Child maltreatment and hypothalamic-pituitary-adrenal axis functioning: A systematic review and meta-analysis

Selina Schär, Ines Mürner-Lavanchy, Stefanie J. Schmidt, Julian Koenig, Michael Kaess

ARTICLE INFO

Keywords:
Child maltreatment
Hypothalamic–pituitary–adrenal axis
Cortisol awakening response
Stress reactivity
Dexamethasone suppression test
Combined dexamethasone-corticotropin releasing hormone test
Corticotropin-releasing hormone test
Hair cortisol
Urinary free cortisol
Meta-analysis
Systematic review

ABSTRACT

Alterations in hypothalamic–pituitary–adrenal (HPA) axis and its effector hormone cortisol have been proposed as one possible mechanism linking child maltreatment experiences to health disparities. In this series of meta-analyses, we aimed to quantify the existing evidence on the effect of child maltreatment on various measures of HPA axis activity. The systematic literature search yielded 1,858 records, of which 87 studies (k = 132) were included. Using random-effects models, we found evidence for blunted cortisol stress reactivity in individuals exposed to child maltreatment. In contrast, no overall differences were found in any of the other HPA axis activity measures (including measures of daily activity, cortisol assessed in the context of pharmacological challenges and cumulative measures of cortisol secretion). The impact of several moderators (e.g., sex, psychopathology, study quality), the role of methodological shortcomings of existing studies, as well as potential directions for future research are discussed.

1. Introduction

Child maltreatment is a widespread phenomenon that affects the lives of millions of children worldwide (Stoltenborgh et al., 2015; Witt et al., 2017). Despite extensive research on the consequences of the experience of child maltreatment, surprising heterogeneity exists across studies with respect to its operational definition (Cicchetti and Toth, 2005; Leeb et al., 2008; Manly, 2005). Researchers, however, generally agree that child maltreatment involves both acts of commission - including physical, sexual, and emotional (psychological) abuse - as well as acts of omission (i.e., any form of neglect) by a parent or other caregiver that results in harm, potential for harm, or threat of harm to a child (usually interpreted as up to 18 years of age; Gilbert et al., 2009; Leeb et al., 2008). Typically, maltreated children experience multi-type maltreatment, suggesting that the different forms of maltreatment often co-occur (Herrenkohl and Herrenkohl, 2009; Vachon et al., 2015). While inconsistencies exist between studies in terms of its definition, extensive research, including findings from prospective and retrospective cohort studies (e.g., Clark et al., 2010; Danese et al., 2009; Dube et al., 2001) as well as twin studies (Kendler et al., 2000; Nelson et al., 2002), has shown that the experience of child maltreatment represents a profound, non-specific risk factor for the development of a broad variety of mental (e.g., Bonoldi et al., 2013; Carr et al., 2013; Dube et al., 2001; Infurna et al., 2016; Liu et al., 2018; Norman et al., 2012; Porter et al., 2020; Teicher and Samson, 2013) as well as physical disorders (e.g., Clemens et al., 2018; Hemmingsson et al., 2014). Importantly, most publications indicate a dose-dependent relationship between the experience of child maltreatment and the risk for health impairments, with those reporting...
more severe experiences or an increasing number of different types of child maltreatment showing stronger associations (Clemens et al., 2018; Dube et al., 2001; Hemmingsson et al., 2014; Kendler et al., 2000; Norman et al., 2012).

One proposed mechanism by which child maltreatment might affect later disease risk involves epigenetic programming - a mechanism known to cause long-lasting changes in gene expression, especially in combination with specific risk alleles, thereby inducing long-lasting changes in biological functioning (i.e., biological embedding; Heim et al., 2019; Jacoby et al., 2016; Smith and Mill, 2011). It is assumed that particularly the expression of several genes relevant to stress regulation might be affected, ultimately leading to the development of a phenotype with core dysfunctions in circuits of the brain involved in the processing of stress and emotion regulation, and related changes in core outflow stress response systems, including - and this will present the focus of the current meta-analysis - the hypothalamic-pituitary-adrenal (HPA) axis (Chen and Baram, 2016; Heim et al., 2019; Jacoby et al., 2016; Koss and Gunnar, 2018; Strüber et al., 2014). In turn, these core dysfunctions might then increase the lifelong risk for the development of a wide range of adverse outcomes later in life (e.g., Heim et al., 2019).

As an important stress response system, HPA axis activity and associated cortisol secretion - i.e., cortisol stress reactivity - serves various functions, depending on the system under investigation (e.g., cardiovascular system, immune system), together being important for survival and restoring homeostasis (adapting to a homeostatic challenge; Nic-olaïdes et al., 2015). Importantly, structures centrally involved in controlling stress-induced HPA axis activity (particularly anticipatory responses to social challenges) include limbic brain regions such as the hippocampus, the amygdala, and the prefrontal cortex (for a detailed review, see Herman et al., 2003), structures also well-known to be affected by chronic stress such as the experience of child maltreatment (e.g., McEwen et al., 2016, 2015; Shirazi et al., 2015; Teicher et al., 2016). Interestingly, an aberrant cortisol stress response has been observed in patients with various mental health problems including borderline personality disorder (BPD), anxiety disorders, depression, and schizophrenia (e.g., Drews et al., 2019; Zorn et al., 2017).

Aside from its central role as a stress response system, HPA axis activity and associated cortisol secretion, in addition to stressors, is also triggered by other regulatory control factors, including circadian signals (for a comprehensive overview, see Spencer and Deak, 2017). Accordingly, in addition to the well-known cortisol stress response, several other measures of HPA axis activity can be distinguished. These include the cortisol awakening response (CAR; Clow et al., 2010; Fries et al., 2009; Spencer et al., 2018), diurnal cortisol (DC; Adam and Kumari, 2009; Segerstrom et al., 2014; Spiga et al., 2014), cortisol assessed following pharmacological challenge tests (i.e., the dexamethasone suppression test (DST), the combined dexamethasone-corticotropin releasing hormone test (Dex-CRH), and the corticotropin-releasing hormone test (CRH); Carroll, 1981; Gold et al., 1986; Watson et al., 2006), as well as cumulative measures of cortisol secretion including 24-hour urinary free cortisol (24-hour UFC; Deutschbein et al., 2011; Moore et al., 1985) and hair cortisol concentrations (HCC; for an overview, see Meyer and Novak, 2012; Stalder and Kirschbaum, 2012). Interestingly, similar to findings related to the cortisol stress response, alterations in these other HPA axis activity measures have been associated with various mental and physical health problems as well, although substantial inconsistencies exist among findings (e.g., Adam et al., 2017; Berger et al., 2016; Chida and Steeple, 2009; Leistner and Menke, 2018; Stalder and Kirschbaum, 2012).

To summarize, a growing number of studies indicate associations between the experience of child maltreatment and epigenetic changes in key genes involved in stress regulation. Interestingly, epigenetic changes have been found in genes important for the regulation of HPA axis activity, such as in the GR gene, the FKBP5 gene and the CRH gene (Hoffmann and Spengler, 2012; Kengel et al., 2013; Palma-Gudiel et al., 2015; Turecki and Meaney, 2016). Together with findings of altered brain morphology, functioning and connectivity, especially in brain regions involved in the regulation of HPA axis activity (e.g., McCrory et al., 2010; Teicher et al., 2016), long-lasting changes in the regulation of this system in individuals exposed to child maltreatment are suggested. If additionally, the manifold biological effects of cortisol are considered (e.g., Sapolsky et al., 2000), a dysregulation in the release of this hormone might increase the susceptibility for a wide variety of negative health outcomes, particularly in combination with other stressful experiences later in life. As discussed above, alterations in several measures of HPA axis activity have indeed been linked to various health conditions including psychopathology. Accordingly, the association between child maltreatment and alterations in cortisol secretion has been widely investigated, including various measures of HPA axis activity. However, findings are inconclusive, with inconsistencies reported (similarly with findings related to psychopathology), specifically regarding the direction of alteration (hyper- versus hyposecretion). To date, three meta-analyses have investigated the association between adverse childhood experiences and aberrant cortisol secretion (Bernard et al., 2017; Bunea et al., 2017; Fogelman and Canli, 2018), whereby two of them focused on early life adversity (ELA) in general and one specifically on child maltreatment. With respect to cortisol stress reactivity, the meta-analysis conducted by Bunea and colleagues (2017) revealed a blunted cortisol stress response to social stressors in those with a history of ELA. In contrast to these findings, the two other meta-analyses failed to show alterations in circadian rhythms, including measures such as the diurnal slope (DSL) as well as the CAR (Bernard et al., 2017; Fogelman and Canli, 2018).

When examining the relationship between child maltreatment and HPA axis activity, careful control of several potential moderators may be of particular importance. As mentioned previously, various mental disorders have been linked to alterations in HPA axis functioning (e.g., Chida and Steptoe, 2009; Stalder et al., 2017; Zorn et al., 2017). The meta-analysis conducted by Zorn et al. (2017), for instance, was able to show that women with major depressive disorder (MDD), women with an anxiety disorder and patients with schizophrenia show a blunted cortisol stress response to psychosocial stressors compared to healthy control participants. The meta-analysis conducted by Drews et al. (2019) similarly found evidence for an overall attenuated cortisol stress response in patients with borderline personality disorder (BPD). Due to the close association between the experience of child maltreatment and psychopathology, these studies, however, cannot rule out the possibility that the observed endocrinological alterations were actually caused by childhood adversity and may have been present even before the development of the respective mental health conditions. Moreover, beyond the epigenetic changes found in genes important for the regulation of HPA axis activity that are likely caused by the experience of chronic stress, such as the experience of child maltreatment, specific polymorphisms of respective genes associated with differences in cortisol secretion independent of child maltreatment have also been found to be related to an increased risk of developing mental disorders (Fan et al., 2021; Ising et al., 2008; Mahon et al., 2013). Accordingly, psychopathology may interfere with or moderate the relationship between child maltreatment and cortisol, and should therefore be considered carefully. In addition, several lines of evidence suggest that women are generally more prone to mental disorders, particularly stress-related disorders, and that sex hormones might account for some of these findings (Li and Graham, 2017). Interestingly, marked sex differences have also been reported with respect to the HPA axis functioning (Foley and Kirschbaum, 2010; Stalder et al., 2016; Stalder and Kirschbaum, 2012; Zänker et al., 2019), indicating that in order to accurately capture the relationship between child maltreatment and cortisol, the influence of sex should be considered as well. Furthermore, there is some support that hormonal activity is elevated at stressor onset (e.g., at the time of child maltreatment) and reduces with passing time (e.g., with increasing age; Miller et al., 2007), implying that the age of the population studied may likewise be important. Finally, ethnic differences in HPA axis functioning.
have been found (Boileau et al., 2019), keeping in mind, of course, that minorities are generally exposed to more adversity (e.g., O’Connor et al., 2020), which in turn may explain this association.

A number of variables related to the measurement of child maltreatment may also be important when investigating the association between child maltreatment and HPA axis functioning. These include the assessment modality employed (e.g., informant versus self-reports), the various approaches to defining the presence of child maltreatment (e.g., records, cut-offs, specifications), the age at maltreatment onset, as well as the chronicity of the maltreatment experiences. For instance, a nationally-representative birth cohort study, the Environmental Risk (E-Risk) Longitudinal Twin Study, demonstrated that retrospective self-report data of child maltreatment were more strongly associated with adult psychopathology than prospective informant reports (Newbury et al., 2018). Chronic maltreatment starting early in life is generally associated with poorer neurocognitive functioning and worse psychopathology (Cowell et al., 2015; Jaffee and Makovich-Fong, 2011; Kaplow and Widom, 2007). In addition, findings suggest that chronic exposure to stress hormones can impact brain structures differentially, depending on the timing and duration of the exposure (Lupien et al., 2009). Thus, effects on HPA axis functioning might vary depending on the age of first child maltreatment experience and/or the chronicity of these experiences.

Furthermore, studies vary widely regarding the assessment of cortisol. Depending on the HPA axis activity measure of interest, some of these variations may account for additional variability in the child maltreatment cortisol relationship. Variables that have been associated with different cortisol findings include, for instance, sample type (blood versus saliva; Spencer and Deak, 2017), slope type, and whether morning samples were collected at awakening for diurnal cortisol (Adam et al., 2017; Ryan et al., 2016), as well as the type of stressor in the context of the cortisol stress reactivity (social-evaluative versus other; e.g., Dickerson and Kemeny, 2004; Zänkert et al., 2019), and differences might emerge depending on how well a task can elicit a cortisol response. In addition, the variation in dosage of dexamethasone in pharmacological stimulation tests (0.5 mg versus 1.0 mg) might account for variability as well (Leistner and Menke, 2018).

Finally, several aspects of methodological quality need to be considered when attempting to quantify the relationship between the experience of child maltreatment and HPA axis functioning. Besides matching the child maltreatment and the control group with respect to age, sex, and psychopathology, and ensuring that no one from the control group was exposed to child maltreatment experiences, these include instructions for sampling and collection, the day and timing of sampling, as well as controlling for specific state covariates such as being sick or experiencing any current stress at the time of testing, with these methodological variables differing to some extent between the various HPA axis activity measures. For a comprehensive overview and corresponding references, see Appendix B Tables B1–B7. Finally, there are certain disease states (e.g., addictions, endocrine diseases), various drugs (e.g., glucocorticoids, psychoactive medications), and sex hormone-dependent variables (e.g., menstrual cycle, oral contraceptive intake, pregnancy) that can strongly influence cortisol levels (e.g., Foley and Kirschbaum, 2010; Granger et al., 2009; Kudielka et al., 2012; Kudielka and Wüst, 1013; Locatelli et al., 2009; Stalder et al., 2016; Zänkert et al., 2019). Considering that participants with experiences of child maltreatment are more likely to suffer from medical conditions, are at increased risk for substance abuse, and more often experience unintended teenage pregnancy (e.g., Hughes et al., 2017), factors related to altered cortisol secretion (e.g. Foley and Kirschbaum, 2010; Stalder et al., 2016; Stalder and Kirschbaum, 2012), careful assessment and matching between groups (i.e., maltreated versus control group) on these variables is of major importance.

Thus, in this comprehensive systematic review and meta-analysis, we aimed to quantify existing evidence on the effect of child maltreatment on cortisol metabolism, including all of the previously mentioned measures of HPA axis activity. In contrast to existing meta-analyses, we were particularly interested in the potential influence of psychopathology in interfering or moderating the effect of child maltreatment on changes in cortisol secretion. Accordingly, psychopathology, and especially the matching of the groups (psychopathology versus no psychopathology) was recorded thoroughly. Furthermore, we were interested in a range of other factors likely to moderate the effect of child maltreatment on cortisol regulation, such as age at the time of study participation, sex, ethnicity, child maltreatment assessment method (informant versus self-report), child maltreatment grouping method (i.e., cut-off scores, child protective services (CPS) records, other), age at the time of child maltreatment onset, chronicity of the child maltreatment experiences, as well as variables related to the assessment of the corresponding HPA axis activity measure. We considered different indices for each HPA axis activity measures, at least including one index of total cortisol production and one index reflecting change in cortisol over time (sensitivity of the system; Puressner et al., 2003). In contrast to the existing meta-analyses, the present investigation sought to determine whether aberrant cortisol secretion patterns following child maltreatment can be observed in both of these largely independent components of HPA axis activity. Finally, a comprehensive quality assessment based on expert guidelines was developed for each activity measure, and several elements of methodological quality and their potential moderating role were investigated.

2. Methods

2.1. Systematic literature search

Articles were identified by searching the electronic databases Pubmed, PsyCINFO, Web of Science (WOS) and the Cumulative Index to Nursing and Allied Health Literature (CINAHL) from their inception to June 2018. The search consisted of titles and abstracts and used the following search string: (“maltreatment” OR “neglect” OR “emotional abuse” OR “sexual abuse” OR “physical abuse” OR “childhood trauma”) AND (“hypothalamic–pituitary–adrenal” OR “HPA” OR “cortisol” OR “adrenocorticotropic hormone” OR “ACTH”). Moreover, the reference lists of the prior meta-analyses on HPA axis functioning and child maltreatment (or ELA) were checked, as were studies proposed by authors who were contacted in the context of data collection.

2.2. Selection criteria

Studies were included if they: (1) involved human participants of all ages; (2) reported on at least one measure of child maltreatment (e.g., self-report [questionnaire, interview], report by an outside source [CPS record, parental report] or self-identification), whereby child maltreatment was defined according to the definition provided by the Centers for Disease Control and Prevention (“any act or series of acts of commission or omission by a parent or other caregiver that results in harm, potential for harm, or threat of harm to a child”; Leeb et al., 2008, p. 11) including the subtypes emotional, physical, sexual abuse, as well as neglect; (3) reported measuring cortisol levels, either as indicator of daily activity (DC, CAR), in response to a stressor (cortisol stress reactivity) or pharmacological challenges (DST, Dex-CRH, CRH), or as a cumulative index (24-hour UFC or HCC). All assessment methods, i.e., saliva, blood (serum, plasma), urine and hair, were eligible. Additionally, several pre-conditions were formulated for the various measures of HPA axis activity: With respect to DC at least two sampling time points, one cortisol assessment in the morning (best with reference to awakening) and one in late afternoon/evening, had to be available (Segerstrom et al., 2014). In case of the CAR, only studies that collected at least two samples, with the first sample anchored to awakening and a second sample between +30 min or +45 min post-awakening, were included (Stalder et al., 2016). With regard to the cortisol stress reactivity, studies needed to collect at least one cortisol baseline measure (before being introduced to a...
stressor) as well as a sample between +20 and +40 min post-stressor onset to capture the peak of the cortisol stress response (Dickerson and Kemeny, 2004). UFC had to be collected over a period of 24 h (Moore et al., 1985) and studies assessing HCC needed to focus on the first 3 cm hair segment (Meyer and Novak, 2012). No restrictions were applied to studies involving pharmacological stimulation tests. Articles were excluded for one or more of the following reasons: (1) evaluated a non-human sample; (2) were not written in English or German; (3) did not contain primary data (e.g., systematic reviews, meta-analysis, book chapters); or (4) were not peer reviewed (e.g., dissertations, mastertheses, conference abstracts). Additionally, studies that included (5) participants with substance abuse (e.g., alcohol, cocaine); (6) in these, conference abstracts). Additionally, studies that included (5) participants with substance abuse (e.g., alcohol, cocaine); (6) individuals who suffered from a medical condition (e.g., endocrine disorder, chronic fatigue, chronic pain); or studies including (7) pregnant women or women in the postpartum period (up to 6 months) were excluded, given the well-known effects of these factors on HPA axis activity (e.g., Stalder et al., 2016; Zinkert et al., 2019). Studies were first screened based on their title and abstract and then further examined in full text in case of suitability.

### 2.3. Data extraction

As mentioned earlier, for each of the HPA axis activity measures (exception: 24-hour UFC and HCC) several outcome indices were defined, including at least one measure for total cortisol production and one measure reflecting changes in cortisol over time. For DC, these included waking (morning) and bedtime levels (for total cortisol) as well as the delta between bedtime and waking cortisol samples (reflecting change over time; DSL). For the CAR, the following assessment time points or indices were extracted: waking cortisol, peak cortisol (expected between +30 and +45 min post awakening), end cortisol (assessed +60 min post awakening), peak reactivity (delta between the peak and the waking cortisol sample) as well as the area under the curve with respect to the ground (AUC \(_g\); reflecting total cortisol production) and the area under the curve with respect to increase (AUC \(_i\); reflecting changes over time; for more details regarding these two formulas, see Pruessner et al., 2003). Similarly, the following assessment time points or indices were extracted for the cortisol stress reactivity, the CRH- and the Dex-CRH-test: baseline cortisol (before being introduced to a stressor; before CRH injection), peak levels, recovery levels (last sample assessed), peak reactivity (defined as delta between peak and baseline levels), AUC \(_g\) and AUC \(_i\). In some of the studies, the timing of the peak differed between the child maltreatment and the control group and the values at the individual peak times were extracted for each group. Whenever a peak occurred in only one of the groups, the value at that peak time was extracted also for the other group. There were also studies in which neither group showed a cortisol response following the perception of a stressor. In this case the values were extracted at the time a response would have been expected (around +30 min post stressor onset). For these few studies, however, no peak reactivity values were extracted or included in this meta-analysis. For the DST, cortisol assessed in the morning prior administration of dexamethasone (which was normally administrated at 11 pm) and cortisol measured the next day, as well as the delta between these two measurement time points were extracted. As we were interested in obtaining all of the defined outcome indices from each study, the authors of all studies containing missing information were contacted with a data request. The data request consisted of an excel file containing all the variables of interest. Since we assumed that the requested additional calculations were relatively time-consuming, one option was to send us the raw data, with which we calculated the desired indices. If the data could not be provided, whenever possible, means and standard deviations were extracted from tables or text. If those data were not available, data were extracted from figures using a web-based digitizer (Bobagti, 2012). Two independent reviewers extracted the data from the corresponding figures and the mean value of both extractions was calculated. If not clearly stated in the text or in the subheading of the figures, we assumed that they represented means and standard errors. Standard deviations were calculated from standard errors or from confidence intervals using the RevMan Calculator provided by the Cochrane group (Drahota and Beller, n.d.). If none of these data sources were available, the study was excluded. In case of multiple publications based on the same cohort, the study with the largest sample size or the one that provided extractable data was included. Whenever possible, non-transformed (raw) cortisol data were extracted. In the case of more than two groups described in the paper, we extracted the data from those two groups that best matched in terms of psychopathology, or if the data of two groups could be combined, weighted means and standard deviations were calculated using the StatsToDo software (https://www.statstodo.com/index.php). If a study included both a clinical and a healthy control group and appropriate measures of child maltreatment were taken in both groups, data were requested or extracted separately for these two subgroups. If the experience of child maltreatment was assessed but grouping was based on other criteria, the authors were asked to (re)group participants based on the presence or absence of child maltreatment or, in case of cut-off scores, in high versus low child maltreatment groups (see Appendix A for details on the respective (re)grouping method of the studies in question). Finally, if a measure did not just assess child maltreatment but other traumatic experiences as well, such as it is the case for the Early Trauma Inventory (ETI; Bremner et al., 2000), authors were requested to group participants including only the subscales which assessed child maltreatment. However, this was not always possible, which is why some of the included studies did not focus on child maltreatment only, but on ELA in general. The few studies to which this applies are marked accordingly. We always asked authors to provide the data including only those without missing cortisol or child maltreatment assessments. Therefore, the data presented in this meta-analysis might not completely correspond to the data displayed in the original studies.

