Longitudinal changes in amygdala, hippocampus and cortisol development following early caregiving adversity

Michelle VanTieghem

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
under the Executive Committee
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2020

Abstract

Longitudinal changes in amygdala, hippocampus and cortisol development following early caregiving adversity

Michelle VanTieghem

Decades of research have shown long-term effects of early caregiving adversity on stress physiology and limbic brain regions, two key biological systems that are implicated in risk for internalizing disorders. Although stress physiology and limbic brain structure undergo significant maturational change during childhood and adolescence, and reciprocally influence each other, the effects of early caregiving adversity on these developmental processes is not well understood. In the current study, we used an accelerated longitudinal design to assess the development of stress physiology, amygdala, and hippocampal volume following early institutional care. Previously Institutionalized (PI; N = 93) and comparison (COMP; N = 161) youth (ages 4-20 years old) completed 1-3 waves of data collection, each spaced approximately 2 years apart, for diurnal cortisol (N = 239, providing a total of 380 diurnal datasets), structural MRI (N = 156, providing a total of 306 scans) and parent-reported internalizing symptoms (N = 133, providing a total of 227 time points). We observed a developmental shift in morning cortisol in the PI group, with blunted levels in childhood and heightened levels in late adolescence. PI history was associated with reduced hippocampal volume and reduced growth of the amygdala, resulting in smaller volumes by adolescence. Results also suggested feed-forward brain-to-hormone mechanisms, such that both amygdala and hippocampal volumes were prospectively associated with morning cortisol levels two years later. Finally, amygdala and hippocampal volumes were independently associated with internalizing scores across the entire sample. These results

indicate that adversity-related physiological and neural phenotypes are not stationary during development but instead exhibit dynamic and interdependent changes from early childhood to early adulthood.

Table of Contents

List of Figures

List of Tables

Acknowledgments

First, I would like to acknowledge the team of individuals who contributed to this longitudinal research project. I am honored to have them all as co-authors on the manuscript that is currently under review for publication*. Marta Korom, for helping to organize the archival data, for performing quality control of hundreds of brain images, and for providing support through each stage of this project. Laurel, Kate, Bonnie, Mor, Jennie, Dylan, Eva, and Dominic, for their 'blood sweat and tears' to collect this longitudinal data over the course of many years, before I was even a graduate student. Jessica Flannery, Tricia Choy, and Christina Caldera, whose collective efforts as DAN lab managers oversaw the study management over the years, as the lab transitioned to a new institution. Niall Bolger, who provided an enormous amount of statistical expertise, guidance, and enthusiasm. He would often meet with me at the end of the day, with a coffee in hand, ready for any new longitudinal modeling problem I threw at him. Finally, I would like to thank my advisor Nim Tottenham. Her empathy, encouragement, and positivity throughout my PhD has been an inspiration and I am so grateful to have her as my graduate advisor.

I also need to thank my cohort-mate Katherine, who has been there since day 1, and my DAN Lab coworkers (past and present) who have all have helped me grow as a scientist and person. I wouldn't have made it through graduate school without such a supportive group to work with every day. Thank you to my friends and family, especially my husband David, who have been there during the highs and lows of the past 6 years and kept me balanced with life outside of graduate school.

Thank you to my committee members Daphna Shohamy, Kim Noble and Moriah Thomason, for your time and support. I would also like to acknowledge the funding that supported my graduate studies, including National Science Foundation, American Psychological Foundation, and National Institutes of Health.

**Please note, this manuscript is posted as a preprint:*

VanTieghem M., Korom, M., Flannery, J., Choy, T., Caldera, C., Gabard-Durnam, L., Goff, B., Gee, D., Humphreys, K., Telzer, E., Shapiro, M., Louie, J. Y., Fareri, D., Bolger, N., Tottenham, N. Longitudinal changes in amygdala, hippocampus, and cortisol development following early caregiving adversity. Preprint: https://psyarxiv.com/yp5h2/

Introduction

Early caregiving adversity (ECA) is associated with increased risk for internalizing disorders, such as anxiety and depression (Green et al., 2010; Kessler et al., 2010). Decades of research suggests that ECA alters key biological systems implicated in psychopathology risk: stress physiology (e.g., HPA axis; (Gunnar & Quevedo, 2008; Miller et al., 2007) and limbic brain regions (e.g., amygdala and hippocampus; McEwen et al., 2016). However, the developmental processes that lead to these adult phenotypes following ECA are not well understood. Although the hypothalamic-pituitary-adrenal (HPA) axis bi-directionally interacts with the amygdala and hippocampus (Herman et al., 2012), they are typically examined in separate studies, leaving open the question as to how they influence each other across development. Secondly, these systems change dramatically across the first two decades of life (Flannery et al., 2017; Herting et al., 2017; Ostby et al., 2009), yet developmental effects of ECA are rarely examined (although see Flannery et al., 2017; Megan R. Gunnar, DePasquale, Reid, & Donzella, 2019; King et al., 2017; Luby, Tillman, & Barch, 2019) in part because of the lack of longitudinal data and the subsequent reliance on cross sectional designs and/or analytic approaches that include age as a covariate and not of explicit interest. These study designs have created the impression that effects of ECA on developing neurobiology are static, limiting our ability to understand the neurodevelopmental sequela of ECA exposure. In the current study, we address these challenges with an accelerated-longitudinal design to characterize the developmental patterns of diurnal cortisol and limbic brain structure following early institutional care, examine bidirectional time-lagged associations between these two systems, and investigate their relevance for internalizing behavior.

Diurnal cortisol

The HPA axis is one of the primary physiological systems involved in regulating responses to environmental stressors (Gunnar & Quevedo, 2007). Diurnal cortisol provides an index of HPA-axis regulation, characterized by high morning levels that normally decline across the day in a circadian fashion (Gunnar & Quevedo, 2007). Across typical development, diurnal cortisol slope becomes steeper (more negative), driven by age-related increases in morning cortisol levels (Flannery et al., 2017). Meanwhile, ECA exposure is typically followed by blunted morning levels in childhood, thus altering in this daily rhythm (Bernard et al., 2015; Dozier et al., 2006; Fisher & Stoolmiller, 2008; Koss et al., 2014; Pitula et al., 2019; Zalewski et al., 2016). Despite this early blunting, initial studies have suggested that adolescence represents a period of recalibration of stress physiology, with normalized morning cortisol (Flannery et al., 2017), higher morning cortisol (Weems & Carrión, 2009), or heightened cortisol awakening response (King et al., 2017; Quevedo et al., 2012) in ECA-exposed relative to non-exposed (comparison) samples. However, these findings have been primarily cross-sectional studies or those longitudinally limited to narrow age-ranges (i.e., 2 year span; King et al., 2017), making it difficult to discern whether the literature reflects developmental changes in adversity-related cortisol phenotypes, or merely differences due to methodological issues (e.g. age distributions, adversity subtypes, analysis methods). The current study uses an accelerated longitudinal design to address these outstanding questions and fully characterize the developmental effects of ECA exposure on diurnal cortisol from early childhood to early adulthood.

Limbic brain volume

Exposure to ECA is also associated with alterations in the amygdala and hippocampus (Tottenham & Sheridan, 2009)—limbic brain regions critically involved in the development of healthy emotion regulation capabilities (Silvers et al., 2016). Reductions in amygdala and hippocampal volume have been consistently observed in adults following ECA exposure (Butterworth et al., 2012; Calem et al., 2017; Dannlowski et al., 2012; Riem et al., 2015; van Velzen et al., 2016) and smaller amygdala and hippocampal volumes have been linked with greater risk of internalizing psychopathology (Gorka et al., 2014; Rao et al., 2010). However, the influence of ECA on limbic brain regions during development has been less clear. While hippocampal differences, if observed, have consistently shown hypotrophy in developmental samples (Hanson et al., 2011, 2014; Hodel et al., 2014; Humphreys et al., 2019; King et al., 2018; Luby, 2013; Noble et al., 2012; Piccolo & Noble, 2018), the literature on amygdala volume is less clear, with some studies showing that childhood adversity is associated with larger volumes (Lupien et al., 2011; Mehta et al., 2009; Roth et al., 2018; Tottenham et al., 2010), other studies showing smaller volumes (Edmiston, 2011; Hanson et al., 2014; Luby, 2013; Noble et al., 2012) and still other studies showing no differences (Hodel et al., 2014; King et al., 2018; Noble et al., 2015; Sheridan et al., 2012). The role of amygdala volume in internalizing problems during development is also mixed, with both larger (Pagliaccio et al., 2014; Roth et al., 2018; Tottenham et al., 2010) and smaller volumes (Merz et al., 2017) associated with increased internalizing symptoms or emotion regulation difficulties. Although the amygdala and hippocampus undergo significant age-related changes (Herting et al., 2017; Ostby et al., 2009; Uematsu et al., 2012; Wierenga et al., 2018), the majority of studies used cross-sectional samples that control for age [although see (Ellwood-Lowe et al., 2018; Merz et al., 2017)] leaving the

possibility that existing discrepancies may be due to developmental effects. In the current study, we address this gap by characterizing ECA-related developmental changes in amygdala and hippocampal volume across a wide age-range and identifying their associations with internalizing scores.

