Published in final edited form as:

J Fam Psychol. 2001 June ; 15(2): 334–346.

Can Questionnaire Reports Correctly Classify Relationship Distress and Partner Physical Abuse?

Richard E. Heyman, Shari R. Feldbau-Kohn, Miriam K. Ehrensaft, Jennifer Langhinrichsen-Rohling, and K. Daniel O'Leary

State University of New York at Stony Brook.

Abstract

Relationship adjustment (e.g., Dyadic Adjustment Scale; DAS) and physical aggression (e.g., Conflict Tactics Scale) measures are used both as screening tools and as the sole criterion for classification. This study created face valid diagnostic interviews for relationship distress and physical abuse, through which one could compare preliminarily the classification properties of questionnaire reports. The DAS (and a global measure of relationship satisfaction) had modest agreement with a structured diagnostic interview; both questionnaires tended to overdiagnose distress compared with the interview. Results for partner abuse reiterated the need to go beyond occurrence of aggression as the sole diagnostic criterion, because men's aggression was more likely than women's to rise to the level of "abuse" when diagnostic criteria (injury or substantial fear) were applied.

Clinical researchers investigating the correlates, etiology, and treatment of relationship distress have long accepted the use of questionnaires to classify couples as "distressed" and "nondistressed" (see Gottman, 1979, 1994, for reviews). In contrast, clinical researchers investigating the correlates, etiology, and treatment of individual psychopathology have used structured clinical interviews with standardized criteria to classify individuals with and without disorders (e.g., the Structured Clinical Interview for *DSM-IV* Axis I Disorders—Clinical version; *SCID*; First, Gibbon, Spitzer, & Williams, 1997). Although such criteria have been criticized for imposing categories on dimensional phenomena and for deriving categories on a less than empirically sound basis (cf. Meehl, 1995), researchers have benefited from standardization of classification criteria and from the external validity provided by common research and clinical criteria.

Despite such controversy, the *Diagnostic and Statistical Manual of Mental Disorders (DSM-IV*; American Psychiatric Association, 1994) represents the consensus opinion among mental health professionals about the criteria for each mental disorder. No such consensus document exists for delineating the criteria for relationship distress. Instead, researchers have adopted a convention that distress is indicated by a score of 97 and below on the Dyadic Adjustment Scale (DAS; Spanier, 1976) or 100 and below on the Marital Adjustment Test (MAT; Locke & Wallace, 1959). Unfortunately, such an approach (a) makes it impossible to determine how often such determinations of distress are correct (Eddy, Heyman, & Weiss, 1991); (b) does not

Correspondence concerning this article should be addressed to Richard E. Heyman, Department of Psychology, State University of New York, Stony Brook, New York 11794-2500. Electronic mail may be sent to richard.heyman@sunysb.edu..

Richard E. Heyman, Shari R. Feldbau-Kohn, Miriam K. Ehrensaft, Jennifer Langhinrichsen-Rohling, and K. Daniel O'Leary, Department of Psychology, State University of New York at Stony Brook.

Miriam K. Ehrensaft is now at the Division of Child and Adolescent Psychiatry, Columbia University. Jennifer Langhinrichsen-Rohling is now at the Department of Psychology, University of South Alabama.

This research was supported by National Institute of Mental Health Grants T32MH19107 and R01MH57779.

take into account measurement error that could render decisions based on a single score unreliable (Jacobson & Truax, 1991); and (c) mirrors Spanier's tautology that relationship adjustment is both a process and an outcome of that process (by making adjustment scales both quantitative measures of the form and content of relationship problems and qualitative measures of clinical relationship distress). Perhaps most problematic, however, is the reification of the cutoff scores themselves. The DAS score of 97 and below appears to have been adopted as the cutoff because it was 1 standard deviation below the mean for Spanier's original sample of factory workers in Eastern Pennsylvania (Spanier, 1976). The origin of the cutoff score of 100 on the MAT is more mysterious, given the brevity of the original report (Locke & Wallace, 1959) and the lack of any paper establishing the psychometric soundness of the cutoff.

To be fair to the developers, despite the 40 years after the publication of the MAT and 25 years after the DAS, establishing the soundness of the cutoff is still difficult, considering that there is still no gold standard diagnostic interview to which they could be compared. Our prior work (Eddy et al., 1991; Heyman, Sayers, & Bellack, 1994) compared the DAS scores of clinic couples with those of community couples (eliminating community couples who declared themselves "unhappy" on the DAS global satisfaction question) and concluded that despite fairly good sensitivity and specificity, spouses classified as distressed on the DAS were truly distressed only 43%–71% of the time. Neither study had a well-operationalized "true diagnosis" criterion variable; the current study is intended to fill that gap by creating face valid diagnostic interviews for relationship distress and physical abuse.

Jacobson and colleagues (e.g., Jacobson & Truax, 1991) have attempted to ameliorate the problem of declaring a treated couple as a treatment success if they score 98 and above on the DAS but a treatment failure if they score 97 or below by suggesting that an error band (based on the reliability of the scale) be placed around adjustment scores. Thus, some couples would be clearly "distressed," some clearly "nondistressed," and some indeterminate (i.e., their scores fall within the cutoff point $\pm$ standard error). Although this suggestion is an improvement, it still fails to allow for the testing of the validity of DAS cutoffs against an independent criterion.

By not adopting more stringent standards (either Jacobson's or others) for classifying couples as distressed or nondistressed, researchers may misclassify couples and thus confound between-groups comparisons. Only a comparison of relationship adjustment measures with clinical criteria will allow us to establish the extent to which adjustment scales misclassify couples.