### 2.4. Coding of study characteristics for moderator analyses

The following details were extracted from each study (1) identifying features (i.e., authors, year of publication, journal), (2) participant characteristics (i.e., sample size, age-range, average age, sex ratio, ethnicity [or race: percentage of Caucasians, non-Caucasians], assessment of psychopathology [clinical sample, healthy controls, mixed, not assessed, as well as the percentage of participants meeting criteria for a current mental disorder]), (3) trauma related information (i.e., measurement method [self-report, informant report, mixed], instrument used, grouping method [cut-off scores, record, other], type of child maltreatment [emotional, physical, sexual abuse, neglect], average age of first child maltreatment report and average duration of child maltreatment), (4) cortisol related information (i.e., type of sample [blood, saliva, urine, hair], measurement unit [e.g., nmol/l, μg/dl], time points of sampling, number of samples, reliability of measure [sampling over one day, two-days, more], minutes to peak, duration of stressor, type of stressor [social-evaluative, other], whether a cortisol response was observed [yes, both groups, only in one group, no], dose of the respective stimulant in case of the pharmacological stimulation tests) and (5) data related information (source of data [paper, provided] and whether the data were (re)grouped or not). Some of these variables were viewed as potential moderators that might account for variability in the child maltreatment cortisol relationship (see moderator analyses).

### 2.5. Risk of bias in individual studies

In order to quantify the risk of bias for each individual study and to examine the potential moderating role of several elements of methodological quality, a quality assessment tool was developed. This quality assessment tool covered the following three key domains: (1) variables associated with the selection of participants (including the measurement of child maltreatment and the matching of the two groups with respect...
to age, gender and psychopathology), (2) variables associated with the measurement of HPA axis activity, and related to, the (3) assessment of important confounders. These quality criteria, particularly those related to the assessment of cortisol and associated confounders, were developed based on expert guidelines and differ to some degree between the various HPA axis activity measures (see Appendix B for corresponding references). The risk of bias assessment for each HPA axis activity measure was conducted by two independent reviewers, with disagreements being resolved through discussion. In case of (re)grouped data or missing statistics (e.g., t-test or Fisher’s exact test), corresponding group comparisons were conducted based on available means and standard deviations using QuickCals from GraphPad (https://www.graphpad.com/quickcals/). As the data of some articles were (re)grouped, the information with respect to some quality items was no longer available at the group level. In this case, a conservative approach was followed and the corresponding point was not awarded (marked accordingly in the corresponding tables of Appendix B). In certain cases, the assessment of a quality item was not meaningful, e.g., scoring the matching between the child maltreatment and the control group with respect to oral contraceptive intake in an all-male sample, and in those cases the corresponding items were coded as NA (not applicable). For each of the three quality domains (selection of participants, appropriate assessment of the corresponding HPA axis activity measure, appropriate control for confounders), a score derived from the mean of all associated items (excluding the NA items) multiplied by 100 was calculated. In addition to these domain-specific scores, we also calculated an overall total score. These scores were then used in corresponding meta-regression analyses.

2.6. Statistical analyses

All analyses were run using R and R studio (version 3.6.2 2019–12-12), packages: metafor, dmetar) and were guided by the online book “A Hands-on Guide” from Harrer et al., (2019). Effect sizes for the primary studies were estimated using the Hedge’s g coefficient corrected for small sample sizes (Hedges, 1982). In order to calculate the overall effect, random-effects models for the different HPA axis activity measures and the various outcome indices were performed, applying the Restricted Maximum-Likelihood (REML) method to estimate the variance of the distribution of the true effect sizes (tau2; Veroniki et al., 2016). Between-study heterogeneity was evaluated focusing on the Cochran’s Q statistics (with a p < 0.05 indicating the presence of statistical heterogeneity), the Higgins’s and Thompson’s I² measure (with I²: 25% = low heterogeneity, 50% = moderate heterogeneity, 75% = high heterogeneity) and the prediction interval (Higgins, 2003; Higgins and Thompson, 2002; IntHout et al., 2016). By means of the find.outliers (meta package) and inf.analysis function (Leave-One-Out-method; dmetar package), studies with extreme effect sizes (outlier studies) and studies exerting a high impact on the overall result (potential influential studies) were identified and excluded in the context of corresponding sensitivity analyses (Harrer et al., 2019; Viechtbauer and Cheung, 2010). Additionally, meta-regression and subgroup analyses (mixed/ fixed-effects model) were conducted to examine the influence of several predefined moderator variables. For some studies, cortisol data were available for a lower number of participants than reported in the original paper, with information on the various moderator variables only available for the original sample. Despite this, these original values were included in corresponding moderator analyses. To our best knowledge, we marked this in the tables describing the characteristics of the included studies. In case of substantial between-study heterogeneity (I² > 50%), meta-regression and subgroup analyses were based on the sensitivity model excluding outlier studies. Finally, in order to evaluate the presence of publication bias, funnel plots were visually inspected and the Egger’s test for funnel plot asymmetry was performed (Egger et al., 1997; Peters et al., 2008).

3. Results

3.1. Search results

The literature search yielded a total of N = 1,858 records of which n = 575 duplicates were removed. Screening of reference lists of existing meta-analyses on HPA axis functioning and child maltreatment (or ELA), as well as studies proposed by authors who were contacted in the context of data collection, yielded an additional n = 9 studies. After title and abstract screening, n = 1,025 articles were discarded because they did not meet inclusion criteria. The remaining n = 267 studies were assessed in full-text. Of these, another n = 120 publications were excluded for the following reasons (1) no appropriate HPA axis measure (n = 52), (2) all participants experienced child maltreatment (n = 8), (3) unusual measure of child maltreatment (n = 6), (4) intervention study with no baseline assessment (n = 3), and (5) samples used in multiple studies (n = 51). Additionally, n = 60 articles had to be excluded due to missing relevant statistics, leaving a total of n = 87 independent studies included in this series of meta-analyses (for full process of study selection, see Fig. 1). Of the n = 87 studies, n = 14 studies included two subgroups, one study contained three subgroups, and n = 18 articles collected data on more than one HPA axis activity measure (with DC and CAR most frequently jointly assessed), leaving a total of k = 132 group comparisons. Since some studies collected data on various HPA axis activity measures, it was possible that an effect size (e.g., cortisol measured at awakening) was included in two different random-effects models relating to two different outcome indices (e.g., morning cortisol in the context of DC and awakening cortisol in the context of the CAR). With respect to the various HPA axis activity measures, n = 23 studies reported on DC (k = 26), n = 22 on the CAR (k = 27), n = 35 on cortisol stress reactivity (k = 39), and n = 19 studies assessed cortisol following pharmacological challenges (DS: n = 11 (k = 17); Dex-CRH: n = 8 (k = 10)). Only two studies examined cortisol after the CRH test, which is why these two studies were combined with the data reported for the Dex-CRH test. With respect to the cumulative measures, n = 8 studies reported on HCC (k = 9) and n = 4 studies on 24-hour UFC. Overall, data from n = 41 studies were provided by the respective authors, of which n = 23 data sets (k = 29) were (re)grouped for the purpose of this meta-analysis (see Appendix A). For three studies including large sample sizes (Hibel et al., 2019; Lovallo et al., 2019; Vreeburg et al., 2009) from which we obtained data, the publications that best described the respective samples and not those that appeared in the initial literature search were chosen as references.

3.2. Synthesis of results

3.2.1. Diurnal cortisol

3.2.1.1. Included studies. In total, our systematic search strategy identified n = 40 studies that assessed waking (or morning) and evening cortisol to measure some aspects of the circadian rhythm of cortisol secretion. Of these, n = 23 studies, including k = 26 comparisons involving a total of n = 5,248 participants were retained for quantitative synthesis. Owing to a lack of statistical information, data from the remaining n = 17 studies that were eligible for inclusion could not be considered. The mean age of the total sample was 26.89 (SD = 14.99) years, the majority of studies included predominantly female subjects (with the percentage of females ranging between 33.1 and 100.0%, M = 65.4%, SD = 24.0%; k = 5 comparisons with a purely female sample), and the percentage of Non-Caucasians ranged between 0.0 and 81.7% (M = 30.0%, SD = 30.7%, k = 14 not reporting on ethnicity). Most studies (k = 16) were conducted with adults only, with fewer studies involving children or adolescents (k = 10). With respect to psychopathology, k = 8 comparisons included healthy participants, k = 5 involved participants all fulfilling diagnostic criteria for a mental...
disorder, \(k = 6\) comparisons comprised participants with at least some fulfilling the diagnostic criteria for a mental disorder (range: 17.5–96.0%; with \(k = 2\) not matched in terms of psychopathology), and \(k = 7\) did not report on psychopathology at all. The majority of studies used self-reports to assess the presence of child maltreatment (\(k = 17\)) and \(k = 9\) comparisons relied on informant reports. The assessment of child maltreatment and the grouping of participants into a child maltreatment and a control group varied across the studies. This refers both to the instruments used as well as to the grouping procedure applied (e.g., cut-off scores, specific definitions, presence of records). Three studies (\(k = 4\)) not only focused on child maltreatment but also included participants with other types of ELA (Carrion et al., 2002; Faravelli et al., 2010; Faravelli et al., 2017). In terms of reliability, the fewest studies assessed cortisol over more than two days (\(k = 5\)). In total, we received data from 13 studies (\(k = 16\)), of which the respective authors of six studies (\(k = 8\)) regrouped or grouped their data based on the available assessment of child maltreatment (or, in case of raw data, the regrouping was performed by us). For further details on the characteristics of the included studies, see Table 1.

### 3.2.1.2. Risk of bias assessment

Studies received an average total score of 51.9/100.0 (SD = 12.7, range: 27.6–76.7). With respect to the selection of participants (\(M = 61.0/100.0, SD = 15.5\)), the majority of studies used an established instrument to assess the experience of child maltreatment and matched their participants with respect to age, sex, and psychopathology (assessed with a gold-standard diagnostic tool). However, less than half of the comparisons (\(k < 13\)) assured that all participants in the child maltreatment group were exposed to child maltreatment, while none of the participants in the control group were, and only four studies (\(k = 5\)) used two different sources to establish the presence of child maltreatment. Relating to the appropriate assessment of DC (\(M = 53.5/100.0, SD = 17.4\)), most studies did report on clear sampling instructions (including prohibitions of certain behaviors before sampling as well as clear information about how to collect, where to place, and how to return samples) and provided details on their test protocol as well as on missing data and/or handling of outliers. However, only few studies provided clear instructions regarding the day of sampling (\(k = 9\)), assured that the time of awakening did not differ between the groups (\(k = 7\)), assessed sampling time adherence (\(k = 7\)), rescheduled sampling if participants were sick (\(k = 8\)), and assured that participants were not under any current extraordinary stress (\(k = 2\)). Finally, as shown by the relatively low scores related to the control of confounding variables (\(M = 40.4/100.0, SD = 24.6\)), less than half of the comparisons (\(k < 13\)) excluded participants with a medical condition or participants working night shifts, assessed smoking, menstrual cycle, oral contraceptive, and medication use (especially medications affecting the central nervous system (CNS)) and thus assured that participants did not differ in these respects, and...
Table 1
Summary characteristics of included studies that reported on diurnal cortisol (DC).

| Study                         | Total Sample | CM Sample | Control Sample | CM Assessment | Cortisol Assessment | Data |
|------------------------------|--------------|-----------|----------------|---------------|---------------------|------|
| N, Age M (SD)                | Sex, Ethn.  | Psychopath. | N, Age M (SD) | Sex, Ethn.  | Psychopath. | Instrument, Method | Grouping, Type | Unit | Time Points | Slope, Rel. Indices | Source | Data (re) |
| Bernard et al., 2010         | 435          | 1.08      | 46.4 72.6, no | 340          | 1.00      | 47.9 1.34 (0.55) | informant (record) | CPS | other (record) | salvina | µg/dl awak., bedtime | NA | 2 | morning, evening, provided, group, paper, table |
| Carrion et al., 2002(6)      | 83(5)        | 10.78     | 41.5 36.6, yes*, PTSD | 87(5)        | 10.70     | 41.2 10.90 (1.60) | self-report (quest.) | PTD | Reaction index | salvina | µg/dl pre breakast, pre bed | NA | 3 | evening, all, paper, table |
| Cicchetti et al., 2010       | 553          | 10.02     | 47.6 81.7, NA | 415(9)       | 10.40     | 288 10.08 (1.89) | informant (mixed) | MCSI (CPS), MMPIC | other (record) | NA  | salvina | log 9am, 4 pm | NA | 3 | morning, provided, group, table |
| Faravelli et al., 2017       | 154(10)      | 93(8)     | 46.3 94.4, no, yes, PTSD | 167(10)      | 96.3     | 100.0 10.00 (1.12) | self-report (mixed) | CECA-Q, FPI | other (record) | salvina | µmol/ l 30 min post, awak., 8 pm | other | 1 | all, provided, group, figure |
| Fisher et al., 2007          | 177          | 4.40      | 46.3 11.0, NA | 117(10)      | 46.2     | 47.0 24 (0.56) | self-report (mixed) | CECA-Q, FPI | other (record) | salvina | µmol/ l 30 min post, awak., 8 pm | other | 1 | all, provided, group, figure |
| Fuchs et al., 2017           | 82(8)        | 35.38     | 100.0 NA NA | 37          | 36.47     | 100.0 45 (5.06) | self-report (quest.) | CTQ | cut-offs | NA  | salvina | mg/ml awak., bedtime | other | 2 | all, provided, group, figure |
| Hibbel et al., 2017          | 248(11)      | 4.91      | 49.6 74.6, NA | 165         | 4.93      | 49.7 83 (1.13) | self-report (mixed) | MCSI (CPS), MMPIC | other (record) | EA, SA, N | salvina | log, between 4 and 6 pm | other | 2 | morning, evening, provided, group, figure |
| Keshin et al., 2017          | 24(11)       | 14.97     | 100.0 NA yes*, mixed | 24          | 15.04     | 100.0 12 (1.34) | self-report (mixed) | CTQ | cut-offs | NA  | salvina | mg/ml awak., bedtime | other | 2 | all, provided, group, figure |
| Klauhs et al., 2009          | 22(10)       | 47.58     | 100.0 NA no | 10          | 47.80     | 100.0 12 (1.34) | self-report (mixed) | CTQ | cut-offs | NA  | salvina | mg/ml awak., bedtime | other | 2 | all, provided, group, figure |
| Kuhlin et al., 2015          | 79(12)       | 12.79     | 40.5 22.0*, yes | 60          | 12.91     | 41.2 19 (2.27) | self-report (mixed) | ECI | other (spec.) | NA  | salvina | mg/ml awak., pre bed | other | 2 | all, provided, group, figure |
| Kunsta et al., 2017          | 57(10)       | 24.02     | 61.4 NA NA | 44(10)      | 24.15     | 68.2 13 (0.45) | self-report | no | instr. used | N  | salvina | mg/ml awak., 8 pm | other | 2 | all, provided, group, figure |
| Kuras et al., 2017           | 61(11)       | 33.80     | 49.2 NA no | 29          | 29.92     | 32 NA NA | self-report | CTQ | cut-offs | NA  | salvina | mg/ml awak., 13 h | other | 1 | all, provided, group, figure |
| Lindley et al., 2004         | 16(11)       | 40.00     | 87.5 6.2, yes, PTSD | 9           | 34.56     | 48.9 7 (9.78) | self-report (quest.) | THS | other (spec.) | PA, SA | salvina | mg/ml 8am, 10 pm | other | 1 | all, provided, group, figure |
| Lopes et al., 2019           | 38(11)       | 40.29     | 100.0 NA yes, MDD | 16          | 41.06     | 100.0 22 (6.60) | self-report (mixed) | CTQ | other (not spec.) | EA, SA, N | salvina | mg/ml 8am, 8 pm | NA  | 1 | morning, all, provided, group, figure |
| Lovallo et al., 2020         | 69(11)       | 23.07     | 55.9 20.3, no | 335         | 23.90     | 63.0 24 (3.02) | self-report (mixed) | C-DIS-IV | other (record) | EA, SA, N | salvina | mg/ml awak., bedtime | other | 1 | all, provided, group, figure |
| Puetz et al., 2017           | 40(11)       | 10.55     | 47.5 35.0, yes, mixed | 17          | 10.88     | 29.0 23 (1.74) | self-report (mixed) | SSR | other (record) | EA, PA, N | salvina | mg/ml 30 min post, awak., 30 min pre sleep | other | 2 | morning, evening, provided, group, table |
| Reichl et al., 2016a         | 25(11)       | 16.25     | 96.0 NA no | 3           | 16.55     | 22 16 (1.12) | self-report (mixed) | CECA | cut-offs | EA, PA, N | salvina | mg/ml awak., pre bed | other | 3 | all, provided, group, paper, figure |
| Reichl et al., 2016b         | 25(11)       | 16.30     | 92.0 NA yes, mixed | 17          | 16.00     | 94.1 18 (1.29) | self-report (mixed) | CECA | cut-offs | EA, PA, N | salvina | mg/ml awak., pre bed | other | 3 | all, provided, group, paper, figure |
| Schreuder et al., 2016        | 114(12)      | 26.75     | 45.6 NA yes, mixed | 16          | 27.81     | 43.8 98 (2.90) | self-report (mixed) | CTQ | cut-offs | EA, PA, N | salvina | mg/ml awak., 10 pm | other | 1 | all, provided, group, figure |
| Smeets et al., 2007          | 13(11)       | 60.50     | 100.0 NA no | 13          | 38.77     | 100.0 9 (11.38) | self-report (mixed) | CTQ | cut-offs | SA  | salvina | mg/ml awak., 8 pm | other | 2 | all, provided, group, figure |
| Studer et al., 2013          | 38(12)       | 37.87     | 90.0 0.0 no | 4           | 42.50     | 100.0 26 (12.04) | self-report (mixed) | CTQ | cut-offs | EA, PA, SA, N | salvina | mg/ml awak., bedtime | other | 1 | provided, group, paper, figure |

(continued on next page)
Table 1

| Study                      | Total Sample | CM Sample | Control Sample | CM Assessment | Cortisol Assessment | Data |
|----------------------------|--------------|-----------|----------------|---------------|---------------------|------|
|                            | N  | Age M (SD) | Sex | Ethnic | Psycho-path. | n  | Age M (SD) | Sex | Ethnic | Psycho-path. | Method | Instrument | Grouping | Type | Type of sample | Unit | Time points | Slope | Rel. Indices | Source | Data (re) grouped |
| van der Vegt et al., 2009  | 529 | 30.83 (1.22) | 33.1 NA | NA     | NA         | 235  | 30.87 (1.24) | 51.9 | NA     | NA         | informant (interview) | self-dev. | other (spec.) | N     | saliva | nmol/ awk., bedtime | NA | 1 morning, evening | puser | no | paper, no table provided, yes |
| Vreeburg et al., 2009<sup>a</sup> | 1131 | 42.67 (12.27) | 66.3 NA | yes, mixed<sup>b</sup> | 645 | 43.92 (11.68) | 70.9 | NA     | NA         | self-report (interview) | NEMESIS CTI | other (spec.) | EA, PA, SA, N | saliva | nmol/ awk., 11 pm wake-to-bed | 1 | all | comparison between PTSD patients with childhood sexual or physical abuse and those reporting no history of childhood sexual or physical abuse. Comparison between MDD patients with early life stress and those without corresponding experiences. This article did not appear in the initial search but was suggested by the respective author as best suited for citation in this meta-analysis. The two groups early physical/sexual abuse and maltreated without early abuse were combined into one group. The two groups CPS-involved, stayed with birth parents and CPS-involved, placed in foster care were combined into one group. This article did not appear in the initial search but was suggested by the respective author as best suited for citation in this meta-analysis. The two groups some neglect and severe neglect were combined into one group. The data are from the Netherlands Study of Depression and Anxiety, a large cohort study; this article did not appear in the initial search but was suggested by the respective author as best suited for citation in this meta-analysis. Patient group: dysthymia, MDD, social phobia, panic with/without agoraphobia, generalized anxiety disorder. No one of the participants fulfilled the diagnostic criteria for dysthymia, MDD, social phobia, panic with/without agoraphobia, or generalized anxiety disorder in the past 6 months. It is not clearly stated in the text or in the subheading of the figure whether means and standard deviations or means and standard errors were presented; we assumed standard errors. |
| Vreeburg et al., 2009<sup>b</sup> | 498 | 42.88 (14.46) | 61.3 NA | no<sup>c</sup> | 120 | 47.62 (12.64) | 73.3 | 369 | 41.33 (14.69) | self-report (interview) | NEMESIS CTI | other (spec.) | EA, PA, SA, N | saliva | nmol/ awk., 11 pm wake-to-bed | 1 | all | The data were regrouped including only the traumatized control subjects and the non-traumatized control subjects (remark: some in the control sample may have experienced other traumatic events). The two groups some neglect and severe neglect were combined into one group. The data are from the Netherlands Study of Depression and Anxiety, a large cohort study; this article did not appear in the initial search but was suggested by the respective author as best suited for citation in this meta-analysis. Patient group: dysthymia, MDD, social phobia, panic with/without agoraphobia, generalized anxiety disorder. No one of the participants fulfilled the diagnostic criteria for dysthymia, MDD, social phobia, panic with/without agoraphobia, or generalized anxiety disorder in the past 6 months. It is not clearly stated in the text or in the subheading of the figure whether means and standard deviations or means and standard errors were presented; we assumed standard errors. |

Note. N = sample size; M = mean; SD = standard deviation; The sex ratio is indicated as percentage of female participants; Ethnic. = ethnicity; The ethnicity ratio is indicated as percentage of non-Caucasians; NA = not assessed; Psychopath. = psychopathology, whereby the following definitions have been used: yes = at least some of the participants met diagnostic criteria for a psychiatric disorder, no = none of the participants met diagnostic criteria for any psychiatric disorder, * = groups are not matched with respect to psychopathology, ◊ = no gold-standard diagnostic tool was named; PTSD = Posttraumatic Stress Disorder; psych. = psychosis; MDD = Major Depressive Disorder; CM = child maltreatment; n = sample size; quest. = questionnaire; CPS = Child Protective Services; MCS = Maltreatment Classification System; MMCI = Maternal Maltreatment Classification Interview; CECA-Q = Childhood Experiences of Care and Abuse Questionnaire; self-dev. = self-developed (the authors used a scale developed by themselves); FPI = Florence Psychiatric Interview; CWA = Child Welfare Agency; CTQ = Childhood Trauma Questionnaire; ETI = Early Trauma Inventory; no instr. used = no instrument used; THS = Trauma History Screen; C-DIS-IV = Computerized version of the Diagnostic Interview Schedule-IV; SSR = Social Services Record; NEMESIS CTI = Netherlands Mental Health Survey and Incidence Study Childhood Trauma Interview; spec. = specification (authors applied a specific definition of CM); N = neglect; EA = emotional abuse; PA = physical abuse; SA = sexual abuse; ◊◊ = The grouping of participants was not just based on CM experiences but also included other traumatic experiences; awk. = awakening; Rel. = reliability, whereby the following definitions have been used: 1 = cortisol assessed over only one day, 2 = cortisol assessed over two days, 3 = cortisol assessed over more than two days; all = morning, evening, delta (evening minus morning value). The two groups CPS-involved, stayed with birth parents and CPS-involved, placed in foster care were combined into one group. This article did not appear in the initial search but was suggested by the respective author as best suited for citation in this meta-analysis. The two groups early physical/sexual abuse and maltreated without early abuse were combined into one group. Comparison between patient groups with and without early trauma. The two groups Multidimensional Treatment Foster Care for Preschoolers (MTFC) and Regular Foster Care (RFC) were combined into one group. This article did not appear in the initial search but was suggested by the respective author as best suited for citation in this meta-analysis. Authors also administered ETI, but grouping was based on CTQ. Percentage of non-Caucasians refers to total sample (N = 127). Deprived adoptees were combined into one group and were compared to the non-deprived UK adoptees. Comparison between PTSD patients with childhood sexual or physical abuse and those reporting no history of childhood sexual or physical abuse. Comparison between MDD patients with early life stress and those without corresponding experiences. This article did not appear in the initial search but was suggested by the respective author as best suited for citation in this meta-analysis. Exclusion of repressed memory group; CM sample: recovered and continuous memory group, control sample: control group. The data are from the Netherlands Study of Depression and Anxiety, a large cohort study; this article did not appear in the initial search but was suggested by the respective author as best suited for citation in this meta-analysis. Patient group: dysthymia, MDD, social phobia, panic with/without agoraphobia, generalized anxiety disorder. No one of the participants fulfilled the diagnostic criteria for dysthymia, MDD, social phobia, panic with/without agoraphobia, or generalized anxiety disorder in the past 6 months. It is not clearly stated in the text or in the subheading of the figure whether means and standard deviations or means and standard errors were presented; we assumed standard errors.
only \( k = 2 \) comparisons assured that participants did not differ with respect to other ELA or adult adversity. It should be noted, however, that several studies would have assessed some of the variables of interest, but since the data of six studies (\( k = 8 \)) were (re)grouped, the corresponding information at the group level was no longer available for all of these studies. For details on individual scoring results of the primary studies as well as a summary of the average risk of bias scores, see Appendix B Table B8 or Table B1 for individual quality items.

