complexity of the model with the sample size[3] and has been shown to have superior out-of-sample prediction relative to stepwise methods; thus, BIC selected models should have greater generalizability to similarly selected samples compared with stepwise methods. The BIC procedure was applied to individual predictors and not classes of predictors.[4]

We explored possible interactions between all predictors (not just those selected by BIC) and treatment condition, gender, and initial severity—all important moderators of outcome in the primary treatment study (Christensen et al., 2004). In conducting the exploratory analyses, we followed the procedures for exploratory analyses outlined in Raudenbush and Bryk (2002). We created a number of scatter plots for each of our predictors against the random effects (i.e., intercepts, slopes, or quadratics) from the

full model. Each of these scatter plots was conditioned by therapy, gender, and initial severity (i.e., separate plots for each of the eight combinations of therapy, gender, and severity). These plots suggested predictors that might interact with one or more of the potential moderators, and these suggested interactions were then tested as additional predictors in the HLM analysis with the BIC predictors. All analyses were conducted with R version 1.9.0 (R Development Core Team, 2004) and made extensive use of the nlme library of functions for HLM (Pinheiro & Bates, 2000).

Results

Basic HLM Analysis

Our initial step in the modeling process was to fit the model presented in Equation 1, a basic three-level model that describes each individual's trajectory of change with intercept, slope, and quadratic components. The basic model estimated an intercept value for all couples of 84.56 DAS points, $SE = 1.03$, $t(710) = 82.2$, $p < .01$. The linear change at the start of treatment[5] is estimated to be 0.36 DAS points per week of therapy, $SE = 0.08$, $t(710) = 4.49$, $p < .01$. Finally, the quadratic (i.e., acceleration or deceleration of change) was nonsignificant, B = −0.002, $SE = 0.002$, $t(710) = −1.02$, *ns*. Thus, the basic HLM analysis demonstrated that couples began quite distressed (~85 DAS points) and improved in a linear fashion at the rate of about 0.36 points per week. Combining this with an average of 36.7 weeks in therapy led to an average change during therapy of 13.2 DAS points. Moreover, analyses of the random effects demonstrated that there was significant variability in each component of the change trajectory: intercept, slope, and quadratic. Because there was significant variability in the quadratic component and because nonsignificant main effects can reveal significant interactions, we retained the quadratic fixed effect in the later analyses.

Blocks of Predictors: Demographic, Intrapersonal, and Interpersonal

The next step in our analysis estimated the joint contribution of each class of predictors (i.e., demographic, intrapersonal, and interpersonal) in explaining variance in each change component (i.e., intercept, slope, and quadratic). Table 1 presents descriptive statistics and a correlation matrix for the predictors. Prior to the HLM analysis, all predictors were centered. Table 2 presents the

---

[3] BIC is calculated as $-2 \times$ (log likelihood of the model) $+ \log(N) \times$ (number of predictors). The first term in the equation is the deviance of the model, which summarizes how well the model fits the data. Thus, each additional predictor is penalized by the log of the sample size. The BIC selection procedure uses a "leave one out" approach; that is, at each step of its model selection, it calculates what the BIC statistic would be if each predictor in the model was excluded and excludes the term that yields the largest drop in BIC. This is similar to calculating the model R^2, excluding each variable one at a time at each iteration of the search. In addition, it will not drop lower order terms (main effects or simple interactions) prior to dropping higher-order terms.

[4] We used the number of couples ($N = 134$) as the sample size in the computation of BIC, though the correct penalty factor for mixed-effects models, which is related to the sample size, is an ongoing area of statistical research (Vaida & Blanchard, 2005).

[5] Because of the presence of the quadratic term, the slope is an estimate of the instantaneous slope at time = 0, which is the start of therapy (see Cohen et al., 2003, for a discussion of interpreting polynomials).

