Child maltreatment under the skin: Basal activity and stress reactivity of the autonomic nervous system and attachment representations in maltreating parents

Reijman, S.

Author: Reijman, Sophie
Title: Child maltreatment under the skin : basal activity and stress reactivity of the autonomic nervous system and attachment representations in maltreating parents
Issue Date: 2015-12-16

CHAPTER 5

STRESS REACTIVITY IN MALTREATING PARENTS AND AT-RISK ADULTS: REVIEW AND META-ANALYSES

Sophie Reijman, Marian J. Bakermans-Kranenburg, Regina Hiraoka, Julie L. Crouch, Joel S. Milner, Lenneke R. A. Alink, and Marinus H. van IJzendoorn. *An adjusted version of this manuscript has been submitted for publication.*

Abstract

We reviewed and meta-analyzed 11 studies (*N* = 524) that examined the association between (risk for) child maltreatment perpetration and basal autonomic activity and 11 studies (*N* = 503) that examined the association between (risk for) child maltreatment and autonomic stress reactivity. We hypothesized that higher basal levels of autonomic activity and increased stress reactivity would be found in maltreating parents and participants at risk for being abusive. The narrative review showed that evidence from significance testing within and across studies was mixed. Results of the first meta-analysis revealed that (risk for) child maltreatment was associated with higher levels of baseline autonomic nervous system (ANS) activity (*g* = 0.24) when measures included indices influenced by both the sympathetic and the parasympathetic branches (such as heart rate and blood pressure). The second meta-analysis yielded no differences in ANS stress reactivity between maltreating/at-risk participants and non-maltreating/low-risk comparison groups. Power analyses showed that most studies reviewed were underpowered. Results are discussed within the framework of allostatic load, and future directions for research are suggested.

Key words: child maltreatment, review, autonomic nervous system, stress, meta-analysis

Introduction

The possibility that dysregulated psychophysiology may serve as a risk factor for child maltreatment has been the topic of long-standing (albeit intermittent) research interest, particularly with respect to the activity of the autonomic nervous system (ANS). The ANS may be considered an endophenotype with relevance for the etiology of child maltreatment because of its role in emotion regulation and behavioral responsiveness (Stemmler, 2004; Sturge-Apple, Skibo, Rogosch, Ignjatovic, & Heinzelman, 2011). Autonomic stress reactivity may provide real-time insight into the nature of maltreating parents' responses to their children. Prior research has suggested increased ANS (re)activity in maltreating parents and individuals at risk for perpetrating child maltreatment; however, inconsistent findings across as well as within studies have been noted (McCanne & Hagstrom, 1996). Furthermore, the last review of the literature on this topic was conducted approximately 20 years ago (McCanne & Hagstrom, 1996), and effect size estimates for the association between physiological (re)activity and (risk for) child maltreatment have not been assessed using meta-analytic methods. To address this gap in the literature, we reviewed the literature and conducted two meta-analyses: one examining the association between (risk for) child maltreatment and baseline ANS activity levels, and another examining the association between (risk for) child maltreatment and ANS stress reactivity.

The ANS is a component of the biological stress system (Stratakis & Chrousos, 1995). It regulates the visceral organs and consists of the parasympathetic and the sympathetic branches whose functions, generally speaking, lead to opposite effects. The parasympathetic division slows down heart rate and stimulates digestion, promoting the conservation and recuperation of energy (i.e., anabolic processes). The sympathetic nervous system increases heart rate and inhibits digestion, mobilizing the body in response to, or in anticipation of, environmental challenges (i.e., catabolic processes; Viamontes & Nemeroff, 2009). In response to stress, parasympathetic influences are typically inhibited, while the sympathetic nervous system is activated. ANS activation helps individuals adapt to changing circumstances and maintain stability through change, a process called allostasis (McEwen & Seeman, 1999). When stressful environmental demands are chronic or too frequent, physiological responses may become dysregulated and have detrimental consequences, a condition referred to as allostatic load (Beckie, 2012; Sterling & Eyer, 1988). Autonomic dysregulation may underlie maladaptive behavior, including inadequate parenting (Sturge-Apple et al., 2011).

Autonomic stress reactivity as a correlate of child maltreatment was first examined in the 1970s (Disbrow, Doerr, & Caulfield, 1977). In their seminal study, Frodi and Lamb (1980) measured heart rate, blood pressure, and skin conductance in abusive and nonabusive mothers as they viewed videos of a crying/smiling infant. The Frodi and Lamb findings sparked additional interest in the role of autonomic stress reactivity and (risk for) child maltreatment, leading to other studies on this topic (e.g., Casanova, Domanic, McCanne, & Milner, 1992; Crowe & Zeskind, 1992; Friedrich, Tyler, & Clark, 1985; Pruitt, & Erickson, 1985; Stasiewicz & Lisman, 1989); Wolfe, Fairbank, Kelly, & Bradlyn, 1983). A narrative review summarizing the findings of these early studies concluded that maltreating/at-risk participants appeared to exhibit increased autonomic reactivity to stressors compared to their nonmaltreating/low-risk counterparts (McCanne & Hagstrom, 1996). This was consistent with the aggressive nature of child physical abuse (Lorber, 2004), the maltreatment subtype on which most of the reviewed studies focused. On the other hand, findings were notably mixed, possibly because studies have varied with respect to sample characteristics (e.g., ranging from maltreating parents to nonparents at risk for child maltreatment), standardized stressors (e.g., ranging from recordings of infant cry sounds to nonchild-related tasks such as solving anagrams), and autonomic measures assessed (e.g., skin conductance, heart rate). Two recent studies found that maltreating and at-risk parents may respond with *less* autonomic arousal to stress than their respective comparison groups (Crouch et al., 2015; Reijman et al., 2014, 2015).

Considering the differences in methodology and findings across studies, and the long lapse since the last review (McCanne & Hagstrom, 1996), the present study was designed to provide an updated narrative review of the literature with a focus on the following research questions: (1) Is (risk for) child maltreatment associated with increased ANS baseline activity? (2) Is (risk for) child maltreatment associated with increased ANS stress reactivity? The qualitative approach of the narrative review offers a descriptive synthesis of the evidence, and allows for an in-depth and critical analysis of methodological discrepancies across studies (Petticrew & Roberts, 2006). On the other hand, narrative reviews cannot quantitatively assess the *combined* effect size over studies. A single study may lack statistical power to detect significant differences between groups, and meta-analysis is a tool to assess overall effects across studies. Therefore, we also conducted two meta-analyses to estimate overall effect sizes for (1) the relationship between (risk for) child maltreatment and autonomic baseline activity, and (2) the association between (risk for) child maltreatment and autonomic stress reactivity. In both meta-analyses, we distinguished between mixed indices of the ANS (i.e., under both sympathetic and parasympathetic influence, e.g., heart rate, respiration rate), sympathetic indices (e.g., skin conductance), and parasympathetic indices (e.g., respiratory sinus arrhythmia).

Based on conclusions from the above mentioned review (McCanne & Hagstrom, 1996), we hypothesized that maltreating and at-risk individuals (compared to nonmaltreating/low-risk individuals) would show (1) greater (generic) ANS and SNS baseline arousal and greater stress reactivity (defined as increase of arousal in response to stress), and (2) lower PNS baseline levels and lower levels of PNS stress reactivity. Furthermore, we examined variables that might explain differences in effect sizes across studies. Identification of moderators may explain divergent results and provide valuable directions for future research. Specifically, we looked at the following variables as potential moderators: (a) parenting status (whether participants were parents or not); (b) maltreatment status (whether participants had been substantiated for maltreatment or had been identified as at risk); (c) maltreatment type (physical abuse only vs inclusion of neglect); (d) presentation of stimulus (auditory, visual, or real-life); (e) the percentage of women in the sample; (f) sample size; and (g) year of publication. Power analyses were performed to evaluate the adequacy of the sample sizes of individual studies.