Cortisol-brain interactions

Animal studies have demonstrated that limbic brain development and the HPA axis are highly coupled during development (Myers et al., 2012). The amygdala and hippocampus are densely innervated with glucocorticoid (GR) and mineralocorticoid receptors (MR), which make them sensitive to circulating levels of cortisol. Because both GR/MR receptor density and the growth rates of amygdala and hippocampus peak during the first year of life (Avishai-Eliner et al., 1996; Gilmore et al., 2012; Payne et al., 2010; Vázquez et al., 2012), early limbic development is particularly susceptible to stress exposure, such as caregiving adversity (Raineki et al., 2019). The amygdala and hippocampus not only receive input from cortisol, but also provide regulatory feedback to the HPA axis via projections to the paraventricular nucleus (PVN) of the hypothalamus, suggesting bidirectional interactions between the HPA-axis and limbic brain regions (Herman et al., 2012). However, how these systems dynamically interact during human development, and whether maturation in one system predicts change in another in a predominant direction, is an open question. Although prior studies in humans have detected cross-sectional relationships between the HPA axis and limbic brain development (Dahmen et al., 2017; Davis et al., 2017; Gee et al., 2013; Liu et al., 2012; Pagliaccio et al., 2013), longitudinal studies have only tested these relationships in one direction, from child cortisol reactivity to adolescent brain function (Burghy et al., 2012; Pagliaccio et al., 2015), leaving open the question whether there may be reciprocal influences of these systems across development. In

the current study, we leverage a longitudinal dataset with diurnal cortisol and subcortical brain volume measurements obtained at 2 time points (~2 years apart) to investigate their longitudinal bidirectional associations across childhood and adolescence.

Current study

In the current study, we used an accelerated longitudinal design to characterize developmental changes in stress physiology and limbic brain development in internationally adopted youth with exposure to institutional caregiving. In general, institutional care is characterized by unstable and sparse caregiving and high infant-to-caregiver ratios (Gunnar et al., 2000). This lack of species-expected contingent caregiving is considered a potent form of adversity for the developing child (Tottenham, 2012). However, this form of ECA exposure is temporally restricted to early life, as youth are subsequently adopted into stable families in the US. Here, we assessed non-linear age-related changes in diurnal cortisol, amygdala, and hippocampal volume following institutional care. Next, we applied cross-lagged structural equation modeling (SEM) to assess longitudinal and bidirectional associations between morning cortisol and limbic brain volume across development. Finally, we investigated the behavioral relevance of these ECA-related phenotypes by examining associations between stress physiology and limbic brain development with parent-reported internalizing symptoms.

Methods

Participants

Previously institutionalized (PI; N = 93) youth with known history of institutional care and comparison (COMP; N = 161) youth between the ages of 4- 20 years are included in the current study (Table 1). PI youth in the current study were internationally adopted in the US, with varying countries of origin and institutional settings (see Table 1). Although information on specific caregiving factors and adversities are unknown in internationally adopted populations in general, we did collect information regarding age of placement and duration of institutional care from PI families when available, and the majority of participants has exposure to institutional care that was limited during the first 3 years of life (Table 1). PI youth were recruited via local international adoption agencies, adoption family networks, posted flyers, and friend referral. Healthy comparison participants (defined as raised by biological parents in the United States and never institutionalized) were recruited via birth records, posted flyers, and friend referral. Comparison participants were pre-screened for prior diagnoses of any behavioral/psychological concerns or learning disabilities. The protocol was approved by the Institutional Review Board at the University of California, Los Angeles. Participants and their parents provided both informed assent and consent.

Study Procedure

The current sample participated in an accelerated longitudinal study, with baseline visits occurring at ages 4-6 years old. By design, this study was not a full longitudinal study (i.e., in order to complete data collection within a 5-year period); a subset of participants was invited to complete 1 or 2 follow-up visits (depending on entry date, not all participants were actually

intended for both follow-up visits). Participants were overenrolled at Time 1 in order to reach follow-up recruitment goals given the anticipated difficulties in scanning this age range (e.g., braces, motion artifact, refusal, and scheduling). The age distribution of baseline and follow-up visits is shown in Figure 1. At each wave of the study, children completed structural MRI scans and diurnal cortisol samples and parent-reported questionnaires were obtained. Table 2 provides the N for each group and each wave for each sub-sample used for analysis (diurnal cortisol, structural MRI, and symptom assessments). Follow-up retention for these sub-samples did not vary by group, age or sex (Supplemental Analyses).

Table 1: Demographic information for all participants included in the current study. PI = Previously Institutionalized, COMP = comparisons. Age placed refers to the age of placement in the institution.

	PI	COMP	Group difference
N	93	161	
Sex (M/F)	31 / 62	78 / 83	$p = 0.017$
IQ at T1, mean (SD)	101.72 (16.68)	111.78 (16.7)	$p < 0.0001$
Age in years, mean (SD), range	9.98 (3.35), 3.92-17.17	9.1 (4.12), 4.08-17.58	$p = 0.065$
Age placed (months), median (SD), range	0.75 (14.31), 0-72		
Age adopted (months), median (SD), range	16.5 (25.7), 0.7-120		
Country of Origin			
Asian	38 (41%)		
Eastern European	50 (54%)		
Unknown/Other	5 (5%)		

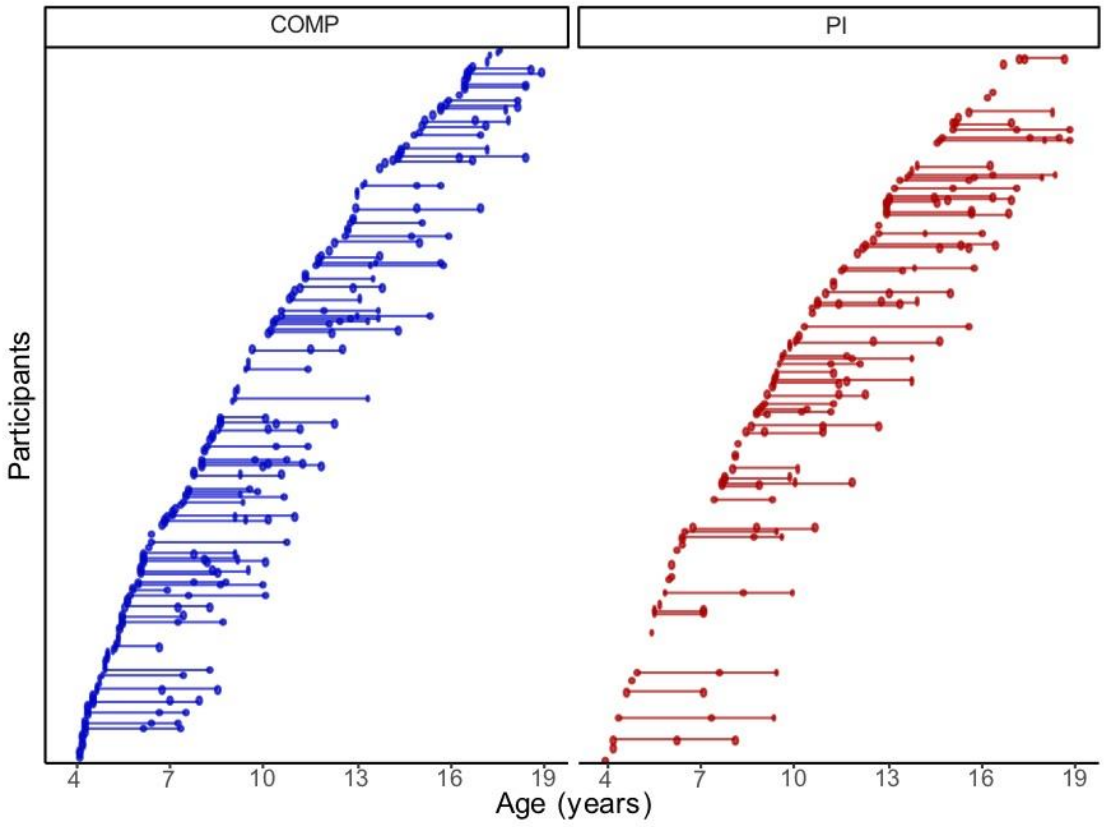

Figure 1: Age of sampling for accelerated longitudinal design. Participants completed 1-3 waves of data collection for diurnal cortisol, structural scans, and/or parent-reported internalizing scores. COMP = comparison, PI = Previously Institutionalized.

Table 2: Demographic information by group and wave for sub-samples of participants who completed diurnal cortisol, structural MRI, and symptom assessments. PI = Previously Institutionalized, COMP = comparisons.

			Sex	Age (years)		
Group	Wave	N	(M/F)	Mean	SD	Range
Diurnal cortisol sample						
	1	141	71 / 70	9.15	4.13	4.08-17.58
COMP	2	56	22 / 34	11.50	4.32	5.25-20.33
	3	30	11 / 19	11.49	3.75	6.67-19.08
	1	82	29 / 53	9.76	3.41	3.92-17.17
PI	2	37	10 / 27	12.08	3.42	6.25-18.25
	3	34	9 / 25	13.66	3.56	7.08-18.83
Structural MRI sample						
	1	70	35 / 35	10.67	3.91	4.25-18.58
COMP	2	69	29 / 40	11.88	4.18	4.83-20.33
	3	41	14 / 27	12.31	3.99	6.67-21.08
	1	45	18 / 27	10.67	2.86	4.58-16.58
PI	2	45	15 / 30	12.41	3.15	6.75-18.25
	3	36	12 / 24	13.90	3.32	7.08-18.83
Symptom analysis sample						
	1	57	28 / 29	10.64	3.92	3.92-17.50
COMP	2	50	20 / 30	11.66	4.28	5.58-20.33
	3	24	9 / 15	12.15	3.92	6.67-19.03
	1	37	16 / 21	10.11	2.88	4.33-16.33
PI	2	32	10 / 22	12.20	3.38	6.75-18.25
	3	26	8 / 18	13.79	3.43	7.08-18/83

Cortisol sampling procedure

Families were instructed to collect salivary cortisol samples on 2 days at 4 target time points per day: wake up, 45 min after wake up, 5pm, and 8pm. They were instructed to collect samples before eating or drinking, or at least 15 min after eating or drinking, and not to collect samples on days they felt ill. Take-home saliva diaries were included for each day of cortisol sampling. Parents were asked to report the following: date, bedtime, illness, medication use, and/or unusual levels of activity. Daily diaries were used for analysis covariates of psychotropic and oral steroid medication use on days of cortisol collection.