Classification for partner abuse has been much more straightforward, although ironically more controversial. The Conflict Tactics Scale (CTS; Straus, 1979) is the standard scale for assessing relationship aggression. Items are arranged from most constructive ("discussed issue calmly") to most abusive ("used a gun or knife on the partner"). All physical assaults, beginning with the mildest ("threw something at the partner"), are classified as physical aggression (Straus, 1979). The CTS's controversial status resides in the consistent finding that women are as aggressive or more aggressive than are men (e.g., Archer, 2000; Straus & Gelles, 1990). If *aggression* connotes physical assault and *abuse* connotes victimization, injury, fear, and subjugation, women are, by far, more frequently abused. For example, in the nationally representative 1985 National Family Violence Survey, severely assaulted women were seven times more likely to seek medical attention than were severely assaulted men (Stets & Straus, 1990). In clinical samples, women's risk of injury was three to seven times that of men (Cantos, Neidig, & O'Leary, 1994; Cascardi, Langhinrichsen, & Vivian, 1992). The adverse mental health consequences of women's victimization include increased rates of depression (Cascardi, O'Leary, Schlee, & Lawrence, 1995) and elevated levels of depressive symptomatology (e.g., Stets & Straus, 1990; Vivian & Langhinrichsen-Rohling, 1994); posttraumatic stress disorder

rates are elevated (between 33% and 47%) in both outpatient treatment and shelter settings (e.g., Cascardi et al., 1995; Dutton, 1992; Housekamp & Foy, 1991; Schlee, Heyman, & O'Leary, 1998). In conclusion, when only behavioral acts are measured, men and women appear to be equally aggressive. When the physical and psychological consequences of aggressive acts are included, women are far more likely to be victims than are men.

Recently, O'Leary and Jacobson (1997) reviewed the partner abuse literature for the *DSM-IV Sourcebook*. They also made recommendations for experimental criteria for the new *DSM-IV* codes (V61.1, Physical Abuse of Adult; 995.81, Physical Abuse of Adult [if focus of attention is on victim]). *DSM-IV* mandates that diagnoses be gender neutral (unless the disorder or problem is intrinsically gender related) posed difficulties for a field that typically views partner aggression as being a form of men's domination and control of women (cf. Yllö, 1993). To set the threshold for diagnosable abuse, O'Leary and Jacobson (1997) proposed criteria that included a consideration of both the occurrence of abuse and its impact (i.e., resulting in injury or significant fear). We have used this interview (Heyman, 1993; see the appendix) in a study of patients in a family practice setting (Pan, Ehrensaft, Heyman, O'Leary, & Schwartz, 1997); this study extends this work by assessing the diagnoses when both spouses are interviewed.

The purpose of this study is to evaluate whether questionnaire reports can correctly classify relationship distress and partner physical abuse when compared with diagnostic interviews. First, in keeping with our prior findings with adjustment measures (Eddy et al., 1991; Heyman et al., 1994), we expect that the DAS will be moderately capable of distinguishing distressed from nondistressed couples. Second, we believe that briefer measures of global satisfaction will perform equally as well. This is worthwhile to test because of both brevity and parsimony (i.e., adjustment measures contain a mélange of constructs, including estimations about the future of the relationship, e.g., Fincham & Bradbury, 1987). Third, given the concrete nature of the CTS, we predict that it will be highly related to interview reports when the occurrence-only criterion is used, thus classifying an equal number of women and men. However, we believe that far more women than men will be classified as victimized when the occurrence and injury–fear criteria are used. Finally, we wish to test the reliability of the diagnoses in a variety of ways (interrater, interpartner, and intermethod, i.e., questionnaire vs. diagnostic interview).

Method

Participants

Seventy-four couples of varying levels of relationship adjustment were recruited through community newspaper advertisements. This study was part of a larger collaborative research effort (e.g., Ehrensaft, Langhinrichsen-Rohling, Heyman, O'Leary, & Lawrence, 1999; Langhinrichsen-Rohling, Schlee, Monson, Ehrensaft, & Heyman, 1998). Although the current study uses a continuous distribution of relationship adjustment, the larger study's recruitment strategy oversampled for distressed couples using the following recruitment groups: 21 happily married couples (H), 20 maritally distressed, reporting husband-to-wife physically aggressive couples (D-A), 15 maritally distressed couples reporting no husband-to-wife physical aggression (D-NA), and 18 not happy but not distressed couples (NHbND). To experimentally constrain heterogeneity, we required all couples to have been married between 1 and 7 years. Couples were paid \$60 for their 3 hr of participation.

Participants were recruited through advertisements in local newspapers in Suffolk County, NY, inviting "couples married between 1 and 7 years to participate in a study of marriage." Because this advertisement yielded a rapid filling of the H group, the announcement was later rerun inviting "couples married 1 to 7 years having problems in their marriage to participate in a

study of marriage.” After couples called to volunteer for the study, spouses were individually interviewed over the telephone at a time they determined was “safe” to respond freely. First, spouses agreed to participate and were read a verbal consent form. Next, spouses gave oral responses to the MAT (Locke & Wallace, 1959; see also Krokoff, 1989) and the CTS (Straus, 1979). Of 168 couples who were screened for the project, 57 were rejected because they did not fit any group, 35 were accepted but were never successfully scheduled, and 74 were accepted and participated.

The H (nonphysically aggressive) group comprised couples with (a) an average MAT score of 115¹ or above, with neither spouse scoring below 100 and (b) no reports of physical aggression in the marriage by either partner. The D-NA group comprised couples with (a) an average MAT score of below 100 and neither spouse scoring above 115 and (b) no CTS reports of physical aggression in the marriage by either partner. The D-A group comprised couples with (a) an average MAT score of below 100, with neither spouse scoring above 115 and (b) at least one spouse reporting at least two CTS acts of husband-to-wife physical aggression during the past year. The NHbND group included couples with an average MAT score of between 100 and 114. If one spouse scored below 100 and one spouse scored above 115 on the MAT, the couple was disqualified.

