

Assessing the Reliability of Retrospective Reports of Adverse Childhood Experiences among Adolescents with Documented Childhood Maltreatment

Ricardo Pinto · Liliana Correia · Ângela Maia

Published online: 15 April 2014
© Springer Science+Business Media New York 2014

Abstract The literature suggests that childhood maltreatment is related to a higher probability of developing psychopathology and disease in adulthood. However, some authors have questioned the reliability of self-reports of maltreatment, suggesting that psychopathology at the time of evaluation affects self-reports. We evaluated the reliability of the self-reports of 79 young adults who were identified in childhood by Child Protective Services by comparing two moments of evaluation. Psychological and physical symptoms were tested to evaluate their interference with the reports. We found good to excellent agreement, with no significant correlation between the changes in self-reported experiences and the changes in physical and psychological symptoms, suggesting that the reliability of reports is not related to the health state at the time of the report.

Keywords Adverse childhood experiences · Reliability · Retrospective reports

Research has demonstrated the relationship between adverse childhood experiences and psychopathology (Herrenkohl, Hong, Klika, Herrenkohl, and Russo 2012; Shaffer, Huston, and Egeland 2008), health risk behaviors (Ramiro, Madrid, and Brown 2010) and disease (Felitti et al. 1998). Most studies rely on retrospective reports from adult samples to collect information about childhood maltreatment. The researchers conducting these studies have used cross-sectional designs with retrospective reports. This approach considers that several abuse survivors were not identified as children (Sedlak et al. 2010) and that the possibility of bias is negligible (Hardt and Rutter 2004). This method has other advantages, such as

the convenience of sampling, a relatively low cost, and short-term results, compared with longitudinal designs.

Given that most studies have used retrospective self-reports of childhood victimization, it is important that researchers know the level of reliability of retrospective self-reports about early childhood experiences (Dube, Williamson, Thompson, Felitti, and Anda 2004). The test-retest method is the most appropriate way to assess the reliability of self-reported trauma experiences (MacKenzie, Podsakoff and Jarvis 2005; Norris and Hamblen 2004) and is preferred to other reliability methods, such as Cronbach's alpha. A measure is considered reliable if a respondent's score is similar during two separate administrations of the same instrument under similar conditions with the same respondents. Assessing test-retest reliability has been the aim of several studies that examined the consistency of responses to sensitive questions, such as those used in studies of childhood victimization (Bernstein, et al. 1994; Dube, Williamson, Thompson, Felitti, and Anda 2004; Wolfe, Scott, Wekerle, and Pittman 2001).

However, relatively few studies have assessed the reliability of retrospective reports of childhood maltreatment, and the results are contradictory. For instance, Fergusson, Horwood, and Woodward (2000) studied the stability of child abuse reports using information from a longitudinal birth cohort study of 1,265 New Zealand young adults, who were questioned at the ages of 18 and 21 about their childhood exposure to physical punishment and sexual abuse. Researchers found that reports were relatively unstable, with kappa values for test-retests of abuse around .45, reflecting fair agreement.

In contrast with these results, a study by Dube et al. (2004) suggested high consistency for reports of adverse childhood experiences in later adulthood. These authors relied on a sample of 658 participants, with a mean age of 64 years. The self-report questionnaire was administered twice, with a 20-month interval. The results of this study showed good to

R. Pinto · L. Correia · Â. Maia (✉)
Escola de Psicologia, Universidade do Minho, Campus de Gualtar,
4710-057 Braga, Portugal
e-mail: angelam@psi.uminho.pt

moderate consistency between the two reports, with kappa values ranging from .46 to .86. In response to these conflicting results, Widom et al. (2004) commented on the Dube et al. (2004) study, noting several concerns about the conclusions on the reliability of retrospective self-reports. Widom et al. (2004) suggested that the findings reflecting moderate to good reliability of retrospective reports were “influenced” by the participants’ age ($M=64$), noting that the memory of one’s life story is well established and fixed at this age. In comparison, the 18 to 21-year-olds in Fergusson et al.’s (2000) study may not have had fixed life stories.

Widom et al. (2004) also hypothesize that current physical or psychological health status influences the recall of prior experiences. Subjects in poor health may have better recall and reporting of early negative experiences, and they may be more likely to interpret their early experiences negatively. The authors based this hypothesis on demonstrations in previous studies of the influence of current mental health on the recollection of past experiences (Raphael and Cloitre 1994; Schraedley, Turner, and Gotlib 2002). However, the literature lacks consensus on this matter. Other studies have shown that reports of adverse childhood experiences are not related to mental health at the time of the report (Brewin, Andrews, and Gotlib 1993; Fergusson, et al. 2000; Monteiro and Maia 2010; Robins, Schoenberg, Holmes, Ratcliff, Benham, and Works 1985).