Method

Literature search and inclusion criteria

We used four search methods in order to retrieve relevant studies. Specifically, we searched the databases Embase, PsycInfo, PubMed, and Web of Science using the following search terms: "*child maltreatment*" *OR* "*child abuse*" *OR* "*child neglect*" *OR* "*physical abuse*" *OR* "*physical neglect*" *OR* "*emotional abuse*" *OR* "*emotional neglect*"*) AND (parent* OR mother* OR father* OR caregiv* OR risk) AND (autonomic OR physiolog* OR cardiovascular OR HR OR* "*blood pressure*" *OR DBP OR SBP OR respirat* OR RR OR HRV OR amylase OR sAA OR sympathetic OR electrodermal OR* "*skin conductance*" *OR SCL OR SCR OR parasympathetic OR vagal OR*

RSA.[1] Second, these same terms were used to assess potentially eligible dissertations and conference proceedings. Third, we searched Web of Science for references to pioneering articles (i.e., Frodi & Lamb, 1980; McCanne & Hagstrom, 1996). Finally, the reference sections of eligible articles and dissertations were checked for additional potentially eligible papers. Eligibility was based on three main inclusion criteria: (1) The sample consisted of parents with substantiated child maltreatment, or participants (parents or nonparents) at high risk for child maltreatment as assessed by a validated instrument (e.g., Child Abuse Potential [CAP] Inventory) or defined as such by the authors based on a substantial number of risk factors; (2) at least one index of the ANS was measured; (3) the physiological measurement included ANS baseline activity and/or stress reactivity. The stress-invoking stimulus could be child-related (e.g., video of a crying infant) or nonchild-related (e.g., having to complete a series of anagrams).

From the 1142 studies obtained through the search of electronic databases, a sample of 150 abstracts was randomly selected in order to establish intercoder reliability with respect to decisions about inclusion in the narrative review and meta-analyses. Two of the authors (RH and SR) independently coded the 150 abstracts as either *not eligible* or *eligible* (i.e., selected for inclusion). When abstracts were potentially eligible but did not provide sufficient information to determine eligibility, the full text articles were retrieved and coded. The two authors reached 100% consensus on studies coded as eligible. Having established adequate intercoder reliability, the remaining abstracts obtained from the literature search were divided between RH and SR for independent coding with respect to inclusion or exclusion. In the case of multiple eligible publications reporting (partly) on the same sample, only the publication with the most available physiological data was included. This ensured that every participant was represented just once in each meta-analysis performed in the present study. For instance, Reijman et al. published two papers on autonomic (re)activity in a largely overlapping sample of maltreating mothers (2014; 2015). The latter included one autonomic measure, namely salivary alpha-amylase, while the former included four, i.e. heart rate, vagal tone, pre-ejection period,

[1] Our initial intent was to include studies on the association between perpetration of child maltreatment and hypothalamic-pituitary-adrenal axis (re)activity, and relevant terms were part of the literature search, but this did not render a sufficient number of studies to be separately meta-analyzed.

and skin conductance, and was therefore selected for inclusion in the meta-analyses. Twelve studies published or submitted for publication between 1977 and 2015 were identified as eligible for inclusion in the present study: 11 included ANS baseline measures and 11 included ANS stress reactivity (10 studies reported both ANS baseline measures and ANS stress reactivity and thus were included in both meta-analyses). Narrative reviews of these 12 studies are provided below (see also Table 1).

When participants were exposed to multiple stressors/stimuli (Casanova et al., 1992; Friedrich et al., 1985; Frodi & Lamb, 1980; Pruitt & Erickson, 1985), we selected one stressor from each study for inclusion in the meta-analyses. This was done for several reasons. First, it ensured that each participant would be represented only once in each meta-analysis. Including multiple effect sizes for samples exposed to multiple stressors would have given more weight to those samples than to others. Alternatively, we could have calculated one combined effect size for the multiple stressors, but this strategy would have made studies less comparable and it would also have made moderator analyses for *presentation of stimulus* impossible. The hierarchy of criteria used to select a single stressor from studies that presented multiple stressors was as follows: a) psychosocial stimuli such as cry sounds were preferred over physiological tasks such as immersing a foot in ice-cold water; b) child-related stimuli were considered more relevant than nonchild-related stimuli; and c) stress-invoking stimuli were selected over nonstress-invoking stimuli. These eligibility criteria led to the inclusion of the stressful film task in the Casanova et al. (1992) study, the audiotaped infant crying in the Friedrich et al. (1985) study, and the video of the crying infant in the Frodi and Lamb (1980) and Pruitt and Erickson (1985) studies.

Narrative review

Parents with substantiated child maltreatment. Six studies included parents who had been substantiated for abuse and/or neglect of their children. Creaven, Skowron, Hughes, Howard, and Loken (2014) recruited 52 mother-child dyads in which the mother had been a perpetrator of child abuse or neglect. Child Protective Services (CPS) records were coded using the Maltreatment Classification System (MCS; Barnett, Manly, & Cicchetti, 1993). The group was compared to 52 mother-child dyads without previous CPS records. Maltreating and nonmaltreating mothers did not differ on age, employment status, child age, or child sex, but maltreating mothers were less educated and had lower household incomes. For a baseline assessment in the lab, dyads were seated together on a couch under dim lights and watched a low-action animation film for five minutes. Heart rate (HR) and respiratory sinus arrhythmia (RSA) were measured in both mother and child. Maltreating mothers showed significantly higher HR and lower RSA at baseline than nonmaltreating mothers. Although dyads' HR and RSA were measured during a joint task, the study did not assess autonomic *stress* reactivity.

The remaining five studies included measurements of autonomic responses to stressful child-related stimuli. Disbrow et al. (1977) recruited 22 physically abusive and 24 neglectful parents via CPS. Of the total sample 63% were mothers. Maltreating parents were matched to a nonmaltreating comparison group on age, education, ethnicity, relationship status (single vs in a couple), and children's age. The comparison group was screened to verify they had not been previously reported to CPS. In the lab, parents watched a videotape of interactions between a mother, father, and child of the same race as themselves. The tape included pleasant and stressful interactions. For the baseline assessment, neutral colors were presented before the start of the tape, as well as in between interaction scenes. No information was reported on whether groups differed in autonomic arousal during baseline. Information on differential reactivity from baseline to the stressful interaction scenes was reported only for HR. The change in HR from baseline to the stressful interaction scenes did not differ significantly for maltreating and comparison parents.

Frodi and Lamb (1980) included 14 physically abusive and 14 comparison mothers. Abusive mothers were recruited through Parents Anonymous and all admitted to having abused at least one of their children. The comparison group was individually matched to the abuse group on age, marital status, social class, number of children, and children's age. Participants watched two videotapes with three 2-min segments each (also used by Pruitt & Erickson, 1985, see below). The first and last segment of each tape showed an infant quiescent but alert. The middle segment of one tape showed the same infant smiling and cooing, while the middle segment of the other tape showed the infant crying. The order of presentation of the two tapes was counterbalanced. HR, skin conductance (SC), and diastolic blood pressure (DBP) were measured during a 2-min rest period and the first and last 30 seconds of each video segment. For reactivity analyses, the last 30 seconds of the first segment showing the infant quiescent were used as a baseline from which change scores were calculated. Abusive and nonabusive mothers did not differ on baseline levels for any of the autonomic measures. However, abusive mothers showed greater HR and SC increases, but smaller DBP increases than nonabusive mothers in response to the crying infant.

Wolfe et al. (1983) also made use of the presentation of videotapes. A group of seven mothers who had been referred to a treatment program by the local child welfare agency after verification of child abuse and seven comparison mothers participated. The groups were individually matched on education, income, number of children, children's age, and parent-reported child behavior problems. After a 5-min resting baseline, mothers watched a 3-min videotape with 12 scenes of mother-child interactions. Some interaction scenes were stressful (e.g., dyadic conflicts) whereas others were not (e.g., mother and child playing together). After that, a 5-min post-task baseline was recorded. HR, respiration rate (RR), and SC responses were measured during baseline and while viewing the interaction scenes. Eight scenes, of which the level of stressfulness was agreed on by more than 65% of mothers, were included in analyses. Although means and standard deviations for autonomic values at baseline and during the stressful scenes were displayed for both abusive and nonabusive mothers, whether the groups differed significantly on autonomic arousal at baseline was not reported. Using baseline levels as a covariate, abusive mothers showed higher SC and RR during the stressful scenes than nonabusive mothers. There were no effects for HR.

Friedrich et al. (1985) and Reijman et al. (2014) used infant cry sounds as a stressor. Friedrich et al. (1985) had a sample of abusive (n = 14), neglectful (n = 13), and comparison mothers (n = 15). Maltreating mothers had been substantiated for abuse or neglect within the past year. The comparison group received financial aid from the county welfare office and during the time they were receiving the assistance no reports of abuse or neglect were filed against them. The three groups did not differ on age, education, income, marital status, or children's age, although abusive and neglectful mothers on average had more children than comparison mothers. Mothers listened to a 9-min audiotape on which 1-min sounds of white noise, a tone, and infant crying were alternated. Results for the cry sound were selected for this review and the meta-analyses (see inclusionary criteria described above). The order of presentation of the segments was counterbalanced, but the cry sound was always preceded by the nonstressful white noise. HR, finger blood volume (FBV), and SC were measured during a 7-min baseline and throughout the presentation of the audiotape. There were no differences between the groups on any of the measures at baseline. For HR and FBV, reactivity to the cry sound was analyzed as the difference between mean values at baseline and during the cry. There were no significant differences among the maltreatment groups for HR reactivity or FBV reactivity. For SC, reactivity was analyzed in two ways: as the increase from the last 10 seconds of white noise to the first 10 seconds of the cry (deflections), and as the total number of seconds SC was higher during the cry than during baseline. There were no differences between groups in their SC deflections, but there was a difference between groups on the number of seconds above baseline. Particularly during the second cry segment, both the abusive and neglectful groups showed more sustained SC increase from baseline as compared to the comparison group.

