

# Accuracy of Adult Recollections of Childhood Victimization: Part 1. Childhood Physical Abuse

Cathy Spatz Widom and Robin L. Shepard  
The University at Albany, State University of New York

Using data from a study with prospective-cohorts design in which children who were physically abused, sexually abused, or neglected about 20 years ago were followed up along with a matched control group, accuracy of adult recollections of childhood physical abuse was assessed. Two hour in-person interviews were conducted in young adulthood with 1,196 of the original 1,575 participants. Two measures (including the Conflict Tactics Scale) were used to assess histories of childhood physical abuse. Results indicate good discriminant validity and predictive efficiency of the self-report measures, despite substantial underreporting by physically abused respondents. Tests of construct validity reveal shared method variance, with self-report measures predicting self-reported violence and official reports of physical abuse predicting arrests for violence. Findings are discussed in the context of other research on the accuracy of adult recollections of childhood experiences.

Numerous articles, books, and reports have described the short- and long-term effects of childhood victimization. Indeed, over the last decade, there has been a dramatic increase in reports linking childhood abuse to a variety of problem behaviors (Widom, 1994). At the same time, many writers have commented on the limitations of research in the field of child abuse and neglect (Beitchman, Zucker, Hood, daCosta, & Akman, 1991; Briere, 1992; Cicchetti & Rizley, 1981; Milner, 1991; Widom, 1988, 1989d; Wolfe & Mosk, 1983). Among these limitations is the predominance of retrospective studies, in which individuals are typically asked about a history of childhood physical abuse in an interview or on a questionnaire designed to elicit this information retrospectively. It is not surprising that considerable controversy exists about the validity of retrospective reports (Briere & Conte, 1993; Della Femina, Yeager, & Lewis, 1990; Herman & Schatzow, 1987; Kruttschnitt & Dornfeld, 1992; Loftus, 1993; Williams, 1994), raising questions about the accuracy of information because of the nature of the data.

Retrospective reports may suffer from a number of problems (Bradburn, Rips, & Shevell, 1987). There may be distortion and loss of information associated with the recollection of

events from a prior time period, especially those from the distant past (Squire, 1989). If asked to recall early childhood events, it is possible that respondents forget or redefine their own behaviors in accordance with later life circumstances and their current situation (Ross, 1989). It is also possible that a person might redefine someone else's behavior in light of current knowledge. Unconscious denial (or repression of traumatic events in childhood) may be at work in preventing the recollection of severe cases of childhood abuse. Also, given society's disapproval of various forms of family violence, adolescents or adults who are asked to provide retrospective accounts of their own childhood experiences might also be sensitive to issues of social desirability when asked to describe their early childhoods.

In Robins' (1966) follow-up of patients in a child guidance clinic, of 71 participants who had been the victim of a "cruel or abusive father," from information from childhood interviews, or reports, or both, only 22 (or 31%) reported it in a follow-up interview conducted 30 years later. In a study of the accuracy of recall of concentration camp survivors over a 40-year period, Wagenaar and Groeneweg (1990) reported generally good recall of certain conditions in the camp but poor recall of many traumatic events.

Della Femina, Yeager, and Lewis (1990) followed up a group of delinquents who had been incarcerated in a Connecticut prison (at approximately age 15) to young adulthood (approximately age 24). Of the original sample of 119 youths, 77 had histories of abuse. Of these, more than half (53%) provided abuse information that was discrepant from records and interviews at the time. "Reasons for denial, whether in adolescence or adulthood, included embarrassment, a wish to protect parents, a sense of having deserved the abuse, a conscious wish to forget the past, and a lack of rapport with the interviewer" (p. 229).

Kruttschnitt and Dornfeld (1992), using physically abused women and their children and women and children from a community sample, assessed concurrent validity and reliability of

---

Cathy Spatz Widom and Robin L. Shepard, School of Criminal Justice, The University at Albany, State University of New York.

This research was supported by grants from the National Institute of Mental Health (MH49467) and the National Institute of Justice (86-IJ-CX-0033, 89-IJ-CX-0007, and 93-IJ-CX-0031). Points of view are those of the authors and do not necessarily represent the position of the United States Department of Justice. The authors express appreciation to Patricia Glynn and Barbara Luntz Weiler for assistance with the analyses for this article. An earlier version of this article was presented at the annual meeting of the American Society of Criminology, Miami, Florida, November 1994.

Correspondence concerning this article should be addressed to Cathy Spatz Widom, School of Criminal Justice, The University at Albany, State University of New York, 135 Western Avenue, Albany, New York 12222.

these preadolescents' (ages 11–12) reports of family violence. Both groups of children significantly underreported the extent to which they were victimized by their mothers during the past year. The accuracy of the childrens' reports of victimization appeared to be influenced both by the salience of the event and by contextual differences (the amount of violence in their environment), whereas reports of violence directed toward their mother were affected by the recall period. Kruttschnitt and Dornfeld suggested that investigations of the reliability and validity of victimization reports is of potential importance in interpreting discrepant findings from studies using different research designs. "For example, the stronger association between physical abuse and delinquency found in retrospective as opposed to prospective studies (Garbarino & Plantz, 1986) could be due to either youngsters exaggerating their victimization histories to obtain sympathy (Lewis, Mallouh, & Webb, 1989) or to youths underreporting ongoing incidents of victimization." (Kruttschnitt & Dornfeld, 1992, p. 137).

In an interesting study, Berger, Knutson, Mehm, and Perkins (1988) examined whether participants who were classified by three different self-report measures of physical abuse labeled themselves as being abused. Of approximately half the college student sample who were classified as having been physically abused as a child, less than 3% labeled themselves as having been abused.