Information about childhood maltreatment is collected by two main methods: self-administered questionnaires, in which respondents complete a paper questionnaire, and in-person interviews, in which questions are administered by a face-to-face interviewer. However, there is a lack of empirical evidence on the impact of in-person interviews vs. self-report questionnaires, and the conclusions differ. Some evidence suggests that subjects may deny experiences during face-to-face interviews, even though those experiences had been revealed by self-administered questionnaires (Koss and Gidycz 1985), whereas other data suggest that subjects deny more experiences using self-report questionnaires compared to face-to-face interviews (Wyatt and Peters 1986). It is recognized that questionnaires provide anonymity and may increase the sense of confidentiality, which facilitates disclosure. Conversely, interviews offer the ability to explore ambiguous responses. Although advantages and disadvantages are associated with both measures, self-administered questionnaires are a more efficient method of data collection because they can be group-administered, do not require the training of interviewers (DiLillo, DeGue, Kras, Di Loreto-Colgan, and Nash 2006), and facilitate responses to sensitive questions and the reporting of socially undesirable behaviors (Tourangeau and Yan 2007).

The principal aim of the present study is to assess the temporal stability of retrospective self-reports, using the self-administered questionnaire employed by Dube et al. (2004),

among young subjects who were identified as having been maltreated in childhood. This study has two main novel contributions. While the majority of the studies have been focusing on reliability reports of sexual and physical abuse (e.g., Fergusson et al. 2000), we assessed the reliability report across 10 categories of childhood maltreatment. The other advantage is the nature of the sample. We evaluated youths who were identified in childhood by Child Protective Services (CPS) because they were exposed to more adverse experiences in their lives than other youths that have not reached CPS status. The CPS cases are inherently more likely to be more severe, chronic, repeated, and occur earlier in life, allowing us to assess the temporal stability of the report of experiences that are officially documented.

Based on Widom et al.’s (2004) hypothesis, we expected low levels of test-retest reliability because the life stories of these adolescents and young adults may not be fixed, in contrast to Dube et al.’s (2004) findings with 64-year-old subjects. We also examined the relation of current health status with reliability reports, as suggested by Widom et al. (2004). We expect positive correlations between the changes in self-reported experiences and the changes in self-reported psychopathology.

Method

Participants

Participants were selected from a sample ($N=136$) in which abused and neglected children were identified by Child Protective Services (CPS) during childhood and assessed later in adolescence. This initial sample included 86 youths who were removed from their homes during childhood and placed in child and youth residential care and 50 youths who remained with their families after identification. During the first evaluation, we asked the 86 institutionalized youths to participate in a second evaluation. We selected these youths for three reasons. First, it is easier to locate youths living in institutions than youths living with families. Second, according to official data, the institutionalized youths were exposed to more serious forms of abuse/neglect and were removed from home to protect them from harm. Third, we excluded youths living with parents to avoid possible interference or pressure from parents to not self-disclose the occurrence of maltreatment between evaluations.

Six months later, of the 86 institutionalized youths, 79 were located and participated in the second evaluation (ages ranging from 14 and 23 years; 41 males, 38 females; M age = 17 years; $SD=2.22$). In all cases, the children lived with their family for at least 5 years before being identified by CPS prior to the age of 13. The participants were identified as victims between 1999 and 2006, and they were removed from their

homes after CPS intervention and placed in child and youth residential care. The retrospective reports of childhood adversity were obtained by self-administered questionnaires during adolescence and early adulthood, at least 4 years after the children's identification by CPS. All participants were from Northern Portugal. See Table 1 for the demographic characteristics of the sample.

Measures

ACE study questionnaire (Felitti et al. 1998). A Portuguese version of this questionnaire was used (Silva and Maia 2008). This is a self-administered paper questionnaire that includes detailed information on ten adverse childhood experiences, organized into two areas: children's experiences and household dysfunction. In the original questionnaire, respondents were asked to focus on experiences occurring during their first 18 years of life. However, given the participants' ages in our study, we asked the participants to focus on experiences occurring during their first 12 years of life. The five categories of children's experiences included emotional abuse, defined by three items (how often did a parent, stepparent, or adult living in your home swear at you, insult you, or put you down?); physical abuse, evaluated with four items (during the first 12 years of life, did a parent, stepparent, or adult living in your home push, grab or slap you, or throw something at you?); and sexual abuse, assessed with four items (e.g., during the first 12 years of life, did an adult, relative, family friend, or stranger, at least 5 years older, ever touch or fondle your body in a sexual way?). The evaluation of

emotional neglect was based on four reverse items (e.g., my family was a source of strength and support), and five additional items evaluated physical neglect (e.g., I did not have enough to eat). The response choices included *never*, *once or twice*, *sometimes*, *often*, or *very often*, with the exception of sexual abuse, for which a dichotomous response (*yes* or *no*) was given.

The evaluation of household dysfunction included questions about how their mother was treated violently, and was assessed with three items (e.g., during the first 12 years of life, how often did your father, stepfather, or mother's boyfriend do any of these things to your mother or stepmother: push, grab, slap, or throw something at her?). The responses for mother treated violently were the same as the five categories of children's experiences, ranging from *never* to *very often*. Household substance abuse was evaluated by two items (e.g., during the first 12 years of life, did you live with anyone who used drugs?). The category mental illness or suicide in family was evaluated by two items (e.g., was a household member depressed or mentally ill?). The other two categories of household dysfunction (parental separation or divorce, and incarcerated household members) were evaluated with one item each (did a household member go to prison?). The responses for these last four categories were dichotomous (*yes* or *no*), and an affirmative response to these questions indicated childhood exposure to each category of household dysfunction.