Sample Characteristics

Sample characteristics for each of the four groups are presented in Table 1. Significant differences were found for number of children (with happy couples having fewer children) and for wives’ education (with the distressed-aggressive group reporting fewer years). The sample’s self-identified racial–ethnic identification was as follows: European American (83.3%), Hispanic (6%), African American (3.3%), Asian American (1.3%), and other (3.3%) ethnicities. There were no significant ethnic differences among groups.

Procedure

An in-person assessment lasting approximately 3 hr was conducted with all participants. The assessment consisted of a questionnaire packet filled out independently by each spouse, a series of individual interviews, and a dyadic interaction task. Data from the questionnaires and the diagnostic interviews were used in this study.

Because the intent of this study was to examine the questionnaires (DAS, Relationship Satisfaction Questionnaire [RSAT], and CTS), as they are normally administered, against a diagnostic interview, we wished to avoid any contamination of questionnaire responses from any other laboratory procedures (i.e., interviews, observed interactions). Thus, all participants completed the DAS, RSAT, and CTS before any other procedures commenced. Although counterbalancing the administration of diagnostic interviews and questionnaires would have eliminated any systematic effect of one on the other, it also would have potentially influenced responses on the questionnaires.

Measures

Modified Conflict Tactics Scale (MCTS; Pan, Neidig, & O’Leary, 1994)—The MCTS adds six additional items to Straus’s (1979) original 18-item CTS. The 24-item self-report inventory assesses the frequency of a variety of functional (e.g., calmly discussing a problem), verbally abusive (e.g., insults or swearing), and physically abusive (e.g., hitting) conflict tactics. The CTS has been used in national surveys of the prevalence of marital aggression

¹Prior work in our lab (O’Leary et al., 1989) found that the average MAT score of couples married 2½ years was approximately 115. Because we wished to define our *happy* group as truly happy (i.e., above average on marital adjustment) and not merely nondistressed, we used this as our cutoff.

(Straus & Gelles, 1990; Straus, Gelles, & Steinmetz, 1980), and it is the most widely used measure of physical abuse in intimate relationships. We used the standard (Straus, 1979), theoretically based, a priori definitions for mild aggression (controlled spouse physically; threw something at partner; pushed, grabbed, or shoved; slapped) and severe aggression (kicked, bit, hit with a fist; choked; beat up; physically forced sex; threatened with gun or knife; used gun or knife). Each item was rated on a frequency scale for its occurrence during the past year on a scale ranging from never to 1 time, 2 times, 3–5 times, 6–10 times, 11–20 times, more than 20 times. Internal consistency of the original CTS is good (Cronbach's $\alpha = .87$ for husbands, .88 for wives; Straus, 1979).

MAT—The MAT is one of the most widely used measures of marital adjustment. It comprises 15 items, 9 of which assess the degree of disagreement on major marital issues. It is sensitive to changes in marital therapy (e.g., O'Leary, 1987), and its convergent validity has been repeatedly demonstrated (e.g., Navran, 1967; Spanier, 1976). Scores can range from 2 to 158, with higher scores indicating higher levels of adjustment; a score of 100 has been the traditional cutoff point for marital distress. The MAT is easier to administer over the telephone than the DAS, and the telephone version has good test–retest reliability with the written version (Krokoff, 1989).

DAS—The DAS is a 32-item self-report inventory designed to measure the severity of relationship discord in intimate dyads. Scores range from 0 to 151, with higher values indicating more favorable adjustment. The items for the DAS were those out of an initial pool of 100 that (a) were normally distributed; (b) discriminated between married and divorced spouses; and (c) loaded highly with one of four factors (Dyadic Consensus; Dyadic Cohesion; Dyadic Satisfaction; and Affectional Expression). Spouses with scores below 98 are classified as discordant (Eddy et al., 1991; Jacobson et al., 1984).

RSAT—The RSAT comprises 11 items that assess satisfaction in various areas of the relationship (e.g., Communication and Openness and Degree of Affection and Caring) and 2 global satisfaction items. Respondents indicate their degree of satisfaction for each item on a scale ranging from 0 (*very dissatisfied*) to 6 (*very satisfied*). Total scores are the sum of the items and range from 0 to 78, with higher scores indicating greater satisfaction. The scale has excellent internal consistency; test–retest reliability; and factorial, content, and convergent validity (Burns & Sayers, 1988; Heyman et al., 1994).

Interview

Design—To evaluate the classification abilities of the DAS and CTS, a comparison measure was needed. Our measure (see the appendix) was patterned after the standard for structured diagnostic interviews: the *SCID* (First et al., 1997). The interview was administered as shown in the appendix: The relationship distress questions preceded the partner abuse questions. (The two were not counterbalanced because we believed participants would feel more at ease answering the more sensitive abuse questions after the relationship distress questions rather than vice versa.)

The marital distress diagnosis was designed to parallel the criteria for major depressive disorder, in which a diagnosis is made when at least one criterion is met for overall depressive feelings (i.e., depressed mood or anhedonia) and several criteria are met for other key symptoms of depression (significant cognitive, physiological, or mood impairment). Spouses were diagnosed as distressed if they met at least one criterion for overall dissatisfaction (perceived unhappiness, pervasive thoughts of divorce, or perceived need for professional help) and at least one criterion for key symptoms (one symptom of significant behavioral, cognitive, or affective impairment). This diagnosis was guided by some of the most established findings in

the marital literature² and represented an attempt to construct a face valid and content valid marital diagnosis instrument.

The partner abuse diagnosis was designed to operationalize O'Leary and Jacobson's (1997) criteria for V61.1 (Physical Abuse of Adult) and 995.81 (Physical Abuse of Adult; if focus of attention is on victim), with the exception that to be classified as "abuse," an act of aggression had to satisfy either Criteria A1 (injury) or A2 (significant fear). When administered to individuals in a family practice setting, this interview demonstrated high interrater reliability (Pan et al., 1997) for diagnosis of abuse.

