

Journal of Interpersonal Violence

<http://jiv.sagepub.com/>

Assessment of the Reliability of the Conflict Tactics Scales: A Meta-Analytic Review

JOHN ARCHER

J Interpers Violence 1999 14: 1263
DOI: 10.1177/088626099014012003

The online version of this article can be found at:

<http://jiv.sagepub.com/content/14/12/1263>

Published by:



<http://www.sagepublications.com>

On behalf of:

American Professional Society on the Abuse of Children

Additional services and information for *Journal of Interpersonal Violence* can be found at:

Email Alerts: <http://jiv.sagepub.com/cgi/alerts>

Subscriptions: <http://jiv.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.com/journalsPermissions.nav>

Citations: <http://jiv.sagepub.com/content/14/12/1263.refs.html>

>> [Version of Record](#) - Dec 1, 1999

[What is This?](#)

This article reports meta-analyses of self-agreement and partners' agreement for physical aggression in relationships, measured by the Conflict Tactics Scales. Evidence from concordance rates was inconclusive, and the limited correlational data indicated high interpartner agreement. Differences between self-reports and partners' reports for men and women were analyzed to address the following hypotheses: Men but not women underreport their own aggression, both sexes underreport their own aggression, and men underreport their victimization. In 18 studies of couples, mean weighted effect sizes showed higher ratings of aggression from partners than from self-reports for both men and women. In 43 studies of unmatched men and women, the mean differences were smaller than for couples but were greater for men than for women. Overall, this evidence indicated systematic underreporting in self-reports by both sexes (Hypothesis 2), which was greater for men among the larger number of studies in which the men and the women did not come from matched couples (Hypothesis 1).

Assessment of the Reliability of the Conflict Tactics Scales

A Meta-Analytic Review

JOHN ARCHER

University of Central Lancashire

The Conflict Tactics Scales (CTS) were designed by Straus (1979) to measure physical aggression between partners by assessing which of various acts they have used in solving conflicts during a stated period of time. The CTS violence scale,¹ which consists of acts of physical aggression, has been the principal method of measuring physical aggression among community samples of married couples or individuals in dating relationships. A general finding from this research is that the proportion of women reporting one or more acts of physical aggression is often greater than or equal to that of men (Bernard & Bernard, 1983; Deal & Wampler, 1986; Foshee, 1996; Morse, 1995). Such findings are apparently inconsistent with feminist accounts of relationship violence: that it is mainly perpetrated by men toward their female partners (Dobash & Dobash, 1977-1978, 1980; Pagelow, 1984; Walker, 1989, 1990).

Author's Note: I thank Sarah Monks for assistance with the meta-analysis and literature search; Blair Johnson and Alice Eagly for answering inquiries about meta-analytic procedures; and Jane Ireland, Norman Birbeck, and Marion Seed for help with the literature search. Address correspondence to John Archer, Department of Psychology, University of Central Lancashire, Preston, PR1 2HE, Lancashire, UK; e-mail: j.archer@uclan.ac.uk.

JOURNAL OF INTERPERSONAL VIOLENCE, Vol. 14 No. 12, December 1999 1263-1289

© 1999 Sage Publications, Inc.

Such discrepancies soon led to criticisms of the CTS by those using different methods and reaching different conclusions.

Criticisms of the CTS have centered on several issues (Dobash, Dobash, Cavanagh, & Lewis, 1998; Dobash, Dobash, Daly, & Wilson, 1992), for example, whether reliance on the specific acts of physical aggression listed on the CTS ignores their meaning and context and whether the impact or consequences of such acts is likely to be different for men and women (Nazroo, 1995; Romkens, 1997). Although these two issues are important, this article will concentrate on a third issue, namely, the reliability of the CTS violence scale, which has been questioned by critics such as Dobash et al. (1992) on the basis of a partial and narrative assessment of the evidence that was then available.

Unlike most general questionnaire measures of aggression (Buss & Durkee, 1957; Buss & Perry, 1992; Gladue, 1991), which are exclusively self-reports, many (but certainly not all) studies using the CTS ask respondents to report on both themselves and their partners, which should enable interpartner agreements to be calculated in a way that is comparable with interobserver reliabilities. This would also correspond to the comparisons of self-reports with reports obtained from friends or acquaintances of the people concerned for general measures of aggression (Buss & Perry, 1992). However, few studies that involved couples provided this type of evidence. Studies involving individuals do not enable couples to be matched and therefore cannot produce agreement statistics.

Values for agreements between partners, in the form of correlations² or concordance rates,³ inform us about the consistency of reporting but not the extent of agreement on the absolute level of aggression. For example, if men consistently underreport every act of physical aggression by half compared to their women partners, the correlation will still indicate complete agreement. Assessment of the difference between self-ratings and partners' ratings of the direction and magnitude of aggression in a sample can inform us about the overall agreement between partners. For example, if the mean frequency of CTS acts is 3.00 according to men's self-reports and 6.00 according to their partners' reports, and the pooled standard deviation from the two measures is 3.00, the effect size (g) will be 1.00 ($g = [6.00 - 3.00]/3.00$). This would be a large difference for the social sciences generally according to Cohen's (1988) criteria. We should note that this value could be obtained from the same population as that given in the previous example, in which the correlation indicated complete agreement.

There are two advantages in calculating the magnitude of difference between self-reports and partners' reports. A first advantage is that unlike a

correlation coefficient, it gives an indication of whether self-reports or partners' reports are overall higher or lower than one another. This is a useful statistic when considering a topic such as physical aggression in relationships that is likely to be subject to both unsystematic inaccuracies in reporting and systematic underreporting or overreporting by one or the other respondent. A second, practical, advantage is that it is possible to calculate effect sizes from the data presented in many existing studies, both those providing data on couples and those providing data on individuals. In the latter case, there will be an additional source of unsystematic variation in that the couples will not be matched. With this proviso, these data can provide useful additional evidence about whether self-reports or partners' reports are consistently higher or lower than one another.

This article presents a systematic examination of the values for male and female partners in studies using the CTS and modifications of it. It is based on data that were used for a separate meta-analysis of sex differences in relationship aggression (Archer, in press). This sample included a few studies using measures of physical aggression⁴ other than the CTS. These were included in this analysis to provide a comparison between the CTS and similar measures.

The relatively few studies that provided correlations or concordance rates are evaluated. Then, the magnitude of the difference between self-reports and partners' reports for men's and women's physical aggression is calculated in the form of effect sizes using studies that obtained values for each member of the same couple and the more numerous studies involving individuals.

The statistics outlined above are used to assess the extent of agreement between perpetrators and recipients of physical aggression and whether there is evidence of systematic bias. Existing studies of relationship aggression (O'Leary & Arias, 1988; Riggs, Murphy, & O'Leary, 1989) suggest a number of possible hypotheses (see Moffitt et al., 1997), the most likely of which are the following:

Hypothesis 1: Men might underreport their own physical aggression in view of widespread disapproval of men hitting women (Arias & Johnson, 1989; Harris & Cook, 1994).

Hypothesis 2: Aggression on the part of both men and women is underreported by perpetrators, since there is also widespread disapproval of any form of women's aggression (White & Kowalski, 1994).

Hypothesis 3: Men might underreport their victimization because of negative attitudes toward male victims of marital violence (George, 1994).

Combinations of these hypotheses are also possible; for example, Hypothesis 1 and Hypothesis 3 or Hypothesis 1 and Hypothesis 2. Table 1

TABLE 1: Summary of the Main Hypotheses

<i>Source of Bias</i>	<i>Predicted Finding</i>
Hypothesis 1: Men underreport their own aggression	Higher values for victim's than perpetrators' reports, for men only
Hypothesis 2: Both sexes underreport their own aggression	Higher values for victim's than perpetrators' reports, for both sexes
Hypothesis 3: Men underreport their victimization	Higher values for perpetrators' than victim's reports, for women only

summarizes the three possible sources of bias and outlines the findings predicted in each case.

METHOD

Sample of Studies

The main search involved PsycLIT(r) on CD-ROM for the years 1976 to mid-1997 using the keywords *marital* or *dating* and *aggression* or *violence* but excluding *sexual*, *rape*, and *pornography*. This search produced 571 titles, which were reduced to those containing usable information by examining the titles and abstracts.

Second, the descendency method was applied to the standard questionnaire measure used in research on relationship aggression, the CTS (Straus, 1979). BIDS⁵ searches were conducted of all subsequent studies that cited this article to find those that fulfilled the selection criteria specified below.

Third, dissertations were searched via Dissertation Abstracts International Online (DISS) using the same keywords as above. This search produced 426 titles and abstracts from 1979 to mid-1997. These were examined for possible relevance to the selection criteria, and those likely to fulfill the selection criteria were subsequently examined on microfiche.