All items for the 10 different examples of childhood adversity can be dichotomized (*yes* or *no*) based on how often the experiences occurred (see Felitti et al. 1998). We used the total score of the items in each category to examine the consistency of the two evaluations, with the exception of the parental separation or divorce and incarcerated household member categories because they were evaluated with one item each and as a dichotomous response. We decided to use the total score in each category instead of a dichotomous format to avoid unbalanced response distributions, which may produce more unstable results (Comrey 1988). The reliability of the ACE Study Questionnaire was tested by Dube et al. (2004) using a kappa statistic for variables coded dichotomously, including: emotional abuse .66 (95 % CI, .55–.76), physical abuse .55 (95 % CI, .47–.63), and sexual abuse .69 (95 % CI, .61–.77). The kappa coefficients for each category of household dysfunction were as follows: mother treated violently .77 (95 % CI, .68–.85), household substance abuse .75 (95 % CI, .68–.81), mental illness in household .51 (95 % CI, .42–.61), incarcerated household members .46 (95 % CI, .27–.65), and parental separation or divorce .86 (95 % CI, .81–.91).

Rotterdam symptom checklist (RSCL; Haes, Van Knippenberg, and Neijt 1990; Portuguese version of Gameiro 1999). This checklist evaluated the occurrence and intensity of 29 physical (e.g., lack of appetite, tiredness, muscular and

Table 1 Participants' Characteristics

Characteristic	Frequencies	
	<i>n</i>	(%)
Sex		
Female	38	48.1
Male	41	51.9
Years of education		
13–15	15	18.9
9–12	52	65.8
4–8	10	12.7
Age in years		
14–16	43	54.4
17–19	27	34.2
20–23	9	11.4
Occupation		
Students	72	91.1
Working	6	7.6
Unemployed	1	1.3

back pain, difficulty sleeping, diarrhea, dry mouth, and nausea) and psychological symptoms (e.g., irritability, depressed mood, difficulty sleeping, despair about the future, tension, anxiety and difficulty concentrating). Each item was scored on 4-point Likert scale ranging from 0 (*never*) to 4 (*very often*), with higher scores indicating greater symptom severity. Current health status was assessed through the total score of physical and psychological symptoms. The internal consistency reliability coefficient in the present sample was ($\alpha=.88$), equivalent to the Portuguese adaptation (Gameiro 1999).

Procedure

We made formal contact with the National Committee for Children and Youth at Risk (CPCJR) for permission to conduct the research, and formal ethical permission was granted by the National Commission for Data Protection (NCDP). Only cases of child abuse and neglect identified by CPS were included in the sample. All participants provided informed consent at each data collection in accordance with procedures approved by the NCDP. We began by examining the official records of children identified by CPS between 1999 and 2006 to identify documented data on experiences of abuse and neglect. The confirmed cases were selected, and the researcher visited the institutions to request permission for data collection. Retrospective reports were obtained in the institutions at two distinct temporal moments (test-retest method) with a 6-month interval. We selected a 6-month time interval in the test-retest method considering that, according to the Portuguese law, the institutionalization (short period) cannot be extended for more than six months, and after this period it would be difficult to find the participants. Although other studies have used a longer time interval (e.g., Dube et al. 2004), a time interval of six months has also been used (Bernstein, et al. 1994).

The current health status and the retrospective self-reports of childhood adversity were obtained under self-administration paper questionnaires. The questionnaires were completed in a private room with the researcher a few meters away and available to assist if necessary. The aims of the study were explained by the researcher to each respondent. To ensure confidentiality in all cases, names and personal data were codified. The use of a code number ensured that information about the child could only be identified by the researcher, who maintained the data from records and questionnaires in secure conditions.

Statistical Analysis

We used Pearson correlations to examine the association between the total adversity reported at the two moments of

evaluation. We used Intraclass Correlations Coefficient (ICC; McGraw and Wong 1996) to estimate the test-retest reliability of the 79 participants' responses to the ACE categories and questions at the first and second evaluations. We used reliability parameters because for repeated measurements on a continuous scale an ICC is the most appropriate measure, as suggested by Vet, Terwee, Knol, and Bouter (2006). ICCs are designated as ≤ 0.40 poor to fair agreement, 0.41–0.60 moderate agreement, 0.61–0.80 good agreement, and 0.81–1.00 excellent agreement. We used a Pearson correlation to test the hypothesis of the relationship of health status to self-reports, examining the correlation between the changes (discrepancies) in self-reported experiences and the changes in self-reported physical and psychological symptoms. In this analysis, the score for total childhood adversity evaluated at the second time point (T2) minus the score at the first time point (T1) was related to the discrepancy in physical and psychological symptoms (score at T2 minus score at T1).