Although not specific to child abuse, other research is relevant here. Yarrow, Campbell, and Burton (1970) compared retrospective reports of child rearing and developmental characteristics with extensive childhood records and found that the correlations between retrospective reports and record information and the absolute levels of agreement were quite low. Furthermore, they noted a tendency to distort the past in the direction of more favorable reports, with mothers overestimating their children's abilities. More recently, Henry, Moffitt, Caspi, Langley, and Silva (1994) compared the extent of agreement between prospective and retrospective measures across multiple content domains (including residence changes, injuries, reading ability, family characteristics, behavior problems, and delinquency) in a large sample of 18 year old youths who had been studied prospectively from birth. They found reasonable correlations for residence changes, reading skill, height and weight, whereas psychosocial variables (i.e., reports about participants' psychological states and family processes) had the lowest levels of agreement between prospective and retrospective measures. Also, despite significant correlations, the absolute level of agreement between the two data sources was low. Henry et al. argued quite convincingly that the levels of agreement in their study should be interpreted as an upper bound because of design characteristics of their study, whereas their conclusions did not differ significantly from those of Yarrow et al. (1970). Two other studies provide additional evidence that retrospective accounts of child-rearing experiences appear to have moderate (Finkel & McGue, 1993) to little reliability (Schwarz, Barton-Henry, & Pruzinsky, 1985).

In a reappraisal of retrospective reports, Brewin, Andrews, and Gotlib (1993) concluded the following:

Obtaining the retrospective recall of childhood events appears, therefore, to be a flawed process that can be shaped by both internal

and external factors. Social influences, childhood amnesia, and the simple fallibility of memory all impose limitations on the accuracy of recall, and fear of the consequences of disclosure may further disadvantage this process. However, provided that individuals are questioned about the occurrence of specific events or facts that they were sufficiently old and well placed to know about, the central features of their accounts are likely to be reasonably accurate. Because the influences on memory serve mainly to inhibit recall or disclosure, it seems fair to conclude that reports confirming events should be given more weight than negative reports. (p. 94)

In discussing studies in which recall was compared with independent records, Brewin et al. (1993) called attention to studies based on events in the first 5 years of a child's life, for which the individual was not likely to have had any direct recollection:

Although the age at which children become able to give verbal reports of significant experiences appears to be around 3 years (Sheingold & Tenney, 1982; Terr, 1988), research with adults has found a falloff in the retrieval of memories for events occurring before the age of 5 years, even when normal retention and forgetting processes are taken into account (Wetzler & Sweeney, 1986). (Brewin et al., 1993, p. 86)

Brewin et al. suggested two possible strategies to enhance the reliability of information obtained retrospectively: to obtain accounts from other sources of information (collaterals) and to use structured investigative methods that minimize unrealistic demands on the participant's memory. They also suggested comparing independent records with memories of events that occurred after the age of 5 years.

This article examines the accuracy of retrospective reports of childhood physical abuse, using a sample of individuals who had officially documented and substantiated cases of childhood victimization and a matched control group. A second article reporting the accuracy of retrospective reports of childhood sexual abuse will follow. Here, the accuracy of self-report measures of childhood physical abuse, the extent to which accuracy varies by the age of the child at the time of the abuse, the predictive efficiency of the measures (using relative improvement over chance), and construct validity are described.

One approach to assessing the power or efficiency of retrospective self-report measures is to calculate the relative improvement over chance (RIOC). Loeber and Dishion (1983) devised this index to represent the improvement over chance as a function of the range of its possible predictive efficiency. Because it is less sensitive to differences in base rates, one of the advantages of this technique is that it makes it possible to compare predictive efficiency of a variety of predictors, potentially over a range of studies. Optimally, this method should identify individuals who were (valid positives) and were not (valid negatives) abused in childhood. Loeber and Dishion argued that the degree that observed values in these cells deviate from random or chance values provides a more accurate assessment of predictive efficiency than is possible by means of a chi-square measure. Errors occur because self-report scales identify individuals who self-report abuse but who were not abused (false positives) and those who were not identified but who were abused (false negatives). Depending on one's priorities, the percentage of false positives and false negatives should be low. For

legal decisions, the percentage of false negatives should be low (Monahan, 1981).

In earlier work using official reports of childhood victimization, Widom (1989c) reported a significant relationship between childhood physical abuse and violent criminal behavior. Here, construct validity will be assessed using predictions about the consequences of physical abuse for subsequent violent behavior. That is, individuals who self-report physical abuse should have higher rates of arrest for violence and self-reported violence than individuals who do not self-report physical abuse.

Despite its strengths, the current design does not permit determination of the extent of false positives. It is not possible to determine whether individuals who self-report childhood abuse but do not have an official record of abuse are reporting accurately or not. The working assumption underlying this research is that these self-reports are valid until some empirical evidence contradicts that assumption. Unfortunately, this is a limitation that affects most research in this field, with the possible exception of some laboratory analogue studies in which behavior and social interactions can be monitored and assessed with more control.

## Method

### Design

The data used in these analyses are part of a research project based on a cohorts-design study (Leventhal, 1982; Schulsinger, Mednick, & Knop, 1981) in which abused and neglected children were matched with nonabused and neglected children and followed prospectively into young adulthood. Characteristics of the design include (a) an unambiguous operationalization of abuse and neglect; (b) a prospective design; (c) separate abused and neglected groups; (d) a large sample; (e) a control group matched as closely as possible for age, sex, race, and approximate social class background; and (f) assessment of the long-term consequences of abuse and neglect beyond adolescence and into adulthood. (For complete details of the original study design and participant selection criteria, see Widom, 1989a).

In the first phase of this research, a large group of children who were abused or neglected, or both, approximately 20 years ago were followed up through an examination of official juvenile and criminal records and compared with a matched control group of children (Widom, 1989b). The rationale for identifying the abused and neglected group was that their cases were serious enough to have come to the attention of the authorities. Only court-substantiated cases of child abuse and neglect were included here. Cases were drawn from the records of county juvenile and adult criminal courts in a metropolitan area in the Midwest during the years 1967 through 1971. Abuse and neglect cases were restricted to those in which children were less than 11 years of age at the time of the abuse or neglect incident. Thus, these are cases of early childhood abuse and neglect.