Table 1
Means and Standard Deviations and Correlation Matrix of Predictors

Variable	M	SD	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. Gender	0.50	0.50	—	−.00	.09	.00	.25	.00	−.11	.09	−.03	−.01	−.02	−.32	−.09	.16	−.06
2. Years married	9.83	7.52		—	.57	−.04	−.01	.15	.04	−.09	−.09	−.01	−.04	.02	−.09	.14	.14
3. Age (years)	42.31	8.59			—	−.03	−.02	−.19	−.00	−.10	−.02	.01	−.02	−.01	−.12	.15	.13
4. Education	17.06	3.11				—	.07	−.05	.15	−.03	−.09	−.13	.06	−.07	−.03	−.04	.01
5. Income	6.80	3.09					—	−.02	−.06	.12	−.04	−.00	−.02	−.07	−.00	.04	−.10
6. Children	0.65	0.48						—	.05	−.07	−.02	−.07	−.09	.02	.06	.15	.02
7. Mental health	61.46	7.83							—	−.55	−.13	−.35	.12	.08	−.18	−.04	−.06
8. Neuroticism	51.61	11.02								—	.24	.27	−.09	−.00	.06	.00	−.05
9. Family distress	55.77	9.10									—	.17	−.02	−.12	−.01	−.05	.01
10. *DSM* diagnosis	0.18	0.38										—	−.00	−.02	.09	.04	.00
11. Good comm.	−2.90	8.95											—	−.05	−.26	−.37	−.05
12. Close	30.97	6.95												—	−.15	.04	.06
13. Steps to divorce	4.07	2.83													—	.28	.19
14. Affect. comm.	63.35	6.85														—	.31
15. Sex. dis.	60.07	9.95															—

Note. For descriptive statistics, gender was coded 0 = wives, 1 = husbands, and children was coded 0 = no children, 1 = children. The income variable is the natural logarithm of the individual's monthly income. Mental health = Mental Health Index of the Compass Outpatient Treatment Assessment System; Family distress = History of Family Distress scale of the Marital Satisfaction Inventory—Revised (MSI–R); *DSM* diagnosis = current *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.) diagnosis based on the Structured Clinical Interview for *DSM–IV* Axis 1 Disorders; Good comm. = Constructive Communication subscale from the Communication Patterns Questionnaire; Close = Closeness and Independence Inventory; Steps to divorce = Marital Status Inventory; Affect. Comm. = Affective Communication scale from the MSI–R; Sex. dis. = Sexual Dissatisfaction scale from the MSI–R.

tests of each block of predictors for each change component along with the amount of variance explained.[6] Considering the three change components, we see that the most variance by far is explained in the intercept (approximately 45% of the initial intercept variability), whereas comparatively little variance is explained in slopes or quadratics. With respect to the blocks of predictors, interpersonal variables explain the most variance in the intercepts (i.e., initial level of marital satisfaction) but little to no variability in the change components of the trajectories. Demographic and intrapersonal predictors explain more modest amounts of variance in total but explain some of the variability in the change components (i.e., slope and quadratic).

Model Selection With BIC

The full model, including all demographic, intrapersonal, and interpersonal variables was then reduced with the BIC variable selection procedure described earlier.[7] The results from the BIC-selected model are presented in Table 3 (within the column "BIC model"). Only 15 predictors (excluding the intercept) were selected by BIC from the full model containing 45 predictors. Moreover, there are comparatively few predictors of the slope and quadratic, which mirrors the findings based on the blocks of predictors.

Looking at predictors of the intercept, we find that higher levels of neuroticism and better mental health as measured by the MHI are associated with less marital distress initially. The unexpected neuroticism finding appears to result from the sizable negative correlation between MHI and neuroticism (r = −.55). When MHI is not included in the model, neuroticism is no longer significant, B = 0.06, *t*(123) = 1.08, *p* = .28; thus, MHI appears to act as a suppressor (see Cohen, Cohen, West, & Aiken, 2003) with respect to neuroticism, meaning that the significant neuroticism finding is due to the variance in neuroticism not shared with the MHI. Given that the MHI measures anxiety in part, it is not straightforward to

assess what this variance represents, and the significant neuroticism finding should be viewed skeptically. The results show that greater desired closeness as measured by the CII ("Close" in Tables 1 and 3) and better communication as measured by the CPQ-CC ("Good comm." in Tables 1 and 3) are associated with less initial marital distress, whereas poorer affective communication ("Affect. comm." in Table 3) and more steps taken toward separation and divorce ("Steps to divorce" in Table 3) are associated with greater initial distress.

In looking at the slope and quadratic,[8] we find that gender is a significant moderator of both slope and quadratic, similar to what was found in the primary outcome analyses (Christensen et al., 2004). Men improve more rapidly in therapy, but this change decelerates over time. Years married is also a significant moderator of the slope, such that couples who have been married longer

---

[6] Readers may be surprised to note that in several instances there are negative values for R^2, which would be impossible in standard regression. Because of the multiple levels of data and multiple error terms in HLM, it is possible to have variance increase following the addition of predictors, which is indicative of poor prediction. For a discussion of explained variance in HLM analyses, see Snijders and Bosker (1999).