Results

A significant positive correlation was found between the total scores for adversity at both evaluation times ($r=.601$, $p<.001$). A significant positive correlation was found between the total score for adversity and physical and psychological symptoms at the first evaluation ($r=.362$, $p<.001$), and the second evaluation ($r=.367$, $p<.001$). However, we did not find significant correlation between the changes in self-reported experiences at T1 and T2, and the changes in self-reported symptoms at T1 and T2 ($r=-.057$, $p=.693$).

Table 2 presents the ICCs for each of the five categories of abuse and neglect, including the questions. We included additional information in analyses, presented by gender and information about response norms and variability, to contribute for the understanding of the results. The highest ICC among the categories was physical abuse, and the lowest ICC was emotional abuse and physical neglect, 95 % CIs .96 [.94, .98], .65 [.46, .78], and .65 [.44, .78], respectively. The highest ICC among the questions was “how hard did they hit?” from physical abuse, and the lowest ICC was “have you touch their body in a sexual way?” and “have oral, anal, or vaginal intercourse with you?” from sexual abuse, 95 % CIs .86 [.78, .91], .38 [.17, .55], .38 [.36, .74], respectively.

Table 3 presents the ICCs for each of the five categories of household dysfunction, including the questions. We also include additional information in analyses as gender and information about response norms and variability. The highest ICC among the categories was for mother treated violently, and the lowest ICC was for incarcerated household member, 95 % CIs .84 [.75, .90] .62 [.46, .74], respectively. The highest ICC among the questions was, “sometimes, often, or very often kicked, bitten, hit with a fist, or hit with something hard?” and

Table 2 Intraclass Correlation Coefficients for Categories and Items of Abuse and Neglect (N=79)

Abuse and neglect by category and item	First assessment <i>M (SD) n (%)</i> ¹	6-month follow-up <i>M (SD) n (%)</i> ¹	Males ICC	Females ICC	Total sample ICC <i>n (%)</i> ²	95 % Confidence interval
Emotional Abuse (How often did a parent or other adult in the household)	0.59 (0.97) 27 (34.2)	0.78 (1.17) 29 (36.7)	.61 ^b	.68 ^c	.65 ^c 18 (22.8)	.46–.78
Insult you or swear?	1.58 (1.38) 19 (24.0)	1.59 (1.40) 23 (29.1)	.73 ^c	.56 ^b	.67 ^c 18 (22.8)	.47–.79
Threaten to hit you or throw something but not do it?	1.37 (1.36) 17 (21.5)	1.38 (1.50) 23 (29.1)	.61 ^b	.82 ^c	.72 ^c 17 (21.5)	.56–.82
Made you afraid of being physically hurt?	0.94 (1.31) 11 (13.9)	0.92 (1.33) 16 (20.3)	.66 ^c	.66 ^c	.67 ^c 13 (16.5)	.47–.79
Physical Abuse (Did a parent or other adult in the household)	0.76 (1.23) 28 (35.4)	0.73 (1.14) 30 (38)	.95 ^c	.97 ^c	.96 ^c 6 (7.6)	.94–.98
Pushed, grabbed or threw you something?	0.90 (1.29) 13 (16.5)	0.80 (1.18) 11 (13.9)	.66 ^b	.81 ^c	.76 ^c 16 (20.3)	.62–.85
Hit you so hard it left marks or bruises?	0.85 (1.32) 14 (17.7)	0.77 (1.34) 18 (22.8)	.58 ^b	.64 ^b	.61 ^c 18 (22.8)	.38–.76
How often did they hit?	0.96 (1.18) 11 (13.9)	0.92 (1.11) 12 (15.2)	.84 ^c	.77 ^c	.81 ^c 11 (13.9)	.69–.88
How hard did they hit?	1.30 (1.44) 22 (27.8)	1.26 (1.41) 25 (31.6)	.77 ^c	.92 ^c	.86 ^c 17 (21.5)	.78–.91
Sexual Abuse (Did an adult or person at least 5 years older ever)	0.54 (1.20) 16 (20.3)	0.45 (1.02) 15 (19.0)	.69 ^c	.79 ^c	.75 ^c 11 (13.9)	.61–.84
Touch or fondle you in a sexual way?	0.18 (0.38) 13 (16.5)	0.17 (0.38) 13 (16.5)	.06	.83 ^c	.45 ^c 12 (15.2)	.25–.61
Have you touch their body in a sexual way?	0.10 (0.30) 7 (8.9)	0.09 (0.29) 7 (8.9)	.45 ^b	.29 ^a	.38 ^c 8 (10.1)	.17–.55
Attempt oral, anal, or vaginal intercourse with you?	0.15 (0.36) 11 (13.9)	0.13 (0.34) 10 (12.7)	.65 ^c	.59 ^c	.62 ^c 7 (8.9)	.46–.74
Actually have oral, anal, or vaginal intercourse with you?	0.11 (0.32) 8(10.1)	0.08 (0.27) 6(7.6)	.37 ^a	.38 ^b	.38 ^a 8 (10.1)	.17–.55
Emotional Neglect	1.30 (1.36) 44(55.6)	1.16 (1.57) 31(39.2)	.65 ^b	.84 ^c	.75 ^c 20 (25.3)	.60–.84
Someone in my family helped me feel special or important.	2.15 (1.49) 29 (36.7)	2.31 (1.42) 21 (26.6)	.54 ^a	.65 ^b	.59 ^c 22 (27.8)	.36–.74
Everyone in my family looked after each other.	2.25 (1.41) 24 (30.4)	2.36 (1.31) 19 (24)	.65 ^b	.80 ^c	.73 ^c 13 (16.5)	.57–.83
Everyone in my family felt close to each other.	2.15 (1.37) 25(31.6)	2.31 (1.41) 24(30.4)	.69 ^c	.66 ^b	.67 ^c 15 (19)	.47–.79
My family was a source of strength and support.	2.29 (1.46) 23 (29.1)	2.39 (1.53) 22 (27.4)	.45 ^a	.88 ^c	.71 ^c 15 (19)	.53–.81
Physical Neglect	0.82 (1.08) 34 (43)	0.73 (1.05) 31 (39.2)	.46 ^a	.80 ^c	.65 ^c 19 (24.1)	.44–.78
I did not have enough to eat.	1.06 (1.43) 14 (17.7)	0.76 (1.06) 9 (11.4)	.35 ⁺	.43 ^a	.50 ^b 17 (21.5)	.21–.69
I knew there was someone to care of and protect me.	2.47 (1.46) 20(25.3)	2.68 (1.43) 19(24.1)	.50 ^a	.64 ^b	.57 ^c 17 (21.5)	.32–.73
My parents were too drunk or disturbed to take care of the family.	0.95 (1.30) 11 (13.9)	0.86 (1.14) 13 (16.5)	.50 ^b	.80 ^c	.76 ^c 10 (12.7)	.62–.85
There was someone to do laundry.	3.11 (1.26) 9 (11.4)	3.15 (1.28) 9 (11.4)	.54 ^b	.54 ^b	.55 ^c 12 (15.2)	.28–.71
There was someone to take me to the doctor if necessary.	2.92 (1.15) 10 (12.7)	2.88 (1.34) 13 (16.5)	.35 ⁺	.60 ^b	.45 ^c 13 (16.5)	.13–.66