Discriminative validity—Discriminative validity for the marital diagnosis was evidenced by significant differences between diagnosed maritally distressed and nondistressed couples on two measures collected for the larger study: (a) the Feelings Questionnaire (O'Leary, Fincham, & Turkewitz, 1983; a self-report inventory of positive affect felt for the spouse, $t[127] = 3.77, p \leq .001$); and (b) the positive behaviors from the Daily Checklist of Marital Activities (Broderick & O'Leary, 1986, a self-report checklist of the frequency of marital behaviors occurring during the past week, $t[77.09] = 6.41, p \leq .001$). Further, diagnosed maritally distressed and nondistressed couples' observed marital discussions differed on codes from the Rapid Marital Interaction Coding System (RMICS; Heyman & Vivian, 1993). As part of the larger study videotaped interactions were collected and RMICS coded. Diagnosed distressed, compared with nondistressed, spouses emitted more Hostile, $t(103) = -3.1, p \leq .005$, and Distress-Maintaining Attribution codes, $t(103) = -3.47, p = .001$. Interrater reliability (Cohen's κ) for these two codes was .59 and .53, respectively. Discriminative validity for the abuse diagnosis was evidenced by significant differences between diagnosed physically abused and nonabused wives on the dominance/isolation items from the Psychological Maltreatment of Women Scale (Tolman, 1989), $t(103) = 6.50, p = .001$.

Interrater agreement—Original interviews were conducted by a doctoral level psychologist or an advanced clinical psychology graduate student. All interviews were videotaped. A random sample of 25% of the interviews were watched and recorded by a research assistant. Reliability among the two coders was excellent across all categories of diagnoses. Agreement between the two coders for a diagnosis of distress was 96% ($\kappa = .92$). The two coders agreed on a diagnosis of partner aggression 91% of the time ($\kappa = .80$). For a diagnosis of partner injury or fear, agreement was 85% ($\kappa = .85$). Agreement for a diagnosis of aggression from one's partner was 93% ($\kappa = .85$). Both coders agreed on a diagnosis of injury 96% of the time ($\kappa = 1$).

Results

We first compare how diagnoses of marital distress from questionnaires correspond to those from interviews.³ Next we report on men's and women's rates of aggression and diagnosed abuse, followed by intra- and interspousal agreement on abuse.

²Items for Criteria A, overall dissatisfaction, were based on a commonsense approach to what constitutes significant distress: overall unhappiness in the relationship, persistent thoughts of divorce, or perceived need for professional help for the relationship. Items in Criteria B–D were based on scientific literature: negative reciprocity and withdrawal (reviewed in Heyman, 2001; Weiss & Heyman, 1997), distressed attributional pattern (Bradbury & Fincham, 1990), low sense of efficacy that the relationship can improve (e.g., Vanzetti, Notarius, & NeeSmith, 1992), intense and persistent levels of anger or sadness (e.g., Gottman, 1999). Overall dissatisfaction accompanied by apathy was included to make the B–D criteria exhaustive and because this is a presenting constellation of some couples (O'Leary, Heyman, & Jongsma, 1998).

³Because the construct of marital distress relies highly on an internal construct (i.e., a spouse's sentiment regarding the marriage), we have chosen not to analyze interpartner agreement. That is, a wife may be unhappy, but her husband may not know (or recognize) this. However, for partner abuse, both partners are rating observable events (i.e., physically aggressive behaviors and their sequelae).

Marital Distress

The DAS distress classification (a score of 97 and below) had fair agreement with the structured interview's distress diagnosis (see Table 2). Compared with the diagnostic interview, the DAS overclassified spouses in our sample as distressed (i.e., about 1 in 3 spouses with DAS scores ≤ 97 were not diagnosed as distressed on the interview). Because the RSAT does not have established cutoffs for distressed, receiver operating characteristic curves were used to determine the optimal cutpoint of 49. As hypothesized, the RSAT, a global measure of marital satisfaction, performed approximately as well as the longer DAS, a multifaceted measure of marital adjustment (see Table 2).

Partner Abuse

Gendered patterns in aggression and diagnosed abuse

As noted earlier, nationally representative studies using the CTS have found women, compared with men, to be equally or more physically aggressive (Straus & Gelles, 1990; Straus et al., 1980); however, physical impact and psychological impact are not the same (e.g., severely assaulted women, compared with men, are seven times more likely to seek medical attention; Stets & Straus, 1990). Responses to the abuse diagnostic interview followed similar patterns. Husbands, compared with wives, reported relatively more physical victimization in the past year (any: 46% vs. 38%; severe: 17% vs. 13%). As shown in Table 3, in keeping with this belief, when we applied diagnostic criteria for being abused (i.e., occurrence of aggression combined with reports of fear or injury), women reported being the victims of diagnosable partner abuse relatively more often than did men (9% for women vs. 5% for men).

Intraspace consistency of physical aggression reports between methods

Because the abuse interview repeated the abuse questions from the CTS (and added impact questions), we tested if participant reports of physical aggression perpetration and victimization were consistent across method (i.e., questionnaire vs. interview). Reports of perpetrating any aggression were concordant (men: 85% agreement, $\kappa = .87$; women: 81% agreement, $\kappa = .82$), as were reports of any victimization (men: 84% agreement, $\kappa = .85$; women: 74% agreement, $\kappa = .77$). However, reports of severe aggression were not as consistent across method (perpetration—men: 36% agreement, $\kappa = .48$; women: 55% agreement, $\kappa = .67$; victimization—men: 60% agreement, $\kappa = .70$; women: 40% agreement, $\kappa = .53$).