Reijman et al. (2014) recruited a sample of maltreating mothers through a mental health clinic, where mothers received therapy focusing on their parenting problems. Incidents of abuse and neglect were coded from CPS records. All mothers were found to be neglectful, while about half were also abusive. Nonmaltreating mothers were recruited from a different subdivision of the same mental health clinic, where their children were in therapy for a developmental or learning disorder. In this group, the Maternal Maltreatment Classification Interview (Cicchetti, Toth, & Manly, 2003) was conducted to verify the absence of maltreatment incidents. Physiological data were available for 42 maltreating and 38 nonmaltreating mothers. The groups did not differ on ethnicity, education, medication intake, number of children, or whether children were clinically diagnosed, but maltreating mothers and their children were significantly younger, more maltreating mothers smoked, and fewer exercised as compared to the nonmaltreating group. These variables (age, smoking, and exercise habits) were controlled for in the analyses. After watching neutral images during a 5-min baseline assessment, they listened to nine 10-sec infant cries of varying pitches. HR, pre-ejection period (PEP), vagal tone (RMSSD), and SC were measured throughout. No differences were found between the groups for any of the autonomic variables at baseline. From baseline to the cry sounds, maltreating and nonmaltreating mothers showed similar HR and RMSSD responses, but there was an effect of maltreatment status on PEP reactivity, with maltreating mothers showing a nonsignificant PEP decrease, while the comparison group showed a nonsignificant PEP increase. Finally, maltreating mothers showed less SC reactivity than nonmaltreating mothers.

Summary. Of these six studies with maltreating parents, one provided evidence supporting the association between child maltreatment and higher levels of baseline autonomic arousal, with lower parasympathetic activation (Creaven et al., 2014). Information on autonomic differences at baseline was not reported by Disbrow et al. (1977) or Wolfe et al. (1983). The three remaining studies found no significant associations between child maltreatment status and baseline levels of autonomic arousal (Friedrich et al., 1985; Frodi & Lamb, 1980; Reijman et al., 2014).

Regarding the association between child maltreatment status and reactivity to stressful stimuli, evidence was mixed as well. In the Frodi and Lamb (1980) and in the Wolfe et al. (1983) studies, effects for two out of three autonomic measures supported the link between child maltreatment and increased stress reactivity. Friedrich et al. (1985) found that abusive and neglectful mothers showed more sustained increases in SC than comparison mothers during a cry sound as compared to baseline, but there were no differences between groups in SC deflections, HR reactivity, or FBV reactivity. In Reijman et al. (2014), only the differential direction of PEP responses to infant crying suggested slightly more sympathetic reactivity in maltreating mothers. However, maltreating mothers showed weaker SC responses than nonmaltreating mothers, indicating less sympathetic reactivity, while there were no significant effects for HR or RMSSD. Finally, autonomic stress reactivity did not distinguish abusive from comparison parents in Disbrow et al. (1977).

Parents and nonparents at risk for child abuse. Six studies assessed the risk for committing child abuse in parents and nonparents. Five of the six studies used a validated instrument designed to assess risk for child physical abuse, namely the Child Abuse Potential Inventory (CAP Inventory; Milner, 1986; Milner & Wimberley, 1979). The CAP Inventory is a self-report questionnaire that consists of 160 statements to which respondents are asked to indicate whether they agree or disagree. It consists of an abuse potential scale (77 items), six factor scales (e.g., distress, rigidity, unhappiness, various interpersonal problems), and three validity scales to detect if respondents answered randomly, faked good (i.e., denied problems), or faked bad (i.e., exaggerated problems). Adequate construct validity, internal consistency, and stability over time have been demonstrated across numerous samples (see Milner, 2004, for a review, but see Voorthuis et al., 2014).

Casanova et al. (1992) recruited 151 parents from day-care and social service agencies. All were screened with the CAP Inventory. Respondents with valid answers were included in the high-risk group if they scored 166 or higher (the signal detection cut-off score), while those who scored below the median norm abuse score of 66 were considered low risk. Fifteen high-risk mothers were individually matched with 15 low-risk mothers on ethnicity, age, marital status, number of children, and children's age. The two groups of mothers were exposed to a series of nonchild-related stimuli, namely a cold pressor task, a stressful film, unsolvable anagrams, and car horn sounds. For each task, HR and SC were measured the minute prior to stimulus onset (baseline), the minute of stimulus presentation, and the minute after stimulus completion. Results for the stressful film were selected for inclusion in this review and the meta-analyses (see inclusion criteria described above). The stressful 1-min film displayed two industrial accidents. There were no differences in HR during baseline, while no information was reported on significant differences between groups for SC baseline levels. The stressful film evoked a stress response on both ANS measures, but no significant differences between high- and low-risk mothers in HR reactivity and SC reactivity (from baseline to film exposure) were found.

Crouch et al. (2015) studied a sample of 48 parents, of which 28 were women. Parents with valid response patterns on the CAP Inventory were classified as high-risk if their CAP abuse score was at or above the signal detection cut-off score of 166, while those with a score below 166 were considered low-risk. The two groups did not differ significantly on age, gender, education, annual household income, marital status, or number of children, but more high-risk parents were African American. Race/ethnicity was not associated with any of the outcome measures. All parents completed a computer task which required them to solve as many anagrams as possible in three minutes. Participants were randomly assigned to either a difficult anagram condition or an easy anagram condition. HR and RSA were measured during a 3-min baseline and during the anagram task. At baseline, high-risk parents showed higher HR and lower RSA than low-risk parents. In response to the anagram task, HR and RSA of high-risk parents did not change, while low-risk parents showed an increase in HR and a decrease in RSA. Difficulty of the anagram task did not moderate patterns of change in HR or RSA over time.

The four remaining studies sampled nonparents and used child-related stimuli. Pruitt and Erickson (1985) recruited 61 nonparents who were 30 years of age or younger. Based on the CAP Inventory, placement in the high- versus low-risk groups was determined by taking the upper and lower 33% of nonweighted abuse scores. Twenty-two participants (14 women) were classified as high-risk (nonweighted abuse score > 9.1) and 22 participants (16 women) were classified as low-risk (nonweighted abuse score $\leq$ 4). No matching of the groups on demographics was reported. Participants were shown two videotapes that were 6 min each. One video showed a 5-month old female infant first quiescent but alert (2 min), smiling and cooing (2 min), then again quiescent (2 min), while the other video showed the same infant quiescent (2 min), crying (2 min), and quiescent (2 min). The same videotapes had been used by Frodi and Lamb (1980; see above). Whether the video with the smiling or the crying infant was shown first was counterbalanced within women/men in the low/high risk groups. HR and SC responses were measured 2 min before and throughout the videotapes. Results were reported in peak HR and peak SC rather than mean levels. Autonomic patterns across the set of videotapes were analyzed, so that baseline activity and reactivity to the crying infant specifically were not included. The authors reported that overall, high-risk participants had significantly higher peak HR and marginally lower peak SC, and showed lower HR variability in response to the videos. There were no significant differences between the low-risk and high-risk groups with respect to SC reactivity in response to the videotapes.

Crowe and Zeskind (1992) screened 284 introductory psychology students for child physical abuse risk using the CAP Inventory. After excluding students whose responses on the CAP Inventory were invalid or incomplete, 30 participants were selected with either high CAP scores (upper 28th percentile of scores; M = 283, SD = 40.7) or low CAP scores (lower 28th percentile of scores; M = 53, SD = 50.4). Both groups consisted of eight men and seven women, and did not differ on age, ethnicity, income, or reported history of abuse. Participants listened to two audio recordings, one with four 10-sec phonated infant cry sounds and one with four 10-sec hyperphonated infant cry sounds. The first tape, containing either phonated or hyperphonated cries, was repeated twice. After a 10-min rest, the remaining tape of phonated/hyperphonated infant cry sounds was played twice. Order of presentation of the phonated/hyperphonated cries was counterbalanced within men/women in the low/high CAP groups. HR and SC were assessed two minutes before stimulus onset and throughout the presentation of the cry sounds. No significant differences at baseline between the high- and low-CAP groups were reported for HR or SC. In response to the cry sounds, the high-CAP group showed marginally greater HR changes than the low-CAP group, but in a negative direction, so that the HR of those at risk for child abuse tended to decrease, while that of the low-CAP group did not. The authors also reported a marginally significant interaction effect of CAP risk status and cry type (phonated vs hyperphonated) on SC responses, such that the high-CAP group showed somewhat higher SC responses to the phonated sounds than the low-CAP group. There were no risk group differences in SC reactivity to the hyperphonated cry sounds.