ICC Intraclass correlation coefficient

¹ number and percentage of participants responding in the affirmative; ² number and percentage of participants responding inconsistently

^a <.05; ^b <.01; ^c <.001

the lowest ICC was, “was a household member depressed or mentally ill?” 95 % CIs .85 [.77, .91] .43 [.22, .60], respectively.

Discussion

Widom et al. (2004) suggested that the findings of Dube et al. (2004), reflecting moderate to good reliability of the retrospective reports, were influenced by the participants’ age (M=64) because the memory of the life story at this age is well

established and fixed. These authors hypothesized that young adult respondents, such as the respondents in the study by Fergusson et al. (2000), would produce poorer reliability because the life story of the participants may not yet be fixed.

The results showed ICC values greater than or equal to .65, representing good to excellent agreement across all 10 categories of adversity. Emotional and physical neglect, categories that were not assessed by Dube et al. (2004), showed good agreement in our study. In fact, the agreement found in this study, with a younger sample, was greater than in Dube et al.’s (2004) results. Curiously, physical abuse had the highest value

Table 3 Intraclass Correlation Coefficients for Categories and Items of Household Dysfunction ($N=79$)

Household Dysfunction by category and item	First assessment $M(SD) n (%)^1$	6-month follow-up $M(SD) n (%)^1$	Males ICC	Females ICC	Total sample ICC $n (%)^2$	95 % Confidence interval
Parental separation or divorce	0.48 (0.50) 37 (46.8)	0.43 (0.50) 33 (41.8)	.79 ^c	.79 ^c	.79 ^c 8 (10.1)	.69–.86
Mother treated violently (Was your mother or stepmother)	1.03 (1.31) 33 (41.8)	1.24 (1.37) 39 (49.4)	.77 ^c	.87 ^c	.84 ^c 16 (20.3)	.75–.90
Pushed, grabbed, slapped, or had something thrown at her?	1.11 (1.48) 27 (34.2)	1.16 (1.38) 34 (43)	.57 ^b	.88 ^c	.82 ^c 15 (19.0)	.71–.89
Kicked, bitten, hit with a fist, or hit with something hard?	1.22 (1.51) 31 (39.2)	1.09 (1.39) 33 (41.8)	.80 ^c	.86 ^c	.85 ^c 12 (15.2)	.77–.91
Ever repeatedly hit over at least a few minutes?	0.97 (1.40) 23 (29.1)	0.97 (1.31) 31 (39.2)	.75 ^c	.84 ^c	.83 ^b 16 (20.3)	.73–.89
Substance abuse	0.73 (0.69) 45 (57)	0.69 (0.66) 44 (55.7)	.80 ^c	.56 ^b	.70 ^c 17 (21.5)	.52–.81
Lived with anyone who used street drugs?	0.15 (0.36) 12 (15.2)	0.13 (0.34) 9 (11.4)	.84 ^c	.48 ^b	.58 ^c 7 (8.9)	.30–.65
Lived with anyone who was a problem drinker or alcoholic?	0.58 (0.50) 45 (57)	0.58 (0.50) 44 (55.7)	.74 ^c	.35 ^a	.49 ^c 23 (29.1)	.30–.64
Mental illness or suicide in family	0.52 (0.72) 29 (36.7)	0.38 (0.67) 21 (26.6)	.67 ^b	.79 ^c	.74 ^c 18 (22.8)	.59–.84
Was a household member depressed or mentally ill?	0.36 (0.48) 27 (34.2)	0.22 (0.42) 17 (21.5)	.44 ^a	.54 ^c	.43 ^c 18 (22.8)	.22–.60
Did a household member attempt suicide?	0.15 (0.36) 12 (15.2)	0.16 (0.37) 12 (15.2)	.49 ^a	.66 ^c	.51 ^b 10 (12.7)	.32–.66
Incarcerated household member	0.30 (0.46) 24 (30.4)	0.26 (0.44) 20 (25.3)	.87 ^c	.60 ^b	.62 ^c 12 (15.2)	.46–.74