[7] An anonymous reviewer suggested that the other four subscales from the NEO might be considered as predictors (i.e., Openness, Agreeableness, Conscientiousness, Extroversion). We ran an additional model including these four variables as predictors of each change component. None of these additional terms were significant in the full model, and none were selected by the BIC procedure. Details of this analysis are available from David C. Atkins.

[8] An anonymous reviewer noted that the quadratic term could be driven by the timing of the final session, which varied across couples. To assess this, we correlated the timing of the final session for each couple with the Empirical Bayes estimate of the quadratic for each couple, which was nonsignificant (r = .02, *p* > .5). Thus, it does not appear that the timing of the final session influences the quadratic term.

Table 2
Tests for Blocks of Predictors

Block	Intercept	Slope	Quadratic
Demographic	$F(6, 126) = 0.56$ $p = .73$ $\Delta R^2 = .01$	$F(6, 698) = 2.09$ $p = .05$ $\Delta R^2 = .05$	$F(6, 698) = 0.88$ $p = .51$ $\Delta R^2 = -.12$
Intrapersonal	$F(4, 122) = 4.02$ $p < .01$ $\Delta R^2 = .05$	$F(4, 692) = 2.36$ $p = .05$ $\Delta R^2 = .04$	$F(4, 692) = 2.61$ $p < .05$ $\Delta R^2 = .13$
Interpersonal	$F(5, 117) = 20.11$ $p < .0001$ $\Delta R^2 = .40$	$F(5, 687) = 0.85$ $p = .51$ $\Delta R^2 = -.05$	$F(5, 687) = 1.16$ $p = .32$ $\Delta R^2 = -.06$
Total R^2	0.45	.04	-.03

improve at a relatively greater rate than those who have been married shorter periods of time. Residual analyses with years married revealed a nonlinear and discontinuous effect. We fit separate slopes for couples married less than 18 years and for those married longer than 18 years, which demonstrated that the positive effect for years married was driven by those married longer than 18 years, B = 0.019, $t(704) = 2.84$, $p < .01$. Finally, there is a weak effect ($p < .06$) that individuals desiring greater closeness decelerate in their change over time.

Exploratory Interactions With Gender, Treatment Condition, and Initial Severity

The exploratory graphs comparing each of the original predictors with treatment condition, gender, and initial severity (moderate vs. severe) suggested several potential interactions between these factors and other predictors. To test these interactions, treatment condition and severity were entered into the model selected with BIC, predicting intercept, slope, and quadratic. The potential interactions revealed from the graphs were then tested in this model. The final results are displayed in Table 3 (within the column "With exploratory interactions").

The results for treatment condition and severity demonstrated similar effects to those previously reported in the primary outcome article (Christensen et al., 2004). Relative to moderately distressed couples, couples who were severely distressed showed a greater deceleration over time in their amount of change. Couples who received TBCT improved more quickly in therapy initially, but their change also slowed down during treatment.[9]

The only predictor that demonstrated significant interactions in the HLM analysis with gender, treatment condition, or severity was sexual dissatisfaction, including an interaction between gender, sexual dissatisfaction, and severity predicting the intercept and significant interactions between treatment condition and sexual dissatisfaction predicting the slope and quadratic. At the intercept, men with high initial severity showed a strong relationship between sexual dissatisfaction and their initial marital satisfaction, whereas women who were either moderately or severely distressed and men with moderate initial severity did not show such a strong relationship.

In examining change over time, those individuals reporting mild to moderate sexual dissatisfaction (i.e., SEX < 60) show similar improvements across the two treatments. However, there is quite a notable difference between the therapy conditions for those couples who are very sexually unhappy in their marriages (i.e., SEX > 70). TBCT couples who are very sexually unhappy improve rapidly early in therapy but slow down and begin to lose some of their gains in marital satisfaction toward the end of therapy; IBCT couples who are very sexually unhappy improve more slowly initially but continue to improve over the duration of therapy, with some evidence of increasing improvement. On average, couples completed treatment at 36.7 weeks ($SD = 7.26$), and approximately one third of couples ($n = 24$ for IBCT, and $n = 20$ for TBCT) finished treatment between 42 and 52 weeks. Thus, although this is an exploratory finding, the effect does not appear to be driven by a few outlying couples.

