This is the accepted manuscript of the article, which has been published in *Psychoneuroendocrinology,* 2020, 115, 104603. https://doi.org/10.1016/j.psyneuen.2020.104603

Salivary Cortisol Reactivity to Psychological Stressors in Infancy: A Meta-Analysis

Ilmari J. A. Puhakka[1,2] & Mikko J. Peltola[2]

[1] Faculty of Education and Culture, Tampere University, Finland

[2] Human Information Processing Laboratory, Psychology, Faculty of Social Sciences, Tampere University, Finland

CORRESPONDING AUTHOR:

Mikko Peltola, Human Information Processing Laboratory, Psychology, Faculty of Social Sciences, 33014 Tampere University, Finland. Tel. +358503186120. E-mail: mikko.peltola@tuni.fi

Abstract

Measurement of salivary cortisol is a practical and non-invasive tool for studying stress reactivity to various types of stressors even in young infants. Whereas studies using physical stressors during the first months of life have found robust cortisol responses to painful stimuli, research with older infants using psychological stressors (e.g., parental separation) has produced mixed findings, limiting our understanding of potential developmental changes in cortisol reactivity across infancy. In the present study, we used meta-analysis to systematically investigate whether psychological stressor paradigms are associated with measurable cortisol responses in infants under 18 months of age and whether the magnitude of the responses is moderated by the type of psychological stressor (i.e., separation, frustration, novelty, or disruption of parental interaction), infant age, and other potential moderators. Across 48 studies (N = 4095, age range: 3-18 months), we found that commonly used psychological stressor paradigms are associated with a small (Hedges' g = .11) increase in salivary cortisol levels in typically developing infants. Stressor type moderated the effect sizes, and when effect sizes in each category were analyzed separately, only the separation studies were associated with a consistent increase in cortisol following the stressor. Age did not moderate the effect sizes either in the full set of studies or within the separate stressor types. These meta-analytic results indicate that the normative cortisol response to psychological stressors across infancy is small and emphasize the need for standardized stressor paradigms to assess cortisol responses systematically across infancy.

Keywords: Cortisol; Infant; Meta-Analysis; Saliva; Stressor

1. Introduction

Given the central role of stress in health and development, there has been a longstanding interest in charting the typical development and individual differences in stress reactivity (Gunnar & Quevedo, 2007). The methodological advancement of measuring cortisol non-invasively from the saliva has been particularly valuable for quantifying activity of the hypothalamic-pituitary-adrenal (HPA) axis in infants who are unable to subjectively report their stress responses. Salivary cortisol reactivity in infancy has been assessed in various experimental settings including physical stressors (e.g., vaccination or blood draw) and psychological stressors designed to trigger negative emotional responses (e.g., separation from the parent or presentation of novel persons or objects). Salivary cortisol reactivity is measured as the difference between a pre-stressor sample and a post-stressor sample collected typically 20–30 min after the stressor presentation.

Research on children's cortisol reactivity indicates that the magnitude of the salivary cortisol response to stressors appears to decrease across infancy and early childhood, and increase again from middle childhood to adolescence (Gunnar, Talge, & Herrera, 2009). In studies measuring infants' cortisol responses to inoculations across the first two years, Davis and Granger (2009) observed increases in cortisol in response to inoculation at 2 and 6 months, but not at 12 and 24 months, while Gunnar, Brodersen, Krueger, and Rigatuso (1996) observed increases at 2, 4, and 6 months, but not at 15 months. It has been suggested that by the end of the first year, the infant's HPA axis becomes buffered as a function of supportive caregiving relationships (see Gunnar & Quevedo, 2007, for a review). The buffered HPA response later in the first year thus represents a functional equivalent of the *hyporesponsive* period of stress reactivity that has been observed in infant rats approximately during postnatal days 4–14, which functions to protect the immature brain from excess glucocorticoids during sensitive periods of brain development (Rosenfeld, Suchecki, & Levine, 1992). In addition to the increasing influence of caregiver support on cortisol reactivity, a major methodological factor that may be associated with the observed pattern of decreased cortisol

reactivity across the first year is that after the first months, a vast majority of studies have used only psychological stressor paradigms to trigger cortisol reactivity. In reviews of the literature, both Gunnar et al. (2009) and Jansen, Beijers, Riksen-Walraven, and de Weerth (2010) noted that while physical stressors generally produced robust increases in cortisol in young infants, studies using psychological stressors in older infants produced more mixed results, with many studies failing to show increases in cortisol at the group level after a psychological stressor. Jansen et al. (2010) averaged effect sizes from multiple studies and estimated that the average effect size of studies using physical stressors to trigger cortisol responses was d = .62 for mild physical stressors and d = 1.10 for painful stimuli. For psychological stressors, results were much more variable with the averaged effect sizes ranging from d = –.15 (novelty paradigms) to d = .37 (separation paradigms).

As psychological stressor paradigms provide opportunities to investigate infants' hormonal responses to stressors that they may also encounter in real life (e.g., separation from the parent), it is important to systematically investigate to what extent psychological stressors presented in a laboratory environment trigger cortisol increases in infants and what factors might affect the magnitude of the responses. In the present study, we used meta-analysis to provide the first quantitative synthesis of studies reporting salivary cortisol responses to psychological stressors in infants between 3 and 18 months of age. Our aim was to analyze whether infants show a measurable cortisol response to psychological stressors and whether the magnitude of the responses is affected by the type of stressor (e.g., separation or novelty), age of the infant, or other potential moderating factors. For individual differences research, it is critical to design paradigms that are able to trigger variable amounts of cortisol reactivity across most infants, rather than relying on the presence or absence of a response (cf. Roos et al., 2017). Therefore, we specifically conducted the meta-analyses on data from normative samples of infants without major risk factors (e.g., preterm birth or parental substance abuse) to derive estimates of cortisol reactivity to psychological stressors in typically developing infants. Furthermore, we focused on studies conducted in the laboratory to

obtain a sufficiently standardized set of studies investigating infants' stress reactivity in a novel but controlled environment.

In addition to the reviews of the literature on infants' salivary cortisol responses to various types of stressors (Gunnar et al., 2009; Jansen et al., 2010), only one meta-analysis has been published previously (Provenzi, Giusti, & Montirosso, 2016). However, Provenzi, Giusti, and Montirosso (2016) focused only on one type of psychological stressor; disruption of parental interaction in the Still-Face paradigm (SFP; Tronick, Als, Adamson, Wise, & Brazelton, 1978), in which the infant's interaction with the parent is disrupted by the parent becoming unresponsive for a brief period. The analysis indicated that, across 10 studies, interaction disruption in the SFP is associated with a small but statistically significant cortisol increase. As the SFP is typically administered to infants only across a narrow age range of 4–8 months, it is difficult to estimate potential age-related changes in cortisol responses from the SFP data alone. Therefore, in addition to studies using the SFP, we included various other types of psychological stressors administered to infants at various ages, such as separation from the parent, exposure to novel objects or persons, and frustration caused by blocking the infant's goal. In addition to the influence of stressor type and age on cortisol responses, we considered other factors that may moderate the effect sizes across studies, such as the material of the saliva collection swab (as synthetic swabs have largely replaced the previously used cotton swabs), time of day of the saliva collection, and the "warm-up" time between the infant's arrival to the laboratory and the baseline saliva collection (i.e., acclimation period), as there are indications that infants' initial cortisol levels decrease if they are given time to familiarize with the novel laboratory environment (Zmyj, Schneider, & Seehagen, 2017).

2. Methods

2.1. Literature Search

The literature search was conducted by using PsycINFO and MEDLINE databases. We searched peer-reviewed articles in the English language available by July 18th, 2019, with the keywords cortisol and child* and/or infant* in the title or abstract (the asterisk indicating that the search contained the word or word fragment). The specific search term was *"(cortisol.ti. or cortisol.ab.) and ((infan$ or child$).ti. or (infan$ or child$).ab.)"*. The search produced a total of 5,689 articles after removal of duplications. First, titles and abstracts of the articles were screened to exclude studies that were clearly outside the scope of the current analyses (studies with adult or non-human participants, and studies using physical stressors), which left 252 articles for full text screening. The full texts were screened using the following inclusion criteria: the study was conducted in a laboratory setting, a psychological stressor was used, the subjects were under 18 months old human infants who were healthy and did not belong to a high-risk group (pre-term birth, prenatal substance exposure, major postnatal health problem, parental psychiatric disorder), and the study included cortisol measurement from saliva before and after presentation of the stressor. From studies comparing control and risk groups, data from the control group was included in the meta-analysis to best reflect cortisol reactivity in typically developing infants. Figure 1 describes the study selection process.

In case of duplicate samples (i.e., the same infant providing data for two or more studies), we chose the largest or most representative study. Articles in which the same infants were assessed at multiple timepoints, the first assessment was included with the exception of one study (Montirosso, Tronick, Morandi, Ciceri, & Borgatti, 2013), where assessments of two separate groups were included (the control group and the assessment of the first stressor exposure of the experimental group). In instances where cortisol means and standard deviations were not reported in the articles,

the authors were contacted for the relevant information. For six articles, the required values were obtained from the systematic review by Jansen et al. (2010). Overall, a total of 49 effect sizes from 48 articles were included in this meta-analysis.

2.2. Moderators

Stressor type was the most relevant categorical moderator in the analysis. A total of 18 different stressors were observed in the set of studies. These stressors were grouped into four categories: 1) *Separation*, including studies in which the infant is separated temporarily from the parent (e.g., the Strange Situation Procedure, SSP; Ainsworth, Blehar, Waters, & Wall, 1978), 2) *Frustration*, e.g., studies using the Arm Restraint Procedure (Stifter & Fox, 1990) in which infants are prevented to move their arms for a short period of time, 3) *Novelty*, in which the infant is presented with a novel situation or object, (e.g., Grunau, Weinberg, & Whitfield, 2004), and 4) *Interaction Disruption*, which includes stressors utilizing the Still-Face paradigm (SFP; Tronick et al., 1978). The chosen categories are similar to the categories used by Gunnar et al. (2009) and Jansen et al. (2010) for the first three. The fourth category (*Interaction Disruption*) was added because the usage of this stressor type has increased notably during the last decade (Provenzi, Giusti, & Montirosso, 2016). The *Interaction Disruption* studies included in the current analyses used either the 3- or 5-episodic SFP. The traditional 3-episodic stressor by Tronick et al. (1978) consists of a play, still-face, and reunion episodes (A-B-A), whereas the modified 5-episodic version includes additional still-face and reunion episodes (A-B-A-B-A; Haley & Stansbury, 2003). In addition to stressor type, the material of the saliva collection instrument (cotton vs. synthetic) and the time of day when the study was conducted (between 8–12 or 12–18) were added as categorical moderators. Participants' mean age (in months) was the most relevant continuous moderator in the analysis. The other continuous moderator was the duration of the acclimation period (in minutes). Descriptive information about the studies included in this meta-analysis is presented in Table 1.