ICC Intraclass correlation coefficient

¹ number and percentage of participants responding in the affirmative; ² number and percentage of participants responding inconsistently

^a <.05; ^b <.01; ^c <.001

of ICC in contrast with Fergusson et al.'s (2000) findings that showed lower reliability. Maybe, this discrepancy is as a result of having used different measures of abuse identification, and different mode of administering the questions. The physical abuse evaluated by the ACE questionnaire focuses on specific and objective parent behaviors, which seems to have contributed to obtain high reliabilities. These findings suggest that focusing in specific behaviors avoids the judgment and interpretation about the experiences, not dependent on reconstructions about what must have occurred. One of the influences for report instability is the form of questioning about childhood maltreatment. More ambiguous questions might lead to different interpretations at different moments in time. The language should be simple and appropriate for the reading level of the scale's target population and understandable by respondents with only a modest education. In addition, it should avoid colloquialisms and other language that varies with age, ethnicity, region, and gender (Clark and Watson 1995). The use of simple language and a focus on specific acts of maltreatment help to overcome potential cultural differences, given that childhood maltreatment and family dysfunction vary from society to society. Another factor that appears to have contributed to high reliabilities was the mode of administering the questions. The problems of reporting (e.g., misreporting, nonresponse or reporting error) in sensitive information have been related to the face-to-face interviewer method because the respondent may be reluctant or embarrassed to report his or her answers to another person

(Tourangeau and Yan 2007). Conversely, under the self-administered questionnaire method, the mere presence of the researcher, as in our study, does not appear to have a substantial effect on the answers (Tourangeau and Yan 2007).

While the findings of the study showed good to excellent agreement across all 10 categories of adversity, the results showed low ICCs in questions of sexual abuse and mental illness. The differences in reliability found with certain questions could be explained by the way those questions were constructed as well as the response options available. The subjectivity and difficulty of interpreting what constitutes mental illness could produce unstable responses. In the area of sexual abuse, it is possible that the phrase "in a sexual way" is ambiguous, and the measurement of each item was limited to "yes or no" responses rather than the full range of response options given for the other forms of abuse, as well as the total score of the items under each category. Certain authors have criticized dichotomous responses, arguing that biases are more problematic with this format than with multiple-choice items (e.g., Comrey 1988). In addition, different reliabilities may arise as a result of differences in the baseline rates of various forms of abuse. Very low baseline-rate events, such as sexual abuse, could result in a measurement with greater instability. This type of abuse is often a major traumatic event, and traumatized children are usually more vulnerable to distortions and errors in recall (Chu, Frey, Ganzel, and Matthews 1999). According to Summit (1983), victims of sexual abuse may respond with self-blame and self-doubt and produce a

series of disclosures of abuse followed by recantations of these disclosures. This author described how sexually abused children disclose abuse and included delay, conflict, unconvincing responses, and retraction of disclosures.

Differences between boys and girls may help to explain the instability of sexual abuse responses. The data showed that the ICC values of categories and questions were equivalent for boys and girls. However, boys and girls diverged on specific categories and questions. The largest discrepancy was observed in the question about being sexually touched. Girls were very consistent in their responses to this question, whereas boys were not. However, both boys and girls were very unstable in responses to questions about being sexually abused. Several studies suggest that boys are more reluctant to disclose sexual abuse than girls (Stroud, Martens, and Barker 2000) and experience different forms of abuse, such as fondling, oral, and anal intercourse (Fontanella, Harrington, and Zuravin 2000). However, the disclosure is also affected by other factors, such as prior disclosure and the relationship to the perpetrator, as well as developmental, cognitive, and socio-emotional factors (Goodman-Brown, Edelstein, Goodman, Jones, and Gordon 2003; London, Bruck, Hopkins, Ceci, and Shuman 2005). In summary, a definitive interpretation of these findings is difficult to achieve. This problem will be the subject of future work requiring the consideration of gender differences in disclosure.