Discussion

The search for predictors of successful marital therapy has been plagued by methodological problems and inconsistent findings, and we believe that the present study is a first step toward a new generation of research on this topic. Overall, perhaps what is most striking is how few predictors there are of the outcome of marital therapy. In this section we interpret the findings in light of the past literature, focusing on both methodological strengths and challenges as well as the application to clinical practice.

Blocks, BIC, and Complex Interactions: What Does It All Mean?

Our tests of classes of predictors revealed that demographic predictors (as a whole) add little to understanding the outcome of marital therapy. They have been studied often, perhaps because of ease of measurement rather than sound theory. In future investigations, we encourage researchers interested in demographic predictors to consider theoretical mechanisms through which these variables might affect marital therapy outcome. We take up this task ourselves with the number of years married below.

Intrapersonal variables are notable for consistently explaining a

---

[9] The minor differences in treatment comparisons reported here versus the primary outcome article (Christensen et al., 2004) are due to the additional predictors and interactions present in the current analysis.

Table 3
Hierarchical Linear Modeling Results for Models With and Without Interactions

Variable	BIC Model			With exploratory interactions		
	B	95% CI	ES	B	95% CI	ES
			Intercept			
Intercept	84.75****	83.13, 86.37	.98	84.87****	83.61, 86.13	.98
Gender	1.88	−0.90, 4.67	.12	1.72	−1.09, 4.53	.11
Years married	−0.12	−0.34, 0.09	.10	−0.05	−0.22, 0.12	.06
Mental health	0.25*	0.05, 0.46	.21	0.25**	0.07, 0.43	.24
Neuroticism	0.16*	0.02, 0.30	.20	0.14*	0.02, 0.27	.20
Close	0.41***	0.19, 0.63	.32	0.37***	0.17, 0.57	.32
Good comm.	0.36****	0.20, 0.52	.38	0.31****	0.16, 0.45	.36
Steps to divorce	−1.20****	−1.70, −0.68	.39	−0.38	−0.85, 0.10	.14
Affect. comm.	−0.46****	−0.67, −0.25	.37	−0.34**	−0.54, −0.13	.29
Severity				−12.09****	−14.94, −9.23	.59
Treatment				−0.28	−2.86, 2.30	.02
Sex. dis.				−0.19**	−0.31, −0.06	.26
Gender × Severity				4.27*	0.07, 8.46	.18
Gender × Severity × Sex. Dis.				−0.45*	−0.90, −0.01	.18
			Slope			
Time	0.36****	0.21, 0.52	.17	0.37****	0.21, 0.52	.17
Gender × Time	0.28*	0.059, 0.494	.09	0.27*	0.054, 0.486	.09
Years Married × Time	0.01**	0.005, 0.023	.11	0.02**	0.006, 0.024	.12
Close × Time	0.01	−0.010, 0.026	.03	0.01	−0.007, 0.029	.04
Severity × Time				0.28	−0.04, 0.59	.06
Treatment × Time				0.32*	0.01, 0.64	.07
Treatment × Sex. Dis. × Time				0.03*	0.01, 0.05	.09
			Quadratic			
$Time^2$	−0.002	−0.006, 0.002	.04	−0.002	−0.006, 0.002	.04
Gender × $Time^2$	−0.007*	−0.012, −0.001	.09	−0.007*	−0.012, −0.001	.09
Close × $Time^2$	−0.0004†	−8.7e-4, 1.5e-5	.07	−0.0005*	−9.3e-4, −3.8e-5	.08
Severity × $Time^2$				−0.009*	−0.017, −0.002	.09
Treatment × $Time^2$				−0.007†	−0.014, 0.001	.07
Treatment × Sex. Dis. × $Time^2$				−0.0006*	−1.2e-3, −3.7e-5	.08

Note. Several extremely small values are shown in scientific notation: The value after the "e" represents the number of positions to move the decimal place. B = fixed-effect regression coefficient; 95% CI = 95% confidence interval; ES = effect size *r*, calculated as $\sqrt{t^2/(t^2+df)}$; Mental health = Mental Health Index of the Compass Outpatient Treatment Assessment System; Close = Closeness and Independence Inventory; Good. comm. = Constructive Communication subscale from the Communication Patterns Questionnaire; Steps to divorce = Marital Status Inventory; Affect. Comm. = Affective Communication scale from the Marital Satisfaction Inventory—Revised (MSI–R); Sex. Dis. = Sexual Dissatisfaction scale from the MSI–R.
† $p < .10$.　* $p < .05$.　** $p < .01$.　*** $p < .001$.　**** $p < .0001$.

significant percentage of variance in each of the change components; on the basis of Cohen's (1992) guidelines for R^2, intrapersonal variables explain a small to medium amount of variance across the change components. Intrapersonal variables have largely been overlooked in the previous prediction literature, yet the broader clinical and research literature has recently emphasized the importance of individual qualities and disorders in marital therapy (Snyder & Whisman, 2004). Thus, future investigations should attend to intrapersonal variables as predictors of outcome.