2.3. Meta-Analytic Procedures

The meta-analyses were performed using the Comprehensive Meta-Analysis program (CMA, version 3.0; Borenstein, Rothstein, & Cohen, 2013). Effect sizes (Hedges' *g*) were calculated for each study as a standardized mean difference between cortisol levels before and after presentation of stressor. Increase in cortisol level was defined as a positive effect size. Hedges' *g* was chosen as the effect size because Cohen's *d* tends to overestimate the population effect size especially when the sample size is smaller than 50 (Cumming, 2012; Lakens, 2013). In terms of interpretation, the effect sizes expressed as Hedges' *g* and Cohen's *d* are directly comparable. Means (*M*), standard deviations (*SD*) and correlations (*r*) were primarily used in calculating effect sizes. In cases where these values were not reported, effect sizes were calculated using the mean difference (one study) or the correlation coefficient (one study). Correlations between the cortisol assessment points were seldom reported in the studies, thus we estimated the correlation to be r = .60 based on the studies which reported the coefficient. In the current set of studies, cortisol was measured 1–2 times before the presentation of stressor and 1–3 times after the stressor. Cortisol assessment points at the baseline and 20–30 minutes after the stressor were chosen to obtain a comparable set of studies reflecting the typical temporal pattern of cortisol reactivity (Gunnar et al., 2009).

First, the main analysis including all studies produced a combined weighted effect size (Hedges' *g*) and its 95% confidence interval (95% CI). Statistical testing for heterogeneity of the effect sizes was carried out using *Q*-tests. The main analysis was carried out using random-effects models because of the variation in study practices (Borenstein, Hedges, Higgins, & Rothstein, 2009; Field & Gillett, 2010). Following this model selection, we wanted to examine variance of the true effects in addition to the combined effect size. Estimates of variance (T^2) and standard deviation (T) of the true effect sizes, and proportion of true variance from observed variance (I^2) were thus calculated. Finally, we calculated a 95% prediction interval (95% PI) for the combined weighted

effect size, which is the estimated distribution of the true effect sizes around the combined weighted effect size. In addition to the *p*-value of the *Q*-test, the certainty of this heterogeneity estimate was examined by calculating 95% confidence intervals for I^2.

Combined weighted effect sizes were calculated also for the subgroups of the categorical moderators. Differences between subgroups were examined with *Q*-tests and these analyses were conducted using mixed-effect models, assuming no common among study variance across subgroups (Borenstein et al., 2009). Proportion of true variance of the observed variance (I^2) was calculated for each subgroup. Finally, we calculated 95% prediction intervals for the combined weighted effect sizes of each subgroup. Meta-regression was used to examine effects of the continuous moderators on the effect sizes. Effects of age and duration of the acclimation period were examined for all effect sizes and additionally for the stressor type subgroups. Using the CMA program, we calculated regression coefficients (*b*) which were tested for statistical significance (*Q*) and calculated the proportion of total between-study variance explained by the model (R^2).

Potential outliers were screened by calculating Fisher's *Z*-values for each effect size and standardizing them. No outliers (standardized *Z*-value >|3.29|; Tabachnick & Fidell, 2001) were observed. We examined potential publication bias graphically by observing funnel plots of the effect sizes against their standard errors. Additionally, we calculated Egger's *t*-test (Egger, Davey Smith, Schneider, & Minder, 1997) to examine statistical asymmetry in the funnel plot and used Rosenthal's fail-safe method (Rosenthal, 1991) to estimate the number of missing studies that would be required to nullify the effect. Finally, we used *p*-curve-analysis (Simonsohn, Nelson, & Simmons, 2014) to test whether the distribution of significant *p*-values was right-skewed towards smaller *p*-values, thus indicating that the published results have evidential value.

3. Results

3.1. Main Analysis

The main analysis (N = 4095) revealed a significant combined weighted effect size that was significantly different from zero (g = .11, SE = .03, CI [.05, .17], p < .001), indicating that psychological stressors cause a relatively small increase on infants' cortisol levels. Heterogeneity across effect sizes was large (Q = 188.78, p < .001) and had an estimated true variance of T^2 = .03, CI [.02, .04], and a true standard deviation of T = .17, CI [.14, .21]. Proportion of true variance in the total observed variance was high (I^2 = 75.10%, CI [66.45, 80.73]). The 95% prediction interval of the combined weighted effect size was [–.24, .46]. Effect sizes of each individual study are presented in Figure 2 with the forest plots grouped by stressor type.

3.2. Moderator Analyses

Combined effect sizes and statistics of the categorical moderators are displayed in Table 2. Stressor category had a moderation effect when all four stressor categories were taken into consideration (Q = 8.13, p = .043). When examined separately, only the *Separation* stressor type had a mean effect size that differed significantly from zero (g = .15, p = .001). The *Interaction Disruption* stressor type was associated with a marginal effect size (g = .16, p = .051), while the mean effect sizes for *Frustration* (g = –.004, p = .91) and *Novelty* (g = .12, p = .14) were not significantly different from zero. Within the Still-Face studies comprising the *Interaction Disruption* stressor type, task type (i.e., 3- or 5-episodic task) did not moderate the effect sizes (Q = 1.55, p = .21). Material of the saliva collection instrument (cotton or synthetic) did not show a moderation effect (Q = .03, p = .87), and the mean effect sizes were of comparable magnitude between studies using synthetic (g = .11, p = .010), and cotton materials (g = .12, p = .009). Finally,

effect sizes did not differ as a function of time of day when the data were collected, $Q = .33$, $p = .57$ (8–12: $g = .09$, $p = .002$; 12–18: $g = .14$, $p = .01$).

Statistics of the continuous moderators are displayed in Table 3. Age did not have a moderation effect on effect sizes ($Q = .21$, $p = .65$) in the whole set of studies or when examining the stressor types separately. However, the stressor types differed in mean age (Figure 2) so that the *Separation* stressor type consisted of older infants (mean age: 13.59 months), whereas the other stressor types were more similar to each other regarding age (mean age: 6.71 [*Frustration*]; 7.59 [*Novelty*]; 5.13 [*Interaction Disruption*] months). Duration of the acclimation period had a positive moderation effect on effect sizes ($Q = 6.36$, $p = .01$), that is, longer acclimation periods were associated with larger effect sizes (Figure 3). Duration of the acclimation period explained 29% of total between-study variance. When examining the stressor types separately, duration of the acclimation period significantly moderated the effect sizes of the *Separation* ($Q = 7.53$, $p = .006$), *Novelty* ($Q = 8.62$, $p = .003$), and *Frustration* ($Q = 5.72$, $p = .02$) stressor types, whereas the moderation effect for the *Interaction Disruption* ($Q = 3.68$, $p = .06$) stressor type was marginal. However, only the *Separation* stressor type had an acceptable model fit in terms of goodness-of-fit values, indicating that the moderation analyses for the other stressor types are not reliable.

3.3. Publication Bias

In the funnel plot, the studies were spread relatively evenly around the combined weighted effect size, which indicates the absence of a major publication bias (Figure 4). Statistical asymmetry in the funnel plot was not detected (Egger's $t = .75$, $p = .46$) and the number of null results required to change the combined weighted effect size to non-significant was notably high (Fail-safe $N = 561$). The *p*-curve-analysis revealed a distinctly right-skewed *p*-curve, p_{full}: $Z = -7.67$, $p < .001$, p_{half}: $Z = -8.48$, $p < .001$ (Figure 5), likewise indicating that the results have evidential value.

4. Discussion

The first aim of the present meta-analysis was to examine to what extent psychological stressor paradigms presented in the laboratory trigger increased salivary cortisol responses in infants under 18 months of age. Cortisol responses were analyzed as the difference in cortisol concentrations in saliva samples collected before and after the stressor presentation. As the number of available studies was fairly large, we were able to analyze potential moderator effects and evaluate the reliability of the included studies with estimates of publication bias. We analyzed whether the magnitude of the cortisol responses is affected by the type of the stressor paradigm, infant age, and procedural moderators such as the acclimation period upon entering the laboratory. To be able to estimate the magnitude of cortisol reactivity to psychological stressors in typically developing infants, the analyses were focused on data from normative samples of infants without major risk factors.

Across 48 effect sizes from 49 studies (N = 4095), we found that commonly used psychological stressor paradigms are associated with a small (Hedges' g = .11) increase in salivary cortisol concentrations in typically developing infants, and no indications of selective reporting were observed. The combined effect size appears to be considerably smaller than the averaged effect sizes of d = .62–1.10 across studies using physical stressors in young infants (Jansen et al., 2010) and the meta-analytic effect size of d = .31 across studies using psychological stressors in adults (Dickerson & Kemeny, 2004). Estimates of heterogeneity and prediction intervals indicated considerable variability in the effect sizes, which made it important to analyze in more detail whether some factors are associated with more pronounced effects on infants' cortisol reactivity.

4.1. Differences Between Stressor Types

We included four most commonly used types of psychological stressors (separation, frustration, novelty, and interaction disruption) in the moderation analysis. The same stressor types were also included in the systematic review by Jansen et al. (2010) with the exception of interaction disruption studies using the Still-Face paradigm (SFP) which have markedly increased during the last decade. When all stressor types were considered together, the moderation test indicated that the stressor paradigms differed in their ability to trigger increases in cortisol. When examining effect sizes in each category separately, only the separation studies were associated with a consistent increase in cortisol following the stressor, with 67% of the separation studies showing a positive effect size, closely paralleling the results of Jansen et al. (2010) (63%). For other stressor types, the combined weighted effect sizes ranged from g = -.004 (frustration) to g = .16 (interaction disruption) and the effect sizes of the individual studies within these stressor types were distributed fairly evenly around zero. The consistent effect observed in studies using parental separation (most commonly the Strange Situation Procedure; Table 1) demonstrates that separation from the attachment figure is one of the major stressors experienced by infants. By indicating cortisol responses to parental separation at the group level, the current result parallels previous findings showing heightened autonomic arousal during the SSP also in avoidantly attached infants not showing overt distress responses to parental separation (Spangler & Grossmann, 1993).