*Physical abuse* cases included injuries such as bruises, welts, burns, abrasions, lacerations, wounds, cuts, bone and skull fractures, and other evidence of physical injury. *Sexual abuse* charges varied from relatively nonspecific charges of "assault and battery with intent to gratify sexual desires" to more specific charges of "fondling or touching in an obscene manner," sodomy, incest, and so forth. *Neglect* cases reflected a judgment that the parents' deficiencies in child care were beyond those found acceptable by community and professional standards at the time. These cases represented extreme failure to provide adequate food, clothing, shelter, and medical attention to children. It should be noted that a person may have experienced more than one type of abuse or neglect, so that types of abuse are not mutually exclusive.

A control group was established with children who were matched on age, sex, race, and approximate family social class during the time period of the study (1967–1971). Children who were under school age at the time of the abuse or neglect were matched with children of the same sex, race, date of birth (+/- 1 week), and hospital of birth through the use of county birth record information. For children of school age, records of more than 100 elementary schools for the same time period were used to find matches with children of the same sex, race, date of birth (+/- 6 months), class in elementary school during the years 1967 through 1971, and home address, preferably within a five-block radius of the abused or neglected child. Overall, there were matches for 74% of the abused and neglected children.

The second phase of the research involved tracing, locating, and interviewing the abused and neglected individuals (20 years after their childhood victimization) and control participants. The follow-up was designed to document long-term consequences of childhood victimization across a number of outcomes (cognitive and intellectual, emotional, psychiatric, social and interpersonal, occupational, and general health).

Two-hour follow-up interviews were conducted between 1989 and 1995. The interview consisted of a series of structured and semistructured questionnaires and rating scales, and a psychiatric assessment. The interviewers were unaware of the purpose of the study, the inclusion of an abused or neglected group, and the participants' group membership. Similarly, the participants were unaware of the purpose of the study. Participants were told that they had been selected to participate as part of a large group of individuals who grew up in that area in the late 1960s and early 1970s. Those who participated signed a consent form acknowledging that they were participating voluntarily.

The findings described here are based on interviews with 1,196 individuals (110 cases of physical abuse, 96 of sexual abuse, 520 of neglect, and 543 controls). Of the original sample of 1,575, 1,292 participants (82%) have been located, and 1,196 have been interviewed (76%). Of the 95 people not interviewed, 39 were deceased, 9 were incapable of being interviewed, and 49 refused to participate (a refusal rate of 3%). Comparison of the current follow-up sample with the original sample indicates no significant differences in terms of percentage male, White, abused or neglected, poverty in childhood census tract, or mean current age. The interviewed group (follow-up sample) is significantly more likely to have an official criminal arrest record than the original sample of 1,575 (42% in the current sample versus 36% in the original sample). However, this is not surprising because people with a criminal history are generally easier to find, in part because they have more "institutional footprints" to assist in locating them.

The mean age of the sample at the time of the interview was 29.23 ( $SD = 3.84$ ), with no differences in age between the abused and neglected group and control participants. The average highest grade of school completed for the sample was 11.47 ( $SD = 2.19$ ), although abused and neglected individuals had completed significantly less school ( $M = 10.99$ ,  $SD = 1.99$ ) than did control participants ( $M = 12.09$ ,  $SD = 2.29$ ). Although two thirds of the control group had completed high school, less than half (48%) of the abused and neglected children at follow-up had done so. Occupational status of the sample was coded according to the Hollingshead Occupational Coding Index (Hollingshead, 1975). Occupational levels of the participants ranged from 1 (laborer) to 9 (professional). Median occupational level of the sample was semiskilled workers, and less than 7% of the overall sample was in levels 7–9 (managers through professionals). More of the control participants were in higher occupational levels than were the abused and neglected participants.

Because the interview asks about a history of childhood abuse and neglect (in a number of different ways), comparisons can be made of self-reported information with information in official case records (recorded at the time of the abuse or neglect experience). Percentage

accuracy of recall is calculated by comparing self-reported retrospective information with official record information from the earlier time period.

### Measures

Two measures were used to retrospectively assess a history of childhood physical abuse: the Conflict Tactics Scale (CTS) and the Self-Report of Childhood Abuse Physical (SRCAP). Both instruments, administered in structured interview formats (available from the authors), are described next.

The CTS was developed by Straus (1979) to assess the amount and severity of family violence. Several subscales have been identified: Reasoning, Verbal Aggression, Minor Violence, Severe Violence, and Very Severe Violence. When they refer to physical abuse, Straus and Gelles (1990) are using the Very Severe Violence (VSV) scale, which includes the following items: "kick, bite, or hit you with a fist"; "beat you up"; "burn or scald you"; "threaten you with a knife or gun"; or "use a knife or gun."

Since the CTS was developed, it has been used primarily to study spousal abuse (Straus, 1988). Fewer studies have used the CTS to assess a history of childhood physical abuse (Brutz & Ingoldsby, 1984; Dembo et al., 1987; Eblen, 1987; Gelles & Edfeldt, 1986; Giles-Sims, 1985; Kruttschnitt & Dornfeld, 1992; Meredith, Abbott, & Adams, 1986). In most of these studies, parents were asked to complete the CTS about their child. Dembo et al. (1987) administered a modified CTS scale (using 6 items) to 145 juvenile detainees housed in a state-operated regional facility (status offender and juvenile delinquents).

Although the CTS has been used to assess physical child abuse, information on criterion and construct validity is not extensive (Straus & Gelles, 1990). Using data from a survey conducted with one-parent households in Sweden to address the concurrent validity of the CTS, Gelles and Edfeldt (1986) reported that factors associated with violence included younger parents and parents with a violent background. Education was not significant. Straus and Gelles (1990) reported that children who experienced severe violence were described by their parents to have higher rates of conduct problems and rule-violating behaviors than those who did not experience severe violence.