Fourth, lists of current studies on aggression under the title *A Guide to the Literature on Aggressive Behavior*, which appears regularly in the journal *Aggressive Behavior*, were examined for the years 1987 to 1997. These lists are derived from extensive keyword searches of ISI Science Citation Index, Social Science Citation Index, and Current Contents. Studies of physical aggression in marital or dating relationships were obtained from the lists

using the titles to assess whether the contents were likely to be within the scope of this review.

Fifth, a hand search of journals covering relationship aggression for the years 1987 to 1997 was conducted, notably *Aggressive Behavior*, *Family Relations*, *Journal of Family Violence*, *Journal of Interpersonal Violence*, *Journal of Marriage and the Family*, *Journal of Personality and Social Psychology*, *Journal of Social and Personal Relations*, and *Violence and Victims*. Again, titles were first examined for studies likely to concern physical aggression in marital or dating relationships, and abstracts of all possibly relevant work were checked.

Finally, as part of a wider search of studies involving sex differences in aggression, PsycINFO 1967-1996 was searched using the keywords *human sex differences* and either *aggressive behavior* or *violence*. Studies involving physical aggression in marital or dating relationships were examined for possible inclusion in this study.

In all of these searches, the criterion used for inclusion was the presence of either or both of the following: first, concordance rates or correlations for the association between self-ratings and partners' ratings of physical aggression against the partner; second, self-reports and partners' reports of men's and women's physical aggression in present or past relationships expressed as either the number or frequency of acts of physical aggression per individual. Studies that involved matched pairs of couples were considered separately from those using samples of men and women who were not selected as couples. Although the main focus of this analysis was the violence scale of the CTS, as indicated above, comparable alternative methods of assessing physical aggression (used in only a minority of studies) were also included to provide a comparison with the more commonly used CTS measure.

Correlations and Concordance Rates

Concordance rates are likely to be unsuitable for meta-analysis. They are sometimes expressed in the form of percentage reliabilities, but it is preferable to use a coefficient that takes account of chance levels of agreement, such as Cohen's kappa (Martin & Bateson, 1993). Few studies have done so; therefore, the information available on concordance rates is presented in a narrative form.

Correlations that were provided were recorded, and the mean value, weighted for the reciprocal of the variance (Hedges & Olkin, 1985), was calculated separately for men and women using DSTAT software (Johnson, 1989).

Effect Size Calculations for Differences Between Partners' Estimates of Aggression

Two meta-analyses were carried out, one involving samples of couples and the other involving data on individuals for men's and women's aggression, assessed by perpetrators and recipients. Studies were included in these meta-analyses if an effect size could be calculated for the difference between self-reports and partners' reports for either the males' physical aggression, the females' physical aggression, or both. The measure of effect size was g , which was obtained, if possible, from interval data. DSTAT software (Johnson, 1989) was used to calculate the g values. Two researchers independently calculated the values, and, in all cases of discrepancies, the values were recalculated.

Analysis of Effect Sizes

For the data on couples and data on individuals, the following were calculated separately for the male and female partners. First, gs were converted to ds^6 for each study. Second, the mean effect size (d) was calculated for each set of studies, weighted by the reciprocal of the variance, which gives more weight to those d values that are more reliably estimated (Hedges & Olkin, 1985). Third, each set was tested for homogeneity of effect sizes across all the component studies by calculating the homogeneity statistic Q_w , which has an approximate chi-square distribution with $k - 1$ degrees of freedom, k being the number of effect sizes. Where significant heterogeneity was found, successive outliers were removed, and the mean d values were recalculated. This process was repeated until Q_w was nonsignificant. All calculations were carried out using D-STAT software (Johnson, 1989).

Variables Coded From Each Study

The following information was coded from each study: (a) source of data (journal article, book or book chapter, dissertation, or other unpublished source), (b) measurement instrument, (c) country, (d) age category, (e) sample, (f) marital status, (g) level of measurement, and (h) reference period. In each case, coding was undertaken separately by two coders, and sources of discrepancies were reevaluated and corrections made to or ambiguities removed from category definitions. Table 2 provides a summary of the study characteristics for the samples of couples and individuals, respectively.

TABLE 2: Study Characteristics

		<i>Couples</i>	<i>Individuals</i>
<i>n</i>	Samples	19	43
	Studies	17	40
Source of data	Journal article	10	33
	Book or book chapter	1	0
	Dissertation	4	5
	Other unpublished source	2	2
Measurement	CTS or modified CTS	16	36
	Frequency of assaults	1	0
	Hit the other	0	1
	Physical aggression	0	3
Country	USA	14	34
	Canada	1	3
	UK	1	2
	New Zealand	1	1
Age category	14-18	0	5
	19-22	3	23
	23-30	3	0
	31-37	6	0
	38-49	1	0
	Wide range or not specified	4	12
Sample	High school	0	4
	College students	3	29
	Community or from military base	7	7
	Treatment program	3	0
	Men receiving counseling for violence	1	0
	Couples referred for husband's violence	3	0
Marital status	Married or cohabiting	14	5 ^b
	Not cohabiting	3	35 ^b
	Mixture	0	1
Level of measurement	Nominal	6	34 ^a
	Interval	11	7 ^a
Reference period	Current or most recent relationship	6	15
	Past year	10	13
	Present and past relationships	1	10
	Past 2 years	0	1
	Past 6 months	0	1

a. One study provided both levels of measurement.

b. One study involved both married and unmarried samples.

The eight sets of study characteristics were used in categorical model analyses, which were carried out to further investigate the sources of heterogeneity in effect sizes within the data sets. In each case, a mean weighted d value for each class was calculated together with the statistic Q_b for a between-classes comparison. Again, calculations were carried out using D-STAT software.

RESULTS

Spousal Concordance Rates

Only three studies provide assessments of spousal concordance for the occurrence or nonoccurrence of CTS acts in the form of Cohen's kappa, which takes account of chance levels of agreement.⁷ Jouriles and O'Leary (1985) reported overall values of .43 for husbands' behavior and .40 for wives' behavior in a sample of 107 couples (a therapy sample and a community sample combined). O'Leary and Arias (1988) found overall values of .36 for husbands' physical aggression and .32 for wives' physical aggression using a sample of 369 couples who were about to marry. Moffitt et al. (1997) found overall values of .46 for husbands' physical aggression and .34 for wives' physical aggression using a sample of 369 couples.

Three studies have reported agreement for individual CTS items. Jouriles and O'Leary (1985) reported values of .15 to 1.00 for specific items (from two samples of 65 and 37 couples).⁸ Schwartz (1994) found values of .33 to .42 for four of the items for men and a value of around 0 for the fifth item using data from 88 couples. Women's values for the same items were lower, from .10 to .22. Moffitt et al. (1997) found values of between .03 and .49 for men and .08 to .58 for women (from a sample of 369 couples).

Correlations

Where there is interval-level data for both self-reports and partners' reports, it is possible to assess whether self-reports and partners' reports show acceptable correlations. As indicated earlier, this assesses overall consistency (but not systematically higher or lower ratings). Table 3 summarizes the six studies that have reported Pearson's correlations. Four are small-scale studies that total 86 couples altogether, and the other two involve 180 and 350 couples, respectively. For men's aggression, the weighted mean $r = .55$ ($d =$

TABLE 3: Summary of Studies Providing Data on Correlations Between Male and Female Partners' Reports

Study	Measure	n	r		Variable (study characteristic) ^a						
			Men's	Women's	1	2	3	4	5	6	7
			Aggression	Aggression							
Browning and Dutton (1986)	Aggregated frequency scores (CTS)	30	.65	.26	1	1	2	1	1	1	2
Archer and Ray (1989)	Mean weighted frequency (CTS)	23	.63	.83	1	1	3	2	2	2	1
Brennan (1990)	Mean weighted frequency (CTS)	23	.49	.12	3	1	3	2	2	2	1
Cantos, Neidig, & O'Leary (1994)	Overall level (CTS)	180	.50	.51	1	1	1	3	3	1	1
Hanley (1996)	Any item (CTS)	10	.33	.09	2	1	2	2	2	2	2
Moffitt et al. (1997)	Aggregated CTS score	350	.58	.54	1	1	2	4	4	2	2

NOTE: CTS = Conflict Tactics Scales.

a. Variable 1 (source of data: 1 = journal article, 2 = dissertation, 3 = other unpublished source), variable 2 (measurement instrument: 1 = Conflict Tactics Scales or modified or earlier version of it), variable 3 (country: 1 = United States, 2 = Canada, 3 = United Kingdom, 4 = New Zealand), variable 4 (age category in means: 1 = 14-18, 2 = 19-22, 3 = 23-30), variable 5 (type of sample: 1 = couples referred to treatment program for assaultive husbands, 2 = college students, 3 = couples referred to treatment program for marital violence or marriage counseling, 4 = community sample), variable 6 (majority marital status: 1 = married or cohabiting, 2 = unmarried and not cohabiting), variable 7 (reference period: 1 = current or most recent relationship, 2 = over the past year).