Laud (1997) also used the infant cry sound as a stress-evoking stimulus. Participants were randomly chosen from a larger pool (N = 199) of unmarried, nonparent, female psychology students that were screened for health (including cardiovascular) and hearing concerns. Based on CAP Inventory abuse scores, 38 respondents were classified as high risk (CAP abuse score ≥ 166) and 34 respondents were classified as low risk (CAP abuse scores ≤ 63). The high-risk and low-risk groups did not differ on ethnicity, age, or education. After a 4-min resting baseline, participants listened to an infant cry sound that lasted eight minutes. HR was recorded throughout the baseline and the cry sound presentations, and systolic and diastolic blood pressures were measured every two minutes. CAP risk groups did not differ on any of the baseline autonomic measures or in their autonomic response from baseline to the cry sounds.

Stasiewicz and Lisman (1989) used the Adult-Adolescent Parenting Inventory (AAPI; Bavolek, Kline, McLaughlin, & Publicover, 1979) to assess child abuse risk in a sample of male, unmarried, nonparent undergraduate students. Participants who obtained scores in the upper 30% of the AAPI distribution of scores were classified as high-risk (n = 16) and those with scores in the lower 38% of the distribution of AAPI scores were classified as low-risk (n = 16). No information on whether the risk groups were demographically matched was reported. After a 6-min resting baseline participants were either exposed to an audio recording of the cry sounds of a medically at-risk infant or the sound of a smoke alarm. Results were reported for the two stressors combined, so that examination of data specific to the infant cry sounds was not possible. The volumes required to evoke similar levels of aversiveness in response to the infant cry sound and the smoke alarm sound had been determined in a pilot study. Infant cries and the smoke alarm sounds were presented for three minutes each and were repeated three times with 2-min breaks between presentations. DBP was assessed as an index of ANS activation. High-risk and low-risk participants did not differ with respect to baseline DBP nor in their DBP response to the infant cry/smoke alarm sounds.

Summary. Results of significance testing in most of the studies with at-risk samples found no significant evidence for a link between risk for child abuse and autonomic activity at baseline (Casanova et al., 1992; Crowe & Zeskind, 1992; Laud, 1997; Stasiewicz & Lisman, 1989), or autonomic reactivity to stressful child- or nonchild-related stimuli (Casanova et al., 1992; Laud, 1997; Stasiewicz & Lisman, 1989). Information on autonomic baseline differences was partially or not explicitly reported in Casanova et al. (1992) and Pruitt and Erickson (1985). Crouch et al. (2015) found that high-risk parents showed higher HR and lower RSA at baseline, although the high-risk group showed less autonomic reactivity to a stressful task than the low-risk group. Crowe and Zeskind (1992) found greater HR reactivity to cry sounds in the high-risk group, but the reactivity constituted a decrease rather than an increase in arousal. In Pruitt and Erickson (1989), high-risk participants showed no HR change in response to a video of a crying infant, while the low-risk group showed a HR decline. There were no other risk group differences in autonomic reactivity to the stressors used in the reviewed studies.

Table 1
Summaries of reviewed studies

Study	*Sample size*	*Parent status*	*Malt status*	*Cut-off scores*	*Malt subtype*	*Autonomic measures*	*Stressor*	*Relevant findings*	
Disbrow et al., 1977	83	parents	substantiated	N/A	abuse and neglect	HR, SC	videos of stressful dyadic interactions	*B*	N/I
								R	N.s.
Frodi & Lamb, 1980	28	parents	substantiated	N/A	abuse	HR, DBP, SC	video crying infant	*B*	N.s.
								R	↑ HR, SC; ↓ DBP
Wolfe et al., 1983	14[a]	parents	substantiated	N/A	abuse	HR, RR, SC	videos of parent-child conflict situations	*B*	N/I
								R	↑ RR, SC; n.s. for HR
Friedrich et al., 1985	42	parents	substantiated	N/A	abuse and neglect	HR, FBV, SC	infant cry sound	*B*	N.s.
								R	N.s. for HR, FBV; ↑ SC sec above baseline

Pruitt & Erickson, 1985	44	nonparents	high-risk (CAPI)	upper 33% (≥ 9.1) lower 33% (≤ 4.0)	physical abuse	HR, SC	video crying infant	*B*	N/I
								R	↓ HR; SC $p > .05$
Stasiewicz & Lisman, 1989	32	nonparents	high-risk (AAPI)	upper 30% lower 30%	abuse	DBP	infant cry sound and smoke alarm	*B*	N.s.
								R	N.s.
Casanova et al., 1992	30	parents	high-risk (CAPI)	> 166 < 66	physical abuse	HR, SC	stressful film	*B*	N.s. for HR; N/I for SC
								R	N.s.
Crowe & Zeskind, 1992	30	nonparents	high-risk (CAPI)	upper 28% (*M*=283, *SD*=40.7) lower 28% (*M*=53, *SD*=50.4)	physical abuse	HR, SC	infant cry sounds	*B*	N.s.
								R	↑ HR (decrease); ↑ SC to phonated sounds
Laud, 1997	72	nonparents	high-risk (CAPI)	≥ 166 ≤ 63	physical abuse	HR, DBP, SBP	infant cry sound	*B*	N.s.
								R	N.s

Creaven et al., 2014	104	parents	substantiated	N/A	abuse and neglect	HR, RSA	N/A	*B*	↑ HR; ↓ RSA
								R	N/A
Reijman et al., 2014	80[b]	parents	substantiated	N/A	abuse and neglect	HR, RMSSD, PEP, SCL	infant cry sounds	*B*	N.s.
								R	↓ SC; ↑ PEP; HR, RMSSD *p*s > .05
Crouch et al., 2015	48	parents	high-risk (CAPI)	> 166 < 166	physical abuse	HR, RSA	anagrams	*B*	↑ HR, ↓ RSA
								R	↓ HR, RSA

[a] For SC results in Wolfe et al., $N = 10$

[b] For PEP results in Reijman et al., $N = 77$

Note. Malt = maltreatment; CAPI = Child Abuse Potential Inventory; AAPI = Adult-Adolescent Parenting Inventory; HR = heart rate; SC = skin conductance; RSA = respiratory sinus arrhythmia; FBV = finger blood volume; DBP = diastolic blood pressure; SBP = systolic blood pressure; RMSSD = root mean square of successive differences; PEP= pre-ejection period; RR = respiration rate; B = Baseline; R= Reactivity; N.s. = Not significant; N/A = does not apply; N/I = no information reported. Relevant findings are reported for maltreating or at-risk populations relative to nonmaltreating or low-risk control groups, at $p < .05$.

Conclusion. Across both sets of studies on parents with substantiated maltreatment and individuals at risk for abuse, only two studies provided evidence (based on significance testing) supporting heightened autonomic activity and lower parasympathetic activation at baseline among maltreating/at-risk individuals (Creaven et al., 2014; Crouch et al., 2015, respectively). These two studies (as well as the seven studies that did not find ANS baseline differences) varied in their sample characteristics (substantiated maltreatment vs at-risk status), maltreatment type (abuse and neglect vs risk for physical abuse), gender ratio (mothers only vs mothers and fathers), sample size (N = 104 vs N = 48), and baseline procedure (watching a video in the presence of their child vs resting in solitude). This state of affairs makes it hard to identify variables that may explain differences in results across studies. Both Creaven et al. and Crouch et al. measured HR and RSA, but studies that did *not* find group differences on ANS baseline activity also included HR and RSA as outcome measures (e.g., Reijman et al., 2014). Synthesis of the findings is further complicated by the fact that several studies did not statistically test (or report) whether the maltreating/at-risk groups differed from their comparison groups on ANS baseline values, despite reporting the respective values. Group differences could not be reviewed in those cases (although effect sizes could potentially be calculated and included in a meta-analysis; see below).