The small combined effect size in stressor paradigms using frustration is in line with the results of Jansen et al. (2010). Compared to other stressor types, the effect sizes of the frustration paradigms had relatively small variation indicating their weak ability to induce cortisol reactivity in infants. However, it should be noted that many of the stressors in this category were part of a stressor battery (e.g., Lab-TAB; Goldsmith & Rothbart, 1999). Thus, even though the post-stressor saliva collection was often scheduled to reflect peak cortisol following the frustration component of the battery, it is possible that other stressors might have influenced the potential cortisol-inducing

effects of the frustration stressor. Regarding studies using novelty as the stressor type, it should be noted that this category consisted of only few studies (k = 5) which differed distinctly from each other, hence any conclusive interpretations regarding this stressor type should await for the accumulation of a larger number of studies using novelty paradigms.

Studies on interaction disruption using the Still-Face paradigm had the largest combined effect size of the four stressor types, but also the highest true variance (Table 2). Our analysis found a smaller combined effect size than Provenzi, Giusti and Montirosso (2016) in their meta-analysis of studies using the SFP (d = .24). There were however two differences in the analytical methods between these two meta-analyses which may have affected the results. First, for the post-stressor measurement point we always used values taken 20-30 min after stressor exposure, whereas Provenzi, Giusti, and Montirosso (2016) selected the peak cortisol value occurring in any saliva measurement taken within 10-40 min after stressor exposure. Secondly, Provenzi, Giusti, and Montirosso (2016) included more than one effect size from individual studies, whereas we included only one effect size per study (excluding Montirosso et al., 2013, for which effect sizes from two separate groups were included). With respect to different types of Still Face paradigms (i.e., 3- or 5-episodic task), we did not find a significant difference between task types. Nevertheless, the 5-episodic stressors had a relatively large combined effect size (g = .25), whereas the combined effect size of the 3-episodic stressors was small, which together indicate that the 5-episodic type should be preferred when studying infants' cortisol responses to interaction disruption using the Still-Face paradigm.

4.2. Age-Related Changes in Cortisol Reactivity

The results of the present meta-analysis did not indicate that salivary cortisol responses decrease from 3 to 18 months of age as age did not significantly moderate the effect sizes either in the full set of studies or within the separate stressor types. However, studies tend to use specific

stressors with certain age groups (Figure 2), which must be taken into consideration when examining age-related changes. For 0–3-month-old infants only mild physical stressors (e.g., inoculation) have been used (Jansen et al., 2010), which limited our analysis to infants older than 3 months of age. After the first months, frustration and interaction disruption are commonly used stressors until 10 months of age. After the first birthday and until 18 months, only stressors using separation were available for the current analyses. Therefore, the variability of psychological stressors across infancy limits our ability to speculate on potential age-related changes in cortisol reactivity to psychological stressors. However, it is interesting to note that in the present dataset, only the separation studies, which were administered with older infants, were associated with consistent cortisol reactivity. This could be considered to be at odds with results of the few studies that presented physical stressors across infancy and showed dampening of cortisol responses to physical stressors after 6 months of age (i.e., at 12-24 months; Davis & Granger, 2009; Gunnar et al., 1996). To determine whether infants show a similar pattern of change to psychological stressors, directly comparable psychological stressors should be applied in a similar manner to a wide age range of infants. Given that separation from the caregivers is a natural and commonly experienced stressor, and separation was the only type of stressor associated with consistent cortisol reactivity, future studies may benefit from using age-appropriate measures of parental separation across infancy as the psychological stressor for investigating the development of cortisol reactivity. While the Strange Situation is not directly applicable for use with younger infants who do not move independently, measurements of reactivity to parental separation during free play assessments could be easily administered.

4.3. Other Moderators

In line with the results of Zmyj et al. (2017), the analysis of studies reporting the duration of time between the infant entering the laboratory and the pre-stressor saliva sample indicated that

cortisol responses to stressors were larger when the infants were given more time to acclimatize to the novel laboratory environment. As information about the length of the acclimation period was provided by less than half of the studies, the results should be interpreted with caution. Nevertheless, as this association suggests that initially high cortisol levels directly upon entering the laboratory may mask some of the effects triggered by the stressor presentation, in future research it will be important to strive for more standardized laboratory protocols including fixed and sufficiently documented acclimation periods. Additional studies on the effects of varying acclimation periods on cortisol levels are also needed to ascertain what amount of time, on average, is enough for cortisol levels to return to a regular level after arrival to the laboratory.

Effect sizes were not affected by the time of day of the saliva collection in the main analysis nor when analyzing stressor types separately. Even though in our analysis the included cortisol measurement points were relatively close to each other (the gap between pre- and post-stressor measurements was 17–56 min), these results are consistent with studies on the development of the circadian rhythm in childhood (Gunnar & Donzella, 2002) in which differences in cortisol levels between the morning and mid-afternoon have not been observed in children until 4 years of age.

Regarding the influence of different saliva collection instruments, our analysis did not indicate that using cotton-based instruments results in systematically lower cortisol levels than when using synthetical saliva swabs, although such differences have been observed in previous studies with adults (e.g., Gröschl & Rauh, 2006). It should be noted that we focused on cortisol *reactivity* and did not compare cortisol *levels* in single measurement points between studies. Therefore, it is possible that differences in cortisol levels caused by collection material are masked because cortisol reactivity accounts only for the difference between cortisol measurements within a study and not between studies using cotton or synthetic instruments. Nevertheless, given that direct comparisons have indicated differences between synthetic and cotton-based instruments (Gröschl &

Rauh, 2006), we recommend systematically using synthetic swabs in studies assessing infants' cortisol.

In the future, it will also be important to investigate with sufficiently large samples whether there are sex differences in infants' cortisol reactivity, because results from previous studies have been mixed (e.g., Davis & Emory, 1995; Kirschbaum, Wüst, & Hellhammer, 1992; Tout, Haan, Campbell, & Gunnar, 2008). In the present meta-analysis, we were able to obtain only information about the sex distributions of some of the studies, thus analyzing sex differences in cortisol reactivity in a reliable way was not possible. Another potentially relevant moderator we were not able to examine for the same reasons as with sex was race/ethnicity. A recent study by Dismukes et al. (2018) indicated that race may moderate cortisol reactivity in 12-month-old infants in the SFP. This result in addition to results from adults (e.g., Chong, Uhart, McCaul, Johnson, & Wand, 2008; Wilcox, Bopp, Wilson, Fulk, & Hand, 2005) underlines the need for more research on potential race differences on infants' cortisol reactivity.

4.4. Conclusions

Our main analysis found that commonly used psychological stressor paradigms are associated with a small increase in salivary cortisol concentrations in typically developing infants. Investigation of a large number of studies also highlighted the large heterogeneity in the magnitude of the effects and also in the procedures of the included studies. Age, time of day, and saliva collection instrument material did not significantly moderate the effect sizes in the full set of studies or within the separate stressor types, but the duration of the acclimation period was found to be associated with the magnitude of cortisol reactivity. Although a clear association between infant age and cortisol reactivity was not found, the most consistent cortisol responses were observed in the separation studies, which were also administered to infants who were older than infants in the other stressor categories. While this finding appears inconsistent with studies showing dampened cortisol

responses to physical stressors in older infants (Davis & Granger, 2009; Gunnar et al., 1996), determining the age-related changes in cortisol reactivity to psychological stressors reliably requires that comparable psychological stressors are applied in a similar manner to a wide age range of infants. Our recommendation would be to use age-appropriate variations of parental separation as the most feasible psychological stressor across infancy.

In conclusion, a somewhat concerning overall outcome of this analysis was that despite the large number of studies, cortisol reactivity to psychological stressor paradigms in infancy appears to be small and inconsistent. This is problematic for research aiming to characterize individual differences in stress reactivity to naturally occurring stressors. Without a clear understanding of cortisol reactivity to psychological stressors in typically developing infants, determining what constitutes an endocrine marker of risk for the development of the stress systems remains very challenging. The contribution of this meta-analysis is to help to elucidate how efficiently the currently used experimental paradigms trigger cortisol reactivity in typically developing infants, which will be important for attempts to design reliable stressor paradigms for research on individual differences in infants' cortisol reactivity to stress.

Acknowledgments

This research was supported by a grant from the Academy of Finland (#307657) to M.J.P. Conflicts of interest: none.

References

(Studies included in the meta-analysis are marked with an asterisk)

Ainsworth, M. D. S., Blehar, M. C., Waters, E., & Wall, S. (1978). *Patterns of attachment: A psychological study of the strange situation*. Hillsdale, NJ: Erlbaum.

*Atkinson, L., Gonzalez, A., Kashy, D. A., Santo Basile, V., Masellis, M., Pereira, J., … Levitan, R. (2013). Maternal sensitivity and infant and mother adrenocortical function across challenges. *Psychoneuroendocrinology*, *38*(12), 2943–2951. https://doi.org/10.1016/J.PSYNEUEN.2013.08.001

*Azar, R., Paquette, D., Zoccolillo, M., Baltzer, F., & Tremblay, R. E. (2007). The association of major depression, conduct disorder, and maternal overcontrol with a failure to show a cortisol buffered response in 4-month-old infants of teenage mothers. *Biological Psychiatry*, *62*(6), 573–579. https://doi.org/10.1016/j.biopsych.2006.11.009

*Beijers, R., Riksen-Walraven, M., Sebesta, K., & de Weerth, C. (2017). Differential associations between behavioral and cortisol responses to a stressor in securely versus insecurely attached infants. *Behavioural Brain Research*, *325*, 147–155. https://doi.org/10.1016/j.bbr.2016.10.008

*Bernard, K., & Dozier, M. (2010). Examining infants' cortisol responses to laboratory tasks among children varying in attachment disorganization: Stress reactivity or return to baseline? *Developmental Psychology*, *46*(6), 1771–1778. https://doi.org/10.1037/a0020660

*Bernard, N. K., Kashy, D. A., Levendosky, A. A., Bogat, G. A., & Lonstein, J. S. (2017). Do different data analytic approaches generate discrepant findings when measuring mother-infant HPA axis attunement? *Developmental Psychobiology*, *59*(2), 174–184. https://doi.org/10.1002/dev.21474

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to Meta-Analysis*. Chichester, UK: John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470743386

Borenstein, M., Rothstein, D., & Cohen, J. (2013). Comprehensive meta-analysis: A computer

program for research synthesis. Englewood, NJ: Biostat.

*Bosquet Enlow, M., King, L., Schreier, H. M., Howard, J. M., Rosenfield, D., Ritz, T., & Wright, R. J. (2014). Maternal sensitivity and infant autonomic and endocrine stress responses. *Early Human Development, 90*(7), 377–385. https://doi.org/10.1016/j.earlhumdev.2014.04.007

Chong, R. Y., Uhart, M., McCaul, M. E., Johnson, E., & Wand, G. S. (2008). Whites have a more robust hypothalamic-pituitary-adrenal axis response to a psychological stressor than blacks. *Psychoneuroendocrinology, 33*(2), 246–254. https://doi.org/10.1016/j.psyneuen.2007.10.014

*Conradt, E., Fei, M., LaGasse, L., Tronick, E., Guerin, D., Gorman, D., … Lester, B. M. (2015). Prenatal predictors of infant self-regulation: The contributions of placental DNA methylation of NR3C1 and neuroendocrine activity. *Frontiers in Behavioral Neuroscience, 9,* 130. https://doi.org/10.3389/fnbeh.2015.00130

Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. New York, NY: Routledge.