Comparing our three different classes of predictors, interpersonal variables were the "big winners," yet their strength was confined to explaining variability in couples' initial distress. However, it is precisely this fact that raises a troubling issue: To what extent are interpersonal predictors proxies for marital satisfaction, our outcome? As Paul Meehl (1990) eloquently and bluntly put it, "In the social sciences . . . everything correlates to some extent with everything else" (p. 204), and this would appear to be especially true for interpersonal predictors of marital satisfaction. Marital satisfaction is a latent construct that taps into the manifold

nature of a good relationship, and couples are likely reporting on interpersonal variables as they assess their relationship satisfaction through the DAS. Given this, we seem to be in the position of indicating that qualities of the relationship (e.g., good communication, good affective communication, and few steps taken toward separation or divorce) are successful at explaining the overall satisfaction in the relationship when both are measured prior to treatment.

From this overview position of considering classes of predictors, the variable selection procedure brings greater specificity as well as confirmation. Of the predictors of initial distress, the MHI from the COMPASS stands out.[10] Individual functioning is reliably

---

[10] Although the intrapersonal block of predictors fared well, few of the individual, intrapersonal predictors were chosen by BIC in the variable selection procedure, discussed below and presented in Table 3. It may be that this group of predictors each have a small additive effect to the prediction of marital satisfaction; considered together, they make a notable contribution, but when considered individually, this effect is not apparent.

associated with the state of the relationship. As we noted earlier, converging lines of research and clinical experience confirm that a relationship takes two individuals, and the qualities and functioning of those individuals are important to the quality and functioning of the relationship. We find it interesting that having a current psychiatric illness (as measured by the SCID) was not predictive of initial distress or change in relationship satisfaction. This fact might provide a clue to the predictive elements of the MHI, which combines subscales of well-being, functioning, and psychopathology. It may be that psychopathology per se is not as important as impaired functioning (e.g., work, social, family) and overall well-being. However, the current study was limited in its ability to examine the impact of psychopathology. Even though a majority of participants in the current sample qualified for a past diagnosis, only 16% qualified for a current diagnosis. With such a small number of currently diagnosable participants, it was impossible to look at the impact of specific disorders. A sample with a larger percentage of diagnoses would allow research to examine the impact of psychopathology in general and of particular disorders on the outcome of marital therapy.

There are far fewer predictors of the change components (i.e., linear and quadratic) of marital satisfaction as opposed to initial distress (i.e., intercept). As we found in the primary outcome analyses, gender moderates change; men improve more rapidly than women initially, but this change slows down over time, with both sexes changing an equal total amount over the course of therapy. Women were significantly more active than their spouses in seeking treatment (Doss, Atkins, & Christensen, 2003), and some men in the study were quite open that they were coming to marital therapy under duress. Thus, one possibility for the gender difference is that men's early boost in marital satisfaction is a reflection that therapy was not as bad as they thought it might be; this interpretation was supported anecdotally from our research therapists. Women may not feel such stigma about therapy and thus might not be expected to show this same pattern. Although this interpretation has some post hoc support from our therapists, it clearly should be explored in future research.

The number of years a couple was married was also a significant predictor of the linear change made in treatment. Longer relationships are more likely to have survived previous periods of stress and discord, and it seems likely that the length of the relationship may represent a measure of commitment, albeit, an indirect measure. Partners in a highly distressed couple must decide between working on the relationship (by themselves or with help) or leaving the relationship. It may be at times such as these that longer relationships have an advantage. Younger couples may find it easier to consider separating and divorcing (e.g., less invested in relationship; prospects for other relationships in the future are greater). As noted earlier, the strongest improvement in therapy was for couples married 18+ years, with little benefit from therapy for those married less than 10 years. Future research might focus on these veterans of marriage to both confirm the finding and explore why this might be the case.