Regarding ANS stress reactivity as a risk factor for child maltreatment, the least equivocal findings were presented by Frodi and Lamb (1980) and Wolfe et al. (1983). Both samples consisted of physically abusive parents, all mothers, who were presented with stress-invoking, child-related videotapes. Common autonomic measures were HR and SC, and abusive mothers showed heightened SC stress reactivity in both studies. Frodi and Lamb (1980) additionally found greater HR reactivity in the abusive group while Wolfe et al. (1983) did not. The remaining three studies with maltreating samples (which included neglectful parents) and none of the studies with at-risk samples showed differential stress reactivity. This tentatively suggests that increased sympathetic reactivity is a risk factor specific to substantiated physical abuse. Operational variation was also present in the set of studies using the CAP Inventory, with cut-off scores being criterion-referenced (i.e., signal detection score of 166) in some studies and norm-referenced (e.g., upper vs lower 33 percentile of sampled scores) in others. Such differences in methodology may help explain the variability of findings observed across studies.

Moreover, small sample sizes may have contributed to instability of results across the studies reviewed (e.g., exaggerated effect sizes, low positive predictive power, increased risk of either Type I or Type II errors; for a discussion of these issues see Button et al., 2013). Use of small samples makes it difficult to draw conclusions from individual studies based on significance testing. A systematic review of effect sizes across studies is needed to determine which (if any) of these issues associated with small effect sizes might be operating in this literature.

Meta-analytic Procedures

Although the narrative review conveys the similarities and differences of the methods and the results of significance testing across studies, it does not quantitatively analyze the strength of the effects observed. Results of the narrative review revealed seemingly contradictory findings (based on significance testing) among as well as within studies (e.g., Frodi & Lamb, 1980; Reijman et al., 2014), an observation that is consistent with that of an earlier review (McCanne & Hagstrom, 1996). Meta-analysis is thus warranted to assess the overall effects for the relation between (risk for) child maltreatment and autonomic baseline activity as well as stress reactivity, and to test whether effects may be moderated by sample or study characteristics.

Moderators

For the meta-analyses, we coded two types of moderators: sample-related and procedure-related. Sample-related moderators were *maltreatment status* (categorical: substantiated maltreatment vs risk for physical abuse) and *percentage of women* in the sample (continuous). Procedural characteristics were *presentation of stressor* (categorical: auditory vs visual vs real-life stimuli), *publication year* (continuous), and *sample size* (continuous). The potential moderators *parenting status* (whether participants were parents or not) and *maltreatment type* (studies that focused on [risk for] physical abuse vs studies that included neglect) were excluded because of their high overlap (83% in both cases) with *maltreatment status* in the current set of studies. Interrater reliability of the coding of moderators was good, with intraclass correlations for continuous moderators ranging from .96 – 1, and kappas for categorical moderators ranging from .85 – 1.

Statistical analyses

We performed meta-analyses on two overall outcomes: the association between (risk for) child maltreatment and ANS baseline activity, and the association between (risk for) child maltreatment and ANS stress reactivity. Within each of these two sets of studies, we conducted several meta-analyses: one for mixed ANS measures (i.e., a combination of mixed indices of both the sympathetic and parasympathetic branches, such as HR, blood pressure, RR); one for the sympathetic nervous system (measured relatively purely as SC or PEP); one for the parasympathetic nervous system (as indicated by RSA/RMSSD); and one for HR alone, since it was included in the large majority of studies. Study outcomes were entered in Comprehensive Meta-Analysis (CMA; Borenstein, Rothstein, & Cohen, 2005). In the case of mean values being reported without standard deviations, we estimated the latter based on the *SD*s for the corresponding autonomic measure in Reijman et al. (2014). *SD*s of SC levels in Pruitt and Erickson (1985) were estimated based on Casanova et al. (1992) because SC was reported in micromhos x 10^6 in both papers. CMA transformed the outcomes into Hedges' *g* effect sizes, which is appropriate for smaller sample sizes such as ours (Cumming, 2012; Lakens, 2013). In line with our hypothesis that (risk for) child maltreatment would be associated with higher levels of (re)activity, effects in that direction were marked positive, while effects in the opposite direction were identified as negative. Reactivity was defined as *increases* in arousal in response to stress, i.e., ANS and SNS increases and PNS decreases. In the case of findings indicated as nonsignificant but without further statistical details, we assigned a zero effect size at p = .50 (Mullen, 1989). These cases are marked with an asterisk in Figure 1. Confidence intervals (CIs) of 95% around the point estimate of every effect size are reported.

Almost all studies had more than one ANS outcome measure. For our analyses on the ANS subsystems (SNS, PNS, mixed ANS), combined effect sizes were calculated for measures belonging to the same subsystem. For instance, for Reijman et al. (2014), a combined effect size for the sympathetic branch was calculated from the measures of SC and PEP. Similarly, for Wolfe et al. (1983), a combined effect size was calculated from HR and RR, which are both mixed indices of the sympathetic as well as the parasympathetic branch and may be considered markers of generic ANS arousal.

Statistics for the combined effect sizes (with 95% CIs) and moderator analyses were drawn from random effect models. Random effect models are based on the assumption that studies differ in their characteristics, and since meta-analytical results are calculated from this assumption, they may be generalized to studies not sampled in the meta-analysis, but belonging to the same population (Hedges & Vevea, 1998). We tested the homogeneity of different sets of effect sizes and moderating effects of categorical variables with the *Q* statistic (Borenstein et al., 2005). Contrast analyses for categorical moderators were conducted only when there were at least two groups with $k \geq 4$ (Bakermans-Kranenburg, Van IJzendoorn, & Juffer, 2003). Continuous moderators were tested in univariate as well as multivariate regression models, since *year of publication* and *sample size* were correlated ($r = .56$, $p = .04$). We also performed a series of cumulative meta-analyses according to *year of publication*, in which the combined effect size with the addition of each new study is calculated, to further inspect time-related trends.

In the case of significant combined effect sizes, funnel plots were inspected for potential publication bias, i.e., the tendency for small studies with nonsignificant or unexpected results to remain unpublished, which would be visually represented by the funnel plot's asymmetrical base. We calculated a fail-safe number to reflect the number of studies with null results necessary to reduce the effect size to a nonsignificant effect. Finally, we conducted power analyses for individual studies based on the combined effect sizes in the program G-Power 3.1 (Faul, Erdfelder, Lang, & Buchner, 2007), to calculate (1) the sample size required to detect the combined effect size, with $\alpha = .05$ and a power of .80, and (2) the power of each study to detect the combined effect size, given their sample size and $\alpha = .05$.

No outliers were found for any of the continuous moderators (standardized *z*-scores < -3.29 or > 3.29; Tabachnik & Fidell, 2001). Checks for outliers in effect sizes were done at the level of analysis, i.e., for sympathetic, parasympathetic, and generic autonomic measures separately, and revealed no outliers.

Results

Child maltreatment and ANS baseline activity

Point estimates and respective CIs of the effect sizes for the outcome measures of each study included in the meta-analysis examining the link between (risk for) child maltreatment and autonomic baseline activity are presented in Figure 1. For the association between (risk for) child maltreatment and baseline activity of the ANS subsystems, we grouped mixed indices of (a) the sympathetic and parasympathetic branches to assess nonspecific ANS activity (i.e., HR, blood pressure, and RR), (b) indices of sympathetic activity only (i.e., SC, PEP), and (c) indices of parasympathetic activity only (i.e., RSA/RMSSD).

For mixed ANS basal activity, the combined effect size was significant, $g = 0.24$, 95% CI [0.05, 0.43], $p < .05$ in a homogeneous set of studies ($k = 11$, $N = 524$; $Q = 11.33$, $p > .05$). The funnel plot was symmetrical, showing no evidence for publication bias. The fail-safe number was 6, indicating that six null results would be necessary to reduce this meta-analytic finding to a nonsignificant effect. Our power analyses showed that a sample size of $N = 432$ would be required to detect the combined effect size $g = 0.24$. The power of the individual studies to detect this effect size ranged from .11 for the study with the smallest sample size to .33 for the largest sample size. The combined effect sizes for the sets of studies examining the association between (risk for) child maltreatment and sympathetic nervous system activity at baseline ($k = 6$, $N = 234$), and parasympathetic baseline activity ($k = 3$, $N = 232$) were not significant ($g = 0.003$ and $g = 0.30$, respectively; see Table 2 for statistical details).

For baseline HR, the set of studies ($k = 10$, $N = 492$) was homogeneous, $Q = 11.81$, $p > .05$. The combined effect size was significant ($g = 0.24$, 95% CI [0.03, 0.45], $p < .05$), indicating that (risk for) perpetration of child maltreatment was associated with higher HR levels at baseline.