Davis, E. P., & Granger, D. A. (2009). Developmental differences in infant salivary alpha-amylase and cortisol responses to stress. *Psychoneuroendocrinology, 34*(6), 795–804. https://doi.org/10.1016/J.PSYNEUEN.2009.02.001

Davis, M., & Emory, E. (1995). Sex differences in neonatal stress reactivity. *Child Development, 66*(1), 14–27. https://doi.org/10.1111/j.1467-8624.1995.tb00852.x

Dickerson, S. S., & Kemeny, M. E. (2004). Acute stressors and cortisol responses: A theoretical integration and synthesis of laboratory research. *Psychological Bulletin, 130*(3), 355–391. https://doi.org/10.1037/0033-2909.130.3.355

*DiCorcia, J. A., Snidman, N. C., & Tronick, E. (2018). Here today, gone tomorrow? Changes in 4-month-olds' physiologic and behavioral responses do not indicate memory for a social stressor. *Frontiers in Psychology, 9,* 128. https://doi.org/10.3389/fpsyg.2018.00128

*Diego, M. A., Field, T., Jones, N. A., Hernandez-Reif, M., Cullen, C., Schanberg, S., & Kuhn, C.

(2004). EEG responses to mock facial expressions by infants of depressed mothers. *Infant Behavior & Development*, *27*(2), 150–162. https://doi.org/10.1016/j.infbeh.2003.10.001

*Dismukes, A., Shirtcliff, E., Jones, C. W., Zeanah, C., Theall, K., & Drury, S. (2018). The development of the cortisol response to dyadic stressors in Black and White infants. *Development and Psychopathology*, *30*(5), 1995–2008. https://doi.org/10.1017/S0954579418001232

Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, *315*(7109), 629–634. https://doi.org/10.1136/bmj.315.7109.629

*Eiden, R. D., Molnar, D. S., Granger, D. A., Colder, C. R., Schuetze, P., & Huestis, M. A. (2015). Prenatal tobacco exposure and infant stress reactivity: Role of child sex and maternal behavior. *Developmental Psychobiology*, *57*(2), 212–225. https://doi.org/10.1002/dev.21284

*Eiden, R. D., Veira, Y., & Granger, D. A. (2009). Prenatal cocaine exposure and infant cortisol reactivity. *Child Development*, *80*(2), 528–543. https://doi.org/10.1111/j.1467-8624.2009.01277.x

*Erickson, S. J., MacLean, P., Qualls, C., & Lowe, J. R. (2013). Differential associations between infant affective and cortisol responses during the Still Face paradigm among infants born very low birth weight versus full-term. *Infant Behavior and Development*, *36*(3), 359–368. https://doi.org/10.1016/J.INFBEH.2013.03.001

*Feldman, R., Singer, M., & Zagoory, O. (2010). Touch attenuates infants' physiological reactivity to stress. *Developmental Science*, *13*(2), 271–278. https://doi.org/10.1111/j.1467-7687.2009.00890.x

Field, A. P., & Gillett, R. (2010). How to do a meta-analysis. *British Journal of Mathematical and Statistical Psychology*, *63*(3), 665–694. https://doi.org/10.1348/000711010X502733

*Frigerio, A., Ceppi, E., Rusconi, M., Giorda, R., Raggi, M. E., & Fearon, P. (2009). The role

played by the interaction between genetic factors and attachment in the stress response in infancy. *Journal of Child Psychology and Psychiatry*, *50*(12), 1513–1522. https://doi.org/10.1111/j.1469-7610.2009.02126.x

*Goldberg, S., Levitan, R., Leung, E., Masellis, M., Basile, V. S., Nemeroff, C. B., & Atkinson, L. (2003). Cortisol concentrations in 12- to 18-Month-Old infants: Stability over time, location, and stressor. *Biological Psychiatry*, *54*(7), 719–726. https://doi.org/10.1016/S0006-3223(03)00010-6

Goldsmith, H. H., & Rothbart, M. K. (1999). Laboratory Temperament Assessment Battery (Lab-TAB), Version 3.1. Wisconsin Temperament and Twin Center, University of Wisconsin, Madison.

*Grant, K.-A., McMahon, C., Austin, M.-P., Reilly, N., Leader, L., & Ali, S. (2009). Maternal prenatal anxiety, postnatal caregiving and infants' cortisol responses to the still-face procedure. *Developmental Psychobiology*, *51*(8), 625–637. https://doi.org/10.1002/dev.20397

Gröschl, M., & Rauh, M. (2006). Influence of commercial collection devices for saliva on the reliability of salivary steroids analysis. *Steroids*, *71*(13–14), 1097–1100. https://doi.org/10.1016/J.STEROIDS.2006.09.007

*Grunau, R. E., Weinberg, J., & Whitfield, M. F. (2004). Neonatal procedural pain and preterm infant cortisol response to novelty at 8 months. *Pediatrics*, *114*(1), e77–e84. https://doi.org/10.1542/peds.114.1.e77

Gunnar, M., & Quevedo, K. (2007). The neurobiology of stress and development. *Annual Review of Psychology*, *58*, 145–173. https://doi.org/10.1146/annurev.psych.58.110405.085605

Gunnar, M. R., Brodersen, L., Krueger, K., & Rigatuso, J. (1996). Dampening of adrenocortical responses during infancy: Normative changes and individual differences. *Child Development*, *67*(3), 877–889. https://doi.org/10.1111/j.1467-8624.1996.tb01770.x

Gunnar, M. R., & Donzella, B. (2002). Social regulation of the cortisol levels in early human

development. *Psychoneuroendocrinology*, *27*(1–2), 199–220. https://doi.org/10.1016/S0306-4530(01)00045-2

*Gunnar, M. R., Larson, M. C., Hertsgaard, L., Harris, M. L., & Brodersen, L. (1992). The stressfulness of separation among nine-month-old infants: Effects of social context variables and infant temperament. *Child Development*, *63*(2), 290–303. https://doi.org/10.1111/j.1467-8624.1992.tb01627.x

*Gunnar, M. R., Mangelsdorf, S., Larson, M., & Hertsgaard, L. (1989). Attachment, temperament, and adrenocortical activity in infancy: A study of psychoendocrine regulation. *Developmental Psychology*, *25*(3), 355–363. https://doi.org/10.1037/0012-1649.25.3.355

*Gunnar, M. R., & Nelson, C. A. (1994). Event-related potentials in year-old infants: Relations with emotionality and cortisol. *Child Development*, *65*(1), 80. https://doi.org/10.1111/j.1467-8624.1994.tb00736.x

Gunnar, M. R., Talge, N. M., & Herrera, A. (2009). Stressor paradigms in developmental studies: What does and does not work to produce mean increases in salivary cortisol. *Psychoneuroendocrinology*, *34*(7), 953–967. https://doi.org/10.1016/j.psyneuen.2009.02.010

*Haley, D. W. (2011). Relationship disruption stress in human infants: A validation study with experimental and control groups. *Stress*, *14*(5), 530–536. https://doi.org/10.3109/10253890.2011.560308

Haley, D. W., & Stansbury, K. (2003). Infant stress and parent responsiveness: Regulation of physiology and behavior during Still-Face and reunion. *Child Development*, *74*(5), 1534–1546. https://doi.org/10.1111/1467-8624.00621

*Hibel, L. C., Trumbell, J. M., Valentino, K., & Buhler-Wassmann, A. C. (2018). Ecologically salient stressors and supports and the coordination of cortisol and salivary alpha-amylase in mothers and infants. *Physiology & Behavior*, *195*, 48–57. https://doi.org/10.1016/j.physbeh.2018.07.024

*Holochwost, S. J., Gariépy, J.-L., Mills-Koonce, W. R., Propper, C. B., Kolacz, J., & Granger, D. A. (2017). Individual differences in the activity of the hypothalamic pituitary adrenal axis: Relations to age and cumulative risk in early childhood. *Psychoneuroendocrinology*, *81*, 36–45. https://doi.org/10.1016/j.psyneuen.2017.03.023

*Huot, R. L., Brennan, P. A., Stowe, Z. N., Plotsky, P. M., & Walker, E. F. (2004). Negative affect in offspring of depressed mothers is predicted by infant cortisol levels at 6 months and maternal depression during pregnancy, but not postpartum. *Annals of the New York Academy of Sciences*, *1032*(1), 234–236. https://doi.org/10.1196/annals.1314.028

Jansen, J., Beijers, R., Riksen-Walraven, M., & de Weerth, C. (2010). Cortisol reactivity in young infants. *Psychoneuroendocrinology*, *35*(3), 329–338. https://doi.org/10.1016/j.psyneuen.2009.07.008

Kirschbaum, C., Wüst, S., & Hellhammer, D. (1992). Consistent sex differences in cortisol responses to psychological stress. *Psychosomatic Medicine*, *54*(6), 648–657. https://doi.org/10.1097/00006842-199211000-00004

*Kuo, P. X., Saini, E. K., Tengelitsch, E., & Volling, B. L. (2019). Is one secure attachment enough? Infant cortisol reactivity and the security of infant-mother and infant-father attachments at the end of the first year. *Attachment & Human Development*, *21*(5), 426–444. https://doi.org/10.1080/14616734.2019.1582595

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4:863*. https://doi.org/10.3389/fpsyg.2013.00863

*Larson, M. C., Gunnar, M. R., & Hertsgaard, L. (1991). The effects of morning naps, car trips, and maternal separation on adrenocortical activity in human infants. *Child Development*, *62*(2), 362–372. https://doi.org/10.1111/j.1467-8624.1991.tb01537.x

*Laurent, H. K., Duncan, L. G., Lightcap, A., & Khan, F. (2017). Mindful parenting predicts

mothers' and infants' hypothalamic-pituitary-adrenal activity during a dyadic stressor. *Developmental Psychology*, *53*(3), 417–424. https://doi.org/10.1037/dev0000258

*Laurent, H. K., Harold, G. T., Leve, L., Shelton, K. H., & Van Goozen, S. H. M. (2016). Understanding the unfolding of stress regulation in infants. *Development and Psychopathology*, 1–10. https://doi.org/10.1017/S0954579416000171