Interspace consistency reports of physical aggression and diagnosed abuse

Several studies have documented mediocre interspousal agreement on physical aggression (e.g., Heyman & Schlee, 1997; Moffitt et al., 1997). Would reports of diagnosable abuse perpetration and victimization evidence similar disagreement? As shown in Table 3, we replicated earlier findings that when one partner reported severe aggression, the other was unlikely to also report it (5/11, or 45%, for male $\rightarrow$ female cases and 4/13, or 31%, for female $\rightarrow$ male cases). As shown in Table 4, we extended these findings: When one partner reported diagnosable levels of abuse, the other was not highly likely to also report diagnosable levels (4/8, or 50%, of the male $\rightarrow$ female cases and 1/6, or 17%, of the female $\rightarrow$ male cases).

Discussion

Our results indicate that the widespread use of the DAS to classify spouses as maritally distressed or nondistressed is inadvisable. The DAS and the RSAT had modest agreement with a structured diagnostic interview, according to Cohen's kappa statistic (which controls for chance agreement). Both questionnaires tended to overdiagnose distress compared with the interview. Given the discriminative validity of the DAS (e.g., Spanier, 1988) and of the

diagnosis (presented above), one cannot automatically dismiss the validity of either one. Further validation work is necessary for both the questionnaire measures and for the diagnostic interview. Because this is the first study to actually test the diagnostic abilities of the questionnaires instead of using an arbitrary cutpoint, the failure to find convergence is cause for concern. It does not provide evidence that one format is more valid than the other.

More disturbing, however, is that adjustment and satisfaction measures are likely to perform even worse in real-world screening situations. We recruited roughly equal numbers of happy and nonphysically aggressive, not happy but not distressed, distressed and aggressive, and distressed and nonaggressive couples. Given the normal distributions of adjustment measures, there would be more couples in the error zone in a randomly drawn sample than there would be in our sample.

The results for partner abuse reflect the common wisdom that men's and women's aggression has differential impacts. Although women's aggression, even severe aggression, is more commonly reported on the CTS, men's perpetration was relatively more common when diagnostic criteria (injury or substantial fear) were applied.

Consistency of reporting, both for intraspouse reports of severe aggression and for interspouse reports of both aggression and diagnosable abuse, is a cause for concern. Further work is necessary to investigate (a) why spouses sometimes report aggression in one format and not the other; (b) why spouses disagree on the occurrence or impact of aggression; and (c) what modifications in methodology can improve the psychometric properties of perpetration and victimization classifications.

Several cautions are necessary when interpreting our results. First, there are no established criteria for marital distress or physical partner abuse. Even the consideration of the establishment of such criteria has been hotly debated (see Kaslow, 1996; O'Leary & Jacobson, 1997). Second, our study must be replicated with a representative sample to most clearly establish the diagnostic performance of the measures. Given the overrepresentation of easily diagnosable participants (i.e., those selected for high and low levels of marital adjustment), it is likely that adjustment measures will perform worse in representative samples than in this study.

Yet, as long as researchers continue to use between-groups designs, the development and use of standard diagnostic criteria for classifying participants is critical (regardless of whether such diagnoses appear in the *DSM-IV*). As our results clearly show, evading the establishment of criteria only continues the use of our current generation of adjustment measures, which although highly efficient as screening tools (Heyman et al., 1994), are highly error prone as diagnostic tools. Establishing criteria would unify classification strategies and allow the development of improved screening tools, which could be optimized for predicting outcomes on the established criteria. Failure to do so will continue to add substantial error variance to relationship studies, obfuscating potentially critical differences. Finally, in regard to partner abuse, diagnostic criteria also are needed. We recommend (a) that impact be taken into account by establishing diagnostic criteria that require that an aggressive incident result in injury or substantial fear and (b) continued work on developing measures that produce, at the very least, more consistent test-retest reliabilities for reports of severe aggression across questionnaire and interview administrations.

Implications for Application and Public Policy

Adjustment measures (e.g., DAS or MAT) and the CTS have been the predominant means of measuring marital adjustment and partner abuse, respectively. Their widespread use also has elicited widespread debate about their utility. Although the *DSM-IV* includes V-code diagnoses

for both partner relational problems and partner abuse, no criteria have yet been disseminated. Our data indicate that the use of adjustment measures as both the measure of and the criteria for distress cannot stand. The use of the CTS as a measure of abuse is inadvisable unless accompanied by injury and impact questions (as have now been incorporated into the CTS2; Straus, Hamby, Boney-McCoy, & Sugarman, 1996). Public policymakers—including politicians, managed care plans, grant funders, and the psychiatric diagnostic establishment—are unlikely to treat “relational problems” as real clinical phenomena without better diagnostic tools. The sole purpose of this article has been to test if our current questionnaire tools are sufficient for diagnostic uses. Because the answer seems to be no, we recommend increased efforts to develop agreed-on diagnostic criteria for relationship distress and abuse and for increased testing of questionnaires against such criteria.