Desired closeness significantly predicted the intercept and, to a much lesser degree, slope and quadratic. Initially, couples with greater desired closeness are less distressed, which may serve as a powerful motivator in therapy. Yet, the findings over time show that couples with greater desired closeness initially improve in therapy and then decline notably in the second half of therapy; couples with less desired closeness initially change quite robustly

in therapy, in a largely linear fashion, ending therapy less distressed than couples who desired greater closeness. Although care must be taken in making too much of an isolated finding, it is tempting to speculate that attachment processes may explain these results. Perhaps couples with greater desires for closeness were more preoccupied with attachment issues. They were more enthusiastic about therapy to begin with but were ultimately disappointed in its impact. Conversely, those couples with lower desires for closeness may have been more securely attached and ultimately performed better in therapy (Davila, 2003).

Finally, our exploratory analyses revealed an intriguing interaction between initial sexual dissatisfaction, change over time, and treatment condition. Couples who are initially very sexually unhappy in their marriages show more rapid improvement with TBCT early in treatment; however, this process slows down and even reverses later in therapy. IBCT couples with similar levels of sexual dissatisfaction show slower but steady improvement over the entire period of therapy. Given our earlier discussion of proxy variables of marital satisfaction, we might wonder whether sexual dissatisfaction simply identifies relationship dissatisfaction in general. Because the effects of sexual dissatisfaction remain significant in the model that contains the initial severity predictor (and its interaction with slope and quadratic), this hypothesis seems unlikely; it appears that there is something specific to sexually dissatisfied couples that interacts with treatment condition.

This finding was not hypothesized and should be regarded as exploratory until it is confirmed in other research. Nonetheless, we offer a few additional details of sexual problems and further anecdotes that may help to shed light on this finding. Very few couples actually complained of sexual dysfunction, and the majority of couples who indicated that sex was a problem area in their relationship complained about lack of frequency and desire, typically because of the poor state of their relationship in general. Therapists in both treatment conditions were free to conduct sex therapy if needed; however, therapists usually addressed general sexual dissatisfaction through the interventions of the given treatment condition.

Couples in TBCT may have gotten an early boost in satisfaction from the initial emphasis on behavior exchange strategies, which encourages mutual positive behavior. Although behavior exchange is focused on small, noncontroversial positive behavior and would not normally encourage sexual activity in a couple plagued with sexual difficulties, the positive actions may have engendered early hope in these sexually dissatisfied couples. However, to alter sexual functioning in a troubled couple may require increased openness, vulnerability, intimacy, and trust between partners. These features of the relationship are not easily influenced by the direct change strategies of TBCT, and sexually dissatisfied couples in TBCT eventually may have become discouraged. However, IBCT was developed with an emphasis on openness, vulnerability, intimacy, and trust, deemphasizing direct change strategies. IBCT puts its faith in spontaneous change that results from greater intimacy between partners, which may also lead to an improved sexual relationship.

Predictors of change in treatment are much more important for the clinician than are predictors of initial status. Predictors of change can inform prognosis and can dictate treatment decisions, such as whether treatment should even be attempted. Predictors of initial status have no such usefulness. Yet, unfortunately, as noted above, we found far more predictors of initial status than of

change. However, as one anonymous reviewer pointed out, this may have resulted from the success of our treatment. In the extreme case, if treatment is perfectly successful in bringing all to normal functioning, then no pretreatment variables can predict change in treatment. A great majority of couples were helped by treatment (Christensen et al., 2004). This help came about at least in part by improving interpersonal variables, the very ones that proved to be potent predictors of initial status but poor predictors of change. Thus, the effectiveness of treatment in changing these variables may also have made them impotent predictors of change.

Strengths and Limitations

A number of strengths of our study bolster our confidence in the results. First, the current study examined the largest sample to date of couples in a marital therapy outcome trial, which leads to greater power to detect true effects and reduced risk of overfitting the data and capitalizing on chance variability. Furthermore, the current study selected seriously and chronically distressed couples who may be representative of the types of couples presenting for marital therapy. In fact, more than half had a current or past comorbid diagnoses such as depressive and anxiety disorders. In addition, the statistical approach provided a number of advantages over more commonly used statistical methods. Whereas most marital therapy studies collect only two waves of data, the use of HLM allowed examination of data collected at multiple times throughout therapy. Rather than splitting our sample in half and analyzing husbands and wives separately, spousal data were combined in a single model that controlled for the interdependence of their data.