1.33, confidence interval [CI] = 1.20/1.45, $p < .0001$). The six studies in the sample showed a homogeneous set of correlations ($Q_w[5] = 8.32$, ns).

For women's aggression, the weighted mean $r = .51$ ($d = 1.18$, $CI = 1.05/1.30$, $p < .0001$). The six studies in the sample showed a heterogeneous set of correlations ($Q_w[5] = 39.26$, $p < .0001$). It required the removal of four outliers—comprising all the small sample studies—to form a homogeneous set consisting of the two larger sample studies, which increases the overall r to .53 ($d = 1.25$, $CI = 1.11/1.38$, $p < .0001$).

Study Characteristics for Main Meta-Analyses of Effect Sizes

Table 2 provides a summary of the study characteristics for the samples of couples and individuals, respectively. The set of couples and of individuals

TABLE 4: Summary of Studies Providing Data on the Overall Discrepancy Between Male and Female Partners' Reports of Men's and of Women's Physical Aggression Expressed As *g* for Couples

Study	n	<i>g</i>		Variable (study characteristic) ^a							
		Men	Women	1	2	3	4	5	6	7	8
Winkler (1981)	29	-.02	-.03	3	1	4	2	3	1	2	1
	26	-.01	-.30	3	1	5	2	3	1	2	1
	8	-.12	.26	3	1	1	2	3	1	2	1
Szinovacz (1983)	103	.00	-.22	1	1	1	9	3	1	1	2
Browning & Dutton (1986)	30	.94	.51	1	1	2	3	5	1	2	2
Okun (1986)	44	.32		2	2	1	2	4	1	2	1
Margolin (1987)	103	-.23	.15	1	1	1	9	3	1	1	2
Marshall (1987b)	15	-.35	-.48	4	1	1	9	2	1	1	3
Brennan (1990)	23	.25	-.10	4	1	3	1	2	2	2	1
Lagrande (1990)	12/11	.73	-.06	3	1	1	3	5	1	2	1
Rollins & Oheneba-Sakyi (1990)	1471	.38	.62	1	1	1	9	3	1	1	2
Maisel (1991)	82	.02	.08	3	1	1	4	3	1	2	2
Cascardi, Langhinrichsen, & Vivial (1992)	93	-.02	.24	1	1	1	3	1	1	1	2
Follette & Alexander (1992)	100	.25	-.23	1	1	1	1	2	2	2	1
Langhinrichsen-Rohling & Vivian (1994)	97	.39	.27	1	1	1	3	1	1	2	2
Bohannon, Dosser, & Lindley (1995)	94	-.28	.06	1	1	1	2	3	1	1	2
Greening (1995)	80	.06	-.77	3	1	1	3	5	1	2	2
Lawrence, Heyman, & O'Leary (1995)	50	.99	.04 ^b	1	1	1	3	1	1	2	1
Moffitt et al. (1997)	350	.25	.18	1	1	6	1	3	2	2	2

NOTE: A positive result indicates that partners' ratings are higher than self-ratings.

a. Variable 1 (source of data: 1 = journal article, 2 = book or book chapter, 3 = dissertation, 4 = other unpublished source), variable 2 (measurement instrument: 1 = Conflict Tactics Scales or modified or earlier version of it, 2 = frequency of assaults), variable 3 (country or ethnic group: 1 = United States, 2 = Canada, 3 = United Kingdom, 4 = Israeli citizens in the United States, 5 = Jewish sample from the United States, 6 = New Zealand), variable 4 (age category in means: 1 = 19-22, 2 = 23-30, 3 = 31-37, 4 = 38-49, 9 = wider age groups or not specified), variable 5 (type of sample: 1 = couples referred to treatment program for marital violence or marriage, 2 = college students, 3 = community or from military base, 4 = sample of men receiving counseling for their spousal violence, 5 = couples referred to treatment program for assaultive husbands), variable 6 (majority marital status: 1 = married or cohabiting, 2 = unmarried and not cohabiting, 9 = mixture), variable 7 (level of measurement: 1 = nominal, frequency or proportions showing one or more acts, 2 = interval, usually from a composite of a frequency scale applied to each act), variable 8 (reference period: 1 = current or most recent relationship, 2 = over the past year, 3 = in present and past relationships).

b. Approximate values calculated from *F* values.

TABLE 5: Meta-Analyses of Studies Summarized in Table 4 to Show the Differences Between Self-Reports and Partners' Reports of Men's and of Women's Physical Aggression Expressed as d Values for Couples

	d (weighted) ^a	CI	p	k	Q_w	p
Men's aggression	.27	.22/.32	<.0001	19	71.28	<.0001
Outliers removed ^b	.13	.05/.21	.002	15	20.4	.12
Women's aggression	.34	.28/.39	<.0001	18	163.1	<.0001
Outliers removed ^c	.08	.00/.16	.06	16	21.1	.13

NOTE: A positive result indicates that partners' ratings are higher than self-ratings; CI = confidence interval; d = mean effect size; k = number of samples included in the analysis; Q_w = homogeneity of effect sizes.

a. Weighted by one for each study in the case of *not weighted* and by the reciprocal of the variance in the case of *weighted*.

b. The following outliers were removed: Rollins and Oheneba-Sakyi (1990); Browning and Dutton (1986); Lawrence, Heyman, & O'Leary (1995); Bohannon, Dosser, & Lindley (1995).

c. The following outliers were removed: Rollins and Oheneba-Sakyi (1990); Greening (1995).

mainly consist of studies published in journals and were conducted in the United States using the CTS, although in other respects the two sets differed. As expected, those samples involving matched couples consisted of people who were married or cohabiting whereas those involving individuals consisted of people who were neither married nor cohabiting, were younger, and were mostly college students. The studies of couples more often involved interval-level data, whereas those of individuals generally involved nominal data.

Differences Between Ratings for Couples

Table 4 summarizes studies of couples used to calculate g values for the two partners' ratings of men's and women's aggression. Table 5 shows the results of the meta-analyses carried out on the g values shown in Table 4 for men's and women's aggression, respectively. Both sexes showed positive overall d values, which indicate that partners' ratings are higher than self-ratings. Comparison of the mean weighted d values for men and women showed no significant difference ($Q_b[1] = 3.29, p = .07$), although values for women were slightly higher than values for men.

For both sexes, the test for the homogeneity of effect sizes (Hedges & Olkin, 1985) indicated that the samples of studies were heterogeneous. It required the removal of four outliers for the men and two outliers for the

women to produce a nonsignificant Q_w value. The resulting d values were much lower in each case, with the value for females being lowered by a greater amount.⁹ The male value was still significantly different from zero, whereas the female value was not (see Table 4).

Categorical Analyses for Differences Between Ratings for Couples

The results of the categorical analyses are presented in this section, but it should be noted that in several cases, the small sample size is likely to render the analyses unreliable. For men and women, there were significant differences according to whether the study was published.¹⁰ For men's aggression, the 11 published studies (mostly journal articles) indicated that partners' reports were significantly higher than self-reports ($d = .29$, $CI = .24/.35$), whereas the 8 unpublished studies (mostly dissertations) showed a d value of .05 ($CI = -.12/.22$; i.e., no significant disagreement). The Q_b value for this comparison was 7.4 ($p = .0075$). For women's aggression, there was an even more striking difference. In published studies, partners' reports were again substantially and significantly higher than self-reports ($d = .40$, $CI = .35/.46$, $k = 10$), whereas in unpublished studies, self-reports were significantly higher than partners' reports ($d = -.25$, $CI = -.42/-.08$, $k = 8$). The Q_b value for this comparison was 50.88 ($p < .0001$). It should be noted that most (6 of 8) unpublished studies involved small samples (<50), whereas most (9 of 11) published sources involved larger samples (>50). The significance of this is that smaller samples tend to produce less reliable findings.

Comparisons that involved nationalities, samples, marital status, source of data, or time period of the assessment indicated no significant differences for men's aggression. There was a significant difference for men between age categories, which indicated more agreement for the 23- to 30-year-old category ($d = -.07$, $CI = -.27/.13$, $k = 5$) than the 19- to 22-year-old ($d = .25$, $CI = .12/.38$, $k = 3$) and 31- to 49-year-old categories ($d = .27$, $CI = .14/.40$, $k = 7$). The overall Q_b was 8.89 ($p = .01$), with the 23- to 30-year-old category being significantly different from the other two age categories (19-22: $Z^2 = 7.1$, post hoc $p = .03$; 31-49: $Z^2 = 7.91$, post hoc $p = .02$).