Following Salimetrics protocol, children under 6 years old used two sorbettes under their tongue for 1 min per sample (Salimetrics, State College, PA). Children 6 years old and older were instructed to chew on a piece of Trident original sugarless gum, which has been shown to not interfere with cortisol collection, to stimulate saliva flow (Dabbs, 1991). Participants were then instructed to place a sorbette underneath their tongue for 60–90 s before placing the sample back in the salivette (Salimetrics, State College, PA). Samples were stored at −20 °C and processed at Dr. Clemens Kirschbaum's Biological Psychology Laboratory at the Technical University Dresden where salivary cortisol concentrations were measured using commercially available chemiluminescence-immunoassays with high sensitivity. The inter-assay coefficient for cortisol was below 8% (Kirschbaum & Hellhammer, 2000). Data were assayed in singlet, so we do not have an intra-assay coefficient, but cortisol values at each time of day were highly stable within-subject across day 1 and day 2 ($p < 0.001$ for all 4 cortisol sample time points).

Structural neuroimaging procedure

High resolution T1-weighted scans were acquired on a Siemen's 3T Trio scanner for waves 1 and 2 (TR = 2170 ms, TE = 4.33 ms, flip angle = 7 degrees, 192 slices, 1.0 x 1.0 x 1.0 mm voxels, FOV = 256 mm, scan time = 8.08 minutes). Wave 3 was collected on a newer Siemen's 3T Trio scanner with the following protocol: TR = 1900, TE = 3.26, flip = 9 degrees, slices = 176, FOV = 250, scan time = 3.83 minutes.

Questionnaires

Parent-reported behavior was assessed using the Revised Child Anxiety and Depression Scale (RCADS; Ebesutani et al., 2011), which is validated for use across wide age-ranges and previously institutionalized populations (Ebesutani et al., 2015). Analyses used T-scored total internalizing scores, which are adjusted for age and sex norms in the population.

Quality Control Procedures

Cortisol data

Although participants were asked to collected cortisol at 4 time-points across the day, samples were categorized into two-time points (morning and evening) for the following reasons. Due to poor compliance and/or sample qualities, 21% of diurnal cortisol samples were missing 1 or more time points in a given day and/or missing time logs, which limited our ability to account for person-centered time across the day. Second, of those samples that included time logs, 24% of morning samples were collected greater than 45 minutes apart. In the absence of accurate time stamps or sleep data, we did not attempt to measure the cortisol awakening response (CAR).

Instead, we aimed to assess change in cortisol concentration across the day from the morning and evening samples.

The diurnal cortisol data was filtered for valid cortisol values (e.g. > 200 nmol/L) and then data points beyond 3SD of the mean within each sampling group were excluded; Group (PI vs. comp), wave (1, 2, 3) and time of day (morning, evening). The final sample included 239 participants (151 comparisons, 88 PIs) with a total of 380 time points of diurnal cortisol collected across 1-3 waves (Table 2). A subsample of these participants was previously published in a cross-sectional analysis (Flannery et al., 2017). Raw cortisol values were used given that linear mixed effects models are more robust to non-linearity in the first level (Maas & Hox, 2004). We controlled for significant batch effects on cortisol values (Table S1) by including batch as a covariate in all statistical models.

Structural MRI data

Freesurfer version 6.0 (Fischl et al., 2002) was used to identify subject-specific segmentations in subcortical brain areas of interest (e.g. amygdala and hippocampus). We used the cross-sectional Freesurfer stream, as the longitudinal pipeline is not recommended for developmental studies with a wide age-range due to whole brain changes (Reuter, 2016). Prior to processing, scan quality assessment was performed by independent raters to identify (a) motion artifact and (b) subcortical segmentation quality. Specifically, motion in the raw T1 images were assessed by two independent raters on a scale of 1 (good) to 4 (poor) based on prior guidelines (Afacan et al., 2016). Next, two independent raters assessed the quality of Freesurfer subcortical segmentation of hippocampus and amygdala on a scale of 1 (good; no visible errors) to 4 (poor; large errors spanning 50 voxels or more). Inter-rater reliability was calculated for motion

(interclass coefficient = 0.65) and subcortical segmentation (interclass coefficient = 0.75). Following exclusion criterion (detailed below), volumes were extracted for each brain area of interest (right, left hippocampus and amygdala). Intracranial volume (ICV) was also extracted as a control measure of global brain volume. Amygdala and hippocampal volumes units were scaled to 100^3 mm and ICV was scaled to $100{,}000^3$ mm for all statistical analyses. Due to a lack of *a priori* hypotheses on laterality, and high correlation between right and left hemispheres for amygdala ($r = 0.73$, $p < .001$) and hippocampus ($r = 0.88$, $p < .001$) volumes, bilateral volumes were used for analyses. Right and left hemispheres were tested separately in follow-up analyses in the Supplemental Analyses.

MRI images with substantial motion artifact (rating of 4 by at least one rater) were excluded from Freesurfer processing (N=13). Additionally, images that failed processing in Freesurfer due to errors or excessive run time (indicative of poor image quality) were excluded from subsequent analysis (N = 2). Of those images with successful segmentation by Freesurfer, N = 4 images were excluded due to substantial errors in hippocampus and/or amygdala segmentation (rating of 4 by at least one rater). Finally, outliers, defined as values beyond 3SD of the mean within each group and wave, were excluded (ICV = 1, amygdala volume = 3, hippocampus volume = 0). After quality control assessment, the final sample included 156 participants (66 PIs, 90 comparisons) with a total of 306 MRI scans collected across 1-3 waves (Table 2). To account for the scanner/protocol differences in wave 3, a dummy covariate for scanner was included in all analyses (Noble et al., 2015). Supplemental analyses were also conducted with only the sub-sample of data collected on the same scanner (i.e., omitted wave 3).

Modeling age-related change

To test how early experience influenced patterns of age-related change, we conducted linear mixed effects models testing for Group X Age interactions. In the event of a significant interaction, we investigated non-linear Age X Group effects using quadratic and piecewise models. For piecewise models, we conducted a data-driven iterative search (using the *optimize* function in R) to determine the optimal breakpoint in age that minimized the deviance of the model fit. The optimal piecewise model was then compared with quadratic and linear models using Akaike Information Criterion (AIC) to determine the best model fit, where lower AIC by at least 2 points indicates a better model fit. Significant interactions were probed with simple slopes in R. All modeling was performed using the lme4 and lmerTest package in R (Bates et al., 2015; Kuznetsova et al., 2017) which uses maximum likelihood estimation and Satterthwaite method of degrees of freedom (Satterthwaite, 1946). For each model, we also calculated the IntraClass Correlation (ICC) using the Sjstats package in R (Daniel Lüdecke, 2018) to assess the reliability of within-subject estimates for each outcome variable.

All analyses controlled for sex due to the higher incidence of females in the PI group (Table 1), which reflects the broader demographic that a higher frequency of children who experience institutional care followed by international adoption are female (Hellerstedt et al., 2008). For all cortisol models, we also controlled for day of collection, wave of collection, batch of sample processing (conducted in 5 batches over 5 years), and medication status (1 = on medication for wave of sample collection, 0 = no medication for that wave). Random intercepts per subject and random effects of wave were included to account for repeated measures within subject. For models of structural brain volume, we included covariates of ICV, motion ratings, and scanner (i.e. dummy coded variable indicating different scanner at wave 3). Random

intercepts per subject were included to account for repeated measurements; random effects of wave were not included due to too few longitudinal data points for model convergence. Descriptive information and Pearson's correlations between covariates and variables of interest are provided in Tables S1 & S2. Group differences in relevant covariates for both sets of analyses are provided in Tables S3 & S4.

Cross-lag autoregressive SEM models

To examine longitudinal and bidirectional effects between cortisol and limbic brain volume across development, we used cross-lag auto-regressive structural equation modeling (SEM) using the Lavaan package in R (Rosseel, 2012). We tested two cross-lag paths: *Path 1* tested the effect of morning cortisol T1 on hippocampal/amygdala volume at T2, controlling for hippocampal/amygdala volume at T1 and covariates (sex, group, age at T2, ICV at T2, and scanner at T2); *Path 2* tested the effect of hippocampal/amygdala volume T1 on morning cortisol at T2, controlling for morning cortisol at T1 and covariates (sex, group, age at T2, and batch at T2). Separate cross-sectional regressions controlled for the effects of covariates (sex, group, age at T1, and batch at T1 / ICV at T1 respectively) on baseline (T1) measures of hippocampal volume and morning cortisol. Covariance between the two T1 variables (representing their correlation after accounting for covariates in the model) and T2 errors (representing the correlation between the remaining error variance of the 2 variables) are also reported. In a parallel set of models, group interactions were added to each cross-lagged path (Supplemental Analyses). Model fit was assessed using nested model comparisons.