3.2.1.3. Meta-analysis. The results of the meta-analyses for the three indices of circadian activity (morning, evening, DSL) revealed no overall differences for morning (Hedges' \( g = -0.02 \), 95% CI [−0.11; 0.06], \( p = 0.586 \)) and DSL cortisol (Hedges' \( g = 0.00 \), 95% CI [−0.11; 0.11], \( p = 0.987 \)) between the child maltreatment and the control sample (this also held true for the corresponding sensitivity analyses; for further details, see Table 2). In contrast, participants in the child maltreatment group had slightly elevated evening cortisol levels compared to their respective control group (Hedges' \( g = 0.10 \), 95% CI [0.03; 0.18], \( p = 0.008 \)) for corresponding forest plots, see Appendix C Figs. 1.1–1.3. Between-study heterogeneity was in the low range for all outcome indices (\( I^2 < 30\% \), Q-statistics all \( p > 0.05 \) and visual inspection of traditional and counter-enhanced funnel plots as well as Egger’s regression test of funnel plot asymmetry implied absence of small-study bias (all \( p > 0.05 \); for funnel plots, see Appendix C Figs. 1.1–1.3).

3.2.1.4. Meta-regression and subgroup analyses. We conducted a number of pre-defined meta-regression and subgroup analyses. The summary results for each outcome index and each moderator examined are shown in Appendix D Table D1. In the following section, results for moderators found to significantly influence the main effects are outlined. Despite low heterogeneity in the effect size estimates between studies reporting on morning cortisol (\( I^2 = 29.1\% \)), the following two continuous moderators influenced the main effect: (1) age at the time of study participation and (2) the sub-domain “appropriate measure of cortisol in the context of DC” of the quality assessment. With respect to the mean age of study participants, studies including older-aged samples reported a tendency for higher morning cortisol (\( \beta = 0.006 \), 95% CI [0.001; 0.010], \( p = 0.014 \), \( R^2 = 73.04\% \); when comparing the child maltreatment group to the control group) compared to younger samples. Concerning the assessment of cortisol, studies with higher quality scores were associated with lower morning cortisol (\( \beta = -0.005 \), 95% CI [−0.010; −0.001], \( p = 0.012 \), \( R^2 = 85.49\% \) in the child maltreatment group compared to the control group. In addition, forming subgroups of studies using informant reports and those relying on self-report data to assess the presence of child maltreatment revealed overall reduced morning cortisol in those studies applying informant reports (Hedge's \( g = -0.114 \), CI [−0.204; −0.024]), whereas a tendency for increased morning cortisol in the child maltreatment compared to the control group was observed in studies relying on self-report information (Hedge's \( g = 0.060 \), 95% CI [−0.026; 0.146]; \( Q_3 = 7.48 \), \( p = 0.006 \)). Since the majority of studies relying on informant reports used the presence of records to group participants, the corresponding subgroup comparison of the different grouping methods applied (records, cut-off scores, other grouping approaches (mainly specifications)) also reached significance (\( Q_2 = 8.32 \), \( p = 0.016 \)). Finally, the subgroup comparison between studies where original data were extracted and those that (re)grouped their data for this meta-analysis also explained some of the between-study heterogeneity, with studies where original data could be extracted implying overall reduced morning cortisol (Hedge's \( g = -0.096 \), 95% CI [−0.177; −0.016]) and those with (re)grouped data pointing to slightly increased morning cortisol levels in the child maltreatment group (Hedge's \( g = 0.087 \), 95% CI [−0.014; 0.189]; \( Q_3 = 7.72 \), \( p = 0.005 \)). However, since the majority of studies (\( k = 8 \)) that relied on informant reports to group participants also belonged to the original data subgroup, interpretation of these findings should be done with caution. With respect to evening cortisol (\( I^2 = 2.0\% \)), studies focusing on other types of ELA showed larger positive effect size estimates than studies focusing on child maltreatment only (\( Q_2 = 12.24 \), \( p < 0.001 \)). Further, studies including original data showed larger positive effect size estimates than studies which provided (re)grouped data (\( Q_2 = 6.47 \), \( p = 0.011 \). It should be noted that by excluding those three studies (\( k = 4 \)) focusing not only on child maltreatment experiences (but also including participants with loss experiences), the initial model on evening cortisol became insignificant (\( p = 0.098 \)). Finally, with respect to DSL cortisol, no moderator was identified that significantly influenced between-study heterogeneity.

3.2.2. Cortisol awakening response

3.2.2.1. Included studies. A total of \( n = 22 \) studies, comprising \( k = 27 \) comparisons, with an overall sample size of \( n = 3,545 \) participants assessed cortisol in response to awakening. The mean age of the total sample was 27.36 (SD = 10.45) years (\( k = 8 \) comparison involved children/adolescents and \( k = 19 \) including adults only), and the majority of studies included mainly female subjects (with the percentage of females ranging between 0.0 and 100.0%, \( M = 65.4\% \), SD = 30.0%; \( k = 7 \) comprised a purely female sample). With respect to ethnicity, most studies included samples composed predominantly of an ethnic majority group with percentages of Non-Caucasians ranging between 0.0 and 100.0% (\( M = 45.7\% \), SD = 42.2%; with \( k = 20 \) not reporting on ethnicity). Concerning psychopathology, \( k = 9 \) comparisons included healthy participants, \( k = 6 \) involved participants all fulfilling diagnostic criteria for a mental disorder, \( k = 6 \) comparisons comprised participants with at least some fulfilling the diagnostic criteria for a mental disorder (with \( k = 4 \) not matched in terms of psychopathology), and \( k = 6 \) did not report on psychopathology at all. The majority of studies (\( n = 18 \), \( k = 23 \)) employed self-reports to assess the presence of child maltreatment and only \( k = 4 \) relied on informant reports. Along with the use of different instruments, the grouping of participants into a child maltreatment and a control group varied, however, with the majority of studies using specific cut-off scores (\( n = 13 \), \( k = 17 \)). Only one of the included studies (Klaus et al., 2018) not only focused on child maltreatment but also included participants with other types of ELA, including death of a close friend or relative, parental separation or divorce, major illnesses or injuries or other traumatic experiences. Three studies (\( k = 4 \)) assessed the CAR over more than two days, and there were several studies (\( n = 8 \), \( k = 9 \)) with cortisol sampled at only two time points (i.e., awakening and +30 min post awakening or +45 min post awakening). In \( n = 4 \) studies peak cortisol values were not observed at the same assessment time points for both groups. Finally, data from 12 studies (\( k = 15 \)) were provided by the respective authors, of which \( k = 10 \) comparisons contained (re)grouped data. For further details on the characteristics of the included studies, see Table 3.

3.2.2.2. Risk of bias assessment. Studies which assessed cortisol in response to awakening received an average total score of 52.8/100.0 (SD = 11.7, range: 35.7–76.7). With respect to the selection of participants (\( M = 65.7/100.0 \), SD = 14.9), the majority of comparisons (\( k > 13 \)) ensured that all participants in the child maltreatment group were exposed to maltreatment, while none of the participants in the control group was, used an established instrument to assess the experience of child maltreatment and matched their participants with respect to age, sex, and psychopathology (assessed with a gold-standard diagnostic tool; \( k = 12 \) in case of self-reports). However, only one study used two different sources to establish the presence of child maltreatment. Concerning the appropriate assessment of cortisol in the context of the CAR (\( M = 56.8/100.0 \), SD = 12.0), most studies reported on clear sampling (\( k = 22 \)) and collection (\( k = 27 \)) instructions, provided information on the day of sampling (\( k = 14 \)), collected at least three samples (with one sample between +30 and +45 min post awakening, \( k = 18 \)) over at least
two days \((k = 14)\), provided information about how samples were collected, stored or analyzed \((k = 27)\), and reported on outliers or missing data \((k = 22)\). However, less than half of the comparisons \((k < 14)\) assessed the time of awakening (thus ensuring that the two groups did not differ in this respect; \(k = 13)\), assessed sampling time adherence \((k = 8)\), reported whether sampling was rescheduled if participants were sick \((k = 7)\), reported on batch analyses \((k = 9)\), and only \(k = 3\) comparisons ensured that participants were not under any current extraordinary stress or whether sampling was rescheduled if participants experienced any stressor during the day of collection. Finally, many of the studies failed to control for several important confounding variables \((M = 38.0/100.0, SD = 22.5)\). For instance, less than half of the comparisons \((k < 14)\) reported whether participants were excluded if pregnant or working night shifts, assessed smoking, menstrual cycle, oral contraceptive, and medication use (especially medications affecting the CNS), thus ensuring that participants did not differ in these respects, and only \(k = 2\) comparisons ensured that participants did not differ with respect to other ELA or adult adversity. Again, several studies would have assessed some of the variables of interest, but since the data of eight studies \((k = 10)\) were \((re)\)grouped, the corresponding information at the group level was no longer available for some of these studies. For details on individual scoring results of the primary studies as well as a summary of the average risk of bias scores, see Appendix B Table B9 or Table B2 for individual quality items.

### 3.2.2.3. Meta-analysis

The pooled effect estimates for the different CAR indices (including corresponding sensitivity analyses) are displayed in Table 4. As shown in the corresponding lines, none of the examined indices suggested a difference \((p > 0.05)\) in cortisol assessed in response to awakening when comparing the child maltreatment and the control group. For corresponding forest plots, see Appendix C Figs. 2.1–2.6. Between-study heterogeneity was in the moderate to high range for some of the outcome indices \((I^2 = 41.8\%–73.1\%)\), exceeding the level of significance \((Q\text{-statistic and } p < 0.05)\) for peak, delta, AUC$_g$, and AUC$_c$ cortisol, suggesting that other variables differing between the included studies might be of importance as well. Visual inspection of traditional and counter-enhanced funnel plots as well as Egger’s regression test of funnel plot asymmetry implied absence of small-study bias for all outcome indices examined \((p > 0.05)\); for funnel plots, see Appendix C Figs. 2.1–2.6.

### 3.2.2.4. Meta-regression and subgroup analyses

The summary results for the pre-defined meta-regression and subgroup analyses for each outcome index and each moderator examined are shown in Appendix D Table D2. In the following, the results for moderators found to significantly influence the main effects are outlined. The subgroup comparison between studies where original data were extracted and those that \((re)\)grouped their data for this meta-analysis explained some of the between-study heterogeneity for awakening cortisol and AUC$_c$ cortisol, with studies where original data could be extracted demonstrating overall reduced morning cortisol (Hedge’s \(g = –0.169\), 95% CI \([-0.323; –0.015]\)) and Hedge’s \(g = –0.611\), 95% CI \([-0.868; –0.355]\), respectively) and those with \((re)\)grouped data pointing to slightly increased morning cortisol levels in the child maltreatment group (Hedge’s \(g = 0.054\), 95% CI \([-0.036; 0.145]\) and Hedge’s \(g = 0.069\), 95% CI \([-0.025; 0.164]\), respectively); \(Q_i = 6.02, p = 0.014\) and \(Q_t = 23.80, p < 0.001\), respectively). Since, as noted before, there is a relatively large overlap between studies reporting on morning cortisol assessed in the context of DC as well as on cortisol assessed in response to awakening \((k = 14)\), this finding was to be expected. Furthermore, with respect to awakening cortisol, age seems to explain some of the between-study variance \((R^2 = 70.27\%)\), but in contrast to morning cortisol (DC), does not represent a significant moderator \((p = 0.338)\). For delta cortisol, we identified the proportion of women in the sample as a significant continuous moderator (for peak cortisol: \(p = 0.071\), for –60 min post awakening cortisol: \(p = 0.096\), and for AUC$_g$ cortisol: \(p = 0.058\), with an increase in the proportion of females being associated with lower cortisol when comparing the child maltreatment and the control sample \((\beta = –0.005, 95\% \text{ CI } [–0.010; –0.000], p = 0.040, R^2 = 0.04\%). Finally, with respect to AUC$_c$ cortisol the sub-domain “appropriate measure of confounders” of the quality assessment explained some of the variance in the effect estimates, with studies with higher quality scores being associated with lower AUC$_c$ cortisol \((\beta = –0.012, 95\% \text{ CI } [–0.019; –0.005], p < 0.001, R^2 = 95.17\%)\) in the child maltreatment group compared to the control group (for awakening cortisol: \(p = 0.086, R^2 = 64.43\%\)).
### Table 3
Summary characteristics of included studies that reported on cortisol assessed in the context of the cortisol awakening response (CAR).

| Study | Total Sample | CM Sample | Control Sample | CM Assessment | Cortisol Assessment | Data |
|-------|-------------|-----------|----------------|---------------|-------------------|------|
| N | Age M (SD) | Sex Ethn. Psycho-path. | n | Age M (SD) | Sex | n | Age M (SD) | Sex | Method Instrument Grouping Type | Type of sample | Unit | Time pointa | Time point peak | Rel. Indices | Source | Data (re) grouped |
| Bicanic et al., 2013 | 89 | 15.89 | 100.0 NA yes*, PTSD | 52 | 16.10 | 100.0 0.37 | 15.60 | 1.65 | 100.0 self-report (interview) | self-dev. | other (spec.) | SA | saliva | nmol/awak., 15, 30, 45, 60 | NA | 1 | AUCp | paper, table | grouped |
| Fuchs et al., 2017 | 82 | 35.38 | 100.0 NA NA | 37 | 36.47 | 100.0 0.45 | 34.48 | 5.06 | 100.0 self-report (quest.) | CTQ | cut-offs | PA, SA | saliva | nmol/awak., 30 | 30 min | 2 | awak., peak | provided | no | group |
| Kaess et al., 2017 | 66 | 14.91 | 42.4 NA NA | 15 | 14.79 | 53.3 51 | 14.95 | (0.42) | 39.2 self-report (quest.) | CTQ | cut-offs | PA, SA | saliva | nmol/awak., 30, 60 min | NA | 2 | awak., AUC | provided | raw | provided |
| Keshin et al., 2014 | 36 | 14.97 | 100.0 NA yes*, mixed | 24 | 15.04 | 100.0 12 | 14.84 | 1.34 | 100.0 informant (record) | medical other (record) | CTQb | cut-offs | SA | saliva | log awak., 30 | 30 min | 3 | awak., peak, delta | paper, table | no | delta |
| Klaassen et al., 2009 | 22 | 47.58 | 100.0 NA no | 10 | 47.80 | 100.0 12 | 47.40 | (1.19) | 100.0 self-report (quest.) | CTQb | cut-offs | PA, N | saliva | nmol/awak., 30, 60 min | 30 min | 2 | awak., peak, AUC & table | paper, figure no | no | paper & table |
| Kraus et al., 2018 | 103 | 34.54 | 10.80 | 0.0 no | 39i | 35.16 | 0.0 | 64 | 34.17 | 0.0 | self-report (quest.) | CTES | cut-offs | PA, SA, other | saliva | nmol/awak., 1 min | 30 min | 2 | awak., peak, delta | paper, figure no | yes, group | provided |
| Kuhlman et al., 2015 | 79 | 12.73 | 40.5.22.6d yes, mixed | 60 | 12.91 | 41.2 19 | 12.16 | 36.8 | informant (quest.) | EIT | other (spec.) | EA, PA | saliva | g/dl awak., 45 min | 45 min | 2 | awak., peak, delta | provided | yes | group |
| Kumsta et al., 2015 | 57 | 24.02 | 61.4 NA NA | 44i | 24.15 | 68.2 13 | 23.55 | 38.5 | no instr. institutionalization N | CTQ | cut-offs | PA, SA, N | saliva | nmol/awak., 30, 45 min | 45 min | 2 | awak., peak, delta | provided | no | group |
| Kusnietz et al., 2017 | 61 | 33.80 | 49.2 NA no | 29 | 32.0 | NA NA | 32 | NA | self-report (quest.) | CTQ | cut-offs | PA, SA, N | saliva | nmol/awak., 30, 60 min | 60 min | 1 | awak., peak, AUC | provided | no | group |
| Li et al., 2015 | 75 | 37.92 | 72.0 46.7 yes*, MDD | 38 | 39.70 | 68.4 37 | 36.10 | (12.30) | 75.7 self-report (quest.) | AUCb | cut-offs | PA, SA | saliva | µg/dl awak., 15, 30, 60 min | 30 min | 1 | awak., peak | provided | no | group |
| Lu et al., 2016a | 35 | 23.80 | 51.4 100.0 yes, MDD | 18 | 23.70 | 44.4 17 | 23.90 | 58.8 | self-report (quest.) | CTQ | cut-offs | PA, SA | saliva | ng/ml awak., 30, 45, 60 min | 30 min | 1 | awak., AUC, AUCp | paper, figure no | yes, group |
| Lu et al., 2016b | 45 | 21.65 | 62.2 100.0 no | 23 | 21.50 | 60.9 22 | 21.80 | 63.6 | self-report (quest.) | CTQ | cut-offs | PA, N | saliva | ng/ml awak., 30, 45, 60 min | 30 min | 1 | awak., AUC, AUCp | paper, figure no | yes, group |
| Mello et al., 2015 | 63 | 10.63 | 47.6 NA NA | 25 | 10.40 | 56.0 38 | 10.79 | 42.1 | self-report (quest.) | CTQ | cut-offs | PA, SA, N | saliva | µg/dl awak., 30, 0, 30 min | 30 min | 1 | awak., peak, delta | provided | yes | group |
| Monte Leone et al., 2015a | 44 | 26.50 | 100.0 yes, MDDb | 24 | 27.90 | 100.0 20 | 21.80 | (7.10) | 100.0 self-report (quest.) | CTQ | cut-offs | PA, SA, N | saliva | nmol/awak., 15, 30, 60 min | 30 min | 1 | awak., peak, delta | provided | no | group |
| Monte Leone et al., 2015b | 36 | 28.90 | 100.0 yes, BN | 22 | 31.00 | 100.0 14 | 25.70 | (9.40) | 100.0 self-report (quest.) | CTQ | cut-offs | PA, SA, N | saliva | nmol/awak., 15, 30, 60 min | 30 min | 1 | awak., peak, delta | provided | no | group |
| Peng et al., 2014a | 58 | 28.61 | 46.6 NA yes, MDD | 28 | 28.87 | 46.4 30 | 28.37 | 46.7 | self-report (quest.) | CTQ | cut-offs | N | saliva | µg/dl awak., 30, 45 min | 30 min | 2 | delta | no | no | delta |
| Peng et al., 2014b | 51 | 28.09 | 47.1 NA no | 22 | 28.37 | 45.5 29 | 27.87 | 48.3 | self-report (quest.) | CTQ | cut-offs | N | saliva | nmol/awak., 30 min | 30 min | 2 | delta | no | no | delta |
| Quevedo et al., 2017 | 55 | 14.91 | 63.6 50.9 yes*, MDDc | 35 | 14.84 | 77.1 20 | 15.04 | 40.0 | self-report (interview) | K-SADS-P | other (spec.) | PA, SA, N | saliva | nmol/awak., 30, 60 min | 30 min | 3 | awak., peak, delta | provided | yes | group |
| Reichl et al., 2016a | 25 | 16.25 | 96.0 NA no | 3 | 16.55 | 100.0 22 | 16.20 | 95.5 | self-report (interview) | CECA | cut-offs | PA, N | saliva | nmol/awak., 30, 60 min | 30 min | 3 | awak., peak, delta | provided | yes | group |
| Reichl et al., 2016b | 24 | 16.32 | 91.7 NA yes, mixed | 16 | 16.65 | 93.8 8 | 15.65 | 87.5 | self-report (interview) | CECA | cut-offs | PA, SA, N | saliva | nmol/awak., 30, 60 min | 30 min | 3 | awak., peak, delta | provided | yes | group |
| Schreuder et al., 2016 | 114 | 26.75 | 45.6 NA yes | 16 | 27.81 | 43.8 98 | 26.58 | 45.9 | self-report (quest.) | CTQ | cut-offs | SA | saliva | µg/dl awak., 30 min | 30 min | 1 | awak., peak, delta | provided | no | group |
| Smeets et al., 2017 | 22i | 40.50 | 110.63 | 0.0 no | 13 | 38.77 | (11.38) | 43.00 | (12.21) | 100.0 self-report (interview) | self-dev. | other (spec.) | SA | saliva | nmol/awak., 15, 30, 45, 60 min | 30 min | 2 | awak., peak, AUCp | provided | no | group |