In this study, CTS items were framed in the context of an introduction that asked respondents about "things that your parents or the people in your family might have done when they had a disagreement with you when you were growing up, that is, up to the time you finished elementary school." This age limit was imposed to ensure that the time period was consistent with official information about the abuse experience. Possible response categories ranged from *never to once, twice, sometimes, frequently, or most of the time*. Dichotomous summary subscale scores were computed using the subscales reported by Straus and Gelles (1990): Reasoning, Verbal Aggression, Minor Violence, Severe Violence, and Very Severe Violence. In addition, a principal-components analysis (oblique promax rotation) was performed to determine the factor structure of the CTS in this study (see Appendix), with the resulting factor structure similar to that reported earlier by Straus (1979).

A second self-report measure of childhood physical abuse (SRCAP) was designed for the purposes of this study to provide an alternative means to retrospectively assess childhood physical abuse. The SRCAP reflects the person's response to the following six items: (a) "beat or really hurt you by hitting you with a barehand or fist"; (b) "beat or hit you with something hard like a stick or baseball bat"; (c) "injure you with a knife, shoot you with a gun, or use another weapon against you"; (d) "hurt you badly enough so that you needed a doctor or other medical treatment"; (e) "physically injure you so that you were admitted to a hospital"; and (e) "beat you when you didn't deserve it." A dichotomous variable was created to indicate whether the person reported hav-

ing had any of these childhood experiences or none. A principal-components analysis (oblique promax rotation) of the six-item SRCAP revealed a two-factor solution explaining 63% of the variance (Cronbach's  $\alpha = .75$ ).

Two measures of violence (official and self-report) were used in the construct validity analyses. The first (official) measure, arrest for violence, is based on information obtained from complete criminal histories collected for these individuals at three levels of law enforcement (local, state, and federal) at two points in time (1986-1987 and 1994). Any arrest for violence refers to arrest as a juvenile or as an adult and includes arrests for the following crimes and attempts: assault, battery, robbery, manslaughter, murder, rape, and burglary with injury. The self-report measure of violence refers to the person's response to seven items embedded in a general crime and delinquency scale completed by the respondent during the in-person interview. These items include "hurt someone badly enough for him or her to require medical treatment"; "threatened to hurt someone if he or she didn't give you money or something else"; "used a weapon to threaten another person"; "forced someone to have sex with you"; "shot someone"; "attacked someone with the purpose of killing him or her"; and "used physical force to get money, drugs, or something else from someone."

### Results

The results are organized into two major sections. The first section focuses on accuracy and describes differences in CTS and SRCAP scores for individuals with official records of physical abuse compared with individuals with sexual abuse or neglect and those with no official records of abuse or neglect (control participants). The results of analyses that were conducted to determine the extent to which there are differences in accuracy associated with the person's age at the time of the abuse experience and predictive efficiency of the measures (RIOC) are also described here. The second section reports on construct validity through a series of multivariate analyses using the two measures of self-reported childhood physical abuse (the CTS and SRCAP) and two measures of violent behavior (official arrest data and self-reported violence information).

#### *Official Reports vs. Retrospective Self-Reports of Childhood Physical Abuse*

Tables 1 and 2 present the accuracy of individual items and summary scales for the CTS, respectively, by type of abuse. Because of the large number of statistical tests in this table and the possibility of alpha inflation, the Bonferroni correction procedure was used. On 8 of the 19 items, respondents who were physically abused (according to official records at the time) differed significantly from individuals who were sexually abused or neglected (according to official records) and from individuals with no official record of having been abused or neglected. In these cases, respondents who had official histories of childhood physical abuse were most likely to self-report histories of physical abuse as measured by the CTS and more likely than individuals with official records of sexual abuse or neglect. In turn, both groups reported more physical abuse than did control participants. For an additional 7 items, physical abuse victims and victims of sexual abuse or neglect also differed significantly from control participants in their responses on the CTS. It is interesting that there were no differences among the

Table 1  
Accuracy of Individual CTS Items by Type of Abuse (in Percentages,  $N = 1,196$ )

Self-reports	Official reports			Significance	
	Physical abuse group	Neglect and sexual abuse group	Control group	Overall	Pairwise
Discuss issue calmly	73.2	80.3	88.0	***	a,b
Get backup information	71.7	75.0	80.1		
Bring someone to settle	39.6	41.1	33.4		
Insult or swear	85.2	75.4	59.7	***	a,b,c
Sulk or refuse to talk	67.6	59.0	49.7	***	a,b
Stomp out of room	57.4	58.3	45.6	***	a,b
Cry	57.4	63.6	59.3		
Do/say spiteful	75.9	71.7	56.4	***	a,b
Threaten to hit/throw	70.8	64.2	50.1	***	a,b
Throw/smash/hit/kick	66.7	57.3	39.4	***	a,b
Throw something at you	52.3	41.9	25.7	***	a,b,c
Push/grab/shove	73.8	61.9	50.0	***	a,b,c
Slap/spank you	88.0	83.5	82.5		
Kick, bite, or hit	55.1	33.0	15.6	***	a,b,c
Hit/try with something	67.3	54.6	40.1	***	a,b,c
Beat you up	49.1	27.2	12.6	***	a,b,c
Burn or scald you	13.9	3.9	1.4	***	a,b,c
Threaten w/knife/gun	15.9	9.1	3.3	***	a,b,c
Use a knife or gun	6.5	3.6	0.4	***	a,b

Note. CTS = Conflict Tactics Scale; a = comparison of physical abuse victims vs. control participants,  $p \leq .05$ ; b = comparison of neglect and sexual abuse victims vs. control participants,  $p \leq .05$ ; c = comparison of physical abuse victims vs. neglect and sexual abuse victims,  $p \leq .05$ .

\*\*\*  $p < .001$  (two-tailed), Bonferroni-corrected chi-square analyses.

three groups for four items ("get backup information," "bring someone in to settle," "cry," and "slap or spank you").