*Lewis, M., & Ramsay, D. (2005). Infant emotional and cortisol responses to goal blockage. *Child Development*, *76*(2), 518–530. https://doi.org/10.1111/j.1467-8624.2005.00860.x

*Lewis, M., Ramsay, D. S., & Sullivan, M. W. (2006). The relation of ANS and HPA activation to infant anger and sadness response to goal blockage. *Developmental Psychobiology*, *48*(5), 397–405. https://doi.org/10.1002/dev.20151

*Luijk, M. P. C. M., Velders, F. P., Tharner, A., van IJzendoorn, M. H., Bakermans-Kranenburg, M. J., Jaddoe, V. W. V., … Tiemeier, H. (2010). FKBP5 and resistant attachment predict cortisol reactivity in infants: Gene–environment interaction. *Psychoneuroendocrinology*, *35*(10), 1454–1461. https://doi.org/10.1016/j.psyneuen.2010.04.012

*Martinez-Torteya, C., Figge, C. J., Gilchrist, M. A., Muzik, M., King, A. P., & Sorenson, M. (2018). Prenatal intimate partner violence exposure predicts infant biobehavioral regulation: Moderation by the brain-derived neurotrophic factor (BDNF) gene. *Development and Psychopathology*, *30*(3), 1009–1021. https://doi.org/10.1017/S0954579418000329

*Montirosso, R., Tronick, E., Morandi, F., Ciceri, F., & Borgatti, R. (2013). Four-month-old infants' long-term memory for a stressful social event. *PLoS ONE*, *8*(12), e82277. https://doi.org/10.1371/journal.pone.0082277

*Nachmias, M., Gunnar, M., Mangelsdorf, S., Parritz, R. H., & Buss, K. (1996). Behavioral inhibition and stress reactivity: The moderating role of attachment security. *Child Development*, *67*(2), 508–522. https://doi.org/10.1111/j.1467-8624.1996.tb01748.x

*O'Connor, T. G., Bergman, K., Sarkar, P., & Glover, V. (2013). Prenatal cortisol exposure

predicts infant cortisol response to acute stress. *Developmental Psychobiology*, *55*(2), 145–155. https://doi.org/10.1002/dev.21007

*Oberlander, T. F., Grunau, R., Mayes, L., Riggs, W., Rurak, D., Papsdorf, M., … Weinberg, J. (2008). Hypothalamic–pituitary–adrenal (HPA) axis function in 3-month old infants with prenatal selective serotonin reuptake inhibitor (SSRI) antidepressant exposure. *Early Human Development*, *84*(10), 689–697. https://doi.org/10.1016/j.earlhumdev.2008.06.008

*Provenzi, L., Giusti, L., Fumagalli, M., Tasca, H., Ciceri, F., Menozzi, G., … Montirosso, R. (2016). Pain-related stress in the Neonatal Intensive Care Unit and salivary cortisol reactivity to socio-emotional stress in 3-month-old very preterm infants. *Psychoneuroendocrinology*, *72*, 161–165. https://doi.org/10.1016/j.psyneuen.2016.07.010

Provenzi, L., Giusti, L., & Montirosso, R. (2016). Do infants exhibit significant cortisol reactivity to the Face-to-Face Still-Face paradigm? A narrative review and meta-analysis. *Developmental Review*, *42*, 34–55. https://doi.org/10.1016/j.dr.2016.07.001

*Reck, C., Müller, M., Tietz, A., & Möhler, E. (2013). Infant distress to novelty is associated with maternal anxiety disorder and especially with maternal avoidance behavior. *Journal of Anxiety Disorders*, *27*(4), 404–412. https://doi.org/10.1016/j.janxdis.2013.03.009

Roos, L. E., Giuliano, R. J., Beauchamp, K. G., Gunnar, M., Amidon, B., & Fisher, P. A. (2017). Validation of autonomic and endocrine reactivity to a laboratory stressor in young children. *Psychoneuroendocrinology*, *77*, 51–55. https://doi.org/10.1016/j.psyneuen.2016.11.023

Rosenfeld, P., Suchecki, D., & Levine, S. (1992). Multifactorial regulation of the hypothalamic-pituitary-adrenal axis during development. *Neuroscience and Biobehavioral Reviews*, *16*(4), 553–568. https://doi.org/10.1016/S0149-7634(05)80196-4

Rosenthal, R. (1991). *Meta-analytic procedures for social research*. Newbury Park, CA: Sage.

*Schuetze, P., Lopez, F. A., Granger, D. A., & Eiden, R. D. (2008). The association between prenatal exposure to cigarettes and cortisol reactivity and regulation in 7-month-old infants.

Developmental Psychobiology, *50*(8), 819–834. https://doi.org/10.1002/dev.20334

*Seehagen, S., Schneider, S., Rudolph, J., Ernst, S., & Zmyj, N. (2015). Stress impairs cognitive flexibility in infants. *Proceedings of the National Academy of Sciences*, *112*(41), 12882–12886. https://doi.org/10.1073/pnas.1508345112

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-curve: A key to the file-drawer. *Journal of Experimental Psychology: General*, *143*(2), 534–547. https://doi.org/10.1037/a0033242

*Spangler, G., & Grossmann, K. E. (1993). Biobehavioral organization in securely and insecurely attached infants. *Child Development*, *64*(5), 1439–1450. https://doi.org/10.1111/j.1467-8624.1993.tb02962.x

*Spangler, G., & Schieche, M. (1998). Emotional and adrenocortical responses of infants to the Strange Situation: The differential function of emotional expression. *International Journal of Behavioral Development*, *22*(4), 681–706. https://doi.org/10.1080/016502598384126

Stifter, C. A., & Fox, N. A. (1990). Infant reactivity: Physiological correlates of newborn and 5-month temperament. *Developmental Psychology*, *26*(4), 582–588. https://doi.org/10.1037/0012-1649.26.4.582

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (Vol. 4). Boston: Allyn & Bacon.

*Thomas, J. C., Letourneau, N., Bryce, C. I., Campbell, T. S., & Giesbrecht, G. F. (2017). Biological embedding of perinatal social relationships in infant stress reactivity. *Developmental Psychobiology*, *59*(4), 425–435. https://doi.org/10.1002/dev.21505

Tout, K., Haan, M., Campbell, E. K., & Gunnar, M. R. (2008). Social behavior correlates of cortisol activity in child care: Gender differences and time-of-day effects. *Child Development*, *69*(5), 1247–1262. https://doi.org/10.1111/j.1467-8624.1998.tb06209.x

Tronick, E., Als, H., Adamson, L., Wise, S., & Brazelton, T. B. (1978). The infant's response to entrapment between contradictory messages in face-to-face interaction. *Journal of the*

American Academy of Child Psychiatry, *17*(1), 1–13. https://doi.org/10.1016/S0002-7138(09)62273-1

*van Bakel, H. J. A., & Riksen-Walraven, J. M. (2004). Stress reactivity in 15-month-old infants: Links with infant temperament, cognitive competence, and attachment security. *Developmental Psychobiology*, *44*(3), 157–167. https://doi.org/10.1002/dev.20001

*Waters, C. S., van Goozen, S., Phillips, R., Swift, N., Hurst, S.-L., Mundy, L., … Hay, D. F. (2013). Infants at familial risk for depression show a distinct pattern of cortisol response to experimental challenge. *Journal of Affective Disorders*, *150*(3), 955–960. https://doi.org/10.1016/j.jad.2013.04.054

*Weisman, O., Zagoory-Sharon, O., & Feldman, R. (2013). Oxytocin administration alters HPA reactivity in the context of parent–infant interaction. *European Neuropsychopharmacology*, *23*(12), 1724–1731. https://doi.org/10.1016/j.euroneuro.2013.06.006

Wilcox, S., Bopp, M., Wilson, D. K., Fulk, L. J., & Hand, G. A. (2005). Race differences in cardiovascular and cortisol responses to an interpersonal challenge in women who are family caregivers. *Ethnicity & Disease*, *15*(1), 17–24.

Zmyj, N., Schneider, S., & Seehagen, S. (2017). Fifteen-month-old infants' cortisol levels decrease after a 30-min-warm-up procedure. *Psychoneuroendocrinology*, *76*, 11–13. https://doi.org/10.1016/j.psyneuen.2016.11.010

TABLE 1. Summary of study characteristics

Study + year	*N*	Age (months)	Gender (% boys)	Stressor	Stressor type	Duration of acclimation period (min)	Material of the saliva collection instrument	Time of study
Atkinson et al. (2013)	277	16.95	47.0	SSP	Separation	5.0	Synthetic	8–12
Azar et al. (2007)	126	4.38	44.0	ARP	Frustration	Not reported	Not reported	8–12
Beijers et al. (2017)	183	12.39	54.1	SSP	Separation	Not reported	Synthetic	12–18
Bernard & Dozier (2010)	28	15.20	41.0	SSP	Separation	Not reported	Cotton	8–12
Bernard et al. (2017)	155	11.80	Not reported	ARP	Frustration	Not reported	Synthetic	12–18
Bosquet Enlow et al. (2014)	35	6.39	54.0	Modified SFP[a]	SF	45.0	Cotton	12–18
Conradt et al. (2015)	128	4.39	50.0	Modified SFP[a]	SF	Not reported	Synthetic	8–12
DiCorcia et al. (2018)	33	3.68	Not reported	Modified SFP[a]	SF	25.0	Synthetic	8–12
Diego et al. (2004)	23	4.12	47.0	Mock facial expressions	Novelty	3.0	Cotton	Not reported
Dismukes et al. (2018)	207	4.00	52.0	SFP	SF	15.0	Synthetic	Not reported
Eiden et al. (2015)	69	9.00	Not reported	Lab-TAB[b]	Frustration	3.0	Synthetic	Not reported
Eiden et al. (2009)	81	7.00	47.0	Lab-TAB[b]	Frustration	Not reported	Cotton	Not reported
Erickson et al. (2013)	24	7.03	62.5	Modified SFP[a]	SF	Not reported	Cotton	Not reported
Feldman et al. (2010)	33	5.90	47.2	SFP	SF	Not reported	Cotton	8–12
Frigerio et al. (2009)	106	14.60	55.3	SSP	Separation	Not reported	Cotton	8–12
Goldberg et al. (2003)	27	14.10	Not reported	SSP	Separation	Not reported	Cotton	8–12
Grant et al. (2009)	71	7.20	57.7	Modified SFP[c]	SF	Not reported	Synthetic	8–18
Grunau et al. (2004)	22	8.00	59.1	Visual novelty paradigm	Novelty	Not reported	Cotton	8–12
Gunnar et al. (1992)	38	9.10	58.0	30-min separation	Separation	Not reported	Cotton	8–12

TABLE 1. (continued)