References

- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.
- Archer J. Sex differences in aggression between heterosexual partners: A meta-analytic review. *Psychological Bulletin* 2000;126:685–689. [PubMed: 10989617]
- Bradbury TN, Fincham FD. Attributions in marriage: Review and critique. *Psychological Bulletin* 1990;107:3–33. [PubMed: 2404292]
- Broderick JE, O’Leary KD. Contributions of affect, attitudes, and behavior to marital satisfaction. *Journal of Consulting and Clinical Psychology* 1986;54:514–517. [PubMed: 3745604]
- Burns, D. D., & Sayers, S. L. (1988). Cognitive and affective components of marital satisfaction: 1. Development and validation of a brief relationship satisfaction scale. Unpublished manuscript, Presbyterian Medical Center of Philadelphia, Department of Psychiatry.
- Cantos AL, Neidig PH, O’Leary KD. Injuries of women and men in a treatment program for domestic violence. *Journal of Family Violence* 1994;9:113–124.
- Cascardi M, Langhinrichsen J, Vivian D. Marital aggression, impact, injury, and health correlates for husbands and wives. *Archives of Internal Medicine* 1992;152:1178–1184. [PubMed: 1599345]
- Cascardi M, O’Leary KD, Schlee KA, Lawrence EE. Major depressive disorder and posttraumatic stress disorder in physically abused women. *Journal of Consulting and Clinical Psychology* 1995;63:616–623. [PubMed: 7673539]
- Dutton MA. Treating battered women in the aftermath stage: Special issue. *Psychotherapy in independent practice: Current issues of clinicians. Psychotherapy in Private Practice* 1992;10:93–98.
- Eddy JM, Heyman RE, Weiss RL. An empirical evaluation of the Dyadic Adjustment Scale: Exploring the differences between marital “satisfaction” and “adjustment. *Behavioral Assessment* 1991;13:199–220.
- Ehrensaft MK, Langhinrichsen-Rohling J, Heyman RE, O’Leary KD, Lawrence E. Feelings controlled in marriage: A phenomenon specific to physically aggressive couples? *Journal of Family Psychology* 1999;13:20–32.
- Fincham FD, Bradbury TN. The assessment of marital quality: A reevaluation. *Journal of Marriage and the Family* 1987;49:797–809.
- First, M. B., Gibbon, M., Spitzer, R. L., & Williams, J. B. W. (1997). *Structured Clinical Interview for DSM-IV Axis I Disorders—Clinician version*. Washington, DC: American Psychiatric Press.
- Gottman, J. M. (1979). *Marital interaction: Experimental investigations*. New York: Academic Press.
- Gottman, J. M. (1994). *What predicts divorce?* Hillsdale, NJ: Erlbaum.
- Gottman, J. M. (1999). *The marriage clinic: A scientifically-based marital therapy*. New York: Norton.
- Heyman, R. E. (1993). *Distress and abuse classification interviews*. Unpublished manuscript, State University of New York at Stony Brook.
- Heyman RE. Observation of couple conflicts: Clinical assessment applications, stubborn truths, and shaky foundations. *Psychological Assessment* 2001;13:5–35. [PubMed: 11281039]
- Heyman RE, Sayers SL, Bellack AS. Global marital satisfaction vs. marital adjustment: Construct validity and psychometric properties of three measures. *Journal of Family Psychology* 1994;8:432–446.

- Heyman RE, Schlee KA. Toward a better estimate of the prevalence of partner abuse: Adjusting rates based on the sensitivity of the Conflict Tactics Scale. *Journal of Family Psychology* 1997;11:332–338.
- Heyman, R. E., & Vivian, D. (1993). *Rapid Marital Interaction Coding System (RMICS): Training manual for coders*. Stony Brook, NY: Authors. Available on the World Wide Web: <http://www.psy.sunysb.edu/marital>
- Housekamp BM, Foy DW. The assessment of posttraumatic stress disorder in battered women. *Journal of Interpersonal Violence* 1991;6:367–375.
- Jacobson NS, Follette WC, Revenstorf D, Baucom DH, Hahlweg K, Margolin G. Variability in outcome and clinical significance of Behavioral Marital Therapy: A reanalysis of outcome data. *Journal of Consulting and Clinical Psychology* 1984;52:497–504. [PubMed: 6470276]
- Jacobson NS, Truax P. Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology* 1991;59:12–19. [PubMed: 2002127]
- Kaslow, F. W. (1996). *Handbook of relational diagnosis and dysfunctional family patterns*. New York: Wiley.
- Krokoff LJ. Predictive validation of a telephone version of the Locke-Wallace Marital Adjustment Test. *Journal of Marriage and the Family* 1989;51:767–775.
- Langhinrichsen-Rohling J, Schlee KA, Monson CM, Ehrensaft MK, Heyman RE. What's love got to do with it? Perceptions of marital positivity in H-to-W aggressive, distressed, and happy marriages. *Journal of Family Violence* 1998;13:197–212.
- Locke HJ, Wallace KM. Short marital adjustment and prediction tests: Their reliability and validity. *Marriage and Family Living* 1959;21:251–255.
- Meehl PE. Bootstrap taxometrics: Solving the classification problem in psychopathology. *American Psychologist* 1995;50:266–275. [PubMed: 7733538]
- Moffitt TE, Caspi A, Krueger RF, Magdol L, Margolin G, Silva PA, Sydney R. Do partners agree about abuse in their relationship? A psychometric evaluation of interpartner agreement. *Psychological Assessment* 1997;9:47–56.
- Navran L. Communication and adjustment in marriage. *Family Process* 1967;6:173–184.
- O'Leary, K. D. (Ed.). (1987). *Assessment of marital discord*. Hillsdale, NJ: Erlbaum.
- O'Leary KD, Barling J, Arias I, Rosenbaum A, Malone J, Tyree A. Prevalence and stability of physical aggression between spouses: A longitudinal analysis. *Journal of Consulting and Clinical Psychology* 1989;57:263–268. [PubMed: 2785126]
- O'Leary KD, Fincham FD, Turkewitz H. Assessment of positive feelings toward spouse. *Journal of Consulting and Clinical Psychology* 1983;51:949–951. [PubMed: 6655111]
- O'Leary, K. D., Heyman, R. E., & Jongsma, A. E. (1998). *The couples therapy treatment planner*. New York: Wiley.
- O'Leary, K. D., & Jacobson, N. S. (1997). Partner relational problems with physical abuse. *DSM-IV sourcebook* (pp. 673–692). Washington, DC: American Psychiatric Association.
- Pan HS, Ehrensaft MK, Heyman RE, O'Leary KD, Schwartz R. Evaluation of the criteria for spouse abuse in a sample of patients at a family practice clinic. *Family Medicine* 1997;29:492–495. [PubMed: 9232411]
- Pan HS, Neidig PH, O'Leary KD. Male-female and aggressor-victim differences in the factor structure of the modified Conflict Tactics Scale. *Journal of Interpersonal Violence* 1994;9:366–382.
- Schlee KA, Heyman RE, O'Leary KD. Group treatment for spouse abuse: Are women with PTSD appropriate participants? Do partners agree about abuse in their relationship? A psychometric evaluation of interpartner agreement. *Journal of Family Violence Psychological Assessment* 1998;1997;139:1–20. 47–56.
- Spanier GB. Measuring dyadic adjustment: New scales for assessing the quality of marriage and similar dyads. *Journal of Marriage and the Family* 1976;38:15–28.
- Spanier, G. B. (1988). *Dyadic adjustment scale manual*. North Tonawanda, NY: Multi-Health Systems.
- Stets, J. E., & Straus, M. A. (1990). Gender differences in reporting marital violence and its medical and psychological consequences. In M. A. Straus & R. J. Gelles (Eds.), *Physical violence in American*