Depending on the particular subscale (see Table 2), from 60% (Very Severe Violence) to 92% (Minor Violence) of the physically abused group reported physical abuse. There was good discriminant validity for the Severe Violence and Very Severe Violence subscales, with all three groups differing significantly from one another. On the other hand, using the Minor Violence subscale, 92% and 86% of the sexually abused and neglected individuals and control participants, respectively, also reported histories of abuse, and the three groups did not differ significantly from one another. Because of these results, the remainder of this article will report on the two subscales of the

CTS (Severe Violence and Very Severe Violence) that appear to have the best discriminant validity.

Table 3 presents the results on the accuracy of the SRCAP measure. With one exception, each of the individual items on the SRCAP and the overall SRCAP score discriminates significantly among the three groups, with the physical abuse group (as defined by official records) reporting the highest percentage, followed by the sexual abuse and neglect group, and then the control group. Almost two thirds (62%) of the group with official records of childhood physical abuse met the criteria using the SRCAP measure, compared with 42% of the sexual abuse and neglect group, and 25% of the control group. These findings provide evidence for the validity of the SRCAP and clear dis-

Table 2  
Accuracy of Overall CTS Scores by Type of Abuse (in Percentages,  $N = 1,196$ )

Self-reports	Official reports			Significance	
	Physical abuse group	Neglect and sexual abuse group	Control group	Overall	Pairwise
Reasoning	83.0	87.9	93.6	***	a,b
Verbal Aggression	95.4	94.1	90.3		
Minor Violence	91.6	91.5	86.3		
Severe Violence	68.9	58.3	43.6	***	a,b,c
Very Severe Violence	59.8	38.9	20.8	***	a,b,c

Note. CTS = Conflict Tactics Scale; a = comparison of physical abuse victims vs. control participants,  $p \leq .05$ ; b = comparison of neglect and sexual abuse victims vs. control participants,  $p \leq .05$ ; c = comparison of physical abuse victims vs. neglect and sexual abuse victims,  $p \leq .05$ .

\*\*\*  $p \leq .001$  (two-tailed), Bonferroni-corrected chi-square analyses.

**Table 3**  
*Accuracy of SRCAP by Type of Abuse (in Percentages, N = 1,196)*

Self-reports	Official reports			Pairwise significance
	Physical abuse group	Neglect and sexual abuse group	Control group	
Beat by parent/didn't deserve	56.2	31.1	19.5	a,b,c
Hit with bare hand	42.2	23.1	11.4	a,b,c
Hit with something hard	31.2	14.9	8.5	a,b,c
Injure with weapon	5.5	3.4	0.6	a,b
Hurt enough—need doctor	26.4	7.1	1.4	a,b,c
Admitted to hospital	13.2	2.3	0.4	a,b,c
SRCAP	62.3	41.5	25.3	a,b,c

*Note.* SRCAP = Self-Report of Childhood Abuse Physical; For overall significance, all  $ps \leq .001$  (two-tailed), Bonferroni-corrected chi-square analyses. a = comparison of physical abuse victims vs. control participants,  $p \leq .05$ ; b = comparison of neglect and sexual abuse victims vs. control participants,  $p \leq .05$ ; c = comparison of physical abuse victims vs. neglect and sexual abuse victims,  $p \leq .05$ .

crimination between externally documented groups of abused and neglected individuals and control participants.

*Age at the time of abuse.* Analyses were conducted to determine whether accuracy of retrospective reports would be higher for individuals who were older at the time of the abuse incident. Taking the suggestion of Brewin et al. (1993), we used 5 years of age at the time of the abuse experience as the cutoff point for the comparisons. The results indicate that there were no differences in recall by age at the time of the abuse experience. Table 4 presents these findings.

*Relative improvement over chance.* Table 5 presents the results of analyses to determine the predictive efficiency of the self-report measures. All three measures identify approximately the same percentage of the sample as valid positives (approximately 5–6%), whereas the actual base rate for the sample is 9.2%. However, differences between the measures appear when looking at other characteristics in Table 5. Both the CTS—Very Severe Violence (CTS–VSV) and the SRCAP identify approximately the same percentages of valid negatives (64% and 60%, respectively) and false positives (28% and 31%), whereas the CTS–SV (Severe Violence) scale identifies a much higher percentage of false positives (47%) and a lower percentage of valid negatives (44%). All three measures identify ap-

proximately 3–4% of the sample as false negatives—people who have official records of having been physically abused but do not self-report a history of abuse according to any of the measures. Overall, the RIOC for the CTS–VSV and the SRCAP is approximately 40–41%, meaning that predictive efficiency is approximately 40% better than what would be expected by chance alone.

#### *Construct Validity of CTS and SRCAP Measures*

Earlier reports found that documented cases of physical abuse were associated with increased risk of arrest for violence (Widom, 1989c). To assess the construct validity of these retrospective self-report measures of childhood physical abuse, self-reported physical abuse scores (using the CTS and SRCAP) were entered into regression equations predicting to arrests for violence and self-reported violence. Analyses were conducted with dichotomous dependent variables (any arrest for a violent crime vs. no arrest for a violent crime and any self-reported violence item). The assumption is that if self-reports of physical abuse (using either the CTS or SRCAP) are valid indicators of childhood physical abuse, then they should predict the same or similar outcomes as found for officially reported physical abuse. Table 6 presents the results of a series of logistic regressions predicting to any violent arrest and any self-reported violence. Pure types of officially reported abuse (physical abuse, sexual abuse, and neglect) were introduced into the regression equations (replicating the earlier work of Widom, 1989c), with controls for sex, age, and race. Officially reported physical abuse is a significant predictor of having an arrest for violence. The self-report measures of childhood physical abuse (CTS–SV, CTS–VSV, and SRCAP) were also introduced into three separate but similar equations. These results indicate that none of the self-report measures of childhood physical abuse predict to arrests for violence.