Despite these strengths, several factors nevertheless may limit the interpretation of the present findings. First, although attempts were made to obtain a diverse sample, couples in the current study were disproportionately White and college educated. Additionally, eligibility in the current study required that the couple be heterosexual, married, and living together and excluded couples who were batterers or whose partners had one of several *DSM–IV* criteria (e.g., substance abuse or dependence, antisocial personality). These selection factors may have contributed to the fact that few demographic factors were significant predictors. Second, aspects of the current analyses suffer from the common method variance problem (Gottman, 1994) in that self-report measures of constructs such as personality and communication were used to explain self-report measures of relationship satisfaction. Other types of data, such as observational coding, could address these issues, but they are not commonly available to clinicians. In addition, although common method variance may have the potential to bias initial values of predictor variables, change over time in marital satisfaction was reliably calculated, and this change cannot easily be attributed to biases due to common method variance. Third, most of the current measures focused on assessing negative qualities of couples, such as conflict, distress, or dissatisfaction. Improvement in prediction may come from using more variables that are reflective of positive qualities of the couple. For example, social support behaviors have been linked to marital quality and changes in marital quality, even after controlling for behaviors observed in standard problem-solving discussions (Pasch & Bradbury, 1998). Fourth, the present study focused solely on acute response to treatment; future research will focus on predictors of posttreatment adjustment, which could be similar to or different

from the findings reported here. Finally, the present study has relatively low power to detect predictors of the quadratic effect of time because of the limited number of repeated measures. Relatively few predictors of the quadratic were found, which could be due to low power.

Conclusions

Because many couples fail to show improvement in marital therapy, improving our ability to predict who will and will not benefit from treatment is critical clinically. The current study advances the understanding of which characteristics influence the amount of change in marital therapy in several important ways. However, the most important finding may be that little predicts successful or unsuccessful outcomes; this may lead us to reconsider the process by which specific variables may moderate treatment outcome. In particular, interpersonal variables that are likely the target of interventions may be poor predictors for that precise reason. We continue to be interested in how couples fare after treatment, whether they maintain their gains, continue to improve, or relapse. We are following the current set of couples for 5 years after treatment termination. Later articles will address the long-term outcome for these couples and the predictors of that outcome so that as a field we may begin to answer the question of "who succeeds in couple therapy?"

References

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington DC: Author.

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

Baucom, D. H., & Aiken, P. A. (1984). Sex role identity, marital satisfaction, and response to behavioral marital therapy. *Journal of Consulting and Clinical Psychology, 52,* 438–444.

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.

Beach, S. R. H., & Broderick, J. E. (1983). Commitment: A variable in women's response to marital therapy. *The American Journal of Family Therapy, 11,* 16–24.

Bennun, I. (1985). Prediction and responsiveness in behavioural marital therapy. *Behavioural Psychotherapy, 13,* 186–201.

Cain, J. A. (1997). *The investigation of predictor variables for successful behaviorally based marital therapy.* Unpublished doctoral dissertation.

Christensen, A., Atkins, D. C., Berns, S. B., Wheeler, J., Baucom, D. H., & Simpson, L. (2004). Integrative versus traditional behavioral couple therapy for moderately and severely distressed couples. *Journal of Consulting and Clinical Psychology, 72,* 176–191.

Christensen, A., & Heavey, C. L. (1999). Interventions for couples. *Annual Review of Psychology, 50,* 165–190.

Christensen, A., & Sullaway, M. (1984). *Communication Patterns Questionnaire.* Unpublished manuscript, University of California, Los Angeles.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112,* 155–159.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences.* Mahwah, NJ: Erlbaum.

Costa, P. T., & McCrae, R. R. (1985). *The NEO Personality Inventory: Manual.* Lutz, FL: Psychological Assessment Resources.

Costa, P. T., & McCrae, R. R. (1989). *NEO PI/FFI manual supplement.* Lutz, FL: Psychological Assessment Resources.

Crane, D. R., & Mead, D. E. (1980). The Marital Status Inventory: Some preliminary data on an instrument to measure marital dissolution potential. *American Journal of Family Therapy, 8*(3), 31–35.

Crane, D. R., Newfield, N., & Armstrong, D. (1984). Predicting divorce at marital therapy intake: Wives' distress and the marital status inventory. *Journal of Marital and Family Therapy, 10,* 305–312.

Crowe, M. J. (1978). Conjoint marital therapy: A controlled outcome study. *Psychological Medicine, 8,* 623–636.

Davila, J. (2003). Attachment processes in couple therapy: Informing behavioral models. In S. M. Johnson & V. E. Whiffen (Eds.), *Attachment processes in couple and family therapy* (pp. 124–143). New York: Guilford Press.