There were no moderating effects of *maltreatment status* ($p > .05$). For the association between (risk for) child maltreatment and mixed ANS baseline activity, effect size estimates were significant for parents with substantiated maltreatment, but not for participants at risk (see Table 2). It bears mentioning that maltreatment status was significantly associated with parenting status, such that participants at risk (as opposed to substantiated) for maltreatment were more often non-parents (e.g., undergraduate students). Regression analyses showed no moderating effects for the percentage of female participants, year of publication, or sample size (ps > .05). Cumulative meta-analyses showed no time-related change in effect sizes.

Child maltreatment and ANS stress reactivity

Point estimates and respective CIs for all outcome measures included in the meta-analysis on the link between (risk for) child maltreatment and autonomic stress reactivity are displayed in Figure 1. The meta-analytical results are summarized in Table 3. The combined effect size for the sets of studies examining the association between (risk for) child maltreatment and mixed ANS stress reactivity ($k = 11$, $N = 503$) was not significant, $g = -0.12$ (see Table 3).

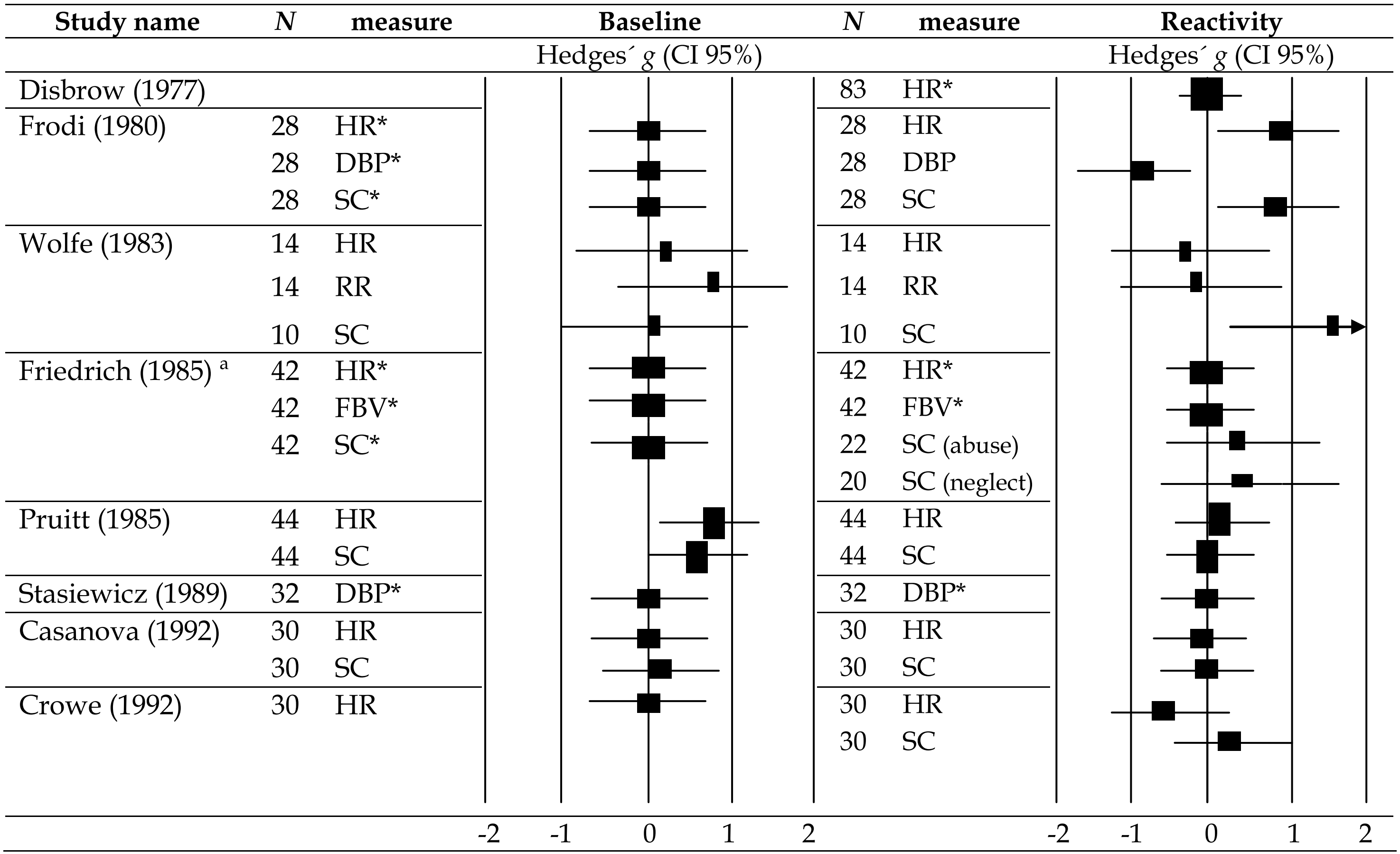
The combined effect size estimating the association between (risk for) child maltreatment and sympathetic reactivity to stressors (k = 8, N = 264) was not significant either, g = 0.27, 95% CI [-0.04, 0.58], p = .09. Finally, there was no effect for parasympathetic reactivity (g = -0.26 for k = 2, N = 128) or HR (g = -0.10 for k = 10, N = 471; see Table 3).

We found no moderating effect of *maltreatment status* or *presentation of stimulus* and there were no significant effect sizes for any of the subgroups (ps > .05). In a multivariate model, only *year of publication* predicted the effect sizes for (risk for) child maltreatment and mixed ANS reactivity, later publications being associated with negative effect sizes (p = .02). No moderating effects could be tested for the combined effect sizes on parasympathetic reactivity due to the low number of studies. In a multivariate model, *year of publication* and *gender ratio* (% women in the sample) predicted effect sizes for the association between (risk for) child maltreatment and HR reactivity (ps < .01). The regression line for *year of publication* showed a change from positive effect sizes to negative effect sizes over the years. This seems mainly due to an early study that found a large positive effect (Frodi & Lamb, 1980) and a recent study that yielded a strong negative effect (Crouch et al., 2015). The regression line for *gender ratio* showed that samples with lower percentages of women were associated with negative effect sizes, while higher percentages of women were associated with smaller negative, null, or positive effects. There were no other moderating effects. However, cumulative meta-analyses showed that for SNS stress reactivity, with each aggregated study after Frodi & Lamb (1980) and Wolfe et al. (1983) the combined effect size further approached a null effect, which is displayed in Figure 2.

Table 2
Combined effect sizes for autonomic baseline activity

	K	*N*	*g*	*95 % CI*	Q^h	Q^c
ANS	11	524	0.24*	0.05, 0.43	11.33	
Maltreatment status						0.37
Substantiated	5	268	0.30*	0.02, 0.59	2.46	
At-risk	6	256	0.18	-0.10, 0.46	8.19	
SNS	6	204	-0.003	-0.27, 0.26	0.29	
Maltreatment status						
Substantiated	4	160	-0.01	-0.33, 0.31	0.04	
At-risk	2	74	0.02	-0.44, 0.47	0.24	
PNS	3	232	0.30	-0.14, 0.75	5.50	
Maltreatment status						
Substantiated	2	184	0.17	-0.38, 0.71	3.43	
At-risk	1	48	0.67	-0.20, 1.54		
HR	10	492	0.24*	0.03, 0.45	11.81	
Maltreatment status						0.13
Substantiated	5	268	0.28	-0.03, 0.59	2.49	
At-risk	5	224	0.20	-0.13, 0.52	8.82	

Note. k = number of studies; N = number of participants; g = Hedges' g effect size; CI = confidence interval; Q^h= homogeneity index; Q^c= contrast index; ANS = autonomic nervous system; SNS = sympathetic nervous system; PNS = parasympathetic nervous system; HR = heart rate. Contrasts were tested for subgroups with $k \geq 4$. * $p < .05$



Study name
N
measure
Baseline
Hedges´ g (CI 95%)
N
measure
Reactivity
Hedges´ g (CI 95%)
Disbrow (1977)
83 HR*
Frodi (1980)
28 HR*
28 DBP*
28 SC*
28 HR
28 DBP
28 SC
Wolfe (1983)
14 HR
14 RR
10 SC
14 HR
14 RR
10 SC
Friedrich (1985) a
42 HR*
42 FBV*
42 SC*
42 HR*
42 FBV*
22 SC (abuse)
20 SC (neglect)
Pruitt (1985)
44 HR
44 SC
44 HR
44 SC
Stasiewicz (1989)
32 DBP*
32 DBP*
Casanova (1992)
30 HR
30 SC
30 HR
30 SC
Crowe (1992)
30 HR
30 HR
30 SC
-2 -1 0 1 2
-2 -1 0 1 2