In the right-hand portion of Table 6, the results of another series of equations are reported, using the dependent variable of self-reported violence instead of arrests for violence. These results indicate that officially reported physical abuse is not a significant predictor of self-reported violence, whereas each of

**Table 4**  
*Recall of Physical Abuse by Age at Time of Abuse (in Percentages)*

Scale	Age at time of physical abuse (years)	
	0–4 (n = 33)	5–11 (n = 73)
	CTS: Severe Violence	72.7
CTS: Very Severe Violence	61.8	58.9
SRCAP	55.9	65.3

*Note.* Official cases of physical abuse only. None of the differences was significant. CTS = Conflict Tactics Scale; SRCAP = Self-Report of Childhood Abuse Physical.

Table 5  
*Relative Improvement Over Chance: Retrospective Reporting of Physical Abuse (in Percentages)*

Self-report measure	Valid positives	False positives	Valid negatives	False negatives	Selection ratio	RIOC	Chi-square	Confidence interval range
CTS: Severe Violence	6.1	46.6	44.4	2.8	52.8	34.1	12.1***	15.7–52.4
CTS: Very Severe Violence	5.4	27.5	63.5	3.6	32.9	40.1	38.6***	26.6–53.7
SRCAP	5.6	30.7	60.4	3.4	36.3	40.8	34.0***	26.6–55.0

Note. CTS = Conflict Tactics Scale; SRCAP = Self-Report of Childhood Abuse Physical; RIOCI = relative improvement over chance.  
 \*\*\*  $p \leq .001$ .

the self-report measures (both CTS and SRCAP) are significant predictors of self-reported violence.

### Discussion

Retrospective self-reports of early childhood physical abuse were assessed through comparison with official cases of physical abuse documented and substantiated through court records. Both self-report measures have reasonable internal reliability, and the CTS factor scores are similar to the original subscales reported by Straus (1979). In many ways, these findings indicate accuracy in retrospective self-reports and good discriminant validity. Individuals who were physically abused, based on official records, retrospectively reported the highest rates of childhood physical abuse in the sample. On the CTS (Severe Violence and Very Severe Violence subscales) and the SRCAP, physically abused individuals reported significantly higher rates of physical abuse than did individuals who had experienced sexual abuse or neglect in childhood and individuals who were part of a matched control group. The extent of remembering (i.e., the percentage of individuals who had been physically abused who reported having been physically abused on one of the measures used here) is in line with previous research. These results also reveal that the extent of reporting a history of childhood physical abuse varied dramatically by the criterion (or measure) used.

Table 6  
*Predicting Violence Using Official and Retrospective Childhood Physical Abuse Measures*

Measure	Any violent arrest			Self-reported violence		
	$\beta$	SE	Odds ratio	$\beta$	SE	Odds ratio
Official						
Physical abuse	.66*	.36	1.94	.01	.34	1.01
Self-reports						
CTS: Severe Violence	.03	.17	1.03	1.09***	.15	2.97
CTS: Very Severe Violence	.03	.18	1.03	.78***	.15	2.18
SRCAP	.16	.18	1.17	.93***	.15	2.55

Note. Logistic regressions predicting to any violent arrest and any self-reported violence, controlling for sex, race, age, and other types of abuse or neglect. CTS = Conflict Tactics Scale; SRCAP = Self-Report Child Abuse Physical. (Further details of these results are available from Cathy Spatz Widom.)

\*  $p < .05$ . \*\*\*  $p < .001$ .

At the same time, there is a problem in underreporting of physical abuse. A substantial group of individuals who were physically abused do not report having been physically abused in childhood. Of the 110 people in the sample who had documented cases of physical abuse in childhood, 60–62% reported abuse using the CTS–VSV and SRCAP. This means that approximately 40% of individuals with documented histories of physical abuse did not report. Whether these people did not report (as suggested by Della Femina et al., 1990) because of embarrassment, a wish to protect parents, a sense of having deserved the abuse, a conscious wish to forget the past, or lack of confidence in or rapport with the interviewer, we do not know. But these findings suggest that a substantial minority would not be included in retrospective self-report assessments of childhood physical abuse. A more lenient criterion (such as the CTS–Minor Violence subscale) would capture most of the physically abused people (see Table 2); however, this criterion also identifies 92% of the sexual abuse and neglect cases and 86% of the control participants as having been physically abused in childhood. Using the CTS–Minor Violence subscale, the rate of false positives (as presented in Table 5) approaches almost half the sample. These findings illustrate that the rate of false positives is directly related to the measure of childhood physical abuse used. For some purposes, such as in clinical settings, a higher false positive rate may be desirable.

Few studies have addressed the accuracy of retrospective recall of childhood physical abuse. However, in studies that have examined these issues, the results indicate that there is a substantial group of individuals who do not remember the physical abuse or do not report it in the context of a questionnaire or interview. We explored one possible explanation associated with the fact that some of these individuals might have been too young at the time of the abuse experience to remember it accurately. These analyses did not reveal differences in accuracy by age at the time of the abuse experience using age 5 as a cutoff. As Yarrow et al. (1970) pointed out many years ago, information that we remember from childhood may be heavily dependent on information told to us in childhood or later, or constructed by a parent, or both. It may well be that children who experience physical abuse (whose cases did not come to the attention of the authorities) would remember their experiences differently. On the other hand, given that these were court-substantiated cases, the amount of underreporting is notable. Unfortunately, by its very nature, family violence occurs “behind closed doors,” and this characteristic makes documenting its occurrence problematic and studying the phenomenon

difficult, because it often hinges on the report of the victim, and there may or may not be physical evidence available.

In terms of construct validity of the self-report measures, these findings are troubling, suggesting substantial shared method variance. Official reports of physical abuse predict to official reports of violence, and self-reports of physical abuse predict to self-reports of violence. Although this article has used official court or police records (both in terms of documented cases of abuse and arrests for violence) as the criteria against which to assess the validity of these retrospective measures, they are not the only criteria possible. For example, medical records of abuse or emergency room contacts might also provide evidence against which to validate physical abuse, although both constitute some level of official processing of these cases. Police records reflect only a sample of all offenses committed by an individual and may include notations of crimes that did not occur as charged. Police have discretion in deciding which subjects to arrest, which arrests to record, and which charges to file. Similarly, official reports of child abuse and neglect are associated with biases, overrepresenting people of low socioeconomic status, and more severe cases of abuse and neglect (Widom, 1988).