(continued on next page)
### Table 3 (continued)

| Study | Total Sample | CM Sample | Control Sample | CM Assessment | Cortisol Assessment | Data |
|-------|--------------|-----------|----------------|---------------|---------------------|------|
| N     | Age M (SD)   | Sex       | Psycho.        | n             | Age M (SD)          | Sex  | Method | Instrument | Grouping | Type | Time | Time | Rel. | Indices | Source | Data (re) grouped |
| Steudte et al., 2013 | 37.87 | (12.04) | | 42.50 | (11.39) | self-report (quest.) | | | | | | | | |
| van der Vegt et al., 2009 | 529 | 30.83 | 33.1 NA NA | 235 | 30.87 | 51.9 | 294 | 30.80 | 18.0 | informant (interview) | self-dev. other (spec.) | | | | | | |
| van Zuider et al., 2011 | 317 | 30.19 | 0.0 NA NA | 152 | 30.20 | 0.0 | 165 | 30.17 | 0.0 | self-report (quest.) | | | | | | |
| Vreeburg et al., 2009a | 996 | 42.35 | 65.9 NA yes, mixedd | 565 | 43.68 | (11.82) | 696 | 43.60 | (12.79) | self-report (interview) | NEMESIS other (spec.) | | | | | |
| Vreeburg et al., 2009b | 431 | 42.71 | 59.4 NA no | 106 | 47.07 | 69.8 | 325 | 41.00 | 56.0 | self-report (interview) | NEMESIS other (spec.) | | | | | |

**Note.** N = sample size; M = mean; SD = standard deviation; The sex ratio is indicated as percentage of female participants; Ethn. = ethnicity; The ethnicity ratio is indicated as percentage of non-Caucasians; NA = not assessed; Psychoopath. = psychopathology, whereby the following definitions have been used: yes = at least some of the participants met diagnostic criteria for a psychiatric disorder, no = none of the participants met diagnostic criteria for any psychiatric disorder, e = groups are not matched with respect to psychopathology; f = no gold-standard diagnostic tool was named; PTSD = Posttraumatic Stress Disorder; MDD = Major Depressive Disorder; AN = Anorexia Nervosa; BN = Bulimia Nervosa; CM = child maltreatment; n = sample size; quest. = questionnaire; self-dev. = self-developed (the authors used a scale developed by themselves); CTQ = Childhood Trauma Questionnaire; CITES = Childhood Traumatic Events Scale; ETI = Early Trauma Inventory; no instr. used = no instrument used; K-SADS-P = Kiddie Schedule for Affective Disorders and Schizophrenia for School-Aged Children Present Version; CIECA = Childhood Experiences of Care and Abuse; SF = short form; NEMESIS CTI = Netherlands Mental Health Survey and Incidence Study Childhood Trauma Interview; spec. = specification (authors applied a specific definition of CM); SA = sexual abuse; PA = physical abuse; EA = emotional abuse; N = neglect; g = The grouping of participants was not just based on CM experiences but also included other traumatic experiences; awak. = awakening; Rel. = reliability, whereby the following definitions have been used: 1 = cortisol assessed over one day, 2 = cortisol assessed over two days, 3 = cortisol assessed over more than two days; AUCg = area under the curve with respect to ground; delta = peak minus awakening levels; AUCi = area under the curve with respect to increase. h = Time points cortisol was sampled. Authors also administered ETI, but grouping was based on CTQ. i = ELS AA/AG and ELS GG were combed into one group and compared to no ELS AA/AG and no ELS GG group. j = Percentage of non-Caucasians refers to total sample (N = 127). k = Deprieved adoptees were combined into one group and were compared to the non-deprived UK adoptees. l = Raw data were provided; participants with sampling adherence of +/- 5 min were included. Patients were matched with respect to MDD but not with respect to PTSD. m = Exclusion of repressed memory group; CM sample: recovered and continuous memory group, control sample: control group. n = The data were regrouped including only the traumatized control subjects and the non-traumatized control subjects (remark: some in the control sample may have experienced other traumatic events). o = The data are from the Netherlands Study of Depression and Anxiety, a large cohort study. The article did not appear in the initial search but was suggested by the respective author as best suited for citation in this meta-analysis. p = Patient group: dysthymia, MDD, social phobia, panic with/without agoraphobia, generalized anxiety disorder. q = No one of the participants fulfilled the diagnostic criteria for dysthymia, MDD, social phobia, panic with/without agoraphobia, or generalized anxiety disorder in the past 6 months. r = It is not clearly stated in the text or in the subheading of the figure what means and standard deviations or means and standard errors were presented; we assumed standard errors.
25 (k = 29) on self-report data, and three studies using both information sources. The most frequently used self-report was the Childhood Trauma Questionnaire (CTQ; n = 15, k = 17), and accordingly, cut-off scores were mostly used to group study participants in these studies. Nevertheless, several other instruments were also employed, resulting again in various grouping approaches. It should be noted that five studies did not focus on child maltreatment only (Hengesch et al., 2018; Ivanov et al., 2011; Kaiser et al., 2018; Otte et al., 2005; Ouellet-Morin et al., 2011) but also included participants with other ELA experiences. By far, the most frequently applied stress task was the Trier Social Stress Test (TSST) or the TSST-C (n = 18, k = 21) and the majority of studies contained some social-evaluative aspects (k = 29; for an overview of the different tasks applied in the various studies, see Appendix E). The average duration of the stressors used was about 19.02 (SD = 16.94) min (k = 27 between 10 and 20 min). In n = 3 studies no cortisol response following the onset of the corresponding stressor was observed. Interestingly, these studies all applied stressors that did not contain any social-evaluative challenges. In k = 29 comparisons a cortisol response was observed in both groups (with different peak times found for k = 5 comparisons), and finally in k = 7 comparisons, the response was observed only in one but not in the other group (k = 6 only in the control sample, k = 1 only in the child maltreatment sample). On average, the time between the onset of a stressor and peak cortisol levels being reached was 29.84 (SD = 15.98) min, with the majority of studies reporting that the peak was reached between +20 and +40 min post stressor onset (k = 26). The vast majority of studies used saliva samples to assess cortisol. Baseline, peak and recovery data were reported by most publications, with considerably fewer studies reporting on AUC or AUCi indices. Finally, the data of n = 19 (k = 20) studies were provided by corresponding authors, with the data of k = 13 comparisons being (re)grouped. For further details on the characteristics of the included studies, see Table 5.

3.2.3.3. Meta-analysis. The pooled effect estimates for the different indices are displayed in Table 6. The results of the sensitivity analyses (where appropriate) are also presented. As shown in the corresponding lines of Table 6, the results of the meta-analyses on baseline cortisol showed no significant overall differences in cortisol levels assessed prior to the onset of the respective stress task between the child maltreatment and the control sample (holding true for the sensitivity analyses). In contrast, the release of cortisol following the perception of a stressor – expressed as peak, recovery, delta, and AUCi cortisol – was lower in the child maltreatment group compared to the control sample (with all pooled effect estimates being in the small range), indicating a blunted cortisol stress reactivity (see Appendix C Figs. 3.1–3.6 for corresponding forest plots). For AUCi, the pooled effect estimate was not statistically significant (p = 0.081). However, when excluding one outlier study (Ivanov et al., 2011), significance was also reached for this outcome index (p = 0.021). Between-study heterogeneity was in the moderate to high range for some of the outcome indices (I² > 50%), exceeding the level of significance (Q-statistics all p < 0.05) for all but AUC cortisol. Visual inspection of traditional and counter-enhanced funnel plots as well as Egger’s regression test of funnel plot asymmetry revealed the absence of small-study bias for baseline, peak, AUCi and AUC cortisol levels (all p > 0.05). However, the Egger’s regression test of funnel plot asymmetry reached significance for delta as well as recovery cortisol levels (p < 0.05), suggesting the presence of small-study bias (see Appendix C Figs. 3.1–3.6 for corresponding funnel plots).

3.2.3.4. Meta-regression and subgroup analyses. We conducted a number of pre-defined meta-regression and subgroup analyses focusing on peak, delta, recovery, AUCi and AUC cortisol. The summary results for each outcome index and each moderator examined are shown in Appendix D Table D3. In the following section, the results for moderators found to significantly influence the main effects are outlined. For delta, recovery, and AUC cortisol, we identified the proportion of women in the sample as a continuous moderator (for AUCi; p = 0.052), with an increase in the proportion of females being associated with lower cortisol secretion following the perception of a stressor when comparing the child maltreatment and the control sample. Additionally, for delta cortisol, the proportion of participants fulfilling diagnostic criteria for a mental disorder significantly moderated the summary effect, with an increase of the proportion being associated with a stronger blunting of the cortisol stress response (β = –0.006, 95% CI [–0.010; –0.002], p = 0.007, R² = 99.20%). This finding, however, should be interpreted with caution, as only two of the studies that included a purely clinical sample (Schalinski et al., 2015; Suzukie et al., 2014) reported on delta cortisol. In addition, and in contrast to the other studies involving a clinical sample (exception Rao & Morris, 2015), these two studies reported relatively strong negative effects. Nevertheless, despite considerable heterogeneity between the studies, all outcome indices showed stronger effects for studies including purely clinical samples and markedly weaker effects for those studies that involved healthy subjects only (see results subgroup analyses). Furthermore, stronger effects were found for studies that observed a cortisol response in just one of the groups (holding true for all outcome indices) compared to studies that found a response in both groups and those that found no response in either of the groups, with the subgroup comparison reaching significance for delta (Q2 = 4.53, p = 0.033) and AUCi (Q2 = 12.33, p = 0.002) cortisol. Comparing studies focusing on child maltreatment experiences only to those involving participants with other types of ELA as well showed that the few studies that also considered other types of ELA overall yielded greater negative effect estimates for all outcome indices, but significant for delta cortisol only (Q2 = 3.95, p = 0.047). However, it should be noted that heterogeneity within these studies varied substantially as the detailed description of the individual quality items can be found in Appendix B Table B10 or Table B3.
between the different outcome indices and thus depended highly on the included studies. Finally, again depending on the outcome index investigated (and thus on the studies included), the different subdomains of the quality assessment appeared to explain part of the variance in the effect estimates between studies, although this effect was investigated (and thus on the studies included), the different sub-
exclusion of outlier studies Lu et al., 2016a and Monteleone et al., 2018b; * exclusion of influential study Monteleone et al., 2018b.

### Table 4

| Outcome indices                    | Random-effects model | Heterogeneity measures |
|-----------------------------------|----------------------|------------------------|
|                                   | N  | k   | Hedges' g | CI (95%) | p   | Q  | p   | Pred. int. | Egger’s Test a |
| Awakening cortisol                | 3342 | 24 | -0.04     | -0.13; 0.05 | 0.334 | 0.01 | 17.3% | 27.82 | 0.223 | -0.24; 0.15 | -0.732 (0.054) |
| Awakening cortisol sensitivity analysis * | 2271 | 22 | -0.05     | -0.14; 0.04 | 0.277 | 0.00 | 0.00 | 19.15 | 0.576 | -0.14; 0.05 | - |
| Peak cortisol                     | 3275 | 23 | -0.02     | -0.19; 0.15 | 0.812 | 0.10 | 58.3% | 52.72 | 0.000 | -0.71; 0.67 | -0.636 (0.254) |
| Peak cortisol sensitivity analysis * | 3204 | 21 | 0.02     | -0.09; 0.13 | 0.701 | 0.02 | 36.5% | 31.48 | 0.049 | -0.28; 0.32 | - |
| Peak cortisol sensitivity analysis | 3239 | 22 | 0.04     | -0.16; 0.17 | 0.596 | 0.04 | 43.5% | 37.16 | 0.016 | -0.38; 0.45 | - |
| 60 min post awakening cortisol    | 2276 | 14 | 0.02     | -0.13; 0.17 | 0.798 | 0.03 | 41.6% | 22.26 | 0.052 | -0.36; 0.40 | -0.095 (0.876) |
| 60 min post awakening cortisol sensitivity analysis * | 2240 | 13 | 0.05     | -0.17; 0.06 | 0.429 | 0.01 | 18.2% | 14.66 | 0.261 | -0.18; 0.28 | - |
| Delta cortisol                    | 2536 | 18 | -0.03    | -0.17; 0.11 | 0.719 | 0.03 | 41.8% | 29.23 | 0.033 | -0.42; 0.36 | -0.562 (0.287) |
| Delta cortisol sensitivity analysis * | 2500 | 17 | 0.02     | -0.06; 0.10 | 0.622 | 0.00 | 0.00 | 15.34 | 0.500 | -0.07; 0.11 | - |
| AUC cortisol                      | 2327 | 15 | -0.13    | -0.41; 0.15 | 0.370 | 0.22 | 73.1% | 51.99 | 0.000 | -1.18; 0.92 | -0.963 (0.243) |
| AUC cortisol sensitivity analysis | 2211 | 12 | -0.18    | -0.46; 0.05 | 0.118 | 0.08 | 61.9% | 28.87 | 0.002 | -0.88; 0.52 | - |
| AUC cortisol                      | 1654 | 9  | -0.07    | -0.48; 0.33 | 0.723 | 0.28 | 70.3% | 26.96 | 0.001 | -1.41; 0.126 | -0.359 (0.719) |

Note. AUCg = area under the curve with respect to ground; AUCa = area under the curve with respect to increase; 95% CI = 95% confidence interval; Pred. int. = prediction interval. * Intercept and (p values) displayed. Awakening cortisol: * exclusion of influential studies Li et al., 2015 and Vreeburg et al., 2009a. Peak cortisol: * exclusion of outlier studies Li et al., 2016a and Monteleone et al., 2018b; * exclusion of influential study Monteleone et al., 2018b. Due to convergence problems associated with the Fisher scoring algorithm, the DerSimonian-Laird estimator was used to estimate $\tau^2$. 60 min post awakening cortisol: * exclusion of outlier study Monteleone et al., 2018b. Delta cortisol: * exclusion of outlier and influential study Monteleone et al., 2018b. AUC cortisol: * exclusion of outlier studies Lu et al., 2016a, Lu et al., 2016b, and Monteleone et al., 2018b. See Appendix C Figs. 2.1–2.6 for corresponding forest and funnel plots.

#### 3.2.4. Pharmacological challenge tests

##### 3.2.4.1. Dexamethasone suppression test

**3.2.4.1.1. Included studies.** Eleven articles, containing $k = 17$ comparisons, involving a total of $n = 2,222$ participants (range: 16–1,112) assessed cortisol in the context of the DST. Of these, $k = 16$ reported on baseline cortisol levels (cortisol assessed before dexamethasone administration; pre-DST), $k = 17$ on cortisol assessed following the administration of dexamethasone (post-DST), and $k = 9$ contained information on delta values (post-DST cortisol minus pre-DST cortisol). The included studies mainly consisted of adults, with only one study involving adolescents. The average age was 33.32 (SD = 8.32) years and studies ranged from 45.6 to 100.0% ($M = 72.6$, $SD = 18.6$) in terms of the proportion of women ($k = 3$ studies with a purely female sample). Five studies ($k = 8$) did not report on the percentage of Non-Caucasians, while the percentage of Non-Caucasians in the remaining studies ranged between 0.0 and 100.0% ($M = 53.4$, $SD = 41.2$). Three out of the $k = 17$ comparisons involved healthy participants and $k = 14$ included participants in whom the proportion of people suffering from a mental illness ranged from 13.2 to 100.0% ($k = 9$ studies involved purely clinical samples and $k = 3$ involved participants where the child maltreatment and the control sample were not matched in terms of psychopathology). Various instruments to assess child maltreatment were applied, all relying on self-report information. The most common self-report used was the CTQ ($n = 5$, $k = 9$). It should be noted that the child maltreatment sample of the study from Faravelli et al. (2010) did not only consist of participants with child maltreatment experiences, but also included several participants with loss experiences. Approximately half of the studies used established cut-off values to group participants in the corresponding child maltreatment and control groups, with the others mostly applying specific definitions. All but three studies ($k = 4$) used 0.5 mg of dexamethasone and the data of four studies ($k = 8$) were re-grouped for this meta-analysis (see Table 7 for more details).

##### 3.2.4.1.2. Risk of bias assessment

Studies assessing cortisol in the context of the DST received an average score of 72.1/100.0 ($SD = 19.0$) for selection of participants, 61.3/100.0 ($SD = 12.1$) for appropriate assessment of cortisol, and 44.7/100.0 ($SD = 20.3$) for adequate controlling for confounders, resulting in an average overall score of 58.1/100.0 ($SD = 13.3$, range: 40.0–76.0). The detailed quality ratings for the individual studies as well as the detailed description of the individual quality items can be found in Appendix B Table B11 or Table B4. None of the studies included used two different sources to establish the presence of child maltreatment, reported whether cortisol was analyzed in one batch and whether participants were excluded when working night shifts, and only one study assessed whether exposure and control groups differed in relation to the experience of other traumatic events during childhood or adulthood. Moreover, less than half of the comparisons reported whether sampling was postponed when participants were sick ($k = 7$), whether dexamethasone intake was checked ($k = 7$), whether participants differed in smoking ($k = 6$), intake of oral contraceptives ($k = 7$), and their use of medication (with CNS effect; $k = 4$).

##### 3.2.4.1.3. Meta-analysis

The results of the meta-analyses yielded no significant overall differences in pre-DST cortisol (Hedges' $g = 0.07$, 95% CI (-0.09; 0.23), $p = 0.402$), post-DST cortisol (Hedges' $g = 0.01$, 95% CI [-0.18; 0.20], $p = 0.936$) nor in delta cortisol values (Hedges' $g = -0.13$, 95% CI [-0.31; 0.06], $p = 0.178$) between participants with child maltreatment experiences and participants without corresponding experiences. Between-study heterogeneity was in the low to moderate range ($r^2 < 45%$) reaching significance for pre- ($Q_{(15)} = 25.21$, $p = 0.047$) and post-DST ($Q_{(8)} = 28.64$, $p = 0.027$) cortisol (see Appendix C Figs. 4.1.1–4.1.3 for corresponding forest plots). Visual inspection of traditional and counter-enhanced funnel plots as well as the results of the Egger's regression tests ($p > 0.316$) implied absence of small-study bias for funnel plots see Appendix C Figs. 4.1.1–4.1.3. Corresponding sensitivity analyses excluding studies with extreme effect sizes and influential studies did not change the overall results (see Table 8 for related statistics).