For women's aggression, there were several significant differences, but again it should be noted that the small sample size in several cases is likely to render the analyses unreliable. There was significantly more disagreement ($Q_b = 10.08$, $p = .0015$) in studies carried out in the United States ($d = .38$, $CI = .32/.44$, $k = 13$) than in those from other countries ($d = .15$, $CI = .02/.28$, $k = 5$). For community samples, there was a substantial effect size in the direction of

higher values for partners' reports than self-reports ($d = .42$, $CI = .36/.48$, $k = 9$), whereas for college samples, there was an effect size in the opposite direction ($d = -.23$, $CI = -.47/.01$, $k = 3$), a difference that was highly significant ($Z^2 = 27.6$, post hoc $p < .0001$). Similarly, three samples of assaultive husbands showed higher values for self-reports than for partners' reports ($d = -.38$, $CI = -.64/-.12$, $k = 3$), which was again significantly different from the value for community samples ($Z^2 = 35.0$, post hoc $p < .0001$). There was more agreement for values calculated from interval data ($d = .035$, $CI = -.07/.12$, $k = 12$) than from nominal data ($d = .49$, $CI = .42/.55$, $k = 6$), a difference that was highly significant ($Q_b = 63.31$, $p < .0001$). There was also more agreement for values based on the current relationship ($d = -.12$, $CI = -.30/.05$, $k = 7$) than for those based on relationships over the past year ($d = .39$, $CI = .33/.44$, $k = 10$), a difference that was highly significant ($Q_b = 29.1$, $p < .0001$).

Differences Between Ratings for Individuals

Table 6 summarizes studies used to calculate g values for individuals rating their own aggression and their partners' aggression in samples not consisting of couples. Again, ratings of men's and women's aggression are shown separately. Table 7 shows the results of the meta-analyses of these studies for men's and women's aggression, respectively.

Both sexes showed low positive overall d values, which indicated that partners' ratings were slightly higher than self-ratings. Comparison of the mean weighted d values for men and women showed a significant difference ($Q_b[1] = 9.49$, $p = .002$), which indicated that the difference between the partners' ratings was greater for men than for women. For both sexes, the samples of studies were heterogeneous. The removal of a small number of outliers to produce a nonsignificant Q_w value did not change the overall d for the men, but it reduced it to a value not significantly different from zero for the women. The outliers for the men included studies in which self-ratings were higher than partners' ratings, whereas those for the women contained values representing higher partners' ratings than self-ratings. Consistent with this analysis, the mean d values not weighted for sample size were not much different from the weighted values for the men but were near zero for the women.

Categorical Analyses for Differences Between Ratings for Individuals

Table 8 shows the results of categorical analyses of the studies, which enabled comparisons between self-ratings and partners' ratings for men's and

TABLE 6: Summary of Studies Providing Data on the Differences Between Male and Female Partners' Reports of Men's and Women's Physical Aggression Expressed as *g* for Individuals

<i>Study</i>	<i>n</i> ^a	<i>g</i>		<i>Variable (study characteristic)^b</i>							
		<i>Men</i>	<i>Women</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>	<i>8</i>
Nisonoff & Bitman (1979)	112/185	-.08	.21	1	2	1	9	3	1	1	1
Irwin (1980)	55/70	.33	-.06	3	1	1	9	3	1	2	1
Bernard & Bernard (1983)	168/293	.52	-.05	1	3	1	9	2	2	1	3
Makepeace (1983)	97/146	-.10	.02	1	1	1	9	2	2	1	2
Henton, Cate, Koval, Lloyd, & Christopher (1983)	351/293	.44	.11	1	1	1	1	1	2	1	3
Sigelman, Berry, & Wiles (1984)	112/384	-.12	.14	1	1	1	2	2	2	1	3
Billingham & Sack (1986)	167/359	.08	-.09	1	1	1	2	2	2	1	1
Deal & Wampler (1986)	109/287	-.36	.85	1	1	1	2	2	2	1	1
O'Keeffe, Brockopp, & Chew (1986)	121/135	.23	-.17	1	1	1	1	1	2	1	3
Arias, Samois, & O'Leary (1987)	95/175	-.09	.35	1	1	1	1	2	2	1	1
DeMaris (1987)	207/277	-.02	-.05	1	1	1	2	2	2	1	1
Marshall (1987a)	34/44	-.14	-.2	2	1	1	9	2	1	1	1
	106/155	-.48	-.02	2	1	1	9	2	2	1	1
	98/93	-.04	-.46	2	1	1	9	2	2	1	1
Marshall (1987b)	77/108	-.33	-.17	2	1	1	9	2	2	1	3
	60/33	-.21	-.28	2	1	1	9	2	2	1	3
Mason & Blankenship (1987)	48/107	-.04	.02	1	1	1	2	2	9	2	2
Stets & Pirog-Good (1987)	126/206	.17	-.17	1	1	1	9	2	2	2	2
Burke, Stets, & Pirog-Good (1988)	207/298	.07	-.05	1	1	1	9	2	2	1	2
Arias & Johnson (1989)	103/99	.13	.28	1	1	1	2	2	2	1	1
Marshall & Rose (1990)	205/249	-.21	-.32	1	1	1	2	2	2	1	3
Polek (1990)	252/140	.10	.05	3	1	1	2	3	2	1	1
Sawin (1990)	550/645	.20	.00	1	1	1	2	2	2	1	3
Stets & Pirog-Good (1990)	442/335	.25	.13	1	1	1	9	2	2	2	2
Worth, Matthews, & Coleman (1990)	31/78	.03	.14	1	1	1	9	2	2	1	3
Follingstad, Wright, Lloyd, & Sebastian (1991)	207/288	.42	-.12	1	1	1	2	2	2	1	3
Gryl, Stith, & Bird (1991)	124/156	.1	.2	1	1	1	2	2	2	1	1
Stets & Henderson (1991)	146/125	.18	-.16	1	1	1	2	3	2	1	2
Thompson (1991)	167/169	.11	-.02	1	1	1	2	2	2	1	4
White & Koss (1991)	2105/2602	-.09	.07	1	1	1	2	2	2	1	2
Pedersen & Thomas (1992)	50/116	.07	.1	1	1	2	2	2	2	1	1
Billingham & Notebaert (1993)	456/834	.06	.03	1	1	1	2	2	2	2	5
Howell (1993)	84/188	.08	.39	3	1	1	2	2	2	1	2
Riggs (1993)	262/391	.16	-.19	1	1	1	2	2	2	1	1
Clark, Beckett, Wells, & Dungee-Anderson (1994)	76/235	-.04	-.12	1	1	1	2	2	2	1	2
Lejeune & Follette (1994)	271/194	.31	.24	1	1	1	2	2	2	2	1

TABLE 6 Continued

Study	n ^a	g		Variable (study characteristic) ^b							
		Men	Women	1	2	3	4	5	6	7	8
Shovlin (1994)	122/225	-.02	.06	3	3	3	2	2	2	1	1
		.07	.03	3	3	3	2	2	2	2	1
Gagne & Lavoie (1995)	56/45	.37	-.29	1	1	2	1	1	2	1	2
Leblanc (1995)	126/302	.11	-.26	3	1	2	2	2	2	1	2
Morse (1995)	177/200	-.03	-.17	1	1	1	9	3	2	1	2
Carrado, George, Loxam, Jones, & Templar (1996)	894/971	.09	.2	1	1	3	9	3	1	1	1
Foshee (1996)	841/854	.51	.25	1	3	1	1	1	2	1	3
Magdol, Moffitt, Caspi, Newman, Fagan, & Silva (1997)	425/436	.12	-.06	1	1	4	2	3	2	1	2

NOTE: A positive result indicates that partners' ratings are higher than self-ratings.

a. The first figure is for the males and the second is for females.

b. Variable 1 (source of data: 1 = journal article, 2 = other unpublished source, 3 = dissertation), variable 2 (measurement instrument: 1 = Conflict Tactics Scales or modified or earlier version of it, 2 = "hit the other," 3 = "physical"), variable 3 (country: 1 = United States, 2 = Canada, 3 = United Kingdom, 4 = New Zealand), variable 4 (age category in means: 1 = 14-18, 2 = 19-22, 9 = wider age groups or not specified), variable 5 (type of sample: 1 = high school students, 2 = college students, 3 = community or from military base), variable 6 (majority marital status: 1 = married or cohabiting, 2 = unmarried and not cohabiting, 9 = mixture), variable 7 (level of measurement: 1 = nominal, frequency or proportions showing one or more acts, 2 = interval, usually from a composite of a frequency scale applied to each act), variable 8 (reference period: 1 = current or most recent relationship, 2 = over the past year, 3 = in present and past relationships, 4 = over the past 2 years, 5 = over the past 6 months).

for women's aggression. Generally, where differences across categories were found, they were larger for men's aggression than women's aggression. One odd finding was that the direction of effect size was reversed in dissertation studies that remained unpublished. This occurred for men's and for women's aggression. Thus, in these five studies, self-ratings of aggression were higher than those of the recipients. In contrast to the same trend for data on couples, these unpublished studies did not tend to involve small samples.