To carry out this analysis, we used a sub-sample of participants with usable cortisol and structural MRI data at least one wave (see Table S5 for sub-sample demographics). To maximize

the sample size, we used Full information maximum likelihood (FIML) to account for missing data, allowing us to include subjects without full coverage (i.e. not all subjects had both cortisol and MRI data at T1 and T2). T1 represented the first wave and T2 represented the final wave of data available, and intervals between T1 and T2 ranged from 2-4 years (mean = 2.73, SD = 1.03). For those participants with usable data at 3 time points, we included their data across the maximal time delay available (e.g. data at wave 1 and wave 3). Based on results showing significant group and age-related differences in morning but not evening cortisol, only morning cortisol values were used for these analyses. Outlier values that exceeded 3SD of the mean of each group at each time point were excluded for analyses. Amygdala and hippocampal volume were too highly correlated to put into the same regression path without causing suppression effects ($r = 0.62$, $p < .001$) and were modeled separately.

Associations with symptoms

A sub-sample of participants with both usable cortisol and scanning data also provided parent-reported internalizing scores (RCADS total internalizing T scores; N = 133, PI = 58, COMP = 75) with a total of 227 data points across 1-3 waves (Table 2). Outlier values beyond 3 SD of the mean for each wave and group were omitted from the analysis. Linear mixed effects models examined whether morning cortisol, amygdala or hippocampal volume was associated with internalizing symptoms across all waves of data collected. Random intercepts were included to account for within-subject repeated measures. Moderations of each variable by group and age were tested, controlling for medication status (recorded on day of cortisol collection), scanner confounds, sex, and ICV. In the event of a significant association, the model was re-run with all predictors of interest (cortisol, amygdala, and hippocampus) to determine if the effects remained significant when controlling for the others.

Results

Developmental Shifts in morning cortisol following early caregiving adversity

A significant three-way interaction between Group, age, and time of day was detected for diurnal cortisol (nmol/L) using linear mixed effects modeling (b = 0.78, t (2,187) = 4.27, p < 0.001, CI = 0.42, 1.14). Because developmental change may not be linear, we then probed non-linear age effects using quadratic and piecewise age terms. A non-linear piecewise age model showed the best model fit (Table 3), with an optimal breakpoint at 13.1 years (CI =11.7, 15.0). with a significant Group X Age X Time of day interaction for ages 13 and older, but not before age 13 (see Table S6 for full model results). To further interrogate these interactions, follow-up models tested piecewise age effects for morning and evening separately.

Morning cortisol: Modeling morning cortisol separately with piecewise Age X Group effects verified an optimal breakpoint at 13.1 years old (Figure 2A). Piecewise Age X Group effects showed differences in age-related change in morning cortisol between groups. Before age 13.1, age moderated the effects of group (b = -0.68, t (301) = -2.12, p = 0.034, CI = -1.29, -0.05; the comparison group showed a significant age-related increase in morning cortisol during this age-range (b = 0.73, t (321) = 4.04, p < 0.001, CI = 0.37, 1.08) but the PI group did not (b = 0.05, t (321) = 0.20, p = 0.841, CI = -0.46, 0.56). After age 13.1, age also moderated the effects of group (b = 2.03, t (349) = 2.89, p < 0.01, CI = 0.65, 3.40), such that age-related increases in morning cortisol were observed in the PI group (b = 2.38, t (313) = 4.07, p < 0.001, CI = 1.23, 3.53) but not the comparison group (b = 0.34, t (313) = 0.84, p = 0.404, CI = -0.47, 1.16). In order to assess significant group differences across the entire age range, we used a regions of significance plot (Preacher, Curran, & Bauer, 2006) which shows that PI group was estimated to

have lower morning cortisol relative to comparisons between ages 9 and 14, and higher morning cortisol after age 17.5 (Figure 2B).

Together, these findings suggest that PI youth switch from blunted levels of morning cortisol during childhood to heightened levels during late adolescence/early adulthood. Importantly, when averaging across age, there was a significant main effect of Group (b = -4.61, t (319) = -2.83, $p < 0.01$, CI = -7.75, -1.43), suggesting that interpretations from datasets of wide age-ranges depends greatly on the analytic strategy (i.e. focuses on group averages vs. age-related changes). All models controlled for batch effects, sex, medications, wave, and day of saliva sample collection (see Table S7 for full model results). Finally, intra-class coefficients (ICC) of the random effects of the model were relatively low (ICC = 0.22), indicating high variability of morning cortisol across waves at the within-subject level.

Evening cortisol: We repeated the piecewise model with evening cortisol only, using the same age-breakpoint of 13.1 years old. Group did not moderate the piecewise effects of age before or after 13.1 years old (see Table S8 for full model results). In contrast to morning cortisol, which showed the best model fit with piecewise age terms (i.e. lower AIC; Table 3), using piecewise age terms for evening cortisol did not provide a better model fit relative to a linear age term (i.e. difference in AIC < 1; Table 3). Together, these results suggest that the Piecewise Age X Group X Time of Day effects observed in the omnibus diurnal cortisol model were driven by age-related shifts in morning cortisol in the PI group.

Table 3: Cortisol model comparisons. Piecewise age models provided the best model fit for both diurnal cortisol and morning cortisol data. AIC = Akaike information criterion.

Model	DF	AIC
Diurnal Cortisol		
Linear Age x Group x Time of Day	17.00	17,955.98
Quadratic Age x Group x Time of Day	21.00	17,970.32
Piecewise Age x Group x Time of Day	**21.00**	**17,942.41**
Morning cortisol		
Linear Age x Group	13.00	9,900.62
Quadratic Age x Group	15.00	9,906.29
Piecewise Age X Group	**15.00**	**9,892.28**
Evening cortisol		
Linear Age x Group	13.00	7,380.11
Quadratic Age x Group	13.00	7,390.83
Piecewise Age x Group	13.00	7,379.13

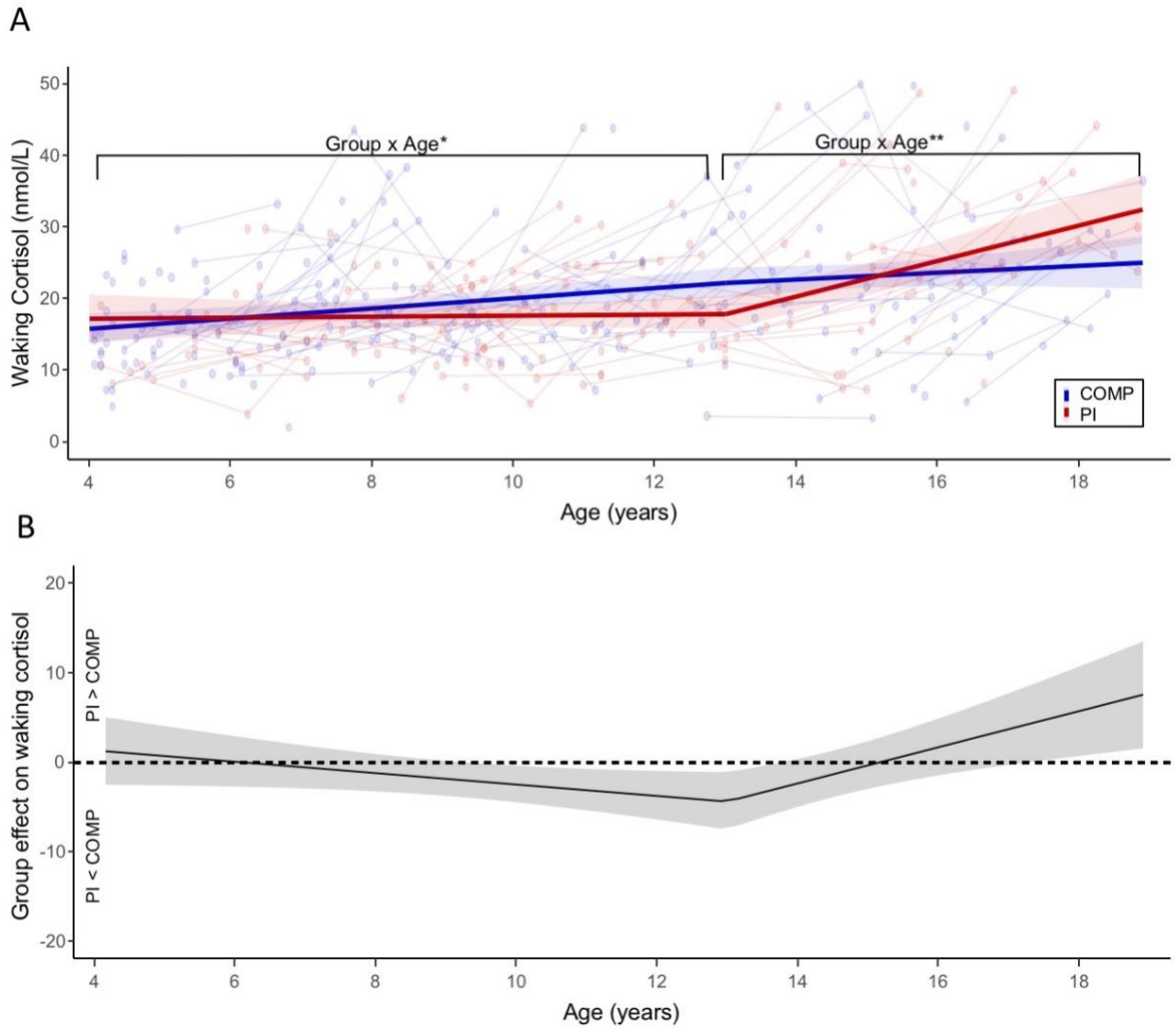

Figure 2: Effects of PI status on morning cortisol depend on age. **(A)** Fitted results of piecewise Age X Group effects on morning cortisol are depicted. Raw cortisol values (lines connecting within-subject observations) are shown with a 95% CI band around the fitted regression lines. **(B**) The region of significant plot is shown, depicting the magnitude of group differences in morning cortisol (PI – COMP) across the entire age-range. When the 95% CI band is above zero, morning cortisol is significantly higher in PI group than the comparison group, and when the 95% CI is below zero, morning cortisol is significantly lower in the PI group than the comparison group.