In this study, we examined the influence of physical and psychological symptoms on self-reports at both evaluations, as suggested by Widom et al. (2004). However, we did not find significant correlation between the changes in self-reported experiences and the changes in self-reported symptoms. These findings are consistent with previous research, which suggests that the reporting of adverse childhood experiences is not influenced by health state at the time of the report (Brewin, Andrews, and Gotlib 1993; Fergusson, et al. 2000; Monteiro and Maia 2010; Robins, Schoenberg, Holmes, Ratcliff, Benham, and Works 1985).

This study has some limitations. The restriction of the sample affects the generalizability of the findings. In addition, the total ACE correlation was .60; only 36 % of the total variance was accounted for by mono-method reports. We cannot explain the high level (64 %) of unshared variance. Moreover, it is possible that this study found no changes in health complaints in adversity reporting due to characteristics of the sample. Young people are generally healthy, whereas substantial variability in health status is apparent at later ages. Future studies should use heterogeneous age samples to further examine the effect of health in reliability reporting.

Despite the limitations of this study, the findings support the retest reliability for most of the experiences of the ACE questionnaire. When the questions are focused on specific behaviors, retrospective reports of childhood adversity are sufficiently stable. Our findings also suggest that the stability

of retrospective self-reports about childhood adversity is not related to the health state at the time of the report. This finding is particularly important because several case-control studies have been performed with early life adversity self-report assessment and its relation to later life outcomes.

Acknowledgments The authors thank the *Fundação para a Ciência e Tecnologia* (grant BD, FCT - SFRH/BD/45414/2008) for financing this project. The authors also express appreciation to the staff of the *Comissão de Protecção das Crianças e Jovens*.

References

- Bernstein, D., Fink, L., Handelsman, L., Foote, J., Lovejoy, M., Wenzel, K., et al. (1994). Initial reliability and validity of a new retrospective measure of child abuse and neglect. *American Journal of Psychiatry*, *151*, 1132–1136.
- Brewin, C. R., Andrews, B., & Gotlib, I. H. (1993). Psychopathology and early experience: a reappraisal of retrospective reports. *Psychological Bulletin*, *113*, 82–89.
- Chu, J. A., Frey, L. M., Ganzel, B. L., & Matthews, J. A. (1999). Memories of childhood abuse: Dissociation, amnesia, and corroboration. *American Journal of Psychiatry*, *156*, 749–755.
- Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, *7*, 309–319.
- Comrey, A. L. (1988). Factor-analytic methods of scale development in personality and clinical psychology. *Journal of Consulting and Clinical Psychology*, *56*, 754–761.
- DiLillo, D., DeGue, S., Kras, A., Di Loreto-Colgan, A. R., & Nash, C. (2006). Participant responses to retrospective surveys of child maltreatment: Does mode of assessment matter? *Violence and Victims*, *21*, 410–424.
- Dube, S. R., Williamson, D. F., Thompson, T., Felitti, V. J., & Anda, R. F. (2004). Assessing the reliability of retrospective reports of adverse childhood experiences among adult HMO members attending a primary care clinic. *Child Abuse & Neglect*, *28*, 729–737. doi:10.1016/j.chiabu.2003.08.009.
- Fergusson, D. M., Horwood, L. J., & Woodward, L. J. (2000). The stability of child abuse reports: A longitudinal study of the reporting behaviour of young adults. *Psychological Medicine*, *30*, 529–544. doi:10.1017/S0033291799002111.
- Felitti, V. J., Anda, R. F., Nordenberg, D., Williamson, D. F., Spitz, A. M., Edwards, V., et al. (1998). Relationship of childhood abuse and household dysfunction to many of the leading causes of death in adults. *American Journal of Preventive Medicine*, *14*, 245–258.
- Fontanella, C., Harrington, D., & Zuravin, S. J. (2000). Gender differences in the characteristics and outcomes of sexually abused preschoolers. *Journal of Child Sexual Abuse*, *9*, 21–40.
- Gameiro, M. (1999). *Sofrimento na doença*. Coimbra: Quarteto Editora.
- Goodman-Brown, T. B., Edelstein, R. S., Goodman, G. S., Jones, D. P. H., & Gordon, D. S. (2003). Why children tell: A model of children's disclosure of sexual abuse. *Child Abuse & Neglect*, *27*, 525–540.
- Haes, J. C., Van Knippenberg, F. J., & Neijt, J. P. (1990). Measuring psychological and physical distress in cancer patients: Structure and application of the Rotterdam Symptom Check List. *British Journal of Cancer*, *62*, 1034–1038.
- Hardt, J., & Rutter, M. (2004). Validity of adult retrospective reports of adverse childhood experiences: Review of the evidence. *Journal of Child Psychology and Psychiatry*, *45*, 260–273. doi:10.1111/j.1469-7610.2004.00218.x.