Although most studies involved the CTS, there was evidence of larger discrepancies among the four studies that used other assessments, which principally asked about fights or physical aggression. When the four studies involving high school students were separated, it was clear that they involved a much larger estimate of men's aggression by partners' ratings than by self-ratings. The trend for women's aggression was in the same direction and produced a significant but smaller difference. The result of removing the high

TABLE 7: Meta-Analyses of Studies Summarized in Table 4 to Show the Differences Between Self-Reports and Partners' Reports of Men's and of Women's Physical Aggression Expressed As d Values for Samples of Individuals

	d^a		CI	p	k	Qw	p
	Not Weighted	Weighted					
Men's							
aggression	.07	.09	.06/.12	<.0001	43	254.3	<.0001
Outliers removed ^b	.06	.09	.06/.12	<.0001	35	47.8	.11
Women's							
aggression	.01	.05	.02/.08	.0001	43	182.0	<.0001
Outliers removed ^c	-.01	.02	-.01/.05	.16	34	47.9	.08

NOTE: A positive result indicates that partners' ratings are higher than self-ratings; CI = confidence interval; d = mean effect size; k = number of samples included in the analysis; Q_w = homogeneity of effect sizes.

a. Weighted by one for each study in the case of *not weighted* and by the reciprocal of the variance in the case of *weighted*.

b. The following outliers were removed: Marshall and Rose (1990); Follingstad, Wright, Lloyd, & Sebastian (1991); Deal and Wampler (1986); Henton, Cate, Koval, Lloyd, & Christopher (1983); Bernard and Bernard (1983); Marshall (1987a, study 2); White and Koss (1991).

c. The following outliers were removed: Riggs (1993); Leblanc (1995); Arias, Samios, & O'Leary (1987); Howell (1993); Marshall (1987a, study 3); Marshall and Rose (1990); Carrado, George, Loxam, Jones, & Templar (1996); Foshee (1996); Deal and Wampler (1986).

school samples was to produce much closer estimates of both men's and women's aggression by perpetrators and recipients.

For men, the degree of discrepancy was similar for both married (or cohabiting) and single samples, whereas for women, the much smaller number of married samples produced a greater discrepancy (i.e., higher estimates by recipients than perpetrators). There was some indication that for men, the discrepancy was larger for interval data than nominal data, which simply records whether one or more than one act has occurred. However, this was not the case for women. For men, a greater discrepancy occurred when the sample were asked about all relationships; asking about the current relationship or a specified time period produced a smaller overall d value. This was not found for women's aggression, where there was greater discrepancy for the current relationship. Variables that did not affect the magnitude of the discrepancies were the country in which the study was located and the age range.

TABLE 8: Categorical Model Analysis of Differences Between Self-Reports and Partners' Reports of Men's and of Women's Physical Aggression Expressed As *d* Values for Samples of Individuals

	<i>Men's Aggression</i>				<i>Women's Aggression</i>			
	<i>d</i>	<i>CI</i>	Q_w	<i>k</i>	<i>d</i>	<i>CI</i>	Q_w	<i>k</i>
Journal articles	.10	.08/.13	219.5***	33	.06	.03/.09	148.1***	33
Dissertations	-.27	-.41/-.13	5.8	5	-.20	-.35/-.06	5.3	5
Other ^a	.09	-.01/.20	2.7	5	.02	-.09/.12	15.1**	5
Q_B (2)	26.37***				13.6***			
CTS	.05	.02/.08	152.1***	39	.03	.00/.06	160.9***	39
Other measures ^b	.39	.31/.46	36.3***	4	.18	.10/.25	8.9	4
Q_B (1)	65.9***				12.3***			
North America	.09	.06/.12	253.2***	40	.04	.01/.07	168.5***	40
Other countries ^c	.09	.16/.04	1.1	3	.11	.04/.18	10.1*	3
Q_B (1)	.00				3.4			
14-22 years	.09	.06/.12	191.3***	28	.05	.02/.08	139.6***	28
Wide age range	.08	.02/.13	62.7***	15	.04	-.01/.09	42.2***	15
Q_B (1)	.35				.29			
High school	.46	.38/.54	4.7	4	.16	.08/.23	15.5**	4
Adults ^d	.04	.01/.07	147.9***	39	.04	.01/.06	157.7**	39
Q_B (1)	101.7***				8.7**			
Married	.07	-.01/.16	4.6	4	.17	.09/.26	4.7	4
Single	.09	.06/.12	249.0***	38	.04	.02/.08	182.0***	38
Q_B (1)	.72				10.1**			
Nominal data	.08	.05/.11	240.4***	37	.05	.02/.08	172.5	37
Interval data	.16	.09/.24	9.1	6	.06	-.01/.13	9.4	4
Q_B (1)	4.8*				.12			
Current Specified	.04	-.01/.09	47.8***	17	.11	.06/.16	92.8***	17
period (2 years)	.02	-.02/.05	36.4**	15	.02	-.02/.05	33.6**	15
All relationships	.28	.23/.33	99.4***	11	.04	-.01/.09	46.0***	11
Q_B (2)	70.7***				9.5**			

NOTE: A positive result indicates that partners' ratings are higher than self-ratings; CI = confidence interval; *d* = mean effect size weighed by sample size; *k* = number of samples included in the analysis; Q_w = homogeneity of effect sizes; Q_B = difference between contrasted categories. a. Conference presentations.

b. Hitting the other or physical aggression, but not using the CTS.

c. United Kingdom and New Zealand.

d. Community samples (*k* = 7) and college students (*k* = 32). These two groups showed nonsignificant differences ($Z^2 = 2.8$).

p* < .05. *p* < .01. ****p* < .001.

DISCUSSION

The few available studies reporting concordance rates found that most overall values were between .36 and .46 for men's aggression and between .32 and .40 for women's aggression. Kappas calculated from specific items indicated variability and considerable disagreement in some cases. However, several qualifications need to be made about any conclusions drawn from existing evidence. First, the sample of studies was small. Second, kappa is usually used to assess the agreement of trained observers or raters (Bakeman & Gottman, 1986) so that we might expect considerably more disagreement when partners are rating their own behavior. Indeed, under these circumstances, agreements of around .40 or above can be considered reasonable. Third, the correction for chance involved in the calculation of kappa is very restrictive when the base rates are low, as they are for some CTS items, and on some occasions for the overall value. For example, if there is perfect agreement that no aggression has occurred, kappa is zero (O'Leary & Arias, 1988).

The correlational evidence indicated that for both sexes, partners' estimates are positively correlated with one another to a reasonable extent, and there is considerable consistency across studies. The values were less than that considered acceptable in most behavioral studies involving interrater agreement (Martin & Bateson, 1993); although, as remarked above, it is doubtful whether the same criterion should be applied to partner agreements and to ratings from trained observers or coders. Indeed, the overall values were comparable with those obtained for spouses' ratings of their partners' personality (McCrae & Costa, 1990). A further point raised by Moffitt et al. (1997) is that Pearson product-moment correlations underestimate the true correlations between the two sets of reports since they are attenuated by measurement error. In their study, latent correlations were calculated using confirmatory factor analysis. These showed substantially larger values than the product-moment correlations (.83 for men and .71 for women). A tentative conclusion from the existing correlational evidence is that agreement based on the frequency or number of acts is greater than expected, especially for men.

As it was pointed out earlier, concordance rates and correlations do not provide any indication of the relative levels of aggression reported by the two partners. They would not detect systematic underreporting or overreporting. This can be illustrated by the study of Browning and Dutton (1986), who found high correlations yet substantially different magnitudes reported by the two raters (see Table 4).