Study + year	*N*	Age (months)	Gender (% boys)	Stressor	Stressor type	Duration of acclimation period (min)	Material of the saliva collection instrument	Time of study
Gunnar et al. (1989)	51	13.00	47.0	SSP	Separation	Not reported	Cotton	8–12
Gunnar & Nelson (1994)	49	11.90	56.0	30-min separation	Separation	5.0	Cotton	8–12
Haley (2011)	16	5.57	61.3	Modified SFP[a]	SF	Not reported	Cotton	12–18
Hibel et al. (2018)	107	5.90	49.0	ARP	Frustration	50.0	Synthetic	12–18
Holochwost et al. (2017)	111	6.00	49.7	Lab-TAB[d]	Frustration	Not reported	Synthetic	12–18
Huot et al. (2004)	71	6.00	Not reported	Lab-TAB[e]	Frustration	15.0	Not reported	Not reported
Kuo et al. (2019)	144	12.00	55.0	SSP	Separation	Not reported	Synthetic	8–18
Larson et al. (1991)	27	9.50	44.4	30-min separation	Separation	15.0	Cotton	Not reported
Laurent et al. (2017)	73	6.00	37.0	SFP	SF	Not reported	Synthetic	8–12
Laurent et al. (2016)	98	10.01	49.0	Uncertain toy approach	Novelty	3.0	Synthetic	12–18
Lewis & Ramsay (2005)	84	5.90	58.3	SFP	SF	Not reported	Cotton	8–18
Lewis et al. (2006)	56	4.00	48.2	Contingency learning situation	Frustration	Not reported	Cotton	8–18
Luijk et al. (2010)	310	14.70	56.5	SSP	Separation	Not reported	Cotton	12–18
Martinez-Torteya et al. (2018)	92	11.86	49.5	SSP	Separation	20.0	Synthetic	12–18
Montirosso et al. (2013)[1]	37	4.46	Not reported	Modified SFP[a]	SF	10.0	Cotton	8–18
Montirosso et al. (2013)[2]	37	4.00	Not reported	Modified SFP[a]	SF	10.0	Cotton	8–18
Nachmias et al. (1996)	77	18.00	51.3	SSP	Separation	5.0	Cotton	8–12
Oberlander et al. (2008)	45	3.00	46.7	Information processing task	Novelty	10.0	Cotton	12–18
O'Connor et al. (2013)	108	16.77	49.0	SSP	Separation	Not reported	Synthetic	8–18

TABLE 1. (continued)

Study + year	*N*	Age (months)	Gender (% boys)	Stressor	Stressor type	Duration of acclimation period (min)	Material of the saliva collection instrument	Time of study
Provenzi et al. (2016)	53	3.31	50.9	Modified SFP[a]	SF	Not reported	Cotton	8–12
Reck et al. (2013)	49	4.19	64.4	SFP	SF	Not reported	Cotton	Not reported
Schuetze et al. (2008)	40	7.00	Not reported	Lab-TAB[b]	Frustration	5.0	Cotton	8–18
Seehagen et al. (2015)	12	14.92	46.2	SSP[g]	Separation	40.0	Synthetic	8–12
Spangler & Grossmann (1993)	32	12.10	58.5	SSP	Separation	Not reported	Cotton	8–12
Spangler & Schieche (1998)	106	12.40	Not reported	SSP	Separation	Not reported	Cotton	8–12
Thomas et al. (2017)	272	6.00	52.2	Lab-TAB[f]	Frustration	Not reported	Synthetic	Not reported
van Bakel & Riksen-… (2004)	85	15.10	52.0	SSP[g]	Separation	46.0	Cotton	8–12
Waters et al. (2013)	152	12.80	Not reported	Emotional setting	Novelty	25.0	Synthetic	12–18
Weisman et al. (2013)	35	5.00	48.6	SFP	SF	45.0	Synthetic	12–18

N = sample size; [1] = control group; [2] = experimental group (1. exposure); SSP = Strange Situation Procedure (Ainsworth et al., 1978); ARP = Arm Restraint Procedure (Stifter & Fox, 1990); SFP = traditional 3-episodic Still-Face paradigm (Tronick et al., 1978).

[a] = 5-episodic SFP (Haley & Stansbury, 2003).

[b] = stressor: puppet-show and the ARP, in which the mother prevents the infant's play (Lab-TAB; Goldsmith & Rothbart, 1999).

[c] = 3-episodic SFP, which includes an additional one-minute separation from the mother.

[d] = stressor: 90 s still-face and the ARP, in which the mother prevents the infant's play (Lab-TAB; Goldsmith & Rothbart, 1999).

[e] = stressor: car seat restraint, 3 loud noises and the ARP.

[f] = stressor: toy removal, toy barrier and the ARP, in which the mother prevents the infant's play (Lab-TAB; Goldsmith & Rothbart, 1999).

[g] = modified SSP: interaction with a robot included.

TABLE 2. Categorical moderators

	k	*N*	*g*	*SE*	95% CI (*g*)	T^2	*T*	I^2	95% CI (I^2)	95% PI (*g*)	*Q-W*	*Q-B*
Stressor type												8.13*
Separation	18	1752	.15***	.05	.01–.24	.02	.15	71.21%	51.22–80.96	–.19–.49	59.06***	
Frustration	10	1088	.00	.04	–.08–.07	.01	.08	47.81%	–9.05–71.36	–.21–.21	17.25*	
Novelty	5	340	.12	.08	–.04–.28	.02	.13	57.07%	–15.74–84.08	–.32–.56	9.32	
SF	15	915	.16	.08	–.00–.32	.08	.28	83.62%	74.31–89.56	–.46–.78	85.48***	1.55
3-ep	7	552	.06	.09	–.12–.23	.04	.20	77.14%	52.26–89.05	–.51–.63	26.24***	
5-ep	8	363	.25*	.13	.00–.50	.11	.33	83.61%	69.27–91.26	–.60–1.11	42.72***	
Material												.03
Cotton	27	1571	.12**	.05	.03–.22	.04	.20	71.25%	56.16–79.66	−.30−.55	90.43***	
Synthetic	19	2327	.11*	.04	.03–.20	.03	.16	80.19%	68.34–86.19	−.25−.47	90.86***	
Time of study												.33
Morning	19	1356	.19**	.06	.10–.31	.05	.23	79.99%	68.00–86.06	−.32−.69	89.97***	
Afternoon	12	1413	.14**	.06	.03–.25	.02	.16	74.89%	52.24–84.29	–.23–.51	43.81***	

k = number of effect sizes; *N* = total sample size; *g* = effect size (Hedges' *g*); *SE* = standard error of effect size; 95% CI (*g*) = 95% confidence interval around effect size; T^2 = estimated true variance of effect sizes; *T* = estimated true standard deviation of effect sizes; I^2 = proportion of true variance; 95% CI (I^2) = 95% confidence ("uncertainty") interval for proportion of true variance; 95% PI (*g*) = 95% prediction interval around effect size; *Q-W* = a statistic testing for the homogeneity within a subgroup; *Q-B* = a statistic testing for the significance of the contrast between subgroups; 3-ep = 3-episodic SFP; 5-ep = 5-episodic SFP.

*$p < .05$.

**$p < .01$.

***$p < .001$.

TABLE 3. Meta-regression statistics for the continuous moderators in the full set of studies and separately by stressor type

Age (months)							
	k	*N*	*b*	*SE*	95% CI	*Q*	R^2
Total	48	4095	.00	.01	–.01–.02	.21	.00
Separation	18	1752	.02	.02	–.06–.02	1.03	.00
Frustration	10	1360	.01	.02	–.02–.05	.60	.00
Novelty	5	340	.02	.02	–.03–.06	.55	.00
SF	15	915	–.06	.07	–.19–.07	.79	.00
Duration of acclimation period (min)							
	k	*N*	*b*	*SE*	95% CI	*Q*	R^2
Total	22	1880	.01	.00	.00–.01	6.36*	.29
Separation	7	619	.02	.01	.00–.03	7.53**	.38
Frustration	5	559	.01	.00	–.00–.01	5.72*	N/A[a]
Novelty	4	318	.01	.00	.00–.02	8.62**	N/A[a]
SF	6	384	–.01	.00	–.02–.00	3.68	N/A[a]

k = number of effect sizes; *N* = sample size; *b* = regression coefficient; SE = standard error of coefficient; 95% CI = 95% confidence interval around the coefficient; *Q* = statistic testing that the coefficient is zero; R^2 = proportion of total between-study variance explained by the moderator; [a] = the model's goodness-of-fit test is non-significant and, therefore, the R^2 values are not valid.

*$p < .05$.