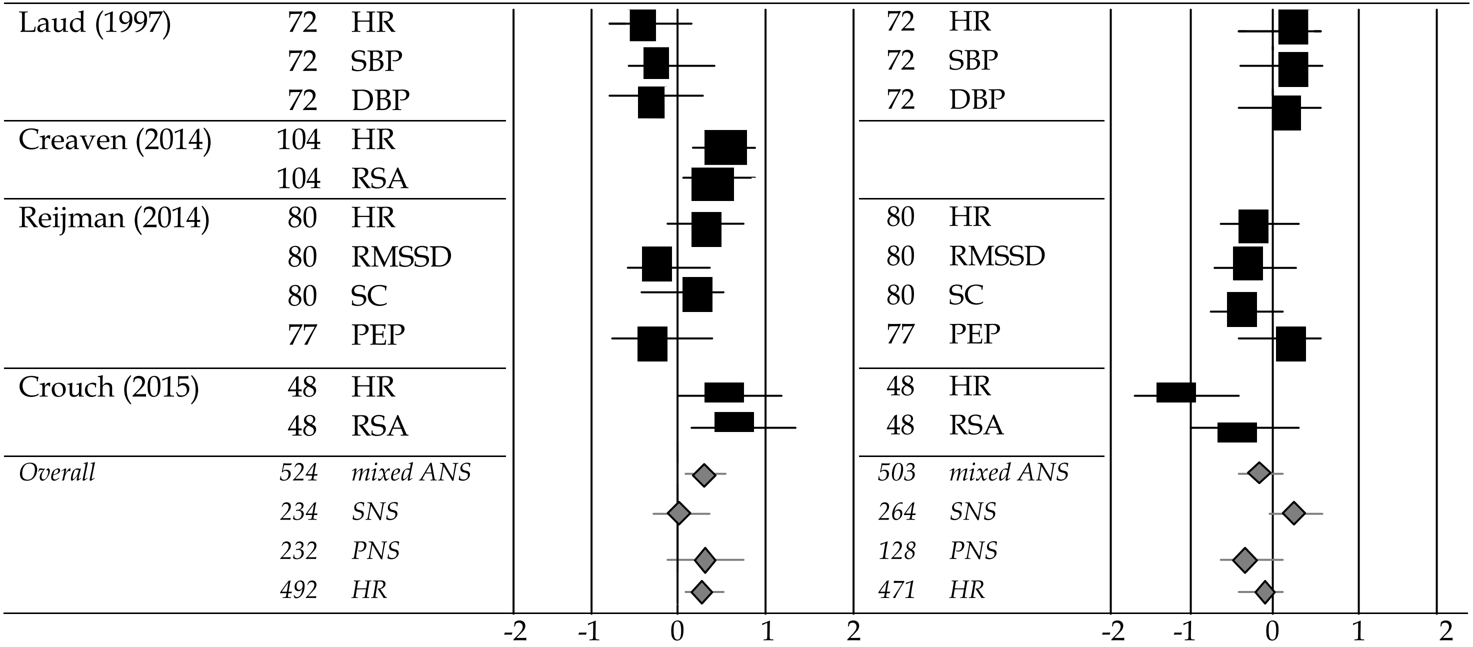


Figure 1. Effect sizes for baseline and reactivity levels for the individual studies. *Note.* HR = heart rate; RSA = respiratory sinus arrhythmia; FBV = finger blood volume; SC = skin conductance; DBP = diastolic blood pressure; SBP = systolic blood pressure; RMSSD = root mean square of successive differences (measure of vagal tone); PEP = pre-ejection period; RR = respiration rate. * Asterisks indicate that effect sizes were based on p = .50 due to lack of statistical details. [a] The sample of Friedrich et al. (1985) consisted of abusive, neglectful, and control mothers. For SC, results for the abusive and neglectful groups were reported separately, so we divided the control group's n by two in order to avoid double representation of participants.

Table 3
Combined effect sizes for autonomic stress reactivity

	K	*N*	*g*	*95% CI*	*Q h*	*Q c*
ANS	11	503	-0.12	-0.32, 0.09	12.61	
Maltreatment status						0.73
Substantiated	5	247	-0.03	-0.28, 0.23	0.94	
At-risk	6	256	-0.22	-0.60, 0.15	10.91	
Presentation of stimulus						0.07
Auditory	5	256	-0.05	-0.30, 0.20	3.29	
Visual	5	199	0.00	-0.28, 029	0.23	
Real-life	1	48	-1.07	-1.72, -0.42		
SNS	8	264	0.27	-0.04, 0.58	9.96	
Maltreatment status						
Substantiated	5	160	0.50	-0.03, 1.02	8.59	
At-risk	3	104	0.07	-0.31, 0.45	0.40	
Presentation of stimulus						0.89
Auditory	4	152	0.13	-0.20, 0.46	1.35	
Visual	4	112	0.47	-0.17, 1.11	7.95	
Real-life	0					
PNS	2	128	-0.26	-0.61, 0.10	0.26	
Maltreatment status						
Substantiated	1	80	-0.19	-0.63, 0.25		
At-risk	1	48	-0.38	-0.98, 0.21		
Presentation of stimulus						
Auditory	1	80	-0.19	-0.63, 0.25		
Visual	0					
Real-life	1	48	-0.38	-0.98, 0.21		
HR	10	471	-0.10	-0.36, 0.16	17.03*	
Maltreatment status						1.16
Substantiated	5	247	0.03	-0.26, 0.31	4.83	
At-risk	5	224	-0.26	-0.71, 0.19	10.84*	
Presentation of stimulus						0.83
Auditory	4	224	-0.09	-0.35, 0.17	2.65	
Visual	5	199	0.09	-0.19, 0.37	4.13	
Real-life	1	48	-1.07	-1.72, -0.42		

Note. k = number of studies; N = number of participants; g = Hedges' g effect size; CI = confidence interval; Q^h= homogeneity index; Q^c= contrast index; ANS = autonomic nervous system; SNS = sympathetic nervous system; PNS = parasympathetic nervous system; HR = heart rate. Contrasts were tested for subgroups with $k \geq 4$. * $p < .05$

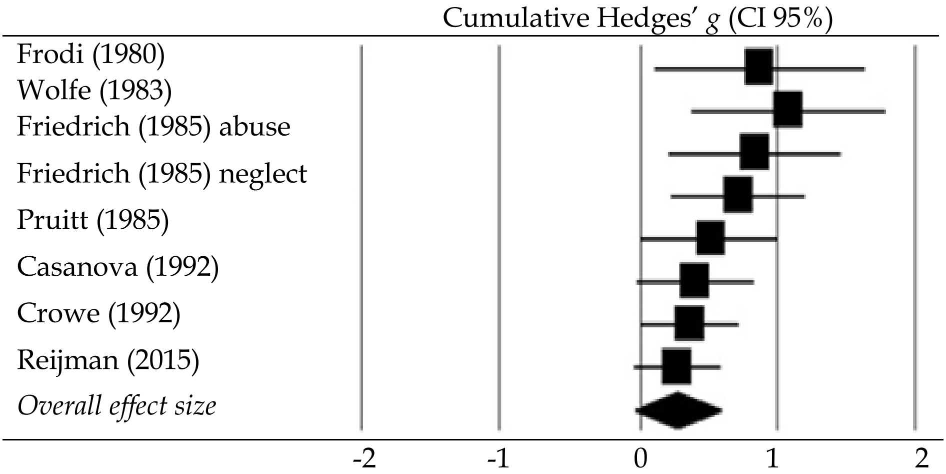


Figure 2. Cumulative effect sizes for sympathetic nervous system reactivity
The sample of Friedrich et al. (1985) consisted of abusive, neglectful, and control mothers. For sympathetic reactivity, results for the abusive and neglectful groups were reported separately, so we divided the control group's *n* by two in order to avoid double representation of participants.

Discussion

Our meta-analyses showed that maltreating parents and participants at risk for child maltreatment exhibited greater autonomic activity at rest ($g = 0.24$) than their respective comparison groups. This is in line with our first hypothesis and with conclusions from an earlier review (McCanne & Hagstrom, 1996). Greater autonomic activity at baseline, as observed in mixed ANS indices (e.g., HR, blood pressure, RR), suggests a chronic state of arousal in maltreating and at-risk participants, even in the absence of stressors. Because these findings are based on mixed ANS indices, it is impossible to say whether they represent impaired parasympathetic vagal regulation, sympathetic influences, or both.