- Herrenkohl, T. I., Hong, S., Klika, J. B., Herrenkohl, R. C., & Russo, M. J. (2013). Developmental impacts of child abuse and neglect related to adult mental health, substance use, and physical health. *Journal of Family Violence*, 28, 191–199. doi:10.1007/s10896-012-9474-9.
- Koss, M. P., & Gidycz, C. A. (1985). Sexual experiences survey: Reliability and validity. *Journal of Consulting and Clinical Psychology*, 53, 422–423. doi:10.1037/0022-006X.53.3.422.
- London, K., Bruck, M., Hopkins, J., Ceci, S. J., & Shuman, D. W. (2005). Disclosure of child sexual abuse: What Does the Research Tell Us About the Ways That Children Tell? *Psychology, Public Policy, and Law*, 11, 194–226.
- MacKenzie, S., Podsakoff, P., & Jarvis, C. (2005). The problem of measurement model misspecification in behavioral and organizational research and some recommended solutions. *Journal of Applied Psychology*, 90, 710–730.
- McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, 1, 30–46. doi:10.1037/1082-989X.1.1.30.
- Monteiro, I. S., & Maia, A. (2010). Family childhood experiences reports in depressed patients: comparison between 2 time points. *Procedia Social and Behavioral Sciences*, 5, 541–547. doi:10.1016/j.sbspro.2010.07.139.
- Norris, F. H., & Hamblen, J. L. (2004). Standardized self-report measures of civilian trauma and PTSD. In J. P. Wilson & T. M. Keane (Eds.), *Assessing psychological trauma and PTSD* (pp. 63–102). New York: Guilford Publications.
- Ramiro, S. L., Madrid, B. J., & Brown, D. W. (2010). Adverse childhood experiences (ACE) and health-risk behaviors among adults in a developing country setting. *Child Abuse & Neglect*, 34, 842–855.
- Raphael, K. G., & Cloitre, M. (1994). Does mood-congruence or causal search govern recall bias? A test of life event recall. *Journal of Clinical Epidemiology*, 47, 555–564.
- Robins, L. N., Schoenberg, S. P., Holmes, S. J., Ratcliff, K. S., Benham, A., & Works, J. (1985). Early home environment and retrospective recall: A test for concordance between siblings with and without psychiatric disorders. *American Journal of Orthopsychiatry*, 55, 27–41.
- Schraedley, P. K., Turner, R. J., & Gotlib, I. H. (2002). Stability of retrospective reports in depression: Traumatic events, past depressive episodes, and parental psychopathology. *Journal of Health and Social Behavior*, 43, 307–316.
- Sedlak, A. J., Mettenberg, J., Basena, M., Petta, I., McPherson, K., Greene, A., et al. (2010). *Fourth National Incidence Study of Child Abuse and Neglect (NIS-4): Report to Congress, Executive summary*. Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families.
- Shaffer, A., Huston, L., & Egeland, B. (2008). Identification of child maltreatment using prospective and self-report methodologies: A comparison of maltreatment incidence and relation to later psychopathology. *Child Abuse & Neglect*, 32, 682–692. doi:10.1016/j.chiabu.2007.09.010.
- Silva, S., & Maia, A. (2008). Versão Portuguesa do Family ACE Questionnaire (Questionário de História de Adversidade na Infância). In A. P. Noronha, C. Machado, L. Almeida, M. Gonçalves, S. Martins, & V. Ramalho (Eds.), *Actas da XIII conferência avaliação psicológica: Formas e contextos*. Psiquibros Edições: Braga.
- Stroud, D., Martens, S. L., & Barker, J. (2000). Criminal investigation of child sexual abuse: A comparison of cases referred to the prosecutor to those not referred. *Child Abuse & Neglect*, 24, 689–700.
- Summit, R. C. (1983). The child sexual abuse accommodation syndrome. *Child Abuse & Neglect*, 7, 177–193.
- Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin*, 133, 859–883. doi:10.1037/0033-2909.133.5.859.
- Vet, C. W., Terwee, C. B., Knol, D. L., & Bouter, L. M. (2006). When to use agreement versus reliability measures. *Journal of Clinical Epidemiology*, 59, 1033–1039.
- Widom, C. S., Raphael, K. G., & DuMont, K. A. (2004). The case for prospective longitudinal studies in child maltreatment research: Commentary on Dube, Williamson, Thompson, Felitti, and Anda (2004). *Child Abuse & Neglect*, 28, 715–722. doi:10.1016/j.chiabu.2004.03.009.
- Wolfe, D. A., Scott, K., Wekerle, C., & Pittman, A. (2001). Child maltreatment: Risk of adjustment problems and dating violence in adolescence. *Journal of American Academy of Child Adolescent Psychiatry*, 40, 282–289.
- Wyatt, G. E., & Peters, S. D. (1986). Methodological considerations in research on the prevalence of child sexual abuse. *Child Abuse & Neglect*, 10(2), 241–251.