Age-dependent effects of early caregiving adversity on amygdala volume

Linear mixed effects modeling revealed a significant Group X Age interaction for bilateral amygdala volume (b = -0.15, t (295) = -2.95, p < 0.01, CI = -0.26, -0.05). Because amygdala growth can exhibit non-linear changes, further analyses tested whether group differences in age-effects were best fit by linear, quadratic, or piecewise models. Relative to linear (AIC = 1049.75) and quadratic age models (AIC = 1023.08), the piecewise age model had the best model fit (AIC = 1015.98), revealing an optimal breakpoint at age 9.5 years (CI: 8.80, 10.90). As shown in Figure 3, we identified a significant Group X Age interaction on amygdala volume before age 9.5 (b = -0.45, t (214) = -3.05, p < 0.01, CI = -0.74, -0.16), such that amygdala volume increased with age in the comparison group (b = 0.65, t (250.99) = 8.26, p < 0.001, CI = 0.50, 0.81), but not in the PI group (b = 0.20, t (250.99) = 1.60, p = 0.11, CI = -0.05, 0.45). After 9.5 years old, there was no Group X Age interaction (b = -0.02, t (283.57) = -0.39, p = 0.7, CI = -0.15, 0.10), and no significant age-related change was observed in the comparison group (b = -0.01, t (295.65) = -0.17, p = 0.868, CI = -0.09, 0.08) or the PI group (b = -0.03, t (295.65) = -0.61, p = 0.543, CI = -0.13, 0.07). This model controlled for the effect of scanners, scan quality by motion assessment, intracranial volume (ICV), and sex (see Table S9 for full model results). A regions of significance plot (Figure 3B) shows that the PI group had significantly smaller amygdala volumes between ages 11 and 16.5 and larger estimated amygdala volumes before age 6.5, although estimations in the youngest and oldest age range should be interpreted with caution due to imbalanced observations between groups. Follow-up analyses in restricted age-ranges from 6-19 showed similar piecewise effects (see Supplemental Analyses). Importantly, ICC estimates of amygdala volume within-subject from the piecewise model are 0.77, indicating high within-subject reliability across waves of the study.

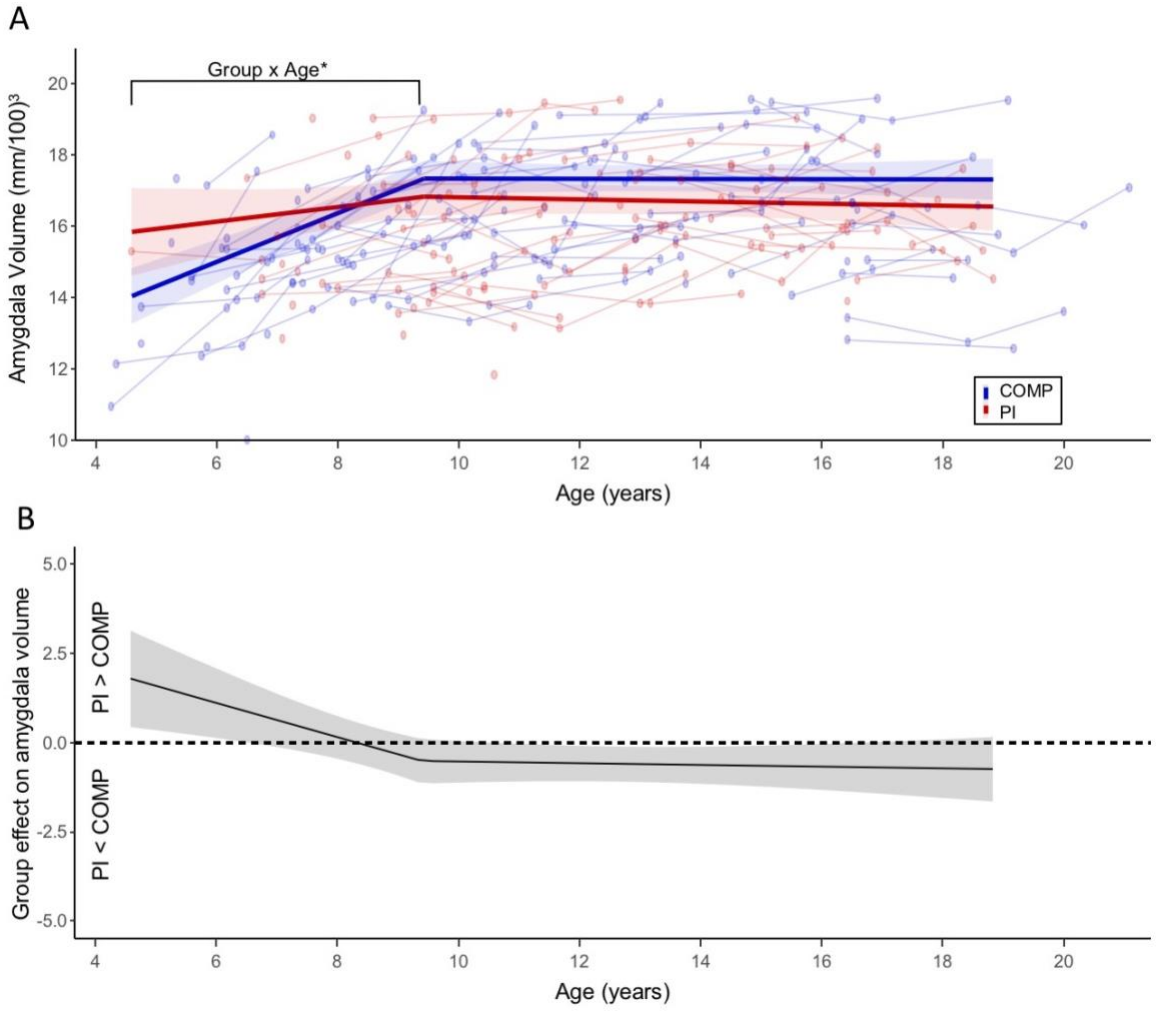

Figure 3. Effects of PI status on amygdala volume depend on age. **(A)** Fitted results of piecewise Age X Group effects on amygdala volume are depicted. Raw data (lines connecting within-subject observations) are shown with a 95% CI band around the fitted regression lines. **(B)** The region of significant plot is shown, depicting the magnitude of group differences in amygdala volume (PI – COMP) across the entire age-range. When the 95% CI band is above zero, amygdala volume is estimated to be significantly larger in PI group than the comparison group, and when the 95% CI is below zero, this indicates the age range when amygdala volume is significantly smaller in the PI group than the comparison group.

Age-invariant effects of early caregiving adversity on hippocampal volume

We observed a significant main effect of group on hippocampus volume (b = -1.78, t (129.82) = -3.24, $p < 0.01$, CI = -2.86, -0.71), such that PI youth had smaller hippocampal volumes relative to comparisons (Figure 4). We also detected a main effect of age (b = 0.23, t (290.15) = 4.03, $p < 0.001$, CI = 0.12, 0.34) on hippocampal volume. In comparison to amygdala volume, PI status did not moderate the effects of age on hippocampal volume (b = -0.10, t (252.55) = -1.09, p = 0.277, CI = -0.29, 0.08). These findings suggest that although early institutional care is associated with reduced size of the hippocampus, this early experience does not impact the rate of age-related growth in hippocampal volume across childhood and adolescence. The observed hippocampal group differences remained significant when controlling for the effects of scanners, scan quality by motion assessment, ICV, and sex (see Table S10 for full model results). Follow-up analyses testing left and right hemispheres separately, and limiting the age-range to 6-16, showed similar group and age main effects on hippocampal volume (see Supplemental Analyses). Importantly, hippocampal volumes showed a within-subject ICC of 0.86, indicating high reliability of hippocampal measurements across waves at the within-subject level.