In any case, the ANS responds to the environment, and cardiovascular measures are particularly sensitive to it (McEwen, 1998). Sustained cardiovascular arousal may be a sign of allostatic load, which could be caused by an environment that is continuously (perceived as) overly demanding or challenging, or by a dysregulated ANS unable to decrease activity in the absence of challenges (Friedman, Karlamangla, Gruenewald, Koretz, & Seeman, 2015; Juster, McEwen, & Lupien, 2010; McEwen, 1998). The notion of allostatic load as applied to child maltreatment is consistent with literature identifying (early) life adversities as risk factors for child abuse and neglect, such as having experienced abuse in childhood, unemployment, single parenthood, and low social support (Stith et al., 2009). Maltreating or at-risk individuals may live in circumstances that are generally more challenging or unpredictable and such circumstances may take a toll on the regulatory function of the ANS. We found no significant effects of (risk for) child maltreatment on pure measures of sympathetic or parasympathetic baseline activity, but only a few studies have included pertinent indices (k = 5 and k = 3, respectively). Contrary to our expectations, we did not find meta-analytical evidence for increased autonomic stress responsiveness as a risk factor contributing to child maltreatment. First of all, this may have to do with the low number of studies included in the meta-analyses. For SNS reactivity, confidence intervals of the effect size (g = 0.27) bordered on zero, indicating that there may have been insufficient power for the effect to reach statistical significance. This seems compelling especially considering the SNS's prominent role in stress reactivity. On the other hand, as revealed by our narrative review, only two early studies (Frodi & Lamb, 1980; Wolfe et al., 1983) found that abusive mothers responded with higher SNS increases to a stressor. The cumulative meta-analyses showed that each aggregated study since then has approximated the overall effect to a null effect. This suggests that the hypothesized association between (risk for) child maltreatment and SNS reactivity may have been subject to the *winner's curse* (Button et al., 2013; Molendijk et al., 2012). In line with the suggestion made in our narrative review, it may be that the large effects found for sympathetic reactivity in early studies with small sample sizes were inflated and hard to replicate. Additional studies might then further reduce the effect, rather than increase the required statistical power to detect significant differences. Figure 1 suggests that research on the physiology of maltreating parents / at-risk adults has shifted from sympathetic to parasympathetic reactivity, with only one study from the last 20 years including sympathetic measures, and all studies including parasympathetic measures dating from the past few years. We emphasize the relevance of including both sympathetic and parasympathetic indices in future studies to clarify the matters discussed above.

The lack of differential autonomic stress reactivity we found seems in contrast with the review by McCanne and Hagstrom (1996). In their definition of autonomic hyperreactivity they included both increased and prolonged autonomic activation during any circumstance, including resting/relaxation (i.e. baseline) and the presentation of stimuli. However, we distinguished between autonomic activation at baseline specifically, and strictly defined autonomic reactivity as the change in ANS activity from baseline to stress. This may at least partly explain the discrepancy in supporting evidence for ANS hyperreactivity as a risk factor for child maltreatment. We chose to focus on these two outcomes because they were most commonly assessed, but some valuable results not represented in our meta-analyses bear mentioning. For instance, parents at risk for child abuse showed possible sensitization to a persistent infant cry sound, as seen by a renewed increase in autonomic arousal after several minutes, whereas low-risk parents did not show signs of sensitization (Laud, 1997). Several studies found that maltreating parents showed similar autonomic responses to child signals of a negative and a positive valence, while nonmaltreating parents distinguished between the two different kinds of stimuli (Disbrow et al., 1977; Frodi & Lamb, 1980). Finally, (risk for) child maltreatment was associated with *overall* higher HR in Disbrow et al. (1977) and Crouch et al. (2015) and higher peak HR in Pruitt & Erickson (1985), i.e., independent of condition (baseline vs stress). Chronic arousal in maltreating parents or at-risk participants suggests there may be a ceiling effect, i.e., high levels of HR activity beyond which they show no further increases in response to stress (Crouch et al., 2015).

We found no evidence for moderating effects of the categorical variables in either meta-analysis, such as whether maltreatment was substantiated or risk for physical abuse was assessed with the CAP Inventory. Combined effect sizes were predominantly homogeneous, suggesting that effects may be similar according to maltreatment status and stimulus presentation, but the small cell sizes preclude any firm conclusions. Multivariate regressions showed that year of publication predicted the effect size for mixed ANS reactivity, and year of publication as well as the percentage of women in the sample predicted the HR reactivity effect size. Later publications and samples with lower percentages of women were associated with a directional change toward negative effect sizes. Again, these findings should be interpreted with caution as strong effects may tilt the regression line disproportionately with this small number of included studies. Notably, the moderating effect of several potentially relevant variables could not be tested, either because the cell size for one of the categories was small even after dichotomization ($k < 4$; e.g., socioeconomic status, whether the stressor was child-related or nonchild-related) or because data were not consistently reported (e.g., ethnicity, participants’ age).

The small number of included studies is one of the limitations of our meta-analyses, and for this reason our findings should be considered exploratory. Homogeneity tests and moderator analyses of small sets of studies might easily lead to type 2 errors. The meta-analytical evidence must therefore be considered with caution. A second limitation is that the two meta-analyses were done on almost the same set of studies, so our findings for ANS baseline activation and ANS stress reactivity were not independent. The fact that these two results were found in almost the same set of studies is consistent with the the notion that the underlying process of allostatic load may result in both chronic autonomic activation and increased reactivity to stressors, as discussed above.

The studies included in our meta-analyses were not without methodological shortcomings. Our narrative review showed a predominant lack of supporting evidence (based on significance testing) for the hyperreactivity hypothesis, which may be due to studies' small samples. As discussed above, small samples may lead to insufficient statistical power for small effects to reach statistical significance, while in other cases it may lead to exaggerated significant effects. Another important shortcoming is that in quasi-experimental designs groups are not equivalent from the onset. When we want to ascribe observed differences in autonomic (re)activity to whether participants are maltreating/at-risk or not, insufficient comparability of groups on potential confounding variables is a threat to internal validity (Cook & Campbell, 1986; Shadish, Cook, & Campbell, 2002). Not all of the reviewed studies matched their groups on variables such as socio-economic status or educational level. Furthermore, when groups differed on a potential confounding variable, this was not always controlled for in analyses. Almost none of the studies controlled for maltreatment experienced by participants in their own youth, a factor that is related to child maltreatment perpetration (e.g., Pears & Capaldi, 2001) as well as autonomic responsiveness (Casanova et al., 1994; Heim et al., 2000). Alternative explanations for observed correlations are thus not ruled out. Finally, all of the studies included in our meta-analyses used a case-control design, precluding causal inferences about the association between autonomic (re)activity and child maltreatment.

Nonetheless, the studies reviewed herein may serve as an impetus to the field, and we hope future research will build on and expand their scope. Although this line of research experienced a 15-year gap in activity after the initial wave of studies, the recent resurgence of studies examining ANS activity in at-risk/maltreating individuals suggests a renewed interest. Recent advances in technology allow for noninvasive assessment of autonomic (re)activity unconfined to laboratory settings. Future research may make use of ambulatory assessments of parents' functioning in their home environment, potentially increasing the ecological validity of findings (De Geus & Van Doornen, 1996; Kupper et al., 2005). More complex operationalizations of child maltreatment would also help advance research in this area. For instance, an expansion of the focus on physical abuse to other types of maltreatment such as emotional abuse and neglect could address relevant questions such as whether different subtypes (or combinations of subtypes) of maltreatment are associated with different autonomic response patterns. Inclusion of degrees of maltreatment severity would allow for a shift from a dichotomous to a more dynamic approach. Finally, randomized experiments using biofeedback or other experimental manipulations of ANS functioning could provide insight into the possible causal role of autonomic activity in perpetration of child maltreatment.

Such additions to the field could further support previous suggestions that maltreating parents may benefit from physiology-based stress regulation (e.g., Casanova et al., 1992; Crouch et al., 2015), but currently the field lacks randomized controlled trials on the effectiveness of such intervention components in maltreating or at-risk populations. A more interactive approach that has been found to be effective is an attachment-based, short-term intervention using video feedback, such as the Video-feedback Intervention to promote Positive Parenting (VIPP; Juffer, Bakermans-Kranenburg, & Van IJzendoorn, 2007). A randomized controlled trial with 67 dyads under surveillance for child maltreatment showed that such an intervention was effective in increasing parental sensitivity (i.e., adequate responding to children's distress; Moss et al., 2011). Future studies could examine whether the effectiveness of similar intervention programs is enhanced by including elements such as biofeedback to improve maltreating parents' stress regulation.