##### 3.2.4.1.4. Meta-regression and subgroup analyses

We conducted a number of pre-defined meta-regression and subgroup analyses focusing on post-DST cortisol only. The summary results for each moderator examined are shown in Appendix D Table D4. The only moderator that
Table 5
Summary characteristics of included studies that reported on cortisol assessed in the context of a stressor (cortisol stress reactivity).

| Study | Total Sample | CM Sample | Control Sample | CM Assessment | Stressor | Duration | Stressor | Cortisol response | Type of Unit | Indices | Source | Data grouped |
|-------|--------------|-----------|----------------|---------------|----------|----------|----------|-------------------|--------------|---------|--------|--------------|
|      | N | Age M (SD) | Sex | Ethn. | Psychopath. | n | Age M (SD) | Sex | n | Age M (SD) | Sex | Method | Instrument | Grouping | Type | Task | Duration | Stressor | Time point | Cortisol | Type of Unit | Datasource | Data grouped |
| Alexander et al., 2018 | 200 | 23.72 (2.85) | 50.0 | 0.0 | no | 32 | 23.81 (3.01) | 56.3 | 168 | 23.70 (2.83) | 48.8 | self-report (quest.) | CTQ | cut-offs | EA, PA, SA | N | TSST | 15 min | 25 min peak yes, both ident. | saliva | nmol/l | all | provided, group | yes |
| Ali and Pruessner, 2012 | 37 | 25.77 (5.37) | 51.4 | NA | no | 20 | 25.95 (5.61) | 50.0 | 17 | 25.56 (5.22) | 52.9 | self-report (quest.) | CTQ | cut-offs | EA, PA, SA | N | TSST | 20 min | 30 min peak yes, both ident. | saliva | nmol/l | all | provided, group | no |
| Buchmann et al., 2014 | 195 | 19.00 (0.00) | 53.9 | NA | yes*, mixed | 15 | 19.00 (0.00) | 53.3 | 180 | 19.00 (0.00) | 53.9 | self-report (quest.) | CTQ | cut-offs | EA, PA, SA | N | TSST | 20 min | 20 min peak yes, both ident. | saliva | nmol/l | all | provided, group | yes |
| Carpenter et al., 2007 | 50 | 29.11 (11.31) | 66.0 | 24.0 | no | 23 | 35.00 (12.90) | 74.0 | 27 | 24.10 (6.60) | 59.0 | self-report (quest.) | CTQ | cut-offs | EA, PA, SA | N | TSST | 25 min | 30 min peak yes, both ident. | saliva | nmol/l | all | provided, group | no |
| Carpenter et al., 2011 | 110 | 30.45 (11.13) | 100.0 | NA | yes*, mixed | 20 | 36.50 (12.50) | 50.0 | 80 | 29.10 (10.40) | 48.2 | self-report (quest.) | CTQ | cut-offs | PA | TSST | 25 min | 45 min peak yes, both ident. | saliva | nmol/l | all | bl, peak end, delta | paper, figures | no |
| Cook et al., 2012 | 175 | 15.36 (1.01) | 51.8 | NA | NA | 86 | 15.39 (1.03) | 55.4 | 89 | 15.30 (0.97) | 48.2 | self-report (quest.) | CTQ | cut-offs | EA, PA, N | TSST-C | 15 min | 33 min NA yes* | saliva | NA | | provided, group | no |
| England-Mason et al., 2017 | 120 | 32.20 (4.26) | 100.0 | NA | NA | 25 | 30.47 (4.36) | 100.0 | 95 | 32.65 (4.14) | 100.0 | self-report (quest.) | CTQ | cut-offs | EA, PA, SA | N | stroop task | 17 min | 37 min peak yes, only in CG | saliva | nmol/l | all | provided, group | yes |
| Fogelman et al., 2016a | 73 | 63.78 (7.39) | 58.9 | 8.5 | no | 24 | 65.00 (9.21) | 66.7 | 49 | 63.18 (6.33) | 55.1 | self-report (quest.) | CTQ | cut-offs | EA, PA, SA | N | TSST | 15 min | 25 min peak yes, both ident. | saliva | nmol/l | all | provided, group | yes |
| Fogelman et al., 2016b | 83 | 23.86 (7.51) | 0.0 | 4.5 | no | 20 | 24.65 (7.51) | 0.0 | 65 | 23.62 (7.30) | 0.0 | self-report (quest.) | CTQ | cut-offs | EA, PA, SA | N | TSST | 15 min | 25 min peak yes, both ident. | saliva | nmol/l | all | provided, group | yes |
| Harkness et al., 2011 | 71 | 15.39 (2.11) | 67.6 | 11.0 | yes*, mixed | 26 | 15.88 (1.71) | 69.2 | 45 | 15.11 (2.26) | 66.7 | self-report (interview) | CECA | cut-offs | EA, PA, SA | N | TSST | 30 min | 30 min peak yes, both ident. | saliva | µg/dl | bl, peak end, delta, AUC<sub>0</sub> | provided, group | yes |
| Heim et al., 2000a | 26 | 29.77 (6.37) | 100.0 | 38.5 | yes*, mixed | 14 | 30.21 (5.46) | 100.0 | 12 | 29.25 (7.52) | 100.0 | self-report (interview) | ETI | other (spec.) | PA, SA | TSST | 20 min | 30 min peak yes, both ident. | blood | nmol/l | peak | paper, table | no |
| Heim et al., 2000b | 23 | 33.35 (6.00) | 100.0 | 13.0 | yes*, MDD<sup>a</sup> | 13 | 32.38 (7.64) | 100.0 | 10 | 34.60 (8.38) | 100.0 | self-report (interview) | ETI | other (spec.) | PA, SA | TSST | 20 min | 30 min peak yes, both ident. | blood | nmol/l | peak | paper, table | no |
| Hengesch et al., 2018 | 44 | 22.15 (2.91) | 56.8 | NA | NA | 22<sup>b</sup> | 22.50 (3.04) | 63.6 | 22<sup>c</sup> | 21.80 (2.80) | 50.0 | mixed | no instr. | other (self-ident., agency) | other (spec.) | PA, SA, loss<sup>c</sup> | CPT + PASAT | 3 min | 25 min peak yes, both ident. | saliva | nmol/l | bl, delta, AUC<sub>0</sub> | paper, table | no |
| Ivanov et al., 2011 | 25 | 10.55 (1.02) | 36.0 | 88.0 | yes, ADHD | 10 | 10.78 (1.12) | 50.0 | 15 | 10.39 (0.95) | 26.7 | mixed | (quest.) | CLES, PtsRI | PA, SA, loss<sup>c</sup> | watching emot. video | 11 min | end of viewing no peak | saliva | µg/dl | bl, peak end, delta, AUC<sub>0</sub>, AUC<sub>C</sub> | provided, group | yes |
| Kaiser et al., 2018 | 55 | 27.62 (6.43) | 100.0 | 43.0<sup>a</sup> | yes, mixed | 42 | 28.43 (6.65) | 100.0 | 13 | 25.00 (5.03) | 100.0 | self-report (interview) | TAQ | cut-offs | PA, SA, WV, peer<sup>c</sup> | MAST | 10.75 min | 28 min peak yes, both ident. | saliva | ng/ml | bl, peak end, delta, AUC<sub>C</sub> | provided, group | yes |
| Kuhlman et al., 2015 | 91 | 12.75 (2.27) | 40.3 | 22.0<sup>a</sup> | yes, mixed | 72 | 12.91 (2.26) | 41.2 | 19 | 12.16 (2.27) | 36.8 | informant (quest.) | SE-CPT | cut-offs | EA, PA, SA | SE-CPT | 5 min | 25 min peak yes, both ident. | saliva | µg/dl | all | provided, group | yes |
| Licalvo et al., 2019<sup>c</sup> | 699 | 23.70 (3.14) | 55.9 | 20.3 | no | 335 | 23.90 (3.26) | 63.0 | 364 | 23.50 (3.02) | 49.5 | self-report (interview) | C-DES-IV | other (spec.) | EA, PA, SA | public speaking, arithmetic | 45 min | 30 min peak yes, both ident. | saliva | µg/dl | bl, peak end, delta | provided, group | yes |
| Luecken et al., 2009 | 76<sup>a</sup> | 18.90 (0.97) | 48.7 | 25.0 | NA | 19 | NA | NA | 48 | NA | self-report (quest.) | CTQ | cut-offs | EA, PA, SA, N | role-play task | 10 min | 10 min peak yes, both ident. | saliva | µg/dl | all | provided, group | yes |

(continued on next page)
Table 5 (continued)

| Study                                    | Total Sample | CM Sample | Control Sample | CM Assessment | Stressor | Duration | Time point peak | Cortisol response | Cortisol Assessment | Data |
|------------------------------------------|--------------|-----------|----------------|---------------|----------|----------|-----------------|-------------------|-------------------|------|
| Luecken and Appelhans, 2006             | 88           | 19.30 (1.63) | 62.5 22.7 NA   | 16            | 19.81 (1.83) | 56.3 19.18 | 63.9 self-report (quest.) | CTQ               | cut-offs           | EA, PA, SA, N |
| Martinson et al., 2016                   | 50           | 19.70 (4.19) | 100.0 11.8 yes*, mixed | 26            | 21.38 (7.59) | 100.0 24 | 18.92 (1.14) | self-report (mixed) | self-dev., SCID-I/P | other (spec.) | SA |
| Mielock et al., 2017                     | 52           | 27.44 (9.47) | 100.0 61.5 yes, MDD | 26            | 26.20 (9.50) | 100.0 26 | 28.70 (9.40) | self-report (interview) | CAI               | cut-offs           | PA, SA, WV |
| O’Connor et al., 2018                    | 145          | 26.93 (9.39) | 60.7 26.8 yes, mixed | 65            | 29.95 (10.15) | 57.8 80 | 24.24 (7.59) | self-report (quest.) | CTQ               | cut-offs           | EA, PA, SA, N |
| Otte et al., 2005                        | 76           | 28.00 (5.00) | 13.2 NA no     | 16            | 27.00 (4.00) | 6.0 60 | 28.00 (5.00) | self-report (interview) | LSC-R             | other (spec.) | PA, SA, bully |
| Ouellet-Morin et al., 2011               | 190          | 12.00 (0.00) | 49.5 8.4 NA    | 64            | 12.00 (0.00) | 54.7 126 | 12.00 (0.00) | self-informant (develop.) | self-informant (develop.) | other (spec.) | PA, SA, bv |
| Ouellet-Morin et al., 2018               | 155          | 24.10 (3.70) | 0.0 NA no      | 56            | 24.00 (3.60) | 0.0 99 | 24.10 (3.70) | self-report (quest.) | CTQ               | cut-offs           | EA, PA, SA, N |
| Pierrehumbert et al., 2009               | 44*          | 33.02 (7.12) | 100.0 NA yes, mixed | 27            | 33.38 (6.01) | 100.0 17 | 32.46 (8.87) | self-report (interview) | ETI               | other (spec.) | EA, PA, SA, N |
| Rao and Morris, 2015                     | 17*          | 18.93 (4.72) | 52.9 35.3 yes, MDD | 7             | 18.40 (3.90) | 42.9 10 | 19.30 (5.40) | self-report (mixed) | CAI               | cut-offs           | PA, SA, WV |
| Schäfke et al., 2013                     | 33*          | 34.60 (10.40) | 100.0 88.0 yes, mixed | 16            | 34.30 (11.65) | 100.0 17 | 34.90 (9.40) | self-report (interview) | ETI               | other (cluster) | EA, PA, SA, SV |
| Schäfke et al., 2014                     | 60           | 52.02 (5.09) | 66.7 NA no     | 30            | 52.57 (5.52) | 66.7 30 | 51.47 (4.64) | self-reported (quest.) | CTQ               | cut-offs           | EA, PA, SA, N |
| Schäfke et al., 2015                     | 73           | 9.54 (1.15)  | 52.1 47.0 NA   | 37*           | 9.40 (1.16) | 54.1 36* | 9.68 (1.14) | self-informant (informant) | informal (record) | other (record) | CPS, CTS |
| Shen et al., 2014                        | 104          | 16.98 (1.18) | 100.0 59.0 NA  | 47            | 16.75 (1.10) | 100.0 57 | 17.13 (1.21) | self-informant (record) | CPS               | other (record) | PA, SA, N |
| Shen et al., 2010                        | 144          | 18.09 (3.47) | 100.0 45.8 NA  | 60            | 18.54 (3.62) | 100.0 84 | 17.76 (3.34) | self-informant (record) | CPS               | other (record) | SA |
| Sullivan et al., 2013                    | 64           | 6.89 (0.95)  | 52.0 88.0 NA   | 30            | 6.98 (0.88)  | 48.0 34 | 6.72 (1.08)  | self-informant (informant) | informal (record) | other (record) | CPS, CTS-PC |
| Sunner et al., 2014                      | 158          | 14.93 (1.39) | 53.8 47.0 NA   | 61            | 15.30 (1.30) | 60.7 97 | 14.70 (1.40) | self-reported (mixed) | self-report (mixed) | CTQ, CEGA | cut-offs & other (specific) |
| Suzuki et al., 2014a                     | 41           | 44.89 (13.06) | 58.5 39.0 no   | 17            | 44.30 (12.50) | 52.9 24 | 45.30 (13.70) | self-reported (quest.) | self-reported (quest.) | CTQ | cut-offs & other (specific) |
| Suzuki et al., 2014b                     | 38           | 51.91 (11.30) | 71.1 0.0 yes, MDD | 20            | 52.10 (12.00) | 80.0 18 | 51.70 (10.80) | self-reported (quest.) | self-reported (quest.) | CTQ | cut-offs & other (specific) |

(continued on next page)
| Study | Total Sample | CM Sample | Control Sample | CM Assessment | Stressor | Cortisol Assessment | Data |
|-------|--------------|-----------|----------------|---------------|----------|---------------------|------|
|       | N | Age M (SD) | Sex | Ethn. | Psycho-path. | n | Age M (SD) | Sex | n | Age M (SD) | Sex | Grouping | Type | Task | Duration | Stressor | Time point | Cortisol | Type of | Unit | Indices | Source | Data (re) grouped |
| Trickett et al., 2014 | 454 | 10.93 (1.16) | NA | no | | 303 | 10.84 (1.16) | NA | | 151 | 11.11 (1.15) | 40.0 | informant | DCF | other | TSST-C | 14 min | 24 min peak | yes | both | ident. | saliva | µg/dl | no | paper, table & figure |
| et al., 2017a | 59 | 34.43 (12.02) | 100.0 | no | | 22 | 36.50 (12.20) | 100.0 | 37 | 33.20 (11.90) | 100.0 | self-report | ETI | other | TSST | 15 min | 40 min peak | yes | both | ident. | saliva | nmol/l | no | paper, table & figure |
| Wingenfeld et al., 2017b | 84 | 35.00 (11.21) | 100.0 | yes, MDD | | 32 | 34.20 (10.50) | 100.0 | 52 | 35.50 (11.70) | 100.0 | self-report | ETI | other | TSST | 15 min | 40 min peak | yes | both | ident. | saliva | nmol/l | no | paper, table & figure |

Note. N = sample size; M = mean; SD = standard deviation; The sex ratio is indicated as percentage of female participants; Ethn. = ethnicity; The ethnicity ratio is indicated as percentage of non-Caucasians; NA = not assessed; Psychopath. = psychopathology, whereby the following definitions have been used: yes = at least some of the participants met diagnostic criteria for a psychiatric disorder, no = none of the participants met diagnostic criteria for any psychiatric disorder. a = groups are not matched with respect to psychopathology; MDD = Major Depressive Disorder; ADHD = Attention Deficit Hyperactivity Disorder; CM = child maltreatment; n = sample size; quest. = questionnaire; CTQ = Childhood Trauma Questionnaire; CECA = Childhood Experiences of Care and Abuse; ETI = Early Trauma Inventory; no instr. used = no instrument used; CLES = Codington Life Events Scale; PSTRI = Posttraumatic Stress Reaction Index; TAQ = Traumatic Antecedents Questionnaire; C-DIS-V = Computerized version of the Diagnostic Interview Schedule-IV; self-dev. = self-developed (the authors used a scale developed by themselves); SCID-I/P = Childhood Adversity Interview; LSC-R = Life Stressor Checklist-Revised; CPS = Child Protective Services; CTS-PC = Parent-Child Conflict Tactics scale; DCFS = Department of Children and Family Services; self. ident. = self-identification (participants self-identified as victims of CM); EA = emotional abuse; PA = physical abuse; SA = sexual abuse; N = neglect; institute. = institutionalization; non-intent. trauma = non-intentional trauma; bully = bullying; b = The grouping of participants was not just based on CM experiences but also included other traumatic experiences; TSST = Trier Social Stress Test; TSST-C = Trier Social Stress Test for Children; CPT = Cold Pressor Test; PASAT = Paced Auditory Serial Addition Task; emot. = emotional; MAST = Maastricht Acute Stress Test; SE-CPT = Socially Evaluated Cold Pressor Task; arithmet. = arithmetic’s; self-disc. = self-disclosure; rel.-build. = relationship-building; PST = Psychosocial Stress Test; perform. = performance; interper. = interpersonal; self-eval. = self-evaluative; ident. = identical; adj. = adjusted; bl = baseline (if possible, value just prior stressor onset was extracted); delta = peak values minus baseline value; AUCg = Area under the curve with respect to ground; AUCr = Area under the curve with respect to increase; • no cortisol peak was observed in both groups, therefore the time point closest to +25 min post stressor onset was extracted. a Community-dwelling older adult sample. b Community-dwelling younger adult sample. c Percentage of non-Caucasians refers to total sample (N = 35). d Percentage of non-Caucasians refers to total sample (N = 35). e CM sample and control sample were matched with respect to MDD but not with respect to PTSD. f Female and male participants with early life adversity were combined into one group and were compared to female and male control participants. g Percentage of non-Caucasians refers to total sample (N = 42). h Percentage of non-Caucasians refers to total sample (N = 127). i This article did not appear in the initial search but was suggested by the respective author as best suited for citation in this meta-analysis. j Demographic data for total sample (age, percentage of females, percentage of non-Caucasians) refers to N = 76. k cortisol data available for N = 67. l Demographic data for total sample and CM and control sample refers to N = 44, cortisol data available for N = 35. m Comparison between participants with MDD and CM and those with MDD without CM. n The data were regrouped, taking into account only the patient group with stress-related disorders. o Female and male maltreated participants were combined into one group and were compared to female and male control participants. p Authors also administered CTQ, but grouping was based on ETI. q It is not clearly stated in the text or in the subheading of the figure whether means and standard deviations or means and MDD without CM were presented; we assumed standard errors.
significantly influenced the main effect was the proportion of women in the respective samples (β = −0.014, 95% CI [−0.022; −0.005], p = 0.003, F² = 95.16%), with an increase of the proportion being associated with lower cortisol levels following the administration of dexamethasone (increased cortisol suppression) when comparing the child maltreatment sample with participants from the control sample. Since only one study was included that focused on ELA in general, the significant subgroup result of different pooled effect estimates for studies focusing on child maltreatment only and the study including also other childhood adversities has to be interpreted with caution. None of the methodological quality criteria significantly influenced the pooled effect estimate.

3.2.4.2. Combined dexamethasone-corticotropin releasing hormone test

3.2.4.2.1. Included studies. In total, our search strategy identified n = 21 studies that measured the responsibility of the pituitary to CRH. Of these, n = 6 studies consisting of k = 8 comparisons reporting on cortisol in the context of the Dx-CRH test and n = 2 studies reporting on cortisol in the context of the CRH test were included (k = 10). Of the included studies, k = 4 comparisons reported on baseline cortisol (cortisol assessed after the administration of dexamethasone, before CRH injection), k = 6 on peak (after the CRH injection) and delta cortisol (peak minus baseline), k = 9 on AUCₚ and k = 6 on AUCₜ cortisol. There was only one study with available recovery data (in all other studies, the peak value corresponded to the last measurement time point). The total sample of the k = 10 comparisons consisted of n = 561 participants (range: 21–230) with a mean age of 31.19 (SD = 12.70) years and an average of 60.0% females (SD = 39.6%), range: 0.0–100.0%; k = 2 comparisons contained a purely male sample and k = 4 a purely female sample). Studies reporting on Dex-CRH cortisol consisted of adult samples only, whereas the two studies focusing on the CRH test were conducted with children or adolescents. Four studies (k = 5) did not report on ethnicity with the other articles ranging between 34.6 and 57.1% in terms of percentage of Non-Caucasians (M = 45.9%, SD = 9.3%). With respect to psychopathology, k = 3 comparisons included healthy participants, k = 3 involved participants all fulfilling diagnostic criteria for MDD, k = 1 comparison included participants with mixed diagnoses, and k = 3 involved participants where the child maltreatment and the control sample were not matched for psychopathology. The authors of the three comparisons in which the subjects were not matched for psychopathology, however, showed that the presence of the specific mental disorder did not affect the cortisol results. Again, various instruments to assess child maltreatment were used, with the majority of studies relying on self-report data (k = 8). The grouping of participants also differed, including the use of cut-off scores, thirds, the presence of a CPS record and the utilization of specific definitions. The data of n = 2 studies were provided by corresponding authors, with the data of one study being re-processed. See Table 9 for more details.

3.2.4.2.2. Risk of bias assessment. Studies assessing the responsivity of the pituitary to CRH received an average total score of 69.8/100.0 (SD = 8.0, range: 58.3–83.3). All studies, or k = 9 out of 10 comparisons used an established instrument to assess the experience of child maltreatment, matched their participants with respect to age and sex, and made efforts to assess psychopathology (selection of participants: M = 76.3/100.0, SD = 12.4). Additionally, most comparisons rescheduled the sampling when participants were sick (k = 8), provided details on their test protocol (k = 10), and on how cortisol was collected, stored and analyzed (k = 10; appropriate measurement of cortisol in the context of the Dex-CRH: M = 62.3/100.0, SD = 14.8). Finally, as shown by the relatively high scores related to the control of confounding variables (M = 69.0/100.0, SD = 10.3), the majority of studies assessed or controlled for a wide variety of potential influential factors. Overall, however, only a few comparisons used two different sources of information to establish the presence of child maltreatment (k = 2), verified the ingestion of dexamethasone (k = 2), reported on whether samples were assessed in one batch (k = 1), or whether participants were excluded in case of working night shifts (k = 2), and only k = 2 comparisons made any effort to ensure that the exposure and control group did not differ with respect to other types of ELA or adult adversity. For details on individual scoring results of the primary studies as well as a summary of the average risk of bias scores, see Appendix B Table B12 or Table B5 for individual quality items.