families: *Risk factors and adaptations to violence in 8,145 families* (pp. 151–165). New Brunswick, NJ: Transaction.

Straus MA. Measuring intrafamily conflict and violence: The Conflict Tactics (CT) Scales. *Journal of Marriage and the Family* 1979;41:75–78.

Straus, M. A., & Gelles, R. J. (1990). *Physical violence in American families*. New Brunswick, NJ: Transaction.

Straus, M. A., Gelles, R. J., & Steinmetz, S. K. (1980). *Behind closed doors: Violence in the American family*. Garden City, NY: Anchor Press.

Straus MA, Hamby SL, Boney-McCoy S, Sugarman DB. The Revised Conflict Tactics Scale (CTS2): Development and preliminary psychometric data. *Journal of Family Issues* 1996;17:283–316.

Tolman RM. The development of a measure of psychological maltreatment of women by their male partners. *Violence and Victims* 1989;4:159–177. [PubMed: 2487132]

Vanzetti NA, Notarius CI, NeeSmith D. Specific and generalized expectancies in marital interaction. *Journal of Family Psychology* 1992;6:171–183.

Vivian D, Langhinrichsen-Rohling J. Are bi-directionally violent couples mutually victimized? A gender sensitive comparison. *Violence and Victims* 1994;9:107–124. [PubMed: 7696192]

Weiss, R. L., & Heyman, R. E. (1997). Couple interaction. In W. K. Halford & H. J. Markman (Eds.), *Clinical handbook of marriage and couples intervention* (pp. 13–41). New York: Wiley.

Yllö, K. (1993). Through a feminist lens: Gender, power, and violence. In R. J. Gelles & D. R. Loseke (Eds.), *Current controversies on family violence* (pp. 47–62). Newbury Park, CA: Sage.

Appendix 1: Distress and Abuse Classification Interviews

Current relationship distress	Criteria				
Now I'm going to ask you some questions about your relationship					
In the last month ...					
... have you been feeling unhappy about your relationship? (Has it been more than half of the days in the last month?)	A. (1) Subjective sense of overall unhappiness in the relationship during the past month, more days than not	?	1	2	3
... have you had thoughts of separation or divorce, or that you'd be better off without your partner? IF YES: How long do these thoughts last?	(2) Thoughts of divorce/separation that are more than transitory	?	1	2	3
Do you think your relationship is in need of professional help?	(3) Perceived need for professional help for the relationship. Excluded are situations in which the relationship itself is not troubled (e.g., sexual problems without accompanying relationship problems, adjusting to infertility in a happy couple)	?	1	2	3
During the past month, when you disagreed about something with your partner ...					
... does it escalate quickly into a fight?	B. (1) Marked escalation of negative behavior or affect (e.g., "little" disputes evolve into screaming matches)	?	1	2	3
(Do little problems escalate into screaming matches?)		?	1	2	3
... do you or your partner leave the room, withdraw or stop talking so that it's impossible to resolve the problem?	(2) Withdrawal from interaction so that resolution is impeded. Withdrawal can either be through leaving a discussion before it is resolved, or through more pervasive disconnectedness that impedes bringing up or resolving problems. NOTE: Constructive use of time-out procedures that include later discussion of the issue would not meet this criterion.	?	1	2	3
During the past month, when your partner has done something you didn't like, did you think that s/he did it on purpose or that s/he had negative intentions? Do you think that it reflects a negative part of his/her personality?	C. (1) Distressed attributional pattern: Negative behaviors of the partner are attributed to negative personality traits, or are perceived to be done voluntarily, intentionally, or with negative intent. Positive behaviors of the partner are attributed to temporary states, or are perceived to be done accidentally, unintentionally, or with hidden negative intentions.	?	1	2	3
Do you think that you can improve your relationship?	(2) Low sense of efficacy that the relationship can improve (without professional help)	?	1	2	3
During the past month, how often have you felt angry or sad about your partner?	D. Interactions with or thoughts about the partner are frequently marked by intense and persistent levels of: (1) Anger or sadness	?	1	2	3