Although further attempts should be made to assess the construct validity of these scales and other measures, researchers should also attempt to use information from more than one source when conducting analyses involving child abuse. As Sternberg et al. (1993) concluded, in a study of the effects of experiencing and witnessing domestic violence that used Israeli children and a control group, "one cannot discuss the effects of domestic violence without considering the source of information, particularly because the levels of agreement among informants were extremely low" (p. 49). The implication is that, if at all possible, researchers need to use multiple sources of information. Failing that, researchers need to recognize that, similar to the biases associated with official reports of childhood abuse and neglect (Newberger, Reed, Daniel, Hyde, & Kotelchuck, 1977), self-reported childhood victimization may contain systematic biases.

Henry et al. (1994) concluded that reliance on retrospective reports about psychosocial variables should be treated with caution. They suggested that "the use of retrospective reports should be limited to testing hypotheses about the relative standing of individuals in a distribution and should not be used to test hypotheses that demand precision in estimating event frequencies and event dates" (p. 92). We support their recommendation to use caution against overly simplistic interpretations that take retrospective reports at face value.

These methodological problems pose significant challenges to researchers in the field. Notwithstanding the real difficulties involved, there is a critical need to develop reliable and valid ways to assess histories of childhood victimization. This research has provided an opportunity to assess the accuracy of retrospectively obtained childhood victimization information, and we hope that these findings will be useful to other researchers and clinicians who are dependent on this information.

## References

- Beitchman, J. H., Zucker, K. J., Hood, J. E., daCosta, G. A., & Akman, D. (1991). A review of the short-term effects of childhood sexual abuse. *Child Abuse and Neglect, 15*, 537-556.
- Berger, A. M., Knutson, J. F., Mehm, J. G., & Perkins, K. A. (1988). The self-report of punitive childhood experiences of young adults and adolescents. *Child Abuse and Neglect, 12*, 251-262.
- Bradburn, N. M., Rips, L. J., & Shevell, S. K. (1987). Answering autobiographical questions: The impact of memory and inference on surveys. *Science, 236*, 157-161.
- Brewin, C. R., Andrews, B., & Gotlib, I. H. (1993). Psychopathology and early experience: A reappraisal of retrospective reports. *Psychological Bulletin, 113*, 82-98.
- Briere, J. (1992). Methodological issues in the study of sexual abuse effects. *Journal of Consulting and Clinical Psychology, 60*, 196-203.
- Briere, J., & Conte, J. (1993). Self-reported amnesia for abuse in adults molested as children. *Journal of Traumatic Stress, 6*, 21-31.
- Brutz, J. L., & Ingoldsby, B. B. (1984). Conflict resolution in Quaker families. *Journal of Marriage and the Family, 46*, 21-26.
- Cicchetti, D., & Rizley, R. (1981). Developmental perspectives on the etiology, intergenerational transmission, and sequelae of child maltreatment. *New Directions for Child Development, 11*, 31-55.
- Della Femina, D., Yeager, C. A., & Lewis, D. O. (1990). Child abuse: Adolescent records vs. adult recall. *Child Abuse and Neglect, 14*, 227-231.
- Dembo, R., Dertke, M., La Voie, L., Borders, S., Washburn, M., & Schmeidler, J. (1987). Physical abuse, sexual victimization and illicit drug use: A structural analysis among high risk adolescents. *Journal of Adolescence, 10*, 13-33.
- Eblen, C. N. (1987). *The influence of stress and social support upon child abuse*. Unpublished doctoral dissertation, Arizona State University.
- Finkel, D., & McGue, M. (1993). Twenty-five year follow-up of child-rearing practices: Reliability of retrospective data. *Personality and Individual Differences, 15*(2), 147-154.
- Garbarino, J., & Plantz, M. (1986). Part I. Review of the literature. In E. Gray, J. Garbarino, & M. Plantz (Eds.), *Child abuse: Prelude to delinquency?* (pp. 5-18). Washington, DC: U.S. Department of Justice, Office of Juvenile Justice and Delinquency Prevention. (Findings of a research conference conducted by the National Committee for Prevention of Child Abuse, April 7-10 1984).
- Gelles, R. J., & Edfeldt, A. W. (1986). Violence towards children in the United States and Sweden. *Child Abuse and Neglect, 10*, 501-510.
- Giles-Sims, J. (1985). A longitudinal study of battered children of battered wives. *Family Relations, 34*, 205-210.
- Henry, B., Moffitt, E. G., Caspi, A., Langley, J., & Silva, P. A. (1994). On the "remembrance of things past": A longitudinal evaluation of the retrospective method. *Psychological Assessment, 6*, 92-101.
- Herman, J., & Schatzow, E. (1987). Recovery and verification of memories of childhood sexual trauma. *Psychoanalytic Psychology, 4*, 1-14.
- Hollingshead, A. B. (1975). *Four factor index of social status*. Unpublished manuscript. Yale University.
- Kruttschnitt, C., & Dornfeld, M. (1992). Will they tell? *Journal of Research in Crime and Delinquency, 29*, 136-147.
- Leventhal, J. M. (1982). Research strategies and methodologic standards in studies of risk factors for child abuse. *Child Abuse and Neglect, 6*, 113-123.
- Lewis, D. O., Mallouh, C., & Webb, V. (1989). Child abuse, delinquency, and violent criminality. In D. Cicchetti & V. Carlson (Eds.), *Child maltreatment* (pp. 707-721). Cambridge, England: Cambridge University Press.
- Loeber, R., & Dishion, T. (1983). Early predictors of male delinquency: A review. *Psychological Bulletin, 94*, 68-99.
- Loftus, E. (1993). The reality of repressed memories. *American Psychologist, 48*, 518-537.
- Meredith, W. H., Abbott, D. A., & Adams, S. L. (1986). Family vio-