3.2.4.2.3. Meta-analysis. The pooled effect estimates for the different indices are displayed in Table 10. For each outcome index the analysis was repeated excluding the studies focusing on the CRH test only. Results of the sensitivity analyses (where appropriate) are also presented. None of the pooled effect estimates were significant, indicating that there is no overall difference in cortisol assessed both before and after the administration of CRH and holding true for all outcome indices (all p > 0.05). Between-study heterogeneity was however in the moderate range (I² > 50%), exceeding the level of significance (Q-statistics all p < 0.05) for peak, delta, AUCₚ, and AUCₜ cortisol. Due to a
| Study                        | Total Sample | CM Sample | Control Sample | Cortisol Assessment | Data |
|-----------------------------|--------------|-----------|---------------|---------------------|------|
| N                           | Age (SD)     | Sex       | Ethn.         | Psychopath.         |      |
| Baes et al., 2014           | 20d          | 38.80 (9.84) | 75.0          | 40.0                |      |
|                              | 13           | 39.50 (9.73) | 76.9          | 9                   |      |
| Carvalho Fernado et al., 2012a | 19   | 27.05 (6.21) | 94.7          | NA                  |      |
|                              | 14           | 27.86 (6.59) | 92.9          | 0.5 mg              |      |
| Carvalho Fernado et al., 2012b | 25  | 32.72 (8.78) | 52.0          | NA                  |      |
|                              | 12           | 34.42 (9.55) | 66.7          | 5                   |      |
| Carvalho et al., 2012c       | 40           | 32.70 (10.39) | 67.5          | NA                  |      |
|                              | 8            | 42.63 (6.41) | 75.0          | 0.5 mg              |      |
| Duval et al., 2004          | 28           | 15.95 (1.93) | 82.1          | NA                  |      |
|                              | 14           | 16.20 (1.90) | 85.7          | 5                   |      |
| Faravelli et al., 2010       | 93d          | 43.60 (15.40) | 50.2          | NA                  |      |
|                              | 32           | NA         | NA            | 61                  |      |
| Lindley et al., 2004         | 16d          | 40.00 (11.17) | 87.5          | 6.2                 |      |
| Lu et al., 2016a             | 35           | 23.80 (3.89) | 51.4          | 100.0               |      |
|                              | 18           | 23.70 (4.13) | 44.4          | 8                   |      |
| Lu et al., 2016b             | 45           | 21.65 (3.46) | 62.2          | 100.0               |      |
|                              | 23           | 21.50 (3.91) | 60.9          | 3                   |      |
| Meeha et al., 201a           | 63a          | 41.76 (13.12) | 68.3          | 87.5                |      |
| Meeha et al., 2011b          | 32a          | 40.63 (12.85) | 71.9          | 90.5                |      |
| Newport et al., 2004a        | 38           | 30.90 (7.00) | 100.0         | 36.8                |      |
|                              | 19           | 33.30 (6.40) | 56.0          | 11                  |      |
| Newport et al., 2004b        | 26           | 32.98 (7.87) | 100.0         | 19.2                |      |
| Schreuder et al., 2016       | 114          | 26.75 (2.85) | 45.6          | NA                  |      |
|                              | 16           | 27.81 (2.90) | 43.8          | 9                   |      |
| Stein et al., 1997           | 40           | 31.47 (6.70) | 70.0          | NA                  |      |
|                              | 19           | 32.20 (6.70) | 70.0          | 5                   |      |

Note: N = sample size; M = mean; SD = standard deviation; The sex ratio is indicated as percentage of female participants; Ethn. = ethnicity; The ethnicity ratio is indicated as percentage of non-Caucasians; NA = not assessed; Psychopath. = psychopathology, whereby the following definitions have been used: yes = at least some of the participants met diagnostic criteria for a psychiatric disorder, no = none of the participants met diagnostic criteria for any psychiatric disorder, = groups are not matched with respect to psychopathology; MDD = Major Depressive Disorder; BPD = Borderline Personality Disorder; PTSD = Posttraumatic Stress Disorder; CM = child maltreatment; n = sample size; quest. = questionnaire; CTQ = Childhood Trauma Questionnaire; SCID-I = Structured Clinical Interview for DSM-IV Axis I Disorders; CECA = Childhood Experiences of Care and Abuse; self-dev. = self-developed (the authors used a scale developed by themselves); THS = Trauma History Screen; ETI = Early Trauma Inventory; NEMESIS CTI = Netherlands Mental Health Survey and Incidence Study Childhood Trauma Interview; self-ident. = self-identification (participants self-identified as victims of CM); spec. = specification (authors applied a specific definition of CM); EA = emotional abuse; PA = physical abuse; SA = sexual abuse; N = neglect; 5 The grouping of participants was not just based on the experience of CM but also included loss experiences; awak. = awakening; Dex.dose = dose of dexamethasone administered to participants. 6 Timepoints of cortisol sampling. 7 Indices: pre = cortisol assessed prior dexamethasone administration; post = cortisol assessed after dexamethasone administration, delta = post values minus pre values. 8 Comparison between depressed patients with and without early life stress. 9 Comparison between patient groups with and without early trauma. 10 Comparison between PTSD patients with childhood sexual or physical abuse and those reporting no history of childhood sexual or physical abuse. 11 The sample represents a highly traumatized, low-income cohort; data provided separately for those patients all fulfilling diagnostic criteria for PTSD and those patients without PTSD. 12 The data are from the Netherlands Study of Depression and Anxiety, a large cohort study; this article did not appear in the initial search but was suggested by the respective author best as suitable for citation in this meta-analysis. 13 Patient group: dysthymia, MDD, social phobia, panic with/without agoraphobia, generalized anxiety disorder. 14 No one of the participants fulfilled the diagnostic criteria for dysthymia, MDD, social phobia, panic with/without agoraphobia, or generalized anxiety disorder in the past 6 months.
limited number of studies (k < 10), however, no subgroup analyses and meta-regressions were performed. Visual inspection of traditional and counter-enhanced funnel plots as well as Egger’s regression test of funnel plot asymmetry suggested the absence of small-study bias (all counter-enhanced funnel plots as well as Egger’s meta-regressions were performed. Visual inspection of traditional and limited number of studies (k < 10), however, no subgroup analyses and meta-regressions were performed. Visual inspection of traditional and counter-enhanced funnel plots as well as Egger’s regression test of funnel plot asymmetry suggested the absence of small-study bias (all counter-enhanced funnel plots as well as Egger’s meta-regressions were performed. Visual inspection of traditional and limited number of studies (k < 10), however, no subgroup analyses and meta-regressions were performed. Visual inspection of traditional and counter-enhanced funnel plots as well as Egger’s regression test of funnel plot asymmetry suggested the absence of small-study bias (all counter-enhanced funnel plots as well as Egger’s meta-regressions were performed. Visual inspection of traditional and limited number of studies (k < 10), however, no subgroup analyses and meta-regressions were performed. Visual inspection of traditional and counter-enhanced funnel plots as well as Egger’s regression test of funnel plot asymmetry suggested the absence of small-study bias (all counter-enhanced funnel plots as well as Egger’s meta-regressions were performed. Visual inspection of traditional and limited number of studies (k < 10), however, no subgroup analyses and meta-regressions were performed. Visual inspection of traditional and counter-enhanced funnel plots as well as Egger’s regression test of funnel plot asymmetry suggested the absence of small-study bias (all counter-enhanced funnel plots as well as Egger’s meta-regressions were performed. Visual inspection of traditional and limited number of studies (k < 10), however, no subgroup analyses and meta-regressions were performed. Visual inspection of traditional and counter-enhanced funnel plots as well as Egger’s regression test of funnel plot asymmetry suggested the absence of small-study bias (all counter-enhanced funnel plots as well as Egger’s meta-regressions were performed. Visual inspection of traditional and limited number of studies (k < 10), however, no subgroup analyses and meta-regressions were performed. Visual inspection of traditional and counter-enhanced funnel plots as well as Egger’s regression test of funnel plot asymmetry suggested the absence of small-study bias (all counter-enhanced funnel plots as well as Egger’s meta-regressions were performed. Visual inspection of traditional and limited number of studies (k < 10), however, no subgroup analyses and meta-regressions were performed. Visual inspection of traditional and counter-enhanced funnel plots as well as Egger’s regression test of funnel plot asymmetry suggested the absence of small-study bias (all counter-enhanced funnel plots as well as Egger’s meta-regressions were performed. Visual inspection of traditional and limited number of studies (k < 10), however, no subgroup analyses and meta-regressions were performed. Visual inspection of traditional and counter-enhanced funnel plots as well as Egger’s regression test of funnel plot asymmetry suggested the absence of small-study bias (all counter-enhanced funnel plots as well as Egger’s meta-regressions were performed. Visual inspection of traditional and limited number of studies (k < 10), however, no subgroup analyses and meta-regressions were performed. Visual inspection of traditional and counter-enhanced funnel plots as well as Egger’s regression test of funnel plot asymmetry suggested the absence of small-study bias (all counter-enhanced funnel plots as well as Egger’s meta-regressions were performed. Visual inspection of traditional and limited number of studies (k < 10), however, no subgroup analyses and meta-regressions were performed. Visual inspection of traditional and counter-enhanced funnel plots as well as Egger’s regression test of funnel plot asymmetry suggested the absence of small-study bias (all counter-enhanced funnel plots as well as Egger’s meta-regressions were performed. Visual inspection of traditional and limited number of studies (k < 10), however, no subgroup analyses and meta-regressions were performed. Visual inspection of traditional and counter-enhanced funnel plots as well as Egger’s regression test of funnel plot asymmetry suggested the absence of small-study bias (all counter-enhanced funnel plots as well as Egger’s meta-regressions were performed. Visual inspection of traditional and limited number of studies (k < 10), however, no subgroup analyses and meta-regressions were performed. Visual inspection of traditional and counter-enhanced funnel plots as well as Egger’s regression test of funnel plot asymmetry suggested the absence of small-study bias (all counter-enhanced funnel plots as well as Egger’s meta-regressions were performed. Visual inspection of traditional and limited number of studies (k < 10), however, no subgroup analyses and meta-regressions were performed. Visual inspection of traditional and counter-enhanced funnel plots as well as Egger’s regression test of funnel plot asymmetry suggested the absence of small-study bias (all counter-enhanced funnel plots as well as Egger’s meta-regressions were performed. Visual inspection of traditional and limited number of studies (k < 10), however, no subgroup analyses and meta-regressions were performed. Visual inspection of traditional and counter-enhanced funnel plots as well as Egger’s regression test of funnel plot asymmetry suggested the absence of small-study bias (all counter-enhanced funnel plots as well as Egger’s meta-regressions were performed. Visual inspection of traditional and limited number of studies (k < 10), however, no subgroup analyses and meta-regressions were performed. Visual inspection of traditional and counter-enhanced funnel plots as well as Egger’s regression test of funnel plot asymmetry suggested the absence of small-study bias (all counter-enhanced funnel plots as well as Egger’s meta-regressions were performed. Visual inspection of traditional and limited number of studies (k < 10), however, no subgroup analyses and meta-regressions were performed. Visual inspection of traditional and counter-enhanced funnel plots as well as Egger’s regression test of funnel plot asymmetry suggested the absence of small-study bias (all counter-enhanced funnel plots as well as Egger’s meta-regressions were performed. Visual inspection of traditional and limited number of studies (k < 10), however, no subgroup analyses and meta-regressions were performed. Visual inspection of traditional and counter-enhanced funnel plots as well as Egger’s regression test of funnel plot asymmetry suggested the absence of small-study bias (all counter-enhanced funnel plots as well as Egger’s meta-regressions were performed. Visual inspection of traditional and limited number of studies (k < 10), however, no subgroup analyses and meta-regressions were performed. Visual inspection of traditional and counter-enhanced funnel plots as well as Egger’s regression test of funnel plot asymmetry suggested the absence of small-study bias (all counter-enhanced funnel plots as well as Egger’s meta-regressions were performed. Visual inspection of traditional and limited number of studies (k < 10), however, no subgroup analyses and meta-regressions were performed. Visual inspection of traditional and counter-enhanced funnel plots as well as Egger’s regression test of funnel plot asymmetry suggested the absence of small-study bias (all counter-enhanced funnel plots as well as Egger’s meta-regressions were performed. Visual inspection of traditional and limited number of studies (k < 10), however, no subgroup analyses and meta-regressions were performed. Visual inspection of traditional and counter-enhanced funnel plots as well as Egger’s regression test of funnel plot asymmetry suggested the absence of small-study bias (all counter-enhanced funnel plots as well as Egger’s meta-regressions were performed. Visual inspection of traditional and limited number of studies (k < 10), however, no subgroup analyses and meta-regressions were performed. Visual inspection of traditional and counter-enhanced funnel plots as well as Egger’s regression test of funnel plot asymmetry suggested the absence of small-study bias (all counter-enhanced funnel plots as well as Egger’s meta-regressions were performed. Visual inspection of traditional and limited number of studies (k < 10), however, no subgroup analyses and meta-regressions were performed.
Table 9
Summary characteristics of included studies that reported on cortisol assessed in the context of the combined dexamethasone-corticotropin releasing hormone (Dex-CRH) test.

| Study | Total Sample | CM Sample | Control Sample | CM Assessment | Cortisol Assessment | Data |
|-------|--------------|-----------|----------------|---------------|---------------------|------|
|       | N Age M (SD) | Sex       | Ethn. Psychopath. | n Age M (SD) Sex | Method Instrument | Grouping Type Type | Type of sample | Unit | Time points | Indices | Dex-dose CRH-dose | Source | Data (re) grouped |
|       |              |           |                  |               |                      |                  |               |      |             |         |                |        |               |
| Carpenter et al., 2009 | 230 29.23 (10.37) 56.5 NA no | 41 35.90 (11.60) 66.0 | self-report (quest.) | CTQ | cut-offs | EA | blood nmol/l | 2:59 pm – 5 pm (6 samples) | bl, peak | 1.5 mg 11 pm | 100 µg 3 pm | paper, figure* | no | grouped |
| Heim et al., 2008a | 21 31.64 (8.76) 0.0 57.1 yes, MDD | 15 32.30 (8.70) 0.0 | self-report (interview) | ETI | other (spec.) | PA, SA | blood µg/dl | 2 pm – 5 pm (9 samples) | peak, delta, AUCtp, AUCip | 1.5 mg 11 pm | 1 mg µg 3 pm | paper, table | no |
| Heim et al., 2008b | 28 30.25 (8.44) 0.0 53.6 yes*, mixed | 14 31.40 (8.00) 0.0 | self-report (interview) | ETI | other (spec.) | PA, SA | blood µg/dl | 2 pm – 5 pm (9 samples) | peak, delta, AUCtp, AUCip | 1.5 mg 11 pm | 1 mg µg 3 pm | paper, table | no |
| Klaassen et al., 2009 | 22 47.58 (11.70) 100.0 NA no | 10 47.80 (12.10) 100.0 | self-report (quest.) | CTQ b | other (thirds) | EA, PA, SA, N | blood nmol/l | 3 pm – 4:45 pm (7 samples) | all | 1.5 mg 11 pm | 100 µg 3 pm | paper, table | no |
| Lee et al., 2012 | 24 33.42 (9.72) 41.7 41.7 yes*, PD | 10 39.10 (11.37) 40.0 | self-report (quest.) | CTQ | other (thirds) | EA, PA, SA | blood µg/dl | 4 pm – 6 pm (7 samples) | all | 1.5 mg 11 pm | 1 mg µg 4 pm | provided, raw | no |
| Spitzer et al., 2018a | 86 35.50 (11.26) 100.0 NA yes, MDD | 35 35.20 (10.90) 100.0 | self-report (mixed) | CTQ, ETI | other (spec.) | PA, SA, N | saliva nmol/l | 1:30 pm – 4:30 pm (7 samples) | delta* | AUCtp, AUCtp | 1.5 mg 11 pm | 100 µg 3 pm | paper, table | no |
| Spitzer et al., 2018b | 58 34.29 (11.85) 100.0 NA no | 21 34.80 (10.90) 100.0 | self-report (mixed) | CTQ, ETI | other (spec.) | PA, SA | saliva nmol/l | 1:30 pm – 4:30 pm (7 samples) | delta* | AUCtp, AUCtp | 1.5 mg 11 pm | 100 µg 3 pm | paper, table | no |
| Watson et al., 2007 | 40f 48.63 (7.34) 47.5 NA yes, mixed | 25 49.12 (6.83) 52.0 | self-report (quest.) | CTQ | cut-offs | EA, PA, SA | blood nmol/l | 3 pm – 6 pm (9 samples) | all | 1.5 mg 11 pm | 100 µg 3 pm | provided, raw | yes |
| De Bellis et al., 1994* | 26 11.65 (2.74) 100.0 42.3 yes*, Dysthymia | 13 11.20 (2.60) 100.0 | informant (record) | CPS | other (record) | N | blood nmol/l | 6 pm – 10 pm (13 samples) | AUCtp, AUCtp | 100 µg 3 pm | no dex adm. | paper, text | no |
| Kaufman et al., 1997* | 26g 9.75 (1.16) 53.8 34.6 yes, MDD | 13 9.60 (1.40) 53.9 | mixed | PSS, medical records | other (not spec.) | PA, SA | blood µg/dl | 5 pm – 8 pm (9 samples) | bl, peak, AUCtp | 1 µg kg 3 pm | provided, raw | no |

**Note.** N = sample size; M = mean; SD = standard deviation; The sex ratio is indicated as percentage of female participants; Ethn. = ethnicity; The ethnicity ratio is indicated as percentage of non-Caucasians; NA = not assessed; Psychopath. = psychopathology, whereby the following definitions have been used: yes = at least some of the participants met diagnostic criteria for a psychiatric disorder, no = none of the participants met diagnostic criteria for any psychiatric disorder, * = groups are not matched with respect to psychopathology; MDD = Major Depressive Disorder; PD = Personality Disorder; CM = child maltreatment; n = sample size; quest. = questionnaire; CTQ = Childhood Trauma Questionnaire; ETI = Early Trauma Inventory; CPS = Child Protective Services; PSS = Psychosocial Schedule for School Aged Children; spec. = specification (authors applied a specific definition of CM); EA = emotional abuse; PA = physical abuse; SA = sexual abuse; N = neglect; bl = baseline (if possible, sample just prior CRH injection was extracted); AUCip = area under the curve with respect to ground; AUCtp = area under the curve with respect to increase. Dex-dose = dose of dexamethasone administered to participants; CRH-dose = dose of corticotropin-releasing hormone administered to participants.

*All extracted values are adjusted for age, gender, and effects of four other maltreatment subtypes. Authors also administered ETI, but grouping was based on CTQ. The low CTQ PD group and the normal control group were combined and compared to the high CTQ PD group. Sample assessed at ~150 min prior CRH injection was not included in analyses (therefore 7 instead of 8 samples; baseline sample at 4 pm). Delta defined as difference between maximum value after CRH injection and the mean of the three baseline measures. The data were regrouped, taking into account only the patient groups. No dexamethasone was administrated (CRH test only). Comparison between the patient groups with and without abuse. 5% were subjected to ongoing EA. Baseline defined as the mean of the three pre CRH infusion samples. It is not clearly stated in the text or in the subheading of the figure whether means and standard deviations or means and standard errors were presented; we assumed standard errors.
ranging between 12.0 and 42.3% (M = 27.6%, SD = 15.2%). With respect to psychopathology, most of the subjects in the child maltreatment group met the criteria for a mental disorder, while the subjects in the control group were mainly healthy controls (n = 3; n = 1 did not report on psychopathology). However, in two of the studies included, the authors were able to demonstrate that the presence of the specific mental disorder did not affect the 24-hour UFC data. Three of the four studies focused exclusively on sexual abuse experiences without collecting information about other types of child maltreatment, and two studies recruited participants solely on the basis of self-identification without using any established measurement to assess child maltreatment. All data were extracted from the respective articles. See Table 12 for further details.

### 3.2.5.2.2. Risk of bias assessment

The risk of bias assessment of the primary studies as well as the average risk of bias scores can be found in Appendix B Table B14. The detailed description of the quality items can be found in the Table B7. On average, a total score of 61.5/100.0 was received (SD = 13.5, range: 42.3–73.1; selection of participants: M = 65.6/100.0, SD = 12.0; appropriate measure of 24-hour UFC: M = 50.0/100.0, SD = 18.4; adequate controlling of confounders: M = 66.5/100.0, SD = 20.7). In all n = 4 studies participants were matched with respect to age and gender. In addition, all four articles provide detailed information on how UFC samples were collected, stored and analyzed and all studies give a relatively good overview of participants’ medication use. The n = 3 studies that focused on sexual abuse did not provide information about other maltreatment experiences, reducing the quality of the grouping into a child maltreatment and a clear control group without any types of child maltreatment experiences. In only one of the four studies it was ensured that the participants did not experience any ongoing significant life stressors, assessed UFC over at least three days, provided batch analysis information, and none of the studies ensured that the participants in the two groups did not differ with respect to other traumatic experiences in childhood or adulthood.

### 3.2.5.2.3. Meta-analysis

Pooling the results of the n = 4 studies (n = 108), the aggregate effect size was Hedges’ g = 0.07, 95% CI [−0.83; 0.98], p = 0.874, suggesting no overall difference in 24-hour UFC in the child maltreatment sample compared to the control sample (the corresponding forest plot is shown in Appendix C Fig. 5.2). There was significant, high heterogeneity in the effect size estimates (Qk = 14.79, p = 0.002, I² = 79.7%), indicating high inconsistencies between studies. No outlier study was detected, but the study from Lemieux et al., (2008) had a high influence on the overall result. Traditional and contour-enhanced funnel plots are also shown in Appendix C Fig. 5.2 and visual inspection of them suggested the absence of small-study bias as did the Egger’s regression test of funnel plot asymmetry (intercept = −6.898, p = 0.428; but attention k < 10). The sensitivity analysis substantially reduced heterogeneity (I² = 79.7% to I² = 8.5%) and yielded a medium significant overall effect (n = 83, Hedges’ g = 0.56, 95% CI [0.11; 1.00], p = 0.014) with participants in the child maltreatment group showing higher 24-hour UFC concentrations compared to the control sample. However, considering the small sample size and the large prediction interval (−2.33; 3.45), it is unclear what the results of future studies will show. In addition, it should be noted that the study excluded in the context of the sensitivity analysis (Lemieux et al., 2008) received the highest average quality score.

### 4. Discussion

This series of meta-analyses, based on a systematic review of the literature, examined the existing evidence on the association between child maltreatment and cortisol metabolism including various measures of HPA axis activity. Measures of interest ranged from cortisol assessed in the context of the circadian rhythm (DC) to cortisol assessed in response to awakening (CAR), in response to the perception of a stressor (cortisol stress reactivity) and pharmacological challenges (DST, Dex-CRH test, CRH test), to cumulative measures of cortisol secretion, namely 24-hour UFC and HCC.