Doss, B., Atkins, D. C., & Christensen, A. (2003). Deciding with their feet? Husbands and wives seeking marital therapy. *Journal of Marital and Family Therapy, 29,* 165–177.

First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (1994). *Structured Clinical Interview for DSM–IV Axis 1 Disorders (SCID-I/P)—patient edition* (Version 2.0). Arlington, VA: American Psychiatric Publishing.

Gottman, J. M. (1994). *What predicts divorce? The relationship between marital processes and marital outcomes.* Hillsdale, NJ: Erlbaum.

Hahlweg, K., Schindler, L., Revenstorf, D., & Brangelmann, J. C. (1984). The Munich marital therapy study. In K. Hahlweg & N. S. Jacobson (Eds.), *Marital interaction: Analysis and modification* (pp. 3–26). New York: Guilford Press.

Harrell, F. E. (2001). *Regression modeling strategies: With applications to linear models, logistic regression, and survival analysis.* New York: Springer.

Heavey, C. L., & Christensen, A. (1991). *Closeness and independence inventory.* Unpublished questionnaire.

Heavey, C. L., Christensen, A., & Zumbotel, D. C. (1996). The Communication Patterns Questionnaire: The reliability and validity of a constructive communication subscale. *Journal of Marriage and Family, 58,* 796–800.

Jacobson, N. S., & Addis, M. E. (1993). Research on couples and couple therapy: What do we know? Where are we going? *Journal of Consulting and Clinical Psychology, 61,* 85–93.

Jacobson, N. S., & Christensen, A. (1996). *Integrative couple therapy.* New York: Norton.

Jacobson, N. S., Follette, W. C., & Pagel, M. (1986). Predicting who will benefit from behavioral marital therapy. *Journal of Consulting and Clinical Psychology, 54,* 518–522.

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

Karney, B. R., & Bradbury, T. N. (1995). The longitudinal course of marital quality and stability: A review of theory, method, and research. *Psychological Bulletin, 118,* 3–34.

Karney, B. R., & Bradbury, T. N. (1997). Neuroticism, marital interaction, and the trajectory of marital satisfaction. *Journal of Personality and Social Psychology, 72,* 1075–1092.

Meehl, P. E. (1990). Why summaries of research on psychological theories are often uninterpretable. *Psychological Reports, 66,* 195–244.

Mendonca, J. D., Lumley, R. N., & Hunt, A. (1982). Brief marital therapy outcome: Personality correlates. *Canadian Journal of Psychiatry, 27,* 291–295.

O'Leary, K. D., & Turkewitz, H. (1981). A comparative outcome study of behavioral marital therapy and communication therapy. *Journal of Marital and Family Therapy, 7,* 159–169.

Pasch, L. A., & Bradbury, T. N. (1998). Social support, conflict, and the development of marital dysfunction. *Journal of Consulting and Clinical Psychology, 66,* 219–230.

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in S and S-Plus.* New York: Springer.

R Development Core Team (2004). *R: A language and environment for statistical computing.* Vienna: R Foundation for Statistical Computing.

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 Publishers.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Newbury Park: Sage.

Sharpley, C. F., & Cross, D. G. (1982). A psychometric evaluation of the Spanier Dyadic Adjustment Scale. *Journal of Marriage and the Family, 44,* 739–741.

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

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

Snyder, D. K., Mangrum, L. F., & Wills, R. M. (1993). Predicting couples' response to marital therapy: A comparison of short- and long-term predictors. *Journal of Consulting and Clinical Psychology, 61,* 61–69.

Snyder, D. K., & Whisman, M. A. (2004). Treating distressed couples with coexisting mental and physical disorders: Directions for clinical training and practice. *Journal of Marital and Family Therapy, 30,* 1–12.

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.

Spanier, G. B. (1989). *Manual for the Dyadic Adjustment Scale.* North Tonawanda, NY: Multi-Health Systems.

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

Spitzer, R. L., Williams, J. B. W., Gibbon, M., & First, M. B. (1994). *Structured Clinical Interview for DSM–IV Personality Disorders (SCID-II).* Washington, DC: American Psychiatric Press.

Vaida, F., & Blanchard, S. (2005). *Conditional Akaike information for mixed effects models. Biometrika, 92,* 351–370.

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

Whisman, M. A., & Jacobson, N. S. (1992). Change in marital adjustment following marital therapy: A comparison of two outcome measures. *Psychological Assessment, 4,* 219–223.