- lence: Its relation to marital and parental satisfaction and family strengths. *Journal of Family Violence*, 1, 299–305.
- Milner, J. S. (1991). Physical child abuse perpetrator screening and evaluation. Physical child abuse [Special issue]. *Criminal Justice and Behavior*, 18, 47–63.
- Monahan, J. (1981). *Predicting violent behavior: An assessment of clinical techniques*. Beverly Hills, CA: Sage Press.
- Newberger, E. H., Reed, R. B., Daniel, J. H., Hyde, J. N., & Kotelchuck, M. (1977). Pediatric social illness: Toward an etiological classification. *Pediatrics*, 50, 178–185.
- Robins, L. N. (1966). *Deviant children grown up*. Baltimore: Williams & Wilkins.
- Ross, M. (1989). Relation of implicit theories to the construction of personal histories. *Psychological Review*, 96, 341–357.
- SAS Institute Inc. (1989). *SAS/STAT User's Guide, Version 6, Fourth Edition, Volume 1*. Cary, NC: SAS Institute Inc.
- Schulsinger, F., Mednick, S. A., & Knop, J. (1981). *Longitudinal research: Methods and uses in behavioral sciences*. Boston: Martinus Nijhoff.
- Schwarz, J. C., Barton-Henry, M. L., & Pruzinsky, T. (1985). Assessing child-rearing behaviors: A comparison of ratings made by mother, father, child, and sibling on the CRPBI. *Child Development*, 56, 462–479.
- Sheingold, K., & Tenney, Y. J. (1982). Memory for a salient childhood event. In U. Neisser (Ed.), *Memory observed* (pp. 201–212). San Francisco: Freeman.
- Squire, L. (1989). On the course of forgetting in very long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 241–245.
- Sternberg, K. J., Lamb, M. E., Greenbaum, C., Cicchetti, D., Dawud, S., Cortes, R. M., Krispin, O., & Lorey, F. (1993). Effects of domestic violence on children's behavior problems and depression. *Developmental Psychology*, 29, 44–52.
- Straus, M. A. (1979). Measuring intrafamily conflict and violence: The Conflict Tactics (CT) Scales. *Journal of Marriage and the Family*, 41, 75–86.
- Straus, M. A. (1988). *Measuring psychological and physical abuse of children with the Conflict Tactics Scales*. Durham: University of New Hampshire, Family Research Laboratory.
- Straus, M. A., & Gelles, R. J. (Eds.). (1990). *Physical violence in American families: Risk factors and adaptations to violence in 8145 families*. New Brunswick, NJ: Transaction.
- Terr, L. (1988). What happens to early memories of trauma? A study of twenty children under age five at the time of documented traumatic events. *Journal of the American Academy of Child and Adolescent Psychiatry*, 27, 96–104.
- Wagenaar, W. A., & Groeneweg, J. (1990). The memory of concentration camp survivors. *Applied Cognitive Psychology*, 4, 77–87.
- Wetzler, S. E., & Sweeney, J. A. (1986). Childhood amnesia: An empirical demonstration. In D. C. Rubin (Ed.), *Autobiographical memory* (pp. 191–201). New York: Cambridge University Press.
- Widom, C. S. (1988). Sampling biases and implications for child abuse research. *American Journal of Orthopsychiatry*, 58, 260–270.
- Widom, C. S. (1989a). Child abuse, neglect and adult behavior: Research design and findings on criminality, violence, and child abuse. *American Journal of Orthopsychiatry*, 59, 355–367.
- Widom, C. S. (1989b). Child abuse, neglect, and violent criminal behavior. *Criminology*, 27, 251–271.
- Widom, C. S. (1989c). The cycle of violence. *Science*, 244, 160–166.
- Widom, C. S. (1989d). Does violence beget violence? A critical examination of the literature. *Psychological Bulletin*, 106, 3–28.
- Widom, C. S. (1994). Childhood victimization and risk for adolescent problem behaviors. In M. E. Lamb & R. Ketterlinus (Eds.), *Adolescent problem behaviors* (pp. 127–164). Elmsford, NJ: Erlbaum.
- Williams, L. M. (1994). Recall of childhood trauma: A prospective study of women's memories of child sexual abuse. *Journal of Consulting and Clinical Psychology*, 62, 1167–1176.
- Wolfe, D. A., & Mosk, M. D. (1983). Behavioral comparisons of children from abusive and distressed families. *Journal of Consulting and Clinical Psychology*, 51, 702–708.
- Yarrow, M. R., Campbell, J. D., & Burton, R. V. (1970). Recollections of childhood: A study of the retrospective method. *Monographs of the Society for Research in Child Development*, 33(5, Serial No. 135).

## Appendix

## Principal-Components Analysis: Oblique Promax Rotation (Conflict Tactics Scale)

Factor 1 (eigenvalue = 5.77, 30.4% of variance)		
i.	threaten to hit or throw something at you	.78
j.	throw, smash, hit or kick something	.58
k.	throw something at you	.68
l.	push, grab or shove you	.76
m.	slap or spank you	.75
n.	kick, bite, or hit you with a fist	.60
o.	hit you or try to hit you with something	.78
p.	beat you up	.53
Factor 2 (eigenvalue = 1.95, 10.3% of variance)		
d.	insult or swear at you	.41
e.	sulk and/or refuse to talk about it	.75
f.	stomp out of the room or house	.76
g.	cry	.47
h.	do or say something to spite you	.55
Factor 3 (eigenvalue = 1.48, 7.8% of variance)		
q.	burn or scald you	.65
r.	threaten you with a knife or gun	.78
s.	use a knife or gun	.82
Factor 4 (eigenvalue = 1.26, 6.7% of variance)		
a.	discuss an issue calmly	.76
b.	get information to back up their side of things	.85
c.	bring in or try to bring in someone to help settle things	.53

*Note.* Cronbach's alpha = .82. Coefficients estimated using SAS procedure PROC FACTOR.

Received January 18, 1996  
 Revision received June 17, 1996  
 Accepted June 19, 1996 ■