**$p < .01$

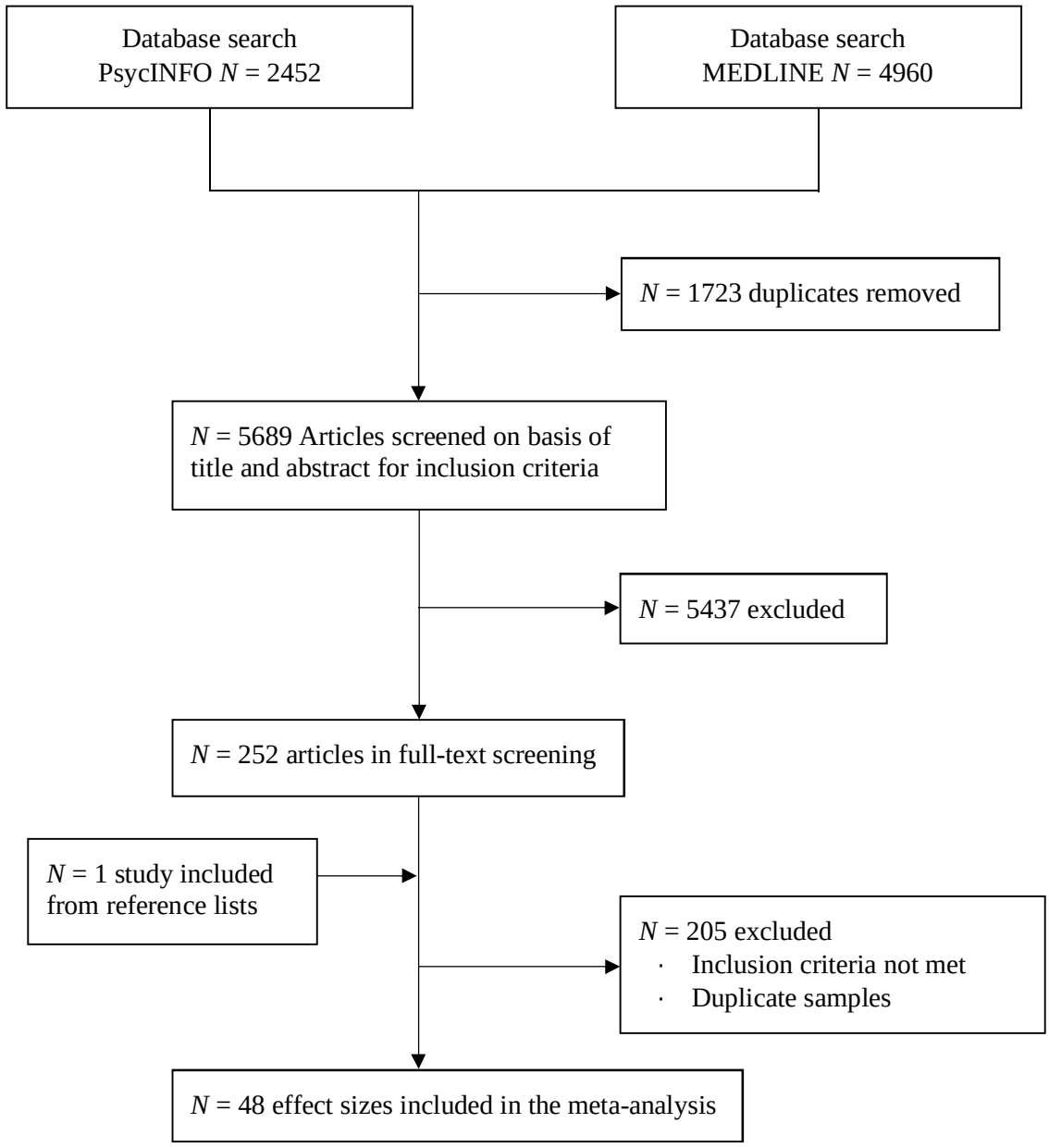
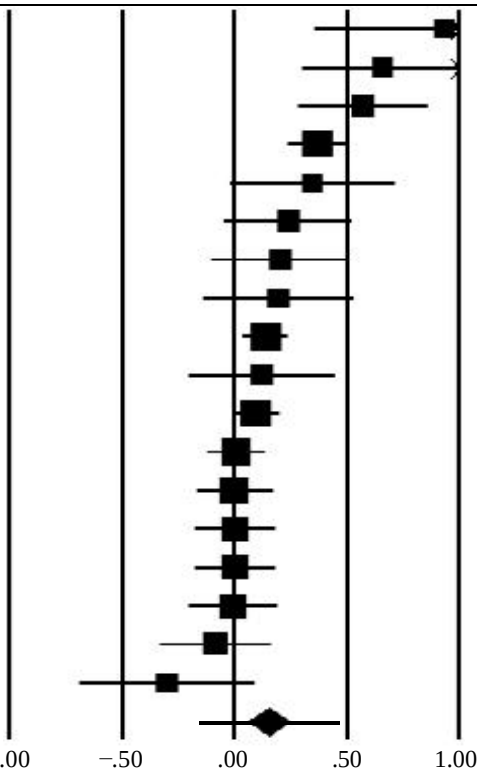


FIGURE 1. Flowchart of the study selection process.

FIGURE 2. Effect sizes and forest plot of studies included in meta-analysis grouped by stressor type

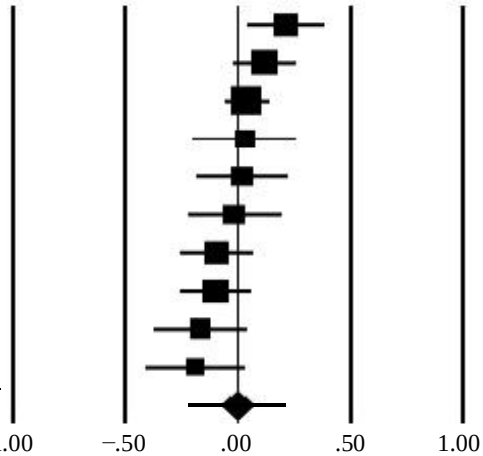
Separation

Study + year	*g*	*p*	*N*	Age
Seehagen et al. (2015)	.933	.002	12	14.92
Larson et al. (1991)	.659	.000	27	9.50
van Bakel & Riksen-... (2004)	.570	.000	85	15.10
Beijers et al. (2017)	.368	.000	183	12.39
Gunnar et al. (1989)	.345	.063	51	13.00
Gunnar et al. (1992)	.239	.098	38	9.10
Spangler & Grossmann (1993)	.204	.192	32	12.10
Goldberg et al. (2003)	.195	.248	27	14.10
Luijk et al. (2010)	.138	.007	310	14.70
Bernard & Dozier (2010)	.119	.470	28	15.20
Atkinson et al. (2013)	.097	.072	277	16.95
Kuo et al. (2019)	.006	.929	144	12.00
O'Connor et al. (2013)	.000	1.000	108	16.77
Spangler & Schieche (1998)	.000	1.000	106	12.40
Martinez-Torteya et al. (2018)	.000	1.000	92	11.86
Nachmias et al. (1996)	−.011	.916	77	18.00
Gunnar & Nelson (1994)	−.089	.480	49	11.90
Frigerio et al. (2009)	−.301	.128	106	14.60
Total	.149	.001	1752	13.59



Frustration

Study + year	*g*	*p*	*N*	Age
Hibel et al. (2018)	.213	.014	107	5.90
Bernard et al. (2017)	.118	.100	155	11.80
Thomas et al. (2017)	.037	.466	272	6.00
Lewis et al. (2006)	.030	.801	56	4.00
Huot et al. (2004)	.017	.871	71	6.00
Eiden et al. (2015)	−.017	.873	69	9.00
Holochwost et al. (2017)	−.095	.260	111	6.00
Azar et al. (2007)	−.100	.208	126	4.38
Eiden et al. (2009)	−.165	.124	81	7.00
Schuetze et al. (2008)	−.190	.099	40	7.00
Total	−.004	.914	1088	6.71

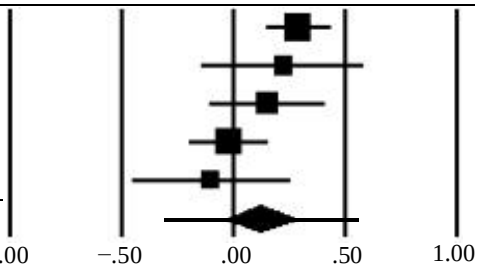


(figure continues on the next page)

FIGURE 2. (continued)

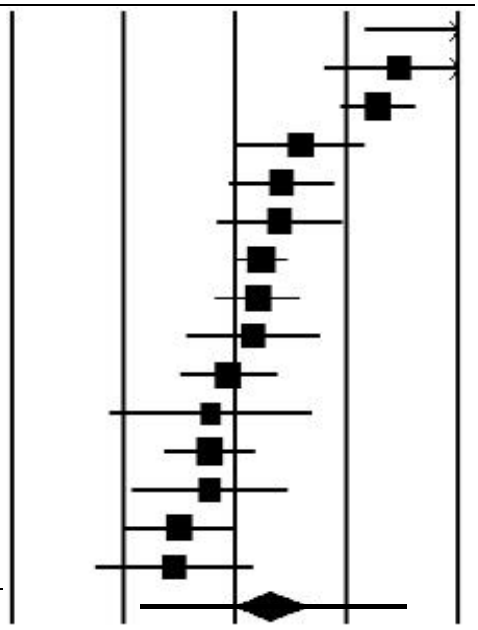
Novelty

Study + year	*g*	*p*	*N*	Age	Hedges' *g*, 95% CI and 95% PI
Waters et al. (2013)	.287	.000	152	12.80	
Grunau et al. (2004)	.220	.236	22	8.00	
Oberlander et al. (2008)	.154	.242	45	3.00	
Laurent et al. (2016)	−.024	.787	98	10.01	
Diego et al. (2004)	−.101	.576	23	4.12	
Total	.121	.141	340	7.59	



Still-Face

Study + year	*g*	*p*	*N*	Age	Hedges' *g*, 95% CI and 95% PI
Haley (2011)	1.124	.000	16	5.57	
Feldman et al. (2010)	.737	.000	33	5.90	
Conradt et al. (2015)	.639	.000	128	4.39	
Montirosso et al. (2013)[2]	.296	.045	37	4.00	
Provenzi et al. (2016)	.209	.088	53	3.31	
Montirosso et al. (2013)[1]	.199	.170	37	4.46	
Dismukes et al. (2018)	.118	.057	207	4.00	
Lewis & Ramsay (2005)	.103	.288	84	5.90	
DiCorcia et al. (2018)	.084	.580	33	3.68	
Laurent et al. (2017)	−.027	.811	73	6.00	
Weisman et al. (2013)	−.108	.643	35	5.00	
Grant et al. (2009)	−.112	.286	71	7.20	
Bosquet Enlow et al. (2014)	−.113	.528	35	6.39	
Reck et al. (2013)	−.252	.049	49	4.19	
Erickson et al. (2013)	−.274	.129	24	7.03	
Total	.157	.051	915	5.13	



Total (combined)	.109	.000	4095	8.89	

g = effect size (Hedges' *g*); 95% CI = 95% confidence interval around the effect size; 95% PI = 95% prediction interval around the effect size; [1] = control group; [2] = experimental group (first exposure). Age is expressed in months. Effect sizes are arranged from largest to smallest from top to bottom. Size of the square represents the weight of the individual study within the stressor category (larger size equals larger weight) and the horizontal line the 95% confidence interval. Width of the diamond shape in the total effect sizes represents a 95% confidence interval and the horizontal line the 95% prediction interval.