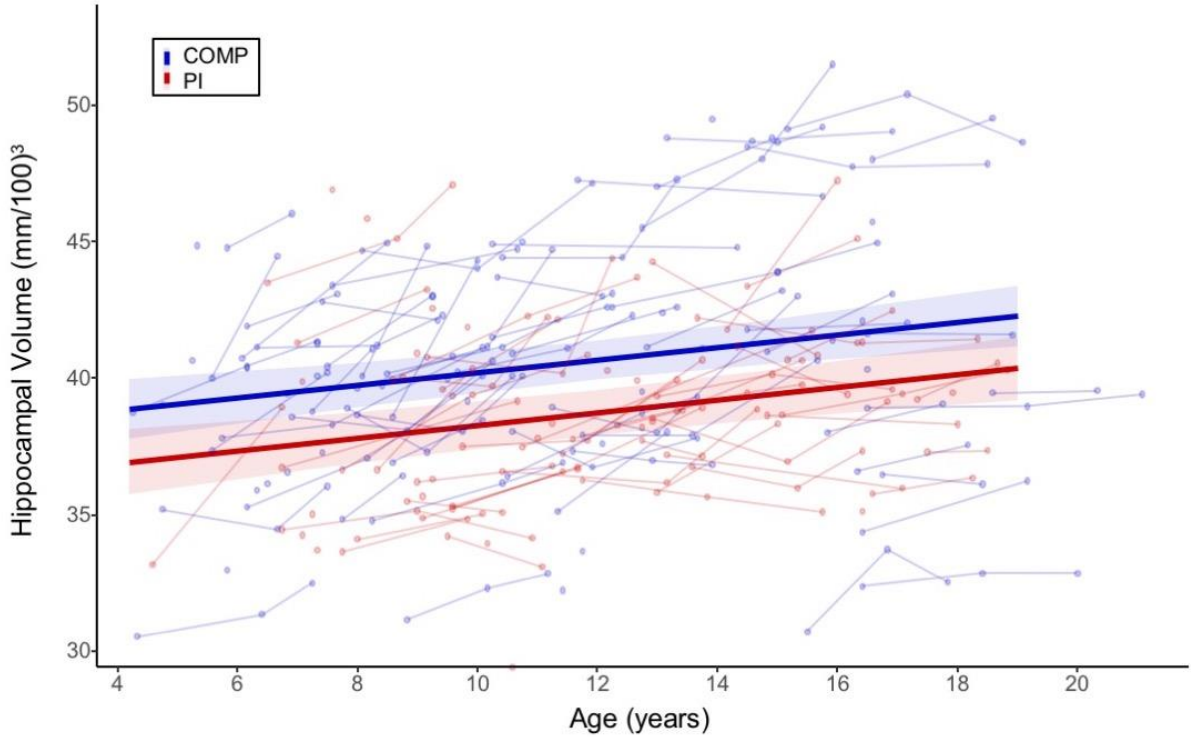

Figure 4. PI status is associated with smaller hippocampal volume across the entire age-range. Main effects of group and age on hippocampal volume are depicted. Fitted lines are shown for PI and comparison groups separately for visualization purposes only. Raw data (lines connecting within-subject observations) are shown with fitted regression lines and 95% CI bands.

Prospective associations between limbic brain volume and morning cortisol

Cross-lagged structural equation modeling was used to investigate bidirectional associations between morning cortisol and limbic brain volume over time. Amygdala volume at T1 significantly predicted morning cortisol at T2 (Figure 5), over and above the effects of baseline cortisol, such that morning cortisol at T2 is better predicted by amygdala volume at T1 than morning cortisol at T1 (see Table S11 for full model results). Importantly, morning cortisol at T1 did not predict amygdala volume at T2. Model fit indices show that the cross-lagged model provided an acceptable model fit (CFI = 0.91, RMSEA = 0.11, SRMR = 0.08, *Chi-square*

= 81.70, df = 21, $p < 0.001$). Further, we conducted nested model comparisons to evaluate the relevance of amygdala T1 coefficient for the model fit. Relative to a model that omitted this coefficient, we detected a significantly better model fit when including amygdala volume at T1 in the cross-lag path predicting morning cortisol at T2 (*Chi-square* (1) = 5.32, $p < 0.05$).

Parallel results were found when modeling bidirectional effects between hippocampus and cortisol (Figure 5), such that hippocampus volume at T1 predicted morning cortisol at T2, but cortisol at T1 did not predict hippocampal volume at T2 (Table S12 for full model results). Model fit indices were acceptable (CFI = 0.93, RMSEA = 0.10, SRMR = 0.05, *Chi-square* = 70.47, df = 21, $p < 0.001$) and nested model comparisons showed that including hippocampal volume at T1 provided a significantly better fit than a model omitting that coefficient from the cross-lag path (*Chi-square* (1) = 7.85, $p < 0.05$).

To visualize these effects and verify SEM results, separate linear regressions were performed with only the significant cross-lagged paths for both amygdala and hippocampal models. Specifically, we regressed morning cortisol at T2 on amygdala / hippocampus T1, controlling for the same set of covariates (morning cortisol at T1, age, group, sex, and ICV). As shown in Figure 6, morning cortisol at T2 was significantly predicted by amygdala volume at T1 ($b = 2.730$, t (63) = 2.803, $p < 0.01$) and hippocampal volume at T1 ($b = 0.931$, t (63) = 2.256, $p = 0.027$).

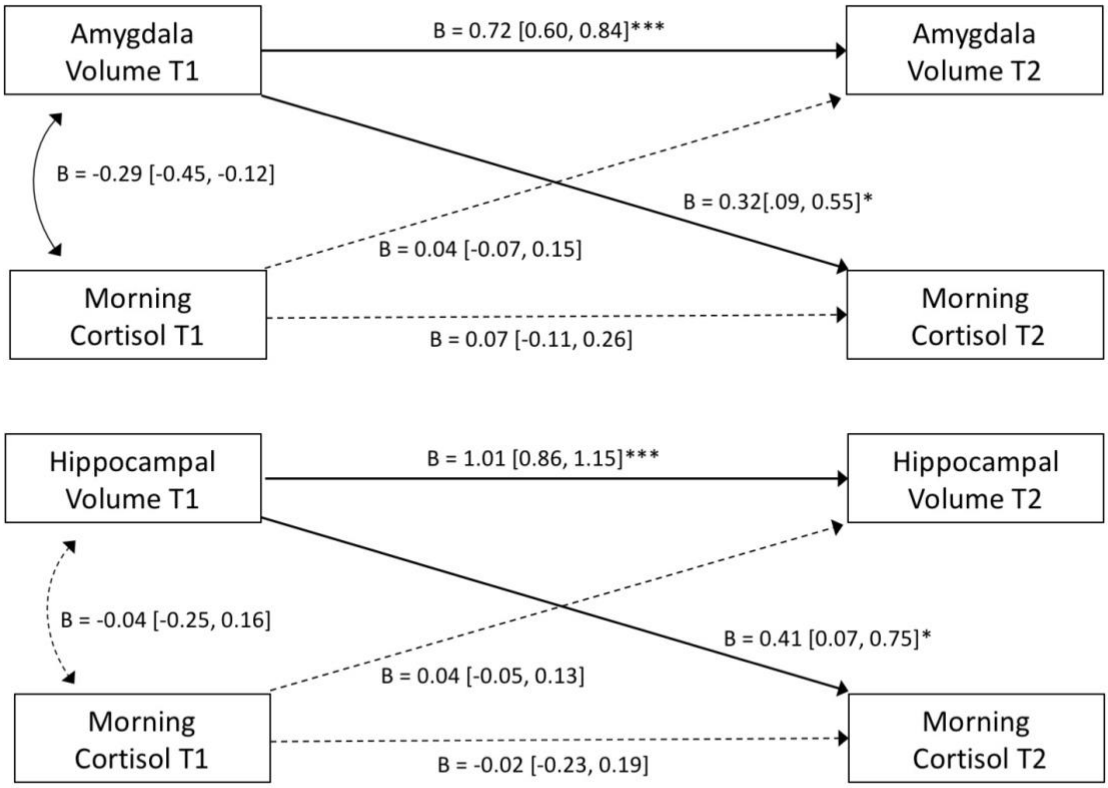

Figure 5. Cross-lagged SEM models show that amygdala and hippocampal volume are prospectively associated with future morning cortisol, controlling for morning cortisol at T1, group, sex, age, and ICV. Standardized model coefficients are provided with 95% CI. Separate models were performed for amygdala and hippocampus and are shown in the same figure for visualization purposes only. *p < 0.05, **p < 0.01, ***p<0.001.