Current relationship distress		Criteria					
(If apathy seems present) Do you feel as strongly about your partner as you did once? Physical aggression (victimization) Many people, at one time or another, get physical with their partners when they're angry. For example, some people threaten to hurt their partners, some push or shove, and some slap or hit. I'm going to ask you about a variety of common behaviors, and I'd like you to tell me if YOUR PARTNER'S done this during the past year: (For each behavior answered yes, ask about • the most serious occurrence in the past year • if subject was bruised or injured in any other way. Code 3 for any injury.) Thrown or smashed or hit or kicked an object ^a Thrown something at you	(2) Apathy			?	1	2	3
		Happened?				Injured?	
	A. (1) Respondent was victimized by partner's physical aggression, resulting in an injury.	Y	N	?	1	2	3
Pushed, grabbed or shoved you		Y	N	?	1	2	3
Slapped, kicked, or bit you		Y	N	?	1	2	3
Hit you with a fist or with an object		Y	N	?	1	2	3
Beat you up		Y	N	?	1	2	3
Threatened you with a gun or knife		Y	N	?	1	2	3
Used a gun or knife		Y	N	?	1	2	3
Physically forced you to have sex when you didn't want to		Y	N	?	1	2	3
Other_____		Y	N	?	1	2	3
IF YES BUT NO INJURY, ask "Some people are afraid that their partners will physically hurt them if they argue with their partners or do something their partners don't like. How much would you say you are afraid of this?" (READ LIST)	A. (2) Respondent was victimized by partner's physical aggression, resulting in significant fear.						
Not at all					1		
A little						2	
Quite a bit							3
Very afraid							3
Physical aggression (perpetration) Now I'd like to ask what YOU've done during the past year: (For each behavior answered yes, ask about • the most serious occurrence in the past year • if subject was bruised or injured in any other way. Code 3 for any injury.) In the last year, have YOU ... Thrown or smashed or hit or kicked an object ^a Thrown something at your partner		Happened?				Injured?	
	A. (1) Respondent committed physical aggression, resulting in an injury to the partner.	Y	N	?	1	2	3
		Y	N	?	1	2	3
Pushed, grabbed or shoved your partner		Y	N	?	1	2	3
Slapped, kicked, or bit your partner		Y	N	?	1	2	3
Hit your partner with a fist or with an object		Y	N	?	1	2	3
Beat your partner up		Y	N	?	1	2	3
Threatened your partner with a gun or knife		Y	N	?	1	2	3
Used a gun or knife on your partner		Y	N	?	1	2	3
Physically forced your partner to have sex when she (he) didn't want to		Y	N	?	1	2	3
Other_____		Y	N	?	1	2	3
IF YES BUT NO INJURY, ask "Some people are afraid that their partners will physically hurt them if they argue with their partners or do something their partners don't like. How much would you say you are afraid of this?" (READ LIST)	A. (2) Respondent committed physical aggression, resulting in significant fear.						
Not at all					1		
A little						2	
Quite a bit							3
Very afraid							3

Current relationship distress	Criteria
-------------------------------	----------

Note. Y = yes; N = no; ? = inadequate information; 1 = absent or false; 2 = subthreshold; 3 = threshold or true

^a Warm-up item; not scored.

Table 1

	Husbands								Wives									
	Happy		Distressed–aggressive		Distressed–nonaggressive		Not happy but not distressed		<i>F^b</i>	Happy		Distressed–aggressive		Distressed–nonaggressive		Not happy but not distressed		<i>F^a</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
Marital status	31.64	4.24	35	8.3	37.47	10.42	34.72	10.18	1.48	30.24	3.82	35.95	6.17	35.07	9.8	31.88	6.43	1.57
Married	4.43	2.16	5.35	3.95	4.10	2.48	4.41	1.50	0.73	4.43	2.16	5.40	4.13	4.03	2.4	4.59	1.67	0.76
Married if children	1.12 ^a	1.67	2.47 ^b	1.31	1.83, ^{ab}	1.19	1.94, ^{ab}	1.03	3.98 [*]	1.17 ^a	1.15	2.26 ^b	1.28	1.67, ^{ab}	0.98	1.82, ^{ab}	0.73	3.62 [*]
Education	13.52	3.40	12.65	3.45	13.80	1.97	14.12	1.80	0.91	14.33 ^a	2.35	12.00	3.35	13.53, ^{ab}	1.80	14.18, ^{ab}	2.67	3.27 [*]
Income	51,105	19,788	36,684	19,310	53,533	21,902	46,750	22,750	2.29	45,000	20,632	41,285	23,305	50,000	25,327	43,562	19,825	0.36

Subscripts indicate where significant differences were found using Tukey's honestly significant difference test.

years of freedom: age (3, 68), years married (3, 69), number of children (3, 61), education (3, 69), and income (3, 65).

years of freedom: age (3, 58), years married (3, 69), number of children (3, 64), and income (3, 62).

05. For remaining F s, p 's were not significant.

Table 2
Comparison of Questionnaire and Interview Classifications of Marital Distress

Questionnaire classification	Interview diagnosis			Cohen's κ
	Not distressed	Distressed	% agreement	
DAS				
Men				
Not distressed		2	82.6	.63
Distressed	37	20		
Women	10		71.4	.43
Not distressed	29	7		
Distressed	13	21		
RSAT				
Men				
Not distressed	37	4	81.8	.60
Distressed	8	17		
Women				
Not distressed	32	7	76.4	.51
Distressed	9	20		

Note. DAS = Dyadic Adjustment Scale; RSAT = Relationship Satisfaction Questionnaire.

Table 3
Concordance of Reports of Perpetration and Victimization of Diagnosed Partner Physical Abuse on the Diagnostic Interview

Perpetrator report	Victim report							
	Male → female abuse ^a				Female → male abuse ^b			
	No abuse		Abuse		No abuse		Abuse	
	No.	%	No.	%	No.	%	No.	%
No abuse	67	89	3	4	69	92	3	4
abuse	1	1	4	5	2	3	1	1

^a $\kappa = .639$.

^b $\kappa = .251$.

Table 4
Concordance of Reports of Perpetration and Victimization of Partner Aggression on the Conflict Tactics Scale

Perpetrator report	Victim report							
	Male → female severe aggression ^a				Female → male severe aggression ^b			
	Not severe		Severe		Not severe		Severe	
	No.	%	No.	%	No.	%	No.	%
Not severe	57	81	4	6	55	80	6	9
Severe	4	6	5	7	3	4	4	6

^a $\kappa = .490$.

^b $\kappa = .398$.