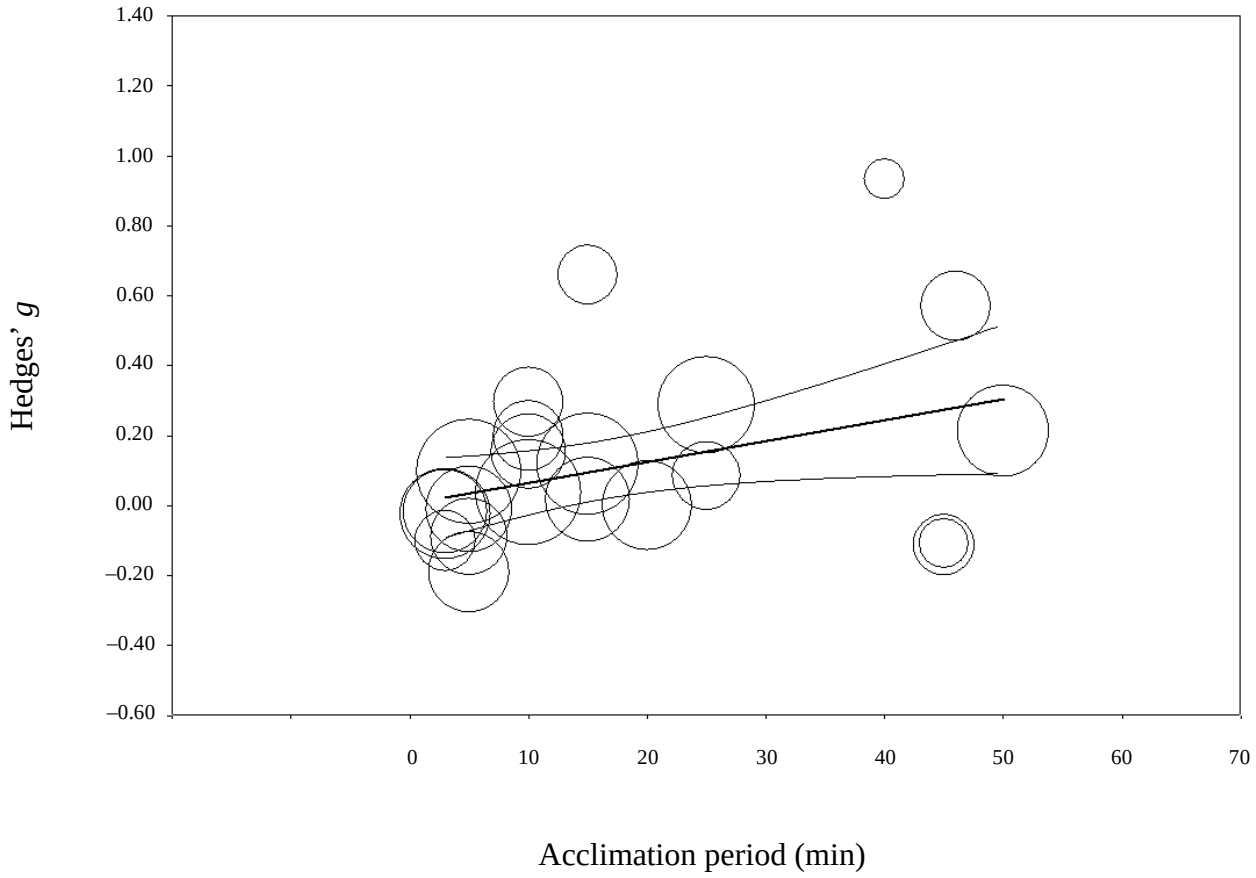


FIGURE 3. Association between the duration of the acclimation period and cortisol reactivity. Each circle represents a study, with a larger radius signifying larger weight (i.e., a smaller standard error). Two lines around the regression line denote the 95% confidence interval for the regression line.

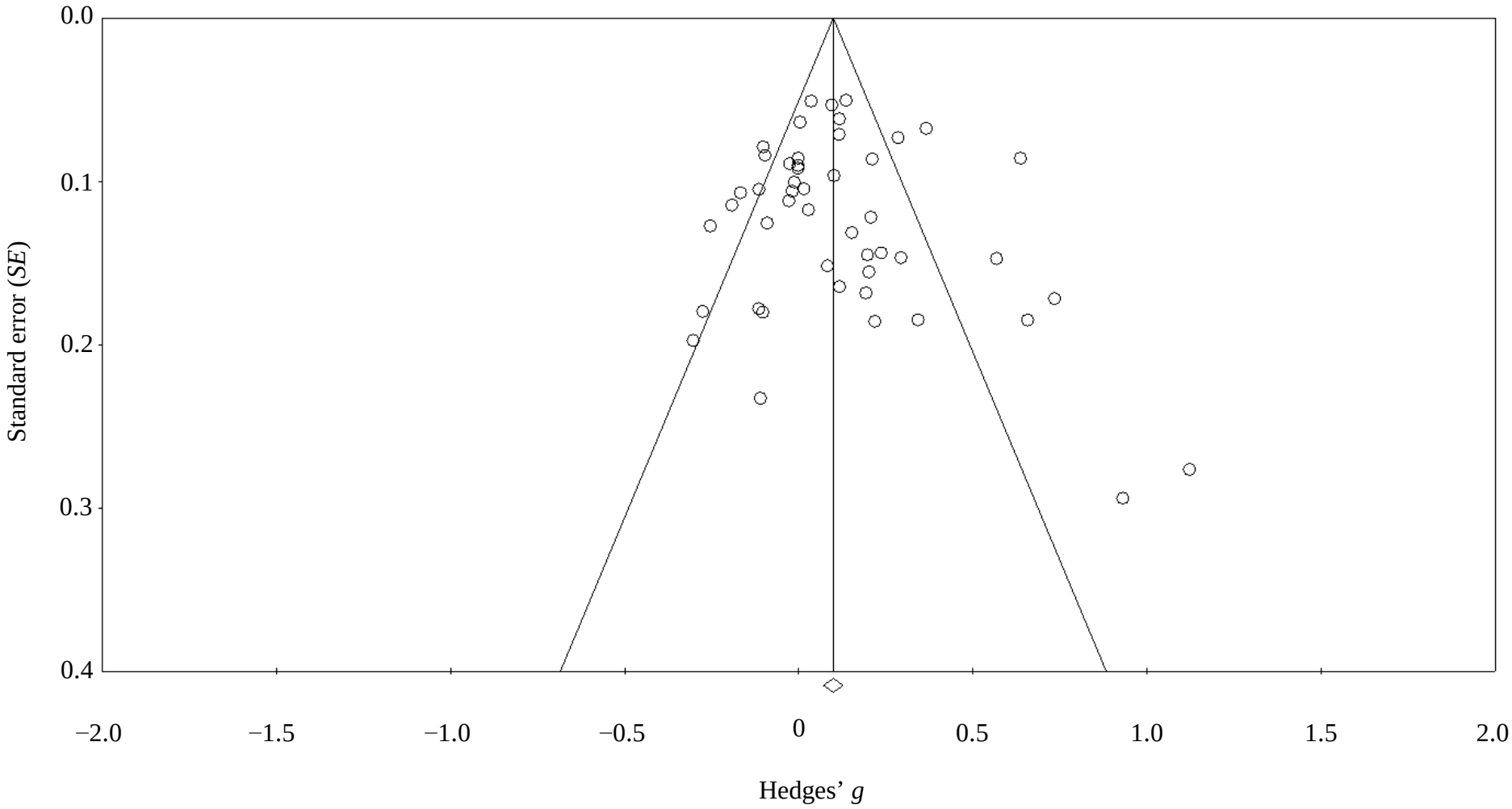


FIGURE 4. Funnel plot of effect sizes (Hedges' *g*) and standard errors (*SE*) of studies included in the main analysis. The diamond shape represents the total mean effect size.

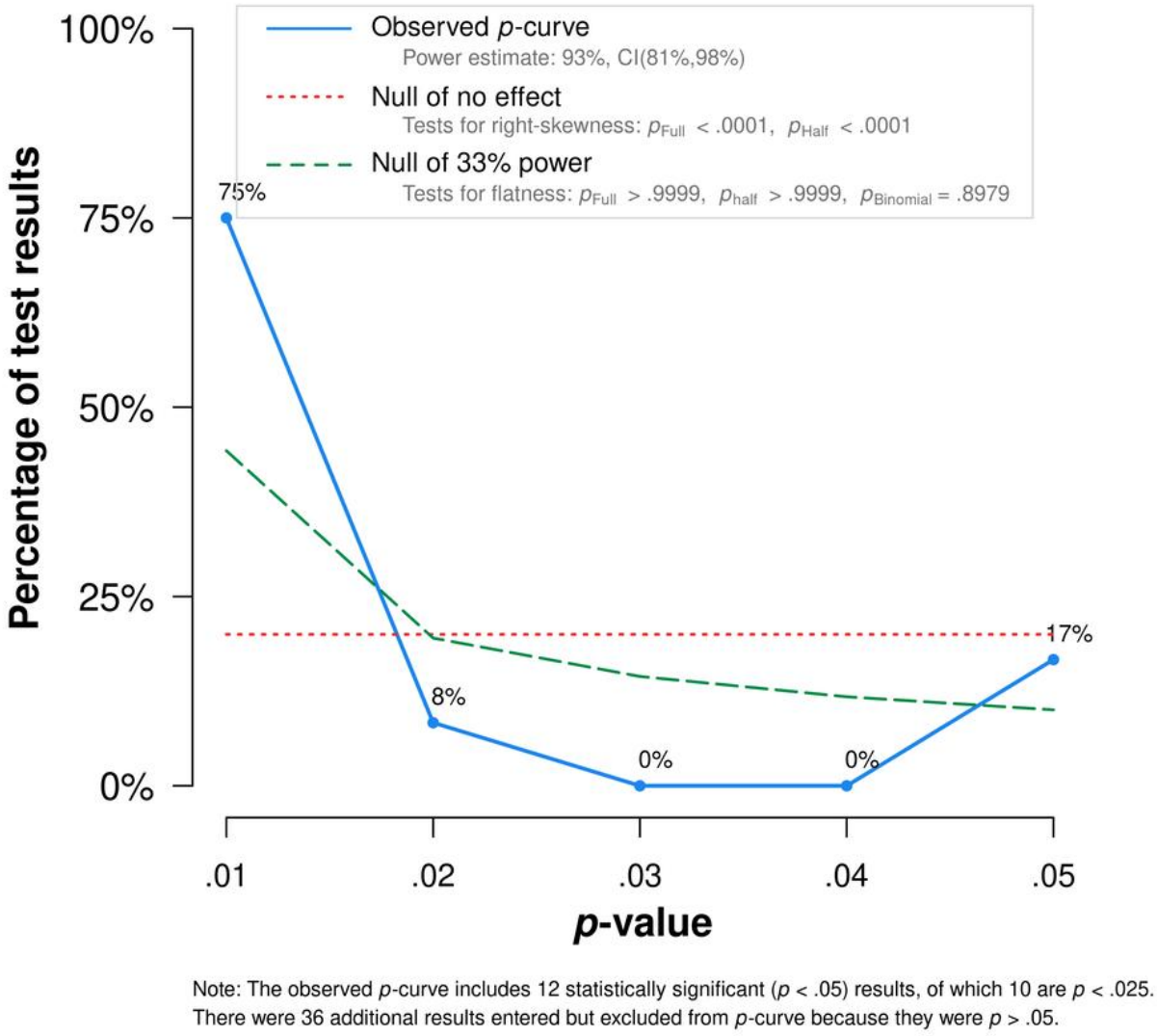


FIGURE 5. *p*-curve obtained from https://www.p-curve.com/ (version 4.06)