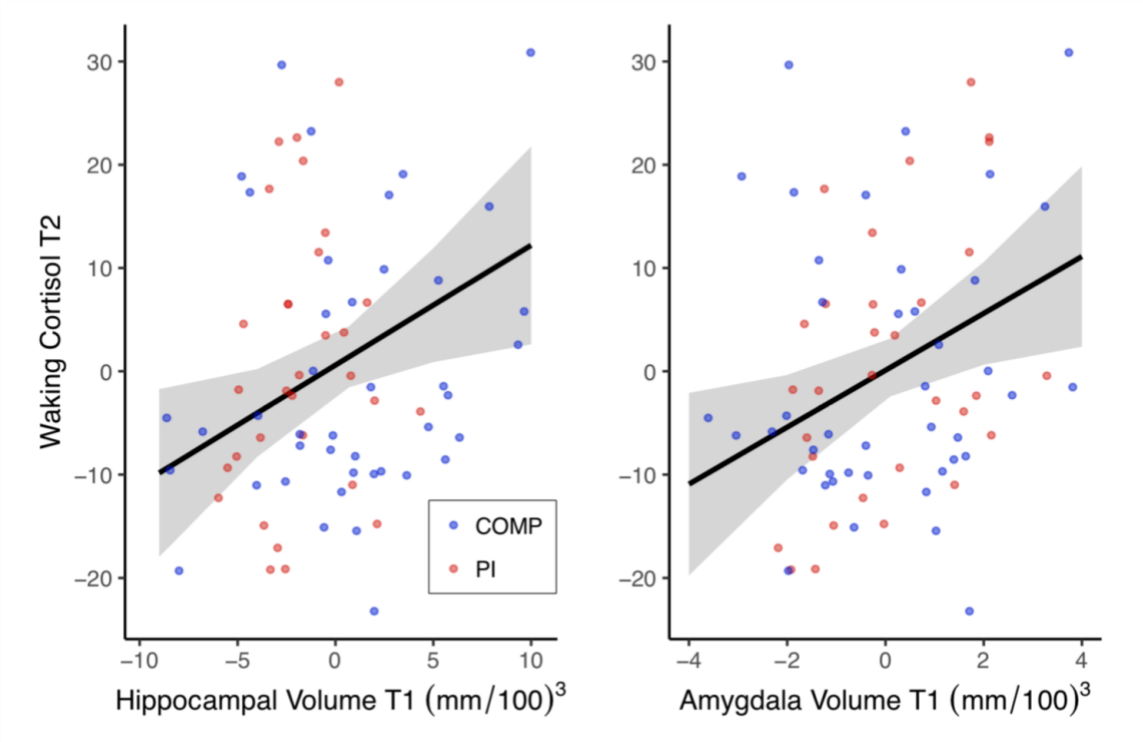

Figure 6: Regressions depicting positive linear relationship between amygdala and hippocampus volume at T1 and morning cortisol at T2, controlling for morning cortisol at T1. Fitted estimates with 95% CI and raw data are shown with mean-centered values.

In separate models, we tested whether PI status moderated the effects of limbic brain volume at T1 on morning cortisol at T2 (see Tables S13 & S14 for full model results). For both amygdala and hippocampal models, no significant group interaction was detected (Amygdala: *b* = 1.67, *z* = 1.14, *p* = 0.256; Hippocampus: *b* = -0.21, *b* = -0.21, *z* = -0.20, *p* = 0.845) and model fit significantly worsened when adding these interaction terms (Amygdala X Group: *Chi-square* (11), = 125.100, p < 0.05; Hippocampus X Group: *Chi-square* (11) = 136.324, p < 0.05). Based on these results, we conclude that early experience does not influence the observed relationship

between limbic brain volume at T1 and morning cortisol at T2. However, it is important to note that the relatively small sample size for these analyses limits the power to detect interaction effects.

Limbic brain volume and internalizing symptoms

PI youth had significantly higher parent-reported internalizing scores relative to comparisons (b = 9.62, t (132) = 6.15, p < 0.001, CI = 6.61, 12.62). While controlling for age and group effects, hippocampal volume was negatively associated with internalizing scores (b = -0.57, t (197) = -2.43, p = 0.016, CI = -1.01, -0.12) such that smaller hippocampal volume was associated with higher symptoms (Figure 7). When accounting for hippocampal effects in the same model, amygdala was independently associated with internalizing symptoms (b = 1.00, t (209) = 2.01, p = 0.046, CI = 0.04, 1.96), such that larger amygdala volume was associated with higher symptoms (Figure 7). This model controlled for group, sex, ICV, scanners, medication, and age (see Table S15 for full model results). These data suggest that although amygdala and hippocampus volumes had similar longitudinal effects on morning cortisol, they had divergent effects on internalizing behaviors.

Morning cortisol did not have any significant main effects on internalizing symptoms (b = 0.004, t (166) = -0.06, p = 0.951, CI = -0.10, 0.09). Follow-up analyses showed a significant Group X Age X Morning cortisol interaction (b = -0.068, t (162.495) = -2.718, p < 0.01, CI = -0.116, -0.02) on internalizing symptoms (see Table S16 for full model results). However, post-hoc analyses that tested the relationship between morning cortisol and internalizing symptoms at 3 different ages (-1SD, mean age, +1 SD) within each group did not reveal significant associations (Table S17). These results suggest that although the relationship between morning

cortisol and symptoms differed as a function of age and group, variance in morning cortisol was not directly associated with symptom scores for any given subset of the sample.

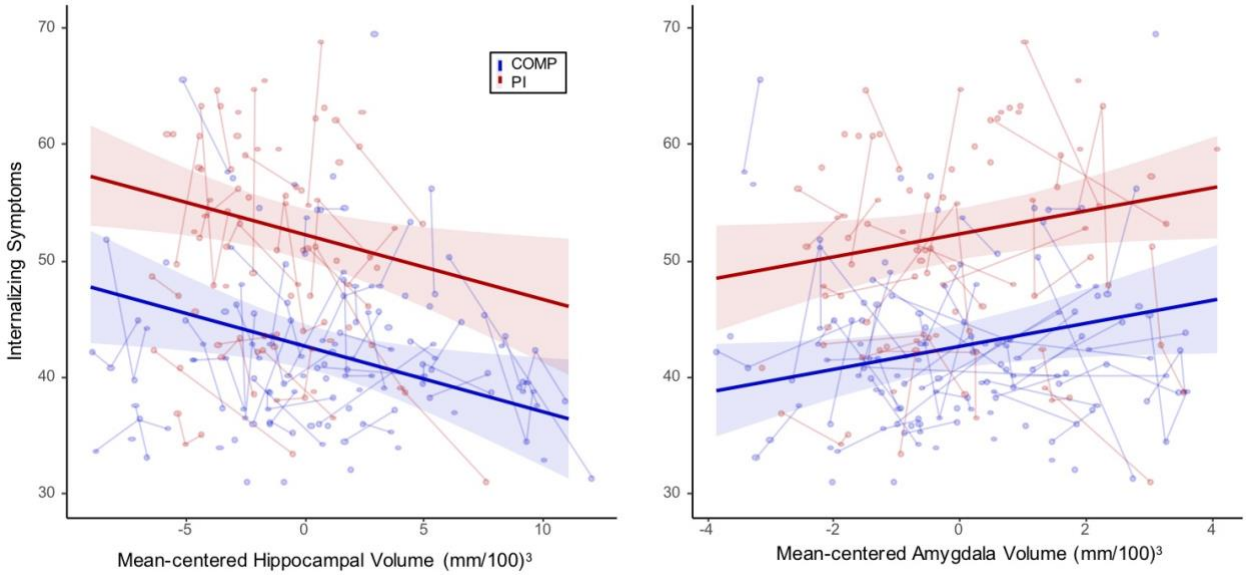

Figure 7. Hippocampal and amygdala volume are associated with internalizing scores. Linear mixed model results show that hippocampal volume is negatively associated with internalizing symptoms, and amygdala volume is positively associated with internalizing symptoms, controlling for group, sex, age, and ICV. Fitted regression lines are depicted with 95% CI bands and raw data (lines connecting within subject observations).

Discussion

The present study demonstrates that stress physiology and limbic brain volumes change dynamically across development and as a function of early caregiving adversity. We showed that the effects of early institutional care on amygdala volume and morning cortisol depend on age, whereas adversity-related reductions in hippocampal volume are age-invariant from childhood to early adulthood. When testing longitudinal cross-lagged relationships between cortisol and limbic brain development, amygdala and hippocampal volumes were prospectively associated with higher morning cortisol levels, suggesting a feed-forward relationship to the HPA-axis during childhood and adolescence. A smaller hippocampus was associated with greater internalizing symptoms across both groups, whereas a larger amygdala was associated with greater internalizing symptoms across both groups. These findings emphasize the importance of age of measurement when interpreting adversity-related phenotypes and in the utility of longitudinal studies to identify periods of development when recalibration of stress physiology may occur.

We identified a developmental shift in ECA-related cortisol phenotypes, such that PI group showed blunted morning cortisol during childhood and heightened levels by late adolescence. These non-linear age effects provide a developmental framework to integrate prior discrepancies in the literature of both blunted and heightened morning cortisol phenotypes (King et al., 2017; Quevedo et al., 2012). The observed pattern of blunted morning cortisol in PI children is consistent with numerous prior studies (Bernard et al., 2015; Dozier et al., 2006; Fisher & Stoolmiller, 2008; Koss et al., 2014; Pitula et al., 2019; Zalewski et al., 2016), and is hypothesized to emerge over time as a result of excess cortisol in response to a chronic stressor (Miller et al., 2007). In contrast, the marked increase in morning cortisol observed in PI

adolescents corresponds with recent work suggesting pubertal recalibration in stress physiology, which has been observed in both morning cortisol levels (Flannery et al., 2017; King et al., 2017; Quevedo et al., 2012) and cortisol reactivity to stressors (DePasquale et al., 2019; Gunnar et al., 2019). Consistent with this idea, secondary analyses showed effects of puberty, independent of age, on morning cortisol in both PI and comparison groups (see Supplement). This adolescent-specific plasticity may allow for positive influences (e.g. time with adoptive family) to recalibrate the HPA-axis for more adaptive functioning following early caregiving adversity, as has been indicated in prior studies (DePasquale et al., 2019; Flannery et al., 2017). However, cortisol was not associated with internalizing scores in the current study. Further within-person longitudinal research is needed to determine whether these developmental changes in stress physiology are associated with adaptive or maladaptive outcomes in PI youth.