Among samples of couples, partners rated both men's and women's incidences of physical aggression higher than self-raters did, and the extent of

this difference was little different for men than for women overall. This supported the second prediction that people of both sexes tend to underreport their own aggression. However, the discrepancy was considerably reduced by the removal of four outliers for the men and two outliers for the women, the resulting value for women being greatly reduced so that it was not significantly different from zero across 16 studies (see Table 5). When men's and women's values were compared, the extent of the disagreement was still not significantly different.

We should note that the data on couples gave little or no support to Hypothesis 1 (only men underreport their own aggression; see Table 1). This is surprising in view of the finding that in more than 75 to 81 studies, the d value for the sex difference in physical aggression was significantly higher in females for self-reports than for partners' reports (Archer, in press), indicating a pattern more consistent with greater underreporting by males than females. We should note, however, that the data on couples analyzed here were a small subset (17 or 18) of this larger sample.

Categorical analysis, although limited by the number of studies available, can be used to identify those studies in which disagreement is more likely. For women, values based on interval data rather than nominal data and on the current relationship rather than relationships over the past year yielded lower discrepancies. For community samples, women's self-reports were considerably lower than partners' reports; whereas in college samples and those involving assaultive husbands, self-reports tended to be higher than partners' reports. Because this was the case for women's aggression only, it is consistent with Hypothesis 3 (men underreport their victimization.) We should note that this pattern only occurred in two specific samples, among men who have themselves been labeled as assaultive and among college students. At present, the empirical base is small and should be treated with caution.

Most of the available evidence came from samples of men and women who were not selected as couples. Therefore, self-ratings and partners' ratings did not refer to the same individuals. We would expect differences derived from unmatched samples to involve more random variations in the values than those from couples but not to involve any additional systematic bias. In fact, the overall weighted d value for the differences between partners' and self-ratings was considerably smaller than that for couples, although it was in the same direction (see Table 7). The d value was significantly larger for the men than for the women in this case, providing some evidence consistent with the first hypothesis. As indicated above, this is more consistent with the finding that the sex difference in aggression is significantly higher in the female direction for self reports than for partners' reports (Archer, in press). However, there was also evidence of a lower level of bias for reports of

women's aggression, indicating some support for Hypothesis 2 in addition to Hypothesis 1. For men, the same pattern was maintained when outliers were removed, whereas for women, there was no difference between self-reports and partners' reports when outliers were removed. We can, therefore, conclude that the evidence for Hypothesis 1 is more robust than that for Hypothesis 2.

Categorical analysis of individuals indicated that there was considerably more agreement in the larger number of studies using the CTS than in those using measures such as asking about fights or physical aggression, especially for the men. This does not support the view that the CTS themselves are the source of unreliability. In fact, the findings suggest that they are better than the less commonly used alternatives. If the minority of studies of dating high school students is removed, the overall effect size is considerably reduced to a more pronounced extent for men's aggression. These two analyses involved an overlap of only one study. Therefore, there would be much greater agreement if samples of high school pupils and studies using measures other than the CTS were omitted. The combined effect of doing so is to reduce the discrepancy for estimates of men's aggression to a greater extent than that for women's aggression. For men, asking about all past and present relationships led to a greater discrepancy between partners' reports and self-reports than did asking about the current relationship or specifying a period of time in the recent past.

The overall conclusion from the two analyses of couples and individuals is that in both cases, there was some support for Hypothesis 1 (self-reports tend to be lower than partners' reports for both sexes.) Among the couples, there was little or no evidence that this occurred to a greater extent for men than for women (Hypothesis 2), but this was found to be the case for the larger sample of individuals. Categorical analyses indicated variation among subsamples such that the discrepancies were higher for some cases and reversed in others. Among two of the samples of couples, self-reports were higher than partners' reports for women's aggression only, a finding consistent with men underreporting their victimization (Hypothesis 3; see Table 1). Agreement for women's aggression among couples was also significantly better when interval-level data rather than nominal-level data were used to calculate the effect sizes and when current relationships were involved as compared to those over the previous year. Both are variables which we should expect to be associated with greater accuracy of reporting. For the samples of individuals, agreement was greater when samples of high school pupils and studies using measures other than the CTS were omitted and when measures were taken from current relationships only.

Combining the limited correlational data with the analyses of differences suggests that some systematic underreporting occurs in self-reports by both

sexes (Hypothesis 2) but that this is greater for men among the larger number of studies where the men and the women did not come from matched couples (Hypothesis 1). This underreporting varies in magnitude so that there are identifiable circumstances in which agreement is more likely and even some where self-reports produce higher values than partners' reports.

The finding that ratings of men's aggression is lower for partners' reports than for self-reports is not surprising in view of the widespread social disapproval of men striking women, which was referred to earlier. The occurrence of the same pattern for women's aggression was consistent with the view that women's aggression is socially undesirable because it conflicts with feminine gender role expectations.

A different way of approaching the issue of reliability is to investigate the association between measures on the CTS and the person's general degree of social desirability when responding to a questionnaire. Sugarman and Hotaling (1997) meta-analyzed seven studies reporting 18 correlations between measures of social desirability and CTS frequency or occurrence scores. Overall, there was a small but significant negative correlation ($r = -.18$). Social desirability of responding was associated with lower CTS scores. This analysis involved measures for men and women and for perpetrators and recipients. Consistent with the present findings for couples, there was no difference between the magnitude of the correlation for men and for women. However, as expected, the association was higher for perpetrators ($r = -.19$) than for recipients ($r = -.12$). These data suggest that even recipients may underreport their victimization, a possibility that could not be investigated using the present data. Thus, although partners may be more accurate than perpetrators, they, too, could be underreporting the real extent of physical aggression.

Riggs, Murphy, and O'Leary (1989) asked people to rate the likelihood that they would report items on the CTS. Not surprisingly, they were more willing to report partners' acts of physical aggression than their own, and the willingness to report their own acts decreased with their severity. This finding is consistent with the observation that the largest negative correlation between social desirability and physical aggression occurred in a study measuring more severe forms (Sugarman & Hotaling, 1997). Riggs et al. also found that men and women did not differ in their willingness to report most CTS acts. Both this study and the meta-analysis of social desirability support the conclusion from the present analysis that both men's and women's self-ratings are likely to be lower than ratings by partners.

This analysis concerned one issue raised by critics of the CTS: the extent of agreement between partners or between men and women from the same sample. However, it is recognized that the use of the CTS raises other issues

that these meta-analyses do not address. One concerns the different meanings of the same acts for men and for women. Physically aggressive acts may have more serious physical and psychological effects when carried out by a man than a woman (Dobash et al., 1992, 1998; Dutton, 1994; Nazroo, 1995; Rhodes, 1992; Romkens, 1997). Nazroo (1995) found more acts of undefendable, intimidating, and dangerous physical aggression among men than women, although an assessment of all acts of physical aggression showed a higher frequency among women than men. This supports the criticism that the CTS measures do not reflect the meaning of the aggressive act for the recipient. Marshall (1994), who used a more comprehensive list of physically aggressive acts than those on the CTS, found evidence that certain acts omitted from the CTS (such as being pinned or held down, shaking, and spanking) were more commonly admitted to by men than women, even though CTS acts were used more by women than men. This study also found that distress among the victims of marital abuse was often most closely related to subtle forms of behavior by the partner as opposed to more obviously harmful acts. However, a further study (Marshall, 1992a, 1992b) involving men and women rating the perceived impact of physical aggression found only a little support for men's and women's acts being perceived in substantially different ways.

These further investigations suggest that future studies should involve additional physically aggressive acts to those on the CTS and include ratings of their consequences and meanings for recipients. However, although the CTS physical aggression measure may have its limitations, data from perpetrator ratings and victim ratings indicate that there is a greater measure of agreement than past critics have suggested.

NOTES

1. Although slight modifications have been made to the Conflict Tactics Scales (CTS) violence scale over the years, the items used in most studies are very similar (see Hertzberger, 1991; Straus, Hamby, Boney-McCoy, & Sugarman, 1996).

2. Based on interval-level data.

3. Based on the occurrence or nonoccurrence of any CTS act for men and women.

4. Theses were mainly specific items found on the CTS such as "hit the other" or a general phrase such as "physical aggression."

5. BIDS is a British electronic information system providing access to ISI databases.

6. *d* provides an estimate of effect size corrected for bias, whereas *g* is the biased estimate (see Hedges & Becker, 1986).

7. Although a number of studies provide information from the participant and the recipient concerning the use or nonuse of any CTS act, these figures cannot be used for calculating kappa because this requires the extent of agreement and disagreement for the presence and absence of the variable concerned.

8. Ignoring values of zero for some infrequent items.
9. The male and female values were not significantly different ($Q_B[1] = .81, p = .37$).
10. This is a standard comparison in meta-analyses to provide one source of information on publication bias.