#### 4.1. Main findings

Consistent with the findings of two previous meta-analyses (Bernard et al., 2017; Fogelman and Canli, 2018) we did not find overall differences in any of the indices related to cortisol secretion in the context of circadian activity (with the exception of evening cortisol) as well as in response to awakening (CAR) when comparing individuals with child maltreatment to those without corresponding experiences. The finding of slightly increased evening cortisol in individuals exposed to child maltreatment was mainly driven by a few studies in which the child maltreatment group also included individuals with loss experiences and thus should be interpreted with caution in the context of this meta-analysis. Individuals with a history of child maltreatment, however, appear to show a blunted cortisol stress response. Though not yet evident before being introduced to a corresponding stressor (baseline cortisol), blunting was seen in indices reflecting total cortisol production (peak, recovery cortisol) as well as in indices expressing changes in cortisol over time (delta, AUC cortisol) following the perception of a stressor. These findings are consistent with the results of a previous meta-analysis examining the effect of ELA on cortisol response to social stress (Bunea et al., 2017), albeit with somewhat smaller effects observed in our meta-analysis. Interestingly, this blunting was not observed in studies where CRH injections (Dex-CRH test) were used to initiate the secretion of cortisol. However, the number of studies on the
Table 11
Summary characteristics of included studies that reported on hair cortisol concentrations (HCC).

| Study                      | Total Sample | CM Sample | Control Sample | CM Assessment | Cortisol Assessment | Data |
|----------------------------|--------------|-----------|----------------|---------------|---------------------|------|
|                            | N    | Age M (SD) | Sex | Ethn. | Psychopath. | n   | Age M (SD) | Sex | Ethn. | Psychopath. | Method | Instrument | Grouping | Type | Unit | Source | Data (re) grouped |
| do Prado et al., 2017      | 57   | 15.39 (1.81) | 57.9 | NA   | no         | 30  | 16.47 (1.25) | 50.0 |        |             | self-report (quest.) | CTQ   | other (percentile) | EA, PA, SA, N | pg/mg | paper, figure | no |
| Fischer et al., 2017       | 135  | 50.31 (14.73) | 71.9 | 9.0   | NA         | 43  | 53.10 (13.90) | 81.0 |        |             | self-report (quest.) | CTQ   | cut-offs | EA, PA, SA, N | pg/mg | provided, group | yes |
| Groer et al., 2016         | 81a  | 46.17 (10.53) | 100.0 | 45.7  | NA         | 27  | 47.20 (10.70) | 100.0 |        |             | self-report (quest.) | self-dev. | other (spec.) | EA, PA, SA, N | mg/mg | paper, figure | no |
| Morris et al., 2017        | 22b  | 24.71 (3.38)  | 100.0 | 18.2  | 7          | 54  | 45.60 (10.50) | 100.0 |        |             | self-report (quest.) | self-report (quest.) | CTQ   | cut-offs | EA, PA, SA, N | pg/mg | provided, group | no |
| Reichl et al., 2016a       | 24   | 16.27 (1.10)  | 91.7  | NA    | no         | 3   | 16.55 (1.34)  | 100.0 |        |             | self-report | CECA cut-offs | EA, PA, SA, N | pg/mg | provided, raw | yes |
| Reichl et al., 2016b       | 25   | 16.27 (1.30)  | 96.0  | NA    | yes, mixed | 17  | 16.56 (1.30)  | 100.0 |        |             | self-report | CECA cut-offs | EA, PA, SA, N | pg/mg | provided, raw | yes |
| Schalinski et al., 2015     | 39c  | 34.95 (10.74) | 100.0 | 87.2  | yes, mixed | 17  | 33.70 (11.70) | 100.0 |        |             | self-report (interview) | self-report | CTQ   | cut-offs | EA, PA, SA, N | pg/mg | provided, group | yes |
| Studt et al., 2013          | 58d  | 39.16 (13.40) | 91.4  | 0     | no         | 10  | 42.20 (11.22) | 100.0 |        |             | self-report | ETI other (clustering)* | cut-offs | EA, PA, SA, N | pg/mg | provided, group | yes |
| White et al., 2017          | 537  | 9.98 (3.13)   | 50.7  | 2.2   | NA         | 245 | 9.86 (3.24)   | 46.1  |        |             | informant (mixed) | MCS (CPS), MMCI | other (record & spec.) | EA, PA, SA, N | pg/mg | provided, group | no |

Note. N = sample size; M = mean; SD = standard deviation; The sex ratio is indicated as percentage of female participants; Ethn. = ethnicity; The ethnicity ratio is indicated as percentage of non-Caucasians; NA = not assessed; Psychopath. = psychopathology, whereby the following definitions have been used: yes = at least some of the participants met diagnostic criteria for a psychiatric disorder, no = none of the participants met diagnostic criteria for any psychiatric disorder; CM = child maltreatment; n = sample size; quest. = questionnaire; CTQ = Childhood Trauma Questionnaire; self-dev. = self-developed (the authors used a scale developed by themselves); CECA = Childhood Experiences of Care and Abuse; ETI = Early Trauma Inventory; MCS = Maltreatment Classification System; CPS = Child Protective Services; MMCI = Maternal Maltreatment Classification Interview; spec. = specification (authors applied a specific definition of CM); EA = emotional abuse; PA = physical abuse; SA = sexual abuse; N = neglect. a Percentage of non-Caucasians refers to total sample (N = 139). b Three outliers in hair cortisol were removed, however it is unclear to which group this applied. c Comparison between women with abuse/neglect without recent interpersonal violence exposure and non-trauma controls. d The data were regrouped, taking into account only the patient group with stress-related disorders. e Grouping was based on k-means clustering method based on ETI sum score excluding the general trauma subscale (remark: the sample represents a refugee’s sample with all having experienced some type of trauma during their life). f The data were regrouped, taking into account only the traumatized control subjects and the non-traumatized control subjects (remark: some in the control sample may have experienced other traumatic events); data provided for a total sample of N = 58 instead of N = 53 as presented in the paper.
### Table 12

| Study | Total Sample | CM Sample | Control Sample | CM Assessment | Cortisol Assessment | Data (no.) | Unit | Source | Notes |
|-------|--------------|-----------|----------------|---------------|---------------------|------------|------|--------|-------|
|       | n | Age (SD) | M | Sex | Ethnicity | Method | Instrument | Type | Grouping | Type | Data (re) |
| a | 26 | 11.65 (2.74) | 13 | 11.20 (2.60) | 12.10 (2.90) | CPS | other | record | no | paper | no |
| b | 44.4 | 10.46 (1.22) | 4 | 10.50 (1.40) | 10.30 (1.10) | PA, µg/day | mixed | interview | no | paper | no |
| c | 17 | 100.0 | NA | 8 | NA | interview | no instrument | self-report | no | paper | no |
| d | 17 | 31.30 (6.13) | 13 | 31.20 (6.00) | 31.30 (6.00) | SA, µg/day | adj. conc. creatinine | no instrument | no | paper | no |
| e | 12 | 12.0 | yes* | MDD | 100.0 | 12 | 31.0 | self-report | no instrument | no | paper | no |

Note. N = sample size; M = mean; SD = standard deviation; CM = child maltreatment; closest possible distance to the original. No significant difference between women with and without a history of childhood sexual abuse on outcomes related to cortisol stress reactivity.

---

**4.2. Between-study heterogeneity and the influence of moderators**

Although no overall differences in cortisol secretion were found for the majority of the HPA axis activity measures except for cortisol assessed in response to a stressor across studies, we generally observed a significant degree of variability in the effect estimates between studies (especially for indices reflecting cortisol secretion in the context of awakening, after a stressor and following the Dex-CRH test), suggesting the likely influence of additional variables in moderating the effect of child maltreatment on cortisol regulation. Before discussing some of the moderators that systematically accounted for between-study heterogeneity, holding true for various of the HPA axis activity measures, it should be kept in mind that the majority of studies were conducted with predominantly young, female adults who belonged to an ethnicity and gender group and in whom child maltreatment experiences were assessed mainly through self-reports. In addition, a considerable number of studies did not report on psychopathology, and studies involving clinical samples were fairly heterogeneous in terms of the predominant mental disorder (e.g., MDD versus posttraumatic stress disorder (PTSD)). Accordingly, our ability to find important moderators or relevant subgroup differences might have been limited.

**4.2.1. Influence of participant related characteristics**

One of the moderators that explained some of the between-study heterogeneity in effect sizes (in line with findings from Bunea et al., 2017; and Zorn et al., 2017) was the proportion of females in the respective sample, with a higher proportion being associated with a stronger blunting of cortisol secretion (CAR, cortisol stress reactivity, and DST). Corresponding sex differences, particularly with respect to stress reactivity, have been repeatedly reported, with men showing higher cortisol levels to psychosocial stress than women (Liu et al., 2017). Factors that influence corticosteroid binding globulin (CBG) levels and thus the level of free cortisol appear to account for some of these gender effects including the use of oral contraceptives and the production of sex steroids throughout the menstrual cycle (e.g., Foley and Kirschbaum, 2010). Both the intake of oral contraceptives and the assessment of the menstrual cycle were not adequately evaluated in many of the studies included and therefore matching of the two groups in these respects was not properly controlled. Apart from the proportion of women in the respective sample, there was little evidence that the remaining participant related characteristics such as age, ethnicity, and participant diagnosis accounted for variability in the effect estimates among primary studies. Interestingly, even though psychopathology did not account for heterogeneity in the child maltreatment cortisol relationship (at least for the majority of outcome indices), a tendency for stronger blunting in clinical samples compared to healthy controls was observed for indices related to cortisol stress reactivity. Nevertheless, an attenuation of the cortisol stress response was observed in participants.
with child maltreatment experiences that at the time of measurement did not report a mental disorder, suggesting that alterations in HPA axis activity may be present prior to the development of mental health issues or independent of psychopathology.

4.2.2. Influence of trauma related information

Another fairly consistent moderator accounting for some of the between-study heterogeneity (DC and CAR, tendency for cortisol stress reactivity) was whether or not data were (re)grouped for the purpose of this meta-analysis, with (re)grouped data showing a tendency towards smaller effects. In some of the studies that provided (re)grouped data, grouping of participants into a child maltreatment and a control group was based on relatively low severity thresholds (particularly for cortisol assessed in the context of awakening and circadian activity), which might account for this finding. Indeed, and in agreement with the observed dose-dependent relationship between child maltreatment and health impairments (e.g., Clemens et al., 2018; Norman et al., 2012), the severity of child maltreatment experiences, though difficult to assess (Jackson et al., 2019), might actually be of particular importance in explaining variability between the association of child maltreatment and HPA axis functioning. Interestingly, several studies, especially those which assessed cortisol in response to a stressor, found a stronger blunting in cortisol secretion following the perception of a stressor with an increase in the severity of child maltreatment (Lovallo et al., 2019; Ouellet-Morin et al., 2018; Trickett et al., 2014; Voellmin et al., 2015).

Unfortunately, our group comparison approach did not allow us to investigate this association systematically. In line with the difficulties in defining child maltreatment and the possibility to rely on various assessment modalities (Cicchetti and Toth, 2005; Manly, 2005), studies generally differed widely in their child maltreatment assessment and grouping approaches. For instance, studies that relied on established self-reports such as the CTQ or the Childhood Experience of Care and Abuse interview (CECA) grouped their participants based on validated cut-off scores, while other studies applied specific definitions (e.g., Heim et al., 2000: “repeated abuse, once a month or more for at least 1 year”), sometimes based on self-developed assessment tools (e.g., Groër et al., 2016; Martinson et al., 2016; Smeets et al., 2007) and still others relied on the presence or absence of a specific record, such as a CPS record (e.g., Bernard et al., 2010; Cicchetti et al., 2010; Hibel et al., 2019). However, neither the assessment (self-report, informant report, mixed) nor the grouping method (cut-offs, other, record) explained variance in the effect estimates of any of the HPA axis activity outcome indices, with the exception of larger effect sizes found in studies focusing on informant reports compared to self-reports for waking/morning cortisol (which is consistent with findings from Bernard et al., 2017). Since there were far fewer studies using informant reports as opposed to self-reports (a pattern also found among studies on the prevalence of child maltreatment; Stoltenborgh et al., 2015), comparison between these approaches might have been inappropriate and the chances of detecting differences accordingly low. Importantly, some of the included studies failed to sufficiently ensure that control participants were not subjected to any type of child maltreatment. The studies that performed poorly in terms of ensuring maltreatment did not take place in the control group were those that relied on specific records, such as CPS records, and studies focusing on one particular type of child maltreatment. In corresponding studies - besides the absence of a record or the corresponding type of maltreatment - no other measures were applied to ensure that control participants had not experienced child maltreatment. This may have influenced our results, since child maltreatment experiences are rather common in the general population, as demonstrated by epidemiological studies (Witt et al., 2017). Neither the role of age at maltreatment onset nor the chronicity of the maltreatment experiences on HPA axis activity could be investigated, as very few studies applied measures that assessed these two factors in the first place - a finding in line with a recent review summarizing research on the operationalization of child maltreatment over the last 10 years (Jackson et al., 2019).

4.2.3. Influence of cortisol related information

In our meta-analysis, neither sample type (blood, saliva), type of stressor (social-evaluative, other), slope type (wake-to-bed, other), whether cortisol was assessed in reference to awakening or not, nor dose of dexamethasone (0.5 mg, 1.0 mg) significantly explained variability in the various effect estimates. The only moderator related to the assessment of cortisol that explained between-study heterogeneity was whether a cortisol response was observed in both, in only one, or in none of the groups following the perception of a stressor, with stronger effects found in those studies that observed a cortisol response in just one of the two groups. This finding might be attributed to the fact that six out of the seven comparisons that observed a response only in one of the two groups, reported an increase in cortisol in the control group only. Overall, there was no evidence to suggest that the two components of cortisol secretion - total cortisol production and change in cortisol over time - which appear to capture different aspects of HPA axis activity (Khoury et al., 2015), are affected differently by child maltreatment. A blunting in cortisol secretion following the perception of a stressor was found in indices reflecting both total and change in cortisol over time. However, as expected, studies generally differed widely with respect to the index or indices reported (e.g., much more studies reported on peak cortisol compared to AUCl cortisol), making comparisons between the different outcome indices difficult. Along with this, depending on the studies included, different moderators emerged, which in turn complicated the interpretation of the corresponding findings.

4.2.4. Influence of several components of methodological quality

Finally, we did not observe a consistent association between study quality as assessed by our self-developed quality assessment tool (which was based on existing recommendations and guidelines) and reported effect sizes. Although, at least for some outcome indices, the quality of the individual studies seemed to explain some of the between-study heterogeneity. For instance, a tendency towards smaller negative differences in cortisol secretion following the perception of a stressor was observed in studies with a higher study quality and thus a lower risk of bias. Importantly, in studies that (re)grouped participants for the purpose of this meta-analysis - although the majority of these studies used an established instrument (an established source of information) to assess child maltreatment experiences - by (re)grouping their study participants, several other aspects of methodological quality could no longer be assessed. Therefore, our conservative approach of not awarding any points in a given case may have induced biases. In addition, two studies were able to achieve the same average total score, but scored in completely different quality items. Since we were not able to value the importance of the various quality items, our ability to find associations with the different effect estimates might thus have been limited. Nevertheless, it should be mentioned that several quality-related aspects were generally well implemented in the majority of the studies (holding true for all HPA axis activity measures), whereas others were insufficiently addressed and controlled in most of the included publications. With respect to the selection of participants, for instance, the two comparison groups were generally matched in terms of sex and age, and most studies used an established measurement to assess child maltreatment. While exposure and the control groups were generally balanced in terms of psychopathology in those studies that reported on psychopathology, a substantial number of studies did not evaluate the presence of mental disorders at all, and therefore information about matching in this respect was unavailable for several studies. This is particularly surprising as, on the one hand, child maltreatment experiences are much more common among individuals with a mental disorder and, on the other hand, psychopathology itself has been repeatedly associated with changes in various HPA axis activity measures, with sometimes opposing findings for different mental disorders (e.g., Adam et al., 2017; Chida and Steptoe, 2009; Leistner and Menke, 2018; Stalder et al., 2017; Zorn et al., 2017). Accordingly, in these studies, the presence of a possible confounding effect of psychopathology cannot be
Table 13
Summary table of main findings from meta-analyses, meta-regression, and subgroup analyses (mixed/fixed-effects models) for the various HPA axis activity measures and related outcome indices.

| Measure                    | Meta-analyses                       | Moderator analyses                      |
|----------------------------|-------------------------------------|----------------------------------------|
|                            | k     | N     | Random-effects models | Between-study heterogeneity (I²) | Significant moderators (meta-regression) | Significant comparisons (subgroup analyses) |
| DC                         | 26    | 5248  | no overall significant effect | low-moderate (29.1%) | tendency for higher morning cortisol in older-aged samples; tendency for lower morning cortisol with higher study quality (in terms of appropriate measure of DC) | reduced morning cortisol in studies (1) that used informant reports, (2) that applied CPS records or cut-offs, (3) where original data were extracted |
| DC: Morning cortisol       | 26    | 5212  | no overall significant effect | low (2.0%) | – | larger positive effect size estimates in studies (1) focusing on other types of ELA; (2) including original data |
| DC: Evening cortisol       | 26    | 5188  | small overall positive effect | low (2.0%) | – | – |
| DSL cortisol               | 17    | 3390  | no overall significant effect | low (23.5%) | – | – |
| CAR: Awakening cortisol    | 27    | 3545  | no overall significant effect | low (17.3%) | – | reduced awakening cortisol in studies where original data were extracted |
| CAR: Peak cortisol         | 22    | 3215  | no overall significant effect | moderate-high (59.8%) | – | – |
| CAR: 60 min post awakening cortisol | 14 | 2276  | no overall significant effect | low-moderate (41.6%) | – | – |
| CAR: Delta cortisol        | 18    | 2536  | no overall significant effect | low-moderate (41.8%) | tendency for lower delta cortisol in studies with higher proportion of females | – |
| CAR: AUCₖ cortisol         | 15    | 2327  | no overall significant effect | moderate-high (73.1%) | tendency for lower AUCₖ cortisol with higher study quality (in terms of appropriate measure of confounders) | reduced AUCₖ cortisol where original data were extracted |
| CAR: AUCᵢ cortisol         | 9     | 1654  | no overall significant effect | moderate-high (70.3%) | k < 10 (no meta-regression performed) | k < 10 (no subgroup analyses performed) |
| Cortisol stress reactivity |                         |                                        |                                        |                                        |                                        |
| Baseline cortisol          | 39    | 4284  | no overall significant effect | moderate-high (69.7%) | no meta-regression performed | no subgroup analyses performed |
| Peak cortisol              | 35    | 3895  | no overall significant effect | high (78.3%) | – | – |
| Recovery cortisol          | 29    | 3407  | small overall negative effect | moderate-high (73.4%) | tendency for lower recovery cortisol in studies with higher proportion of females | – |
| Delta cortisol             | 24    | 2678  | small overall negative effect | low-moderate (46.1%) | tendency for lower delta cortisol in studies (1) with higher proportion of females; (2) with a higher proportion of participants fulfilling diagnostic criteria for a mental disorder | stronger negative effect estimates in studies (1) that included clinical samples, (2) involved participants with other types of ELA, (3) that observed a cortisol response in only one of the two comparison groups |
| AUCₖ cortisol              | 20    | 1614  | no overall significant effect | moderate-high (62.5%) | – | – |
| AUCᵢ cortisol              | 23    | 1992  | small overall negative effect | low-moderate (31.4%) | tendency for higher AUCᵢ cortisol in studies with higher study quality (in terms of selection of participants) | stronger negative effect estimates in studies that observed a cortisol response in only one of the two comparison groups |
| DST                        | 17    | 2222  | no overall significant effect | low-moderate (40.5%) | no meta-regression performed | no subgroup analyses performed |
| Pre-DST cortisol           | 16    | 2201  | no overall significant effect | low-moderate (40.5%) | – | – |
| Post-DST cortisol          | 17    | 2214  | no overall significant effect | low-moderate (44.1%) | tendency for lower post-DST cortisol in studies with higher proportion of females | increased post-DST cortisol in studies focusing on other types of ELA |
| Delta                      | 9     | 1888  | no overall significant effect | low-moderate (34.0%) | k < 10 (no meta-regression performed) | k < 10 (no subgroup analyses performed) |
| Dex-CRH test               | 10    | 561   | low (0.0%) | k < 10 (no meta-regression performed) | k < 10 (no subgroup analyses performed) | (continued on next page)
related to participant selection, and as already indicated in the context of the assessment of child maltreatment, several studies inadequately ensured that none of the control participants were exposed to any type of child maltreatment, and only a handful of studies used two different sources to evaluate the presence of child maltreatment. Numerous studies collected, stored, and analyzed samples, provided details on their test procedures, and thus ensured that none of the control participants were exposed to any type of child maltreatment, and only a handful of studies used two different sources to evaluate the presence of child maltreatment. Several studies inadequately ensured that none of the control participants were exposed to any type of child maltreatment, and only a handful of studies used two different sources to evaluate the presence of child maltreatment. Researchers assessed cortisol levels in children and adults, and thus ensured that none of the control participants were exposed to any type of child maltreatment, and only a handful of studies used two different sources to evaluate the presence of child maltreatment. Researchers assessed cortisol levels in children and adults, and thus ensured that none of the control participants were exposed to any type of child maltreatment, and only a handful of studies used two different sources to evaluate the presence of child maltreatment.
disease risk (e.g., Dube et al., 2001). Nevertheless, ELA and especially the experience of multiple adverse childhood experiences have been related to various health conditions later in life (e.g., Clark et al., 2010; Danese et al., 2009; Hughes et al., 2017). In addition, the meta-analysis conducted by Bunea et al. (2017), although slightly smaller effects were observed compared to studies focusing on child maltreatment only, showed that ELA was similarly associated with a blunted cortisol stress response. Thus, considering that experiencing adversity during childhood is rather the rule than the exception (Merrick et al., 2019), for the vast majority of studies, it cannot be ruled out that control participants have experienced other forms of ELA, which in turn may have influenced the results of this systematic review. Considering these methodological shortcomings, including the limitations associated with this series of meta-analyses, our ability to establish a consistent link between the experience of child maltreatment and HPA axis functioning may indeed be compromised. Finally, it should be noted that we decided to focus on peer-reviewed papers only to allow for a transparent and replicable search of the literature. Appropriate statistical methods (e.g., funnel plots and Egger’s regression tests) were applied to evaluate and control for publication bias. Nevertheless, the inclusion of grey literature could have counteracted the problem of including data that are not fully representative of the evidence as a whole.

4.3. Interpretation of the findings in the context of developmental programming of the HPA axis

Nevertheless, taking the above constraints into account, we found evidence of an altered cortisol stress response in individuals exposed to child maltreatment as compared to control participants. The null findings with regard to the other HPA axis activity measures (keeping in mind the various methodological shortcomings as one potential explanation) could also indicate that alterations causing aberrant cortisol secretion are less apparent at the level of the pituitary or adrenal glands, but are rather expressed in brain regions involved in stress processing (e.g., limbic brain areas including the hippocampus, the amygdala, and the prefrontal cortex) and in the connectivity of these brain regions to the hypothalamus (Herman et al., 2003). In line with this idea, a review summarizing findings on the neuronal control of chronic stress adaptation, suggests that changes in HPA axis regulation following severe stress exposure might be traced back to long-term changes in the limbic input to neurons controlling stress responsiveness (Herman, 2013). Additionally, it is well known that limbic brain areas including the hippocampus, the amygdala and the prefrontal cortex widely express GRs, and therefore it is not surprising that acute and chronic stress appear to significantly affect synaptic physiology and connectivity in these regions (e.g., Myers et al., 2014). In contrast, structures involved primarily in the regulation of cortisol release in the context of circadian signals or following awakening (i.e., the suprachiasmatic nucleus, Spiga et al., 2014) might be less affected. Thus, corresponding alterations in HPA axis activity measures that are not primarily activated by stress perception (e.g., DC, CAR, HCC, UFC) might only become apparent when cortisol is measured during periods of high life stress - when stress processing actually becomes relevant for these activity measures as well. Interestingly, findings of a longitudinal study evaluating stress exposure across the lifespan on HPA axis functioning at age 37 provide some support for this assumption (Young et al., 2021). In this study, in accordance with the theory of developmental programming of biological systems - the biological embedding model (e.g., Heim et al., 2019; Heindel et al., 2015) - individuals with adversities experienced during early or middle childhood showed a blunted cortisol response to a modified version of the TSST. This blunting of cortisol secretion following the perception of this stressor was independent of whether or not participants were experiencing current life stress. Additionally, similar cortisol stress response patterns were seen in participants with high and low cumulative stress, if these cumulative stress exposures did not involve early life stress (Young et al., 2021). These findings thus support the notion that when attempting to explain differences in the cortisol stress reactivity, it is not so much stress in general, but early childhood stress in particular that seems to be critical (supporting the biological embedding model). In contrast, flatter DSL profiles were only observed in those individuals who experienced ELA and were currently subjected to high levels of stress (Young et al., 2019). While these DSL results remain consistent with the biological embedding model, they also provide support for the assumption that alterations causing aberrant cortisol secretion likely relate to circuits of the brain involved in the processing of stress, and accordingly, meaningful differences in HPA axis activity measures that in terms of their activation do not per se require the experience of stress, are only to be expected when stress processing actually is involved (i.e., under high current life stress; see also Kuhlman et al., 2016).