REFERENCES

- Archer, J. (in press). Sex differences in aggression between heterosexual partners: A meta-analytic review. *Psychological Bulletin*.
- Archer, J., & Ray, N. (1989). Dating violence in the United Kingdom: A preliminary study. *Aggressive Behavior, 15*, 337-343.
- Arias, I., & Johnson, P. (1989). Evaluations of physical aggression among intimate dyads. *Journal of Interpersonal Violence, 4*, 298-307.
- Arias, I., Samios, M., & O'Leary, K. D. (1987). Prevalence and correlates of physical aggression during courtship violence. *Journal of Interpersonal Violence, 2*, 82-90.
- Bakeman, R., & Gottman, J. M. (1986). *Observing Interaction: An Introduction to Sequential Analysis*. Cambridge: Cambridge University Press.
- Bernard, M. L., & Bernard, J. L. (1983). Violent intimacy: The family as a model for love relationships. *Family Relations, 32*, 283-286.
- Billingham, R. E., & Notebaert, N. L. (1993). Divorce and dating violence revisited: multivariate analysis using Straus's Conflict Tactics subscores. *Psychological Reports, 73*, 679-684.
- Billingham, R. E., & Sack, A. R. (1986). Courtship violence and the interactive status of the relationship. *Journal of Adolescence Research, 1*, 315-325.
- Bohannon, J. R., Dossier, D. A., & Lindley, S. E. (1995). Using couple data to determine domestic violence rates: An attempt to replicate previous work. *Violence and Victims, 10*, 133-141.
- Brennan, M. B. (1990). *Violence in the dating relationships of a sample of British students*. Unpublished B.Sc. dissertation, Lancashire Polytechnic, Preston, UK.
- Browning, J., & Dutton, D. (1986). Assessment of wife assault with the Conflict Tactics Scale: Using couple data to quantify the differential reporting effect. *Journal of Marriage and the Family, 48*, 375-379.
- Burke, P. J., Stets, J. E., & Pirog-Good, M. A. (1988). Gender identity, self esteem, and physical and sexual abuse in dating relationships. *Social Psychology Quarterly, 51*, 272-285.
- Buss, A. H., & Durkee, A. (1957). An inventory for assessing different types of hostility. *Journal of Consulting Psychology, 21*, 343-349.
- Buss, A. H., & Perry, M. (1992). The aggression questionnaire. *Journal of Personality and Social Psychology, 63*, 452-459.
- Cantos, A. L., Neidig, P. H., & O'Leary, K. D. (1994). Injuries to women and men in a treatment program for domestic violence. *Journal of Family Violence, 9*, 113-124.
- Carrado, M., George, M. J., Loxam, E., Jones, L., & Templar, D. (1996). Aggression in British heterosexual relationships: A descriptive analysis. *Aggressive Behavior, 22*, 401-415.
- Cascardi, M., Langhinrichsen, T., & Vivian, D. (1992). Marital aggression: Impact, injury, and health correlates for husbands and wives. *Archives of Internal Medicine, 152*, 1178-1184.
- Clark, M. L., Beckett, J., Wells, M., & Dungee-Anderson, D. (1994). Courtship violence among African American college students. *Journal of Black Psychology, 20*, 264-281.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

- Deal, J. E., & Wampler, K. S. (1986). Dating violence: the primacy of previous experience. *Journal of Personal and Social Relationships*, 3, 457-471.
- DeMaris, A. (1987). The efficacy of a spouse abuse model in accounting for courtship violence. *Journal of Family Issues*, 8, 291-305.
- Dobash, R. E., & Dobash, R. P. (1977-1978). Wives: The "appropriate" victims of marital violence. *Victimology: An International Journal*, 2, 426-442.
- Dobash, R. E., & Dobash, R. P. (1980). *Violence Against Wives: A Case Against the Patriarchy*. London: Open Books.
- Dobash, R. P., Dobash, R. E., Cavanagh, K., & Lewis, R. (1998). Separate and interacting realities: A comparison of men's and women's accounts of violence against women. *Violence Against Women*, 4, 382-414.
- Dobash, R. P., Dobash, R. E., Daly, M., & Wilson, M. (1992). The myth of sexual symmetry in marital violence. *Social Problems*, 39, 71-91.
- Dutton, D. G. (1994). Patriarchy and wife assault: The ecological fallacy. *Violence and Victims*, 9, 167-182.
- Follette, V. M., & Alexander, P. C. (1992). Dating violence: current and historical debates. *Behavioral Assessment*, 14, 39-52.
- Follingstad, D. R., Wright, S., Lloyd, S., & Sebastian, J. A. (1991). Sex differences in motivations and effects in dating violence. *Family Relations*, 40, 51-57.
- Foshee, V. A. (1996). Gender differences in adolescent dating prevalence, types and injuries. *Health Education Research*, 11, 275-286.
- Gagne, M-H., & Lavoie, F. (1995). La violence physique et la maltraitance affective dans les fréquentations chez un groupe d'adolescent(e)s. [Physical violence and emotional ill treatment in the relationships of a group of adolescents]. *Canadian Journal of Counselling/Revue Canadienne de Counseling*, 29, 22-36.
- George, M. J. (1994). Riding the donkey backwards: Men as the unacceptable victims of marital violence. *The Journal of Men's Studies*, 3, 137-159.
- Gladue, B. A. (1991). Qualitative and quantitative sex differences in self-reported aggressive behavioral characteristics. *Psychological Reports*, 68, 675-684.
- Greening, D. E. (1995). Affects and cognitions of convicted batterers and their intimate partners (Doctoral dissertation, Wayne State University). *Dissertation Abstracts International*, 57, 114A.
- Gryl, F. E., Stith, S. M., & Bird, G. W. (1991). Close dating relationships among college students: differences by use of violence and by gender. *Journal of Social and Personal Relationships*, 8, 243-264.
- Harris, R. J., & Cook, C. A. (1994). Attributions about spouse abuse: It matters who batterers and victims are. *Sex Roles*, 30, 553-565.
- Hedges, L. V., & Becker, B. J. (1986). Statistical methods in the meta-analysis of research on gender differences. In J. S. Hyde & M. C. Linn (Eds.), *The psychology of gender: Advances through meta-analysis* (pp. 14-50). Baltimore, MD: Johns Hopkins University Press.
- Hedges, L. V., & Olkin, I. (1985). *Statistical Methods for Meta-Analysis*. San Diego, CA: Academic Press.
- Henton, J., Cate, R., Koval, J., Lloyd, S., & Christopher, S. (1983). Romance and violence in dating relationships. *Journal of Family Issues*, 4, 465-482.
- Hertzberger, S. D. (1991). The Conflict Tactics Scales. In D. J. Keyser & R. C. Sweetland (Eds.), *Test Critiques* (Vol. 8). Kansas City, MO: Test Corporation of America.
- Howell, S. B. (1993). Personality correlates of physical violence in dating relationships (self-esteem, shame) (Doctoral dissertation, George Mason University, 1993). *Dissertation Abstracts International*, 54, 1670B.

- Irwin, J. K. (1980). Interpersonal needs orientations and violence in the conjugal dyad (Doctoral dissertation, Hofstra University, 1980). *Dissertation Abstracts International*, 41, 4742B.
- Johnson, B. T. (1989). *Software for the Meta-Analytic Review of Research Literatures*. Hillsdale, NJ: Lawrence Erlbaum.
- Jouriles, E. N., & O'Leary, K. D. (1985). Interspousal reliability of reports of marital violence. *Journal of Consulting and Clinical Psychology*, 53, 419-421.
- Lagrande, S. (1990). An exploratory study investigating the viability of couples' group therapy as a treatment alternative for men court-mandated to treatment for assaultive behaviors toward their intimate partners (spouse abuse) (Doctoral dissertation, The Fielding Institute, Alaska, 1989). *Dissertation Abstracts International*, 51, 434B.
- Langhinrichsen-Rohling, J., & Vivian, D. (1994). The correlates of spouses' incongruent reports of marital aggression. *Journal of Family Violence*, 9, 265-283.
- Lawrence, E., Heyman, R. E., & O'Leary, K. D. (1995). Correspondence between telephone and written assessments of physical violence in marriage. *Behavior Therapy*, 26, 671-680.
- Leblanc, S. D. (1995). *Abuse in dating relationships: An urban/rural comparison*. Unpublished master's thesis, Acadia University.
- Lejeune, C., & Follette, V. (1994). Taking responsibility: Sex differences in reporting dating violence. *Journal of Interpersonal Violence*, 9, 133-140.
- Magdol, L., Moffitt, T. E., Caspi, A., Newman, D. L., Fagan, J., & Silva, P. A. (1997). Gender differences in partner violence in a birth cohort of 21-year-olds: bridging the gap between clinical and epidemiological approaches. *Journal of Consulting and Clinical Psychology*, 65, 68-78.
- Maisel, R. L. (1991). Causal attributions about marital conflicts: their relationship to marital violence and marital distress (Doctoral dissertation, California School of Professional Psychology, Berkeley/Alameda, 1991). *Dissertation Abstracts International*, 52, 3300B.
- Makepeace, J. M. (1983). Life event stress and courtship violence. *Family Relations*, 32, 101-109.
- Margolin, G. (1987). The multiple forms of aggressiveness between marital partners: How do we identify them? *Journal of Marital and Family Therapy*, 13, 77-84.
- Marshall, L. L. (1987a, July). *Gender differences in the prediction of courtship abuse from family of origin violence, anxiety proneness and recent positive and negative stress*. Paper presented at the Family Violence Research Conference, Durham, NH.
- Marshall, L. L. (1987b, May-June). *Partnership abuse: Comparison of two studies using college students*. Paper presented at the Iowa Conference on Personal Relationships, North Texas State University.
- Marshall, L. L. (1992a). Development of the Severity of Violence Against Women Scales. *Journal of Family Violence*, 7, 103-121.
- Marshall, L. L. (1992b). The Severity of Violence Against Men Scales. *Journal of Family Violence*, 7, 189-203.
- Marshall, L. L. (1994). Physical and psychological abuse. In W. R. Cupach & B. H. Spitzberg (Eds.), *The dark side of interpersonal communication* (pp. 281-311). Hillsdale, NJ: Erlbaum.
- Marshall, L. L., & Rose, P. (1990). Premarital violence: The impact of family of origin violence, stress, and reciprocity. *Violence and Victims*, 5, 51-64.
- Martin, P., & Bateson, P. P. G. (1993). *Measuring Behaviour: An Introductory Guide* (2nd ed.). Cambridge: Cambridge University Press.
- Mason, A., & Blankenship, V. (1987). Power and affiliation motivation, stress, and abuse in intimate relationships. *Journal of Personality and Social Psychology*, 52, 203-210.
- McCrae, R. R., & Costa, P. T. (1990). *Personality In Adulthood*. New York: Guilford.

- Moffitt, T. E., Caspi, A., Krueger, R. F., Magdol, L., Margolin, G., Silva, P. A., & Sydney, R. (1997). Do partners agree about abuse in their relationship? A psychometric evaluation in interpartner agreement. *Psychological Assessment*, 9, 47-56.
- Morse, B. J. (1995). Beyond the Conflict Tactics Scale: Assessing gender differences in partner violence. *Violence and Victims*, 10, 251-272.
- Nazroo, J. (1995). Uncovering gender differences in the use of marital violence: The effect of methodology. *Sociology*, 29, 475-494.
- Nisonoff, L., & Bitman, I. (1979). Spouse abuse: Incidence and relationship to selected demographic variables. *Victimology: An International Journal*, 4, 131-140.
- O'Keeffe, N. K., Brockopp, K., & Chew, E. (1986). Teen dating violence. *Social Work*, 31, 465-468.
- Okun, L. (1986). *Woman Abuse: Facts Replacing Myths*. Albany: State University of New York Press.
- O'Leary, K. D., & Arias, I. (1988). Assessing agreement of reports of spouse abuse. In G. T. Hotaling, D. Finkelhor, J. T. Kirkpatrick, & M. A. Straus (Eds.), *Family abuse and its consequences* (pp. 218-227). Newbury Park, CA: Sage.
- Pagelow, M. D. (1984). *Family Violence*. New York: Praeger.
- Pedersen, P., & Thomas, C. D. (1992). Prevalence and correlates of dating violence in a Canadian university sample. *Canadian Journal of Behavioural Science*, 24, 490-501.
- Polek, D. S. (1990). Repetition of the interpersonal behavior patterns of physically aggressive families of origin in young adults' dating behavior patterns (Doctoral dissertation, University of South Carolina, 1990). *Dissertation Abstracts International*, 51, 3143B.
- Rhodes, N. R. (1992). The assessment of spousal abuse: An alternative to the Conflict Tactics Scale. In E. C. Viano (Ed.), *Intimate violence: Interdisciplinary perspectives* (pp. 27-35). Washington, DC: Hemisphere.
- Riggs, D. S. (1993). Relationship problems and dating aggression: A potential treatment target. *Journal of Interpersonal Violence*, 8, 18-35.
- Riggs, D. S., Murphy, C. M., & O'Leary, K. D. (1989). Intentional falsification in reports of interpartner aggression. *Journal of Interpersonal Violence*, 4, 220-232.
- Rollins, B. C., & Oheneba-Sakyi, Y. (1990). Physical violence in Utah households. *Journal of Family Violence*, 5, 301-309.
- Romkens, R. (1997). Prevalence of wife abuse in the Netherlands: Combining quantitative and qualitative methods in survey research. *Journal of Interpersonal Violence*, 12, 99-125.
- Sawin, L. L. (1990). The relationship of personality, sex role attitudes, and childhood family violence to the use of physical violence and other argument strategies by dating partners (Doctoral dissertation, University of Texas at Austin, 1990). *Dissertation Abstracts International*, 51, 5076B.
- Schwartz, H.A.G. (1994). Coercion in the lives of young adults: child abuse, dating violence and willingness to use physical discipline (Doctoral dissertation, University of Iowa). *Dissertation Abstracts International*, 55, 4007A.
- Shovlin, J. M. (1994). The relationship among adult attachment style, physical aggression and psychological abuse in dating relationships (Doctoral dissertation, Old Dominion University). *Dissertation Abstracts International*, 54, 4408B.
- Sigelman, C. K., Berry, C. J., & Wiles, K. A. (1984). Violence in college students' dating relationships. *Journal of Applied Social Psychology*, 5, 530-548.
- Stets, J. E., & Henderson, D. A. (1991). Contextual factors surrounding conflict resolution while dating: results from a national study. *Family Relations*, 40, 29-36.
- Stets, J. E., & Pirog-Good, M. A. (1987). Violence in dating relationships. *Social Psychology Quarterly*, 50, 237-246.

- Stets, J. E., & Pirog-Good, M. A. (1990). Interpersonal control and courtship aggression. *Journal of Social and Personal Relationships*, 7, 371-394.
- Straus, M. A. (1979). Measuring intrafamily conflict and violence: The Conflict Tactics (CT) Scales. *Journal of Marriage and the Family*, 41, 75-88.
- Straus, M. A., Hamby, S. L., Boney-McCoy, S., & Sugarman, D. B. (1996). The revised Conflict Tactics Scales (CTS2): Development and preliminary psychometric data. *Journal of Family Issues*, 17, 283-316.
- Sugarman, C. B., & Hotaling, G. T. (1997). Intimate violence and social desirability: A meta-analytic review. *Journal of Interpersonal Violence*, 12, 275-290.
- Szinovacz, M. E. (1983). Using couple data as a methodological tool: The case of marital abuse. *Journal of Marriage and the Family*, 45, 633-644.
- Thompson, E. H. (1991). The maleness of violence in dating relationships: An appraisal of stereotypes. *Sex Roles*, 24, 261-278.
- Walker, L.E.A. (1989). Psychology and violence against women. *American Psychologist*, 44, 659-702.
- Walker, L.E.A. (1990). Response to Mills and Mould. *American Psychologist*, 45, 676-677.
- White, J. W., & Koss, M. P. (1991). Courtship violence: incidence in a national sample of higher education students. *Violence and Victims*, 6, 247-256.
- White, J. W., & Kowalski, R. M. (1994). Deconstructing the myth of the nonaggressive woman: A feminist analysis. *Psychology of Women Quarterly*, 18, 487-508.
- Winkler, I. R. (1981). Communication styles and marital satisfaction for Israeli and American couples in America (Doctoral dissertation, University of Iowa, 1980). *Dissertation Abstracts International*, 42, 396B.
- Worth, D. M., Matthews, P. A., & Coleman, W. R. (1990). Sex role, group affiliation, family background, and courtship violence in college students. *Journal of College Student Development*, 31, 250-254.

John Archer is a professor of psychology at the University of Central Lancashire, Preston. He was educated at the Universities of Wales (B.Sc., Zoology) and Bristol (Ph.D., Psychology). His earlier research concerned hormones and behavior in animals and the social and developmental psychology of gender. His current interests are human aggression, grief and loss, and evolutionary psychology. He has written more than 100 journal articles and book chapters and a number of edited and authored books, the most recent of which are Male Violence (Routledge, 1994) and The Nature of Grief (Routledge, 1999).