4.4. Conclusion and future directions

Taking into account all the findings and difficulties in the context of this series of meta-analyses, including the unbalanced recruitment of study participants in the primary studies (e.g., predominantly young, female adults who belonged to an ethnic majority group and in whom child maltreatment experiences were assessed mainly through self-reports), the considerable number of studies that did not report on psychopathology, the limitations related to the assessment of child maltreatment (i.e., the use of various definitions and our inability to investigate the role of age at onset and the chronicity of the maltreatment experiences), and the various constraints related to the assessment of the various HPA axis activity measures (i.e., the inadequate control of state factors and confounding variables and limitations related to the reliability of the cortisol outcome measures), it becomes apparent that, on the one hand, a comprehensive conclusion about the functioning of the HPA axis in individuals who have been exposed to child maltreatment cannot be drawn at this time point, and on the other hand, our ability to find important moderators or relevant subgroup differences might have been limited. Nevertheless, child maltreatment appears to be associated with a blunted rather than an exaggerated activity when considering cortisol secretion following the perception of a stressor (while a tendency was also shown for HCC), and several moderators including the proportion of females in the sample, psychopathology, and the study quality (to name a few) have been identified to account for some of the observed between-study heterogeneity. Considering that cortisol, when secreted in excess (e.g., during prolonged stress exposure like it is the case for child maltreatment), can have a variety of deleterious effects (Feelders et al., 2012), particularly in the brain (Sapolsky, 1999; Sapolsky et al., 2000), a corresponding downregulation may indeed serve as an adaptive function protecting the body from these various adverse effects, an idea that has been subsumed under the so-called “attenuation hypothesis” (e.g., Kaess et al., 2018; Trickett et al., 2010). While probably adaptive in the first place, there is growing evidence linking not only an exaggerated but increasingly also a blunted cortisol stress response to various adverse behavioral and health outcomes (Carroll et al., 2017; de Rooij, 2013; Turner et al., 2020). Cortisol has various important anti-inflammatory and immune suppressive functions (Sapolsky et al., 2000) and several studies have shown that an attenuated cortisol stress reactivity (irrespective of cause) is associated with a stronger proinflammatory immune response (Buske-Kirschbaum et al., 2010; Janusek et al., 2017; Schwaiger et al., 2016). Interestingly, a growing number of studies suggest that inflammatory processes may precede the onset of, or be involved in the development of various types of mental disorders (Kivimäki et al., 2014; Molhem et al., 2017; Slavich et al., 2020). Accordingly, future studies should not only pay more attention to the potential moderating influence of current life stress, especially if interested in HPA axis activity measures that are not primarily regulated by stress perception alone, but, if interested in the consequences arising from an altered HPA axis activity, studies specifically should examine how an alerted cortisol secretion might be related
to dysfunctions in other biological systems. In addition, by investigating the potentially moderating role of genes and epigenetic changes, knowledge of which individuals are most susceptible to the long-term consequences of child maltreatment (or ELA in general) may be further enhanced (e.g., Heim et al., 2019). However, reliable and reproducible results are only to be obtained if future studies more consistently rely on measurement tools that capture the assessment of various types of ELA, their onset, their chronicity and, in particular, these tools should permit the assessment of the perceived severity of the corresponding experiences. Related to a growing number of studies showing different neurobiological consequences of deprivation and threat experiences (e.g., Colich et al., 2020), a more fine-grained analysis of child maltreatment or adversity in general could further improve our understanding of the functioning of the HPA axis in individuals exposed to corresponding experiences. However, in order to obtain reliable and valid HPA axis activity measures, future studies must focus more consistently on cortisol assessment guidelines, which provide important information regarding various state and confounding variables, as well as information on the reliability of the corresponding outcome activity measures (e.g., Adam and Kumari, 2009; Allen et al., 2017; Foley and Kirschbaum, 2010; Kudielka et al., 2012; Stalder et al., 2016; Stalder and Kirschbaum, 2012; Zänkert et al., 2019).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We would like to thank all authors and associated research groups who responded to our requests, who suggested further studies for inclusion in our meta-analysis, who sent us additional information on their studies to clarify upon inclusion, and, in particular, we would like to thank all study authors who provided additional data for the purpose of this meta-analysis (alphabetically by last name): Alexander Nina, Ali nor Daryl B., Otte Christian, Ouellet-Morin Isabelle, Peckins Melissa K., Matthew C., Negriff Sonya, Nelson Elizabeth, Noorden Martijn, Oews Herbert L., Mehta Divya, Melhem Nadine M., Mello Andrea F., Kuzminskaite Erika, Lee Royce J., Linares Oriana, Lindley Steven E., Adam, E.K., Quinn, M.E., Tavernier, R., McQuillan, M.T., Dahlke, K.A., Gilbert, K.E., 2017. Diurnal cortisol slopes and mental and physical health outcomes: A systematic review and meta-analysis. Psychoneuroendocrinology 83, 25–41. https://doi.org/10.1016/j.psyneuen.2017.05.018.

References

References with an asterisk (*) were included in this series of meta-analyses

Adan, E.K., Kumari, M., 2009. Assessing salivary cortisol in large-scale, epidemiological research. Psychoneuroendocrinology 34 (10), 1423–1436. https://doi.org/10.1016/j.psyneuen.2009.06.011.

Adan, E.K., Quinn, M.E., Tavernier, R., McQuillan, M.T., Dahlke, K.A., Gilbert, K.E., 2017. Diurnal cortisol slopes and mental and physical health outcomes: A systematic review and meta-analysis. Psychoneuroendocrinology 83, 25–41. https://doi.org/10.1016/j.psyneuen.2017.05.018.

Alexander, N., Kirschbaum, C., Wankert, M., Stauch, B.J., Staehler, T., Steudte-Schmiedgen, S., Muehlhan, M., Miller, R., 2018. Glucocorticoid receptor gene methylation moderates the association of childhood trauma and cortisol stress reactivity. Psychoneuroendocrinology 90, 68–75. https://doi.org/10.1016/j.psyneuen.2018.01.020.

Ali, N., Pruessner, J.C., 2012. The salivary alpha amylase over cortisol ratio as a marker to assess dysregulations of the stress systems. Physiol. Behav. 106, 65–72. https://doi.org/10.1016/j.physbeh.2011.10.003.

Allen, A.P., Kennedy, P.J., Dockray, S., Cryan, J.F., Dinan, T.G., Clarke, G., 2017. The Trier Social Stress Test: Principles and practice. Neurobiol. Stress 6, 113–126. https://doi.org/10.1016/j.mnsr.2016.11.001.

Baes, C.V.W., Martins, C.M.S., Tofoli, S.M.d.C., Juruena, M.F., 2014. Early Life Stress in Depressed Patients: HPA Axis Response to GR and MR Agonist. Front. Psych. 5 https://doi.org/10.3389/fpsyg.2014.00002.

Berger, M., Kneuer, A.K., Romainik, M., Malouf, P., Amming, G.P., Sarnyai, Z., 2015. Cortisol awakening response in patients with psychosis: Systematic review and meta-analysis. Neurosci. Biobehav. Rev. 2019.54 https://doi.org/10.1016/j.neubiorev.2019.07.017.

Bernard, K., Butzin-Dozier, Z., Rittenhouse, J., Dozier, M., 2010. Cortisol Production Patterns in Young Children Living With Birth Parents vs Children Placed in Foster Care Following Involvement of Child Protective Services. Arch. Pediatr. Adolesc. Med. 164 https://doi.org/10.1001/archpediatrics.2010.54.

Bernard, K., Frost, A., Bennett, C.B., Lindhiem, O., 2017. Maltreatment and diurnal cortisol regulation: A meta-analysis. Psychoneuroendocrinology 78, 57–67. https://doi.org/10.1016/j.psyneuen.2017.01.005.

Bicicain, I.A.E., Postma, R.M., Smitena, G., De, Roos, C., Olff, M., Van Wesel, F., Van de Putte, E.M., 2013. Salivary cortisol and dehydroepiandrosterone sulfate in adolescent rape victims with post traumatic stress disorder. Psychoneuroendocrinology 38, 408–415. https://doi.org/10.1016/j.psyneuen.2012.06.015.

Boileu, K., Barbeau, K., Sharma, R., Bielajew, C., 2019. Ethnic differences in diurnal cortisol profiles in healthy adults: A meta-analysis. Brit. J. Health Psychol. 24 (4), 806–827. https://doi.org/10.1111/bjhp.12380.

Bonoldi, I., Simeone, E., Rocchetti, M., Codjoe, L., Rossi, G., Gambi, F., Balottin, U., Cavernazzi, E., Politi, P., Fusar-Poli, P., 2013. Prevalence of self-reported post-traumatic stress disorder in adolescents and young adults: A systematic review and meta-regression. J. Psychopharmacol. 27 (11), 1210–1217. https://doi.org/10.1177/0269768812468895.

Brenner, J.D., Vermetten, E., Mazure, C.M., 2000. Development and preliminary psychometric properties of an instrument for the measurement of childhood trauma: The Early Trauma Inventory. Depress. Anxiety 12, 1–12. https://doi.org/10.1002/1520-6394(200012)12:1<827.e;3.0.CO;2-V;845. https://doi.org/10.1016/j.euroneuro.2013.12.001.

Buckmann, A.F., Holz, N., Boeker, R., Blessey, D., Ritschel, M., Witt, S.H., Schmidt, M.H., Esser, G., Banaschewski, T., Brandeis, D., Zimmermann, U.S., Laucht, M., 2014. Moderating role of FKBP5 genotype in the impact of childhood adversity on cortisol stress response during adulthood. Eur. Neuropsychopharmacol. 24, 837–845. https://doi.org/10.1016/j.euroneuro.2014.03.011.

Bunea, L.M., Szentagotai-Tatar, A., Miu, A.U., 2017. Early-life adversity and cortisol response to social stress: A meta-analysis. Transl. Psych. 7, 1274. https://doi.org/10.2147/s19131878-011785.

Buske-Kirschbaum, A., Ebracht, M., Hellhammer, D.H., 2010. Blunted HPA axis responsiveness to stress in atopic patients is associated with the acuity and severity of allergic inflammation. Brain Behav. Immun. 24, 1353–1367. https://doi.org/10.1016/j.bbi.2010.06.013.

Carpenter, L.L., Carvalho, J.P., Tyrka, A.R., Wier, L.M., Aello, M.F., Aello, M.F., Anderson, G.M., Wilkinson, C., Price, L.H., 2007. Decreased Adrenocorticotropic Hormone and Cortisol Responses to Stress in Healthy Adults Reporting Significant Childhood Maltreatment. Biol. Psych. 62, 1080–1087. https://doi.org/10.1016/j.biopsych.2007.05.002.

Carpenter, L.L., Shattuck, T.T., Tyrka, A.R., Geracioti, T.D., Price, L.H., 2011. Effect of childhood physical abuse on cortisol stress response. Psychopharmacology 214, 1267–1275. https://doi.org/10.1007/s00213-010-2007-4.

Carpenter, L.L., Tyrka, A.R., Ross, N.S., Khoury, L., Anderson, G.M., Price, L.H., 2009. Effect of Childhood Emotional Abuse and Age on Cortisol Responsivity in Adults. Biol. Psych. 66, 69–75. https://doi.org/10.1016/j.biopsycho.2009.02.030.
S. Schar et al.

Frontiers in Neuroendocrinology 66 (2022) 100987

Experience and Implications for Prevention — 25 States, 2015–2017. MMWR Morb. Mortal Wkly Rep. 68. 999–1005. https://doi.org/10.15585/mmwr.mm6844e1.

Meyer, P.S., Novak, M.A., — 2014. A Novel Biomarker of Hypothalamic-Pituitary-Adrenocortical Activity. Endocrinology 154, 4120–4127. https://doi.org/10.1210/en.2012-1226.

Mielock, A.S., Morris, M.C., Rao, U., 2017. Patterns of cortisol and alpha-lipoyl activity to psychosocial stress in maltreated women. J. Affect. Disord. 209, 46–52. https://doi.org/10.1016/j.jad.2016.11.009.

Miller, G.E., Chen, Z., Zhou, E.S., 2007. If it goes up, must it come down? Chronic stress and the hypothalamic-pituitary-adrenocortical axis in humans. Psychol. Bull. 133, 25–45. https://doi.org/10.1037/0033-2909.133.1.25.

Monteleone, A.M., Monteleone, P., Volpe, U., De Riso, F., Fico, G., Giuliano, R., Nigro, M., Maj, M., 2018. Impaired cortisol awakening response in eating disorder vulnerable with childhood trauma exposure: evidence for a dose-dependent effect of the traumatic load. Psychol. Med. 48, 952–960. https://doi.org/10.1017/S0033291717002049.

Moore, A., Atikin, R., Burke, C., Gaskell, S., Groom, H., Holder, G., Selby, C., Wood, P., 1985. Cortisol Assay: Guidelines for the Provision of a Clinical Biochemistry Service. Ann. Clin. Biochem. Int. J. Biochem. Labo. Med. 22 (5), 435–454. https://doi.org/10.1177/000453858502200501.

Morrís, M.C., Abelson, J.L., Mielock, A.S., Rao, U., 2017. Psychobiology of cumulative stress: hair cortisol as a risk marker for stress exposure in women. Stress 20, 350–354. https://doi.org/10.1080/10253890.2017.1340450.

Myers, B., McKevitt, J.M., Herman, J.P., 2014. Glucocorticoid actions on synapses, circuits, and behavior: Implications for the energetics of stress. Front. Neuroendocrinol. 35 (2), 180–196. https://doi.org/10.1016/j.yfrne.2013.12.003.

Nelson, E.C., Heath, A.C., Madden, P.A.F., Cooper, M.L., Dinwiddie, S.H., Buchock, K.K., Glowinski, A., McLaughlin, T., Dunne, M.P., Statham, D.J., Martin, N.G., 2002. Association Between Self-reported Childhood Sexual Abuse and Adverse Psychosocial Outcomes: Results from a Twin Study. Arch. Gen. Psych. 59 (2), 139. https://doi.org/10.1001/archpsych.59.2.139.

Newbury, J.B., Arsenault, L., Moffitt, T.E., Caspi, A., Danese, A., Baldwin, J.R., Fisher, H.L., 2018. Measuring childhood maltreatment to predict early-adult psychopathology: Comparison of prospective interview-reports and retrospective self-reports. J. Psychiatr. Res. 96, 57–64. https://doi.org/10.1016/j.jpsychires.2017.09.020.

Newport, D.J., Heim, C., Bonsall, R., Miller, A.H., Nemeroff, C.B., 2004. Pituitary-adrenal responses to traumatic stress in adult survivors of childhood abuse. Biol. Psych. 55, 10–20. https://doi.org/10.1016/j.biopsych.2003.06.092.

Nicolaides, N.C., Kyrazi, E., Lamprokostopoulou, A., Chrousos, G.P., Charmandari, E., 2015. Stress, the Stress Hormones and the Role of Glucocorticoids. Neuroendocrinology Modulation 22 (1-2), 6–19. https://doi.org/10.1016/j.nlm.2013.06.027.

Norman, R.E., Byambaa, M., De, R., Butchart, A., Scott, J., Yos, T., Tomlinson, M., 2012. The Long-term Health Consequences of Child Physical Abuse, Emotional Abuse, and Neglect: A Systematic Review and Meta-Analysis. PLoS Med. 9 (11), e1001349. https://doi.org/10.1371/journal.pmed.1001349.

O’Connor, D.B., Green, J.A., Ferguson, E., O’Carroll, R.E., O’Connor, R.C., 2018. Effects of childhood trauma on cortisol levels in suicide attempters and ideators. Psychoneuroendocrinology 88, 9–16. https://doi.org/10.1016/j.psyneuen.2017.11.004.

O’Connor, M., Molen, S., Becares, L., Burgen, G., Williams, D.R., Priest, N., 2020. Inequalities in the Distribution of Childhood Adversity From Birth to 11 Years. Acad. Pediatrics. 20 (5), 609–618. https://doi.org/10.1016/j.acap.2019.12.004.

Otto, C., Neylan, T.C., Pole, M., Metzler, T., Best, S., Henn-Haase, C., Yehuda, R., Marmar, C.R., 2005. Association between childhood trauma and catecholamine monoamine oxidase a activity in depressed patients with and without childhood neglect. BMC Psych. 14, 45. https://doi.org/10.1186/1471-244X-14-45.

Ouellet-Morin, I., Odgers, C.L., Danese, A., Bowes, L., Shakoor, S., Papadopoulos, A.S., Caspi, A., Moffitt, T.E., Arsenault, L., 2011. Blunted Cortisol Responses to Stress Signal Social and Behavioral Problems Among Maltreated/Bullied 12-Year-Old Children. Biol. Psych. 70, 1016–1023. https://doi.org/10.1016/j.biopsych.2011.06.017.

Ouellet-Morin, I., Robitaille, M., Langevin, S., Cantave, C., Brendgen, M., Lupien, S.J., 2018. Enduring effect of childhood maltreatment on cortisol and heart rate responses to stress: The moderating role of family environment. Dev. Psychopathol. 30, 447–458. https://doi.org/10.1017/s0954579418000099.

Ouellet-Morin, I., Odgers, C.L., Danese, A., Bowes, L., Shakoor, S., Papadopoulos, A.S., Caspi, A., Moffitt, T.E., Arsenault, L., 2011. Blunted Cortisol Responses to Stress Signal Social and Behavioral Problems Among Maltreated/Bullied 12-Year-Old Children. Biol. Psych. 70, 1016–1023. https://doi.org/10.1016/j.biopsych.2011.06.017.

Ouellet-Morin, I., Robitaille, M., Langevin, S., Cantave, C., Brendgen, M., Lupien, S.J., 2018. Enduring effect of childhood maltreatment on cortisol and heart rate responses to stress: The moderating role of family environment. Dev. Psychopathol. 30, 447–458. https://doi.org/10.1017/s0954579418000099.

Ouellet-Morin, I., Odgers, C.L., Danese, A., Bowes, L., Shakoor, S., Papadopoulos, A.S., Caspi, A., Moffitt, T.E., Arsenault, L., 2011. Blunted Cortisol Responses to Stress Signal Social and Behavioral Problems Among Maltreated/Bullied 12-Year-Old Children. Biol. Psych. 70, 1016–1023. https://doi.org/10.1016/j.biopsych.2011.06.017.

Piera, A.M., Favor, J., Varese, F., 2020. Childhood adversity and borderline personality disorder: a meta-analysis. Acta Psychiatr. Scand. 141 (1), 6–20. https://doi.org/10.1111/acps.13118.
S. Schär et al.

Cohort Study. Arch. Gen. Psych. 66, 617. https://doi.org/10.1001/archgenpsychiatry.2009.50. Watson, S., Gallagher, P., Smith, M.S., Ferrier, I.N., Young, A.H., 2006. The dex/CRH test—is it better than the DST? Psychoneuroendocrinology 31 (7), 889–894. https://doi.org/10.1016/j.psyneuen.2006.03.001. Watson, S., Owen, B.M., Gallagher, P., Mean, A.J., Young, A.H., Ferrier, I.N., 2007. Family history, early adversity and the hypobalamic-pituitary-adrenal (HPA) axis. Mediation of the vulnerability to mood disorders. Neuropsychiatr. Dis. Treat. 3, 647–653.

White, L.O., Ising, M., von Klitzing, K., Sierau, S., Michel, A., Klein, A.M., Andreas, A., Keil, J., Quintrero, I., Müller-Myhok, B., Uhr, M., Gausche, R., Manly, J.T., Crowley, M.J., Kirschbaum, C., Stalder, T., 2017. Reduced hair cortisol after maltreatment mediates externalizing symptoms in middle childhood and adolescence. J. Child Psychol. Psychiat. 58 (9), 998–1007. https://doi.org/10.1111/jcpp.12700.

Wingenfeld, K., Kuehl, L.K., Boeker, A., Schultebraucks, K., Ritter, K., Hellmann-Regen, J., Otte, C., Spitzer, C., 2017. Stress reactivity and its effects on subsequent food intake in depressed and healthy women with and without adverse childhood experiences. Psychoneuroendocrinology 80, 122–130. https://doi.org/10.1016/j.psyneuen.2017.03.014.

Witt, A., Brown, R.C., Plener, P.L., Brähler, E., Fegert, J.M., 2017. Child maltreatment in Germany: prevalence rates in the general population. Child Adolesc. Psych. Mental Health 11 (1). https://doi.org/10.1186/s13034-017-0185-6.

Young, E.S., Doom, J.R., Farrell, A.K., Carlson, E.A., Englund, M.M., Miller, G.E., Gunnar, M.R., Roisman, G.I., Simpson, J.A., 2021. Life stress and cortisol reactivity: An exploratory analysis of the effects of stress exposure across life on HPA-axis functioning. Dev. Psychopathol. 33 (1), 301–312. https://doi.org/10.1017/S0954579419001779.

Young, E.S., Farrell, A.K., Carlson, E.A., Englund, M.M., Miller, G.E., Gunnar, M.R., Roisman, G.I., Simpson, J.A., 2019. The Dual Impact of Early and Concurrent Life Stress on Adults’ Diurnal Cortisol Patterns: A Prospective Study. Psychol. Sci. 30 (5), 739–747. https://doi.org/10.1177/0956797619833664.

Zankert, S., Bellingrath, S., Wüst, S., Kudielka, B.M., 2019. HPA axis responses to psychological challenge linking stress and disease: What do we know on sources of intra- and interindividual variability? Psychoneuroendocrinology 105, 86–97. https://doi.org/10.1016/j.psyneuen.2018.10.027.

Zorn, J.V., Schür, R.R., Boks, M.P., Kahn, R.S., Joels, M., Vinkers, C.H., 2017. Cortisol stress reactivity across psychiatric disorders: A systematic review and meta-analysis. Psychoneuroendocrinology 77, 25–36. https://doi.org/10.1016/j.psyneuen.2016.11.036.