The PI group showed altered amygdala volume growth, which also resulted in different group effects depending on the age of measurement. Comparisons showed significant age-related amygdala growth during childhood which leveled off around age 9-10, while the PI group showed relatively stable amygdala volumes from early childhood to early adulthood. Therefore, although the PI group had larger estimated amygdala volumes at the youngest ages, due to the lack of age-related growth during childhood, they exhibited smaller amygdala volumes relative to comparisons by adolescence. These findings place the existing human literature in a developmental context, suggesting that prior conflicting findings of hypotrophy and hypertrophy of the amygdala may reflect differences in age sampling. In particular, we note that the majority of cross-sectional studies reporting larger amygdala included younger participants (e.g. childhood, early adolescence; Buss et al., 2012; Lupien et al., 2011; Roth et al., 2018; Tottenham et al., 2010; but see Mehta et al., 2009). Further, a recent study with dense sampling between

ages 4-6 also showed larger amygdala volume in the context of parental insensitivity (Lee et al., 2019). In contrast, the majority of studies reporting smaller amygdala volume have a wider age-range, including participants in late adolescence and young adulthood (Noble et al., 2012; Edminston et al., 2011; but see Hanson et al., 2014; Luby, 2013) or only detected smaller amygdala volume in adolescence (Merz et al., 2017).

Prior hypotheses have suggested that early increases in amygdala volume, as shown in animal models (Castillo-Gómez et al., 2017; Eiland et al., 2012; Guadagno et al., 2018; Raineki et al., 2019; Vyas et al., 2002), may sensitize the amygdala to future stressors, resulting in amygdala atrophy later on (Teicher et al., 2016). However, the apparent reduced amygdala volume in the current study of PI adolescents is not due to decrease of amygdala volume over time, but instead reflects a lack of age-related growth in PI children relative to comparisons. As such, it is possible instead that the period of amygdala growth observed in childhood is shifted earlier in time by exposure to ECA, and results in a lower ceiling of possible maximum volume. Further research is needed to characterize amygdala volume changes across early life, beginning at the time of adversity exposure, to further determine how ECAs alter the developmental timing of amygdala growth.

In the current study, PI status was associated with reduced hippocampal volume across all ages, controlling for group differences in intracranial volume, as well as sex and scanner effects. These results are consistent with prior research showing adversity-related reductions in hippocampal volumes in samples of varying age-ranges (Calem et al., 2017; Dannlowski et al., 2012; Gorka et al., 2014; Hanson et al., 2011, 2014; Hodel et al., 2014; Humphreys et al., 2019; L.S. King et al., 2018; J. Luby, 2013; J. L. Luby et al., 2019; Noble et al., 2012; Piccolo & Noble, 2018; Rao et al., 2010; Riem et al., 2015; M. H. Teicher et al., 2012). The fact that these

reductions are stable by 4 years and persist throughout development suggests that ECA exposure has an enduring impact on hippocampal volume, in line with work suggesting that early life is a sensitive period for hippocampal development (Andersen et al., 2008; Humphreys et al., 2019). These data are also consistent with animal models, showing reduced volume following chronic stress exposure due to atrophy of dendritic branching of the hippocampus (Magarin˜os & McEwen, 1995; Vyas et al., 2002). However, not all human developmental studies have identified smaller hippocampal volume following ECAs (Lupien et al., 2011; Mehta et al., 2009; Sheridan et al., 2012; Tottenham et al., 2010); discrepancies that may be related to adversity type (King et al., 2018), timing (Humphreys et al., 2019), or ratio of males to females (Tottenham et al., 2010), as the group effect in the current study was driven by males (see Supplemental Analyses). We also detected independent effects of puberty on hippocampal volume (see Supplemental Analyses) consistent with prior work (Goddings et al., 2014; Wierenga et al., 2018), but these effects were not moderated by ECA exposure, providing further confidence that PI group showed typical rates of hippocampal growth, despite their smaller volume. Together, these findings suggest that although ECA is associated with reduced volume in both amygdala and hippocampus during adolescence, they reach these phenotypes via different developmental mechanisms.

The longitudinal design of the present study also allowed us to examine how limbic brain development and HPA-axis function influence each other over time. We show that baseline amygdala/hippocampal volumes predict future morning cortisol phenotypes, whereas baseline morning cortisol did not predict future amygdala and hippocampal volumes (2-4 years later). These findings build on prior human work showing positive associations between morning cortisol and hippocampal volume (Dahmen et al., 2017) and suggest that these systems are

coupled during development, such that that morning cortisol patterns are driven by earlier limbic brain development. Notably, although amygdala and hippocampal volumes were highly stable within subject, cortisol measurements had low within-subject reliability across waves, further indicating that diurnal cortisol is potentially more malleable across development. In non-human animal models, mineralocorticoid- and glucocorticoid receptors in the amygdala and hippocampus are involved in regulating the diurnal rhythm (Bradbury, 1994; Reul & de Kloet, 1985; ter Heegde et al., 2015) and the current data may indicate this relationship is present in human development as well. Although baseline morning cortisol was not associated with future amygdala or hippocampal volume in the current study, this does not preclude the possibility that at younger ages, closer to the adversity exposure, stress-related HPA axis activation may directionally influence early hippocampal and amygdala development, as has been observed in animal models of ECA exposure (Hatalski et al., 1998; Meaney et al., 1996; Plotsky et al., 2005; Raineki et al., 2019; Santiago et al., 2017).

In the current study, smaller hippocampal volume and larger amygdala volume was associated with higher levels of internalizing symptoms. The negative association between hippocampal volume and symptoms of anxiety and depression coincides with decades of research with both adversity-exposed and clinical samples (Bremner et al., 2000; Gorka et al., 2014; Koolschijn et al., 2013; Pagliaccio et al., 2014; Sheline et al., 1996; Vythilingam et al., 2002). However, prior research on amygdala volume and internalizing behaviors during development is mixed, with some studies showing significant positive associations (De Bellis et al., 2000; Pagliaccio et al., 2014; Roth et al., 2018; Tottenham et al., 2010), negative associations (Merz et al., 2017), or no associations (Koolschijn et al., 2013). Importantly, the significant positive effect of amygdala volume on internalizing problems was only observed when

controlling for the effects of hippocampal volume, suggesting that these brain areas have diverging influences on these behaviors, but these effects may be masked statistically when tested independently from each other. Further, these brain-behavior associations were not moderated by age or group in this sample, suggesting that despite the dynamic effects of development and early caregiving history on limbic brain volume, amygdala and hippocampus show relatively stable associations with behavioral outcomes.

There are several limitations to be noted in this study. First, the data in this study was collected as part of a large-scale study, whose design intentionally did not include longitudinal follow-ups for all participants. As a result, we were not able to model within-subject changes in cortisol or brain volume (Madhyastha et al., 2018) which may potentially reveal different associations at the single-subject level. We also note the imbalance between PI and comparison groups at younger and older ages in the MRI data. Secondary analyses conducted in a restricted age range (6-19 years) yielded similar results for both hippocampal and amygdala volume, providing confidence that the observed effects were not driven by leverage points in the younger age-ranges. However, given evidence that discrepancies in volume studies may also arise from methodological choices such as processing pipelines (Lyden et al., 2016), future research is needed to extend/replicate these findings, particular with higher rates of sampling at the extreme ages. Similarly, due to the limited sample size of participants with two usable data points of both cortisol and limbic brain volume, we were unable to test whether their bidirectional relationships differed as a function of age using cross-lagged structural equation modeling. It should be noted that while the cross-lagged structural equation modeling showed longitudinal associations between limbic brain volume and morning cortisol, we cannot deduce causal relationships from these models. Instead, they provide evidence to suggest directional coupling between these

stress-responsive systems over time. Continued translational work in animal models are needed to assess whether manipulations in amygdala and hippocampus development at specific ages influences future diurnal cortisol and vice versa.

In the current study, we recruited children with known history of institutional caregiving. Although the conditions and quality of institutions may vary, in general they are characterized as a form of psychosocial deprivation (Nelson, 2007) due to lack of species-expected caregiving (Tottenham, 2012). Internationally adopted youth are then adopted into stable caregiving settings, which allows for the unique opportunity to examine development following discrete adversity exposure with known timing and duration, in contrast to other forms of caregiving adversities that are often chronic (e.g., physical abuse; (Warmingham et al., 2019). In secondary analyses (see Supplement), consistent with previous studies (Gunnar et al., 2001; Kumsta et al., 2017), we detected dose-dependent effects of institutional care on morning cortisol (irrespective of age), but not limbic brain volume. It is important to note that dose-dependent effects of ECA on stress physiology and limbic neurobiology may change over development or wane over time, and in the current study, we were under-powered to test such interactions. Additional longitudinal research is needed to further assess how timing and chronicity of adversity exposure might influence the developmental trajectories of limbic brain development and stress physiology.

In conclusion, the present study demonstrated that ECA-related changes in stress physiology and limbic neurobiology depend on the developmental stage. Although we observed smaller amygdala and hippocampal volumes in PI adolescents, these volume reductions emerged via different developmental mechanisms (i.e. smaller volume vs. reduced growth). Further,

although the effects of ECA on amygdala and hippocampal volume were relatively stable after age 10 years, stress physiology continued to show dramatic developmental shifts during adolescence. By capitalizing on longitudinal data across wide age-ranges, we can gain greater resolution into the developmental sequelae of early caregiving adversity, with critical implications for identifying sensitive windows of development during which potential recalibration of stress physiology can occur.

References

Afacan, O., Erem, B., Roby, D. P., Roth, N., Roth, A., Prabhu, S. P., & Warfield, S. K. (2016). Evaluation of motion and its effect on brain magnetic resonance image quality in children. *Pediatric Radiology*, *46*(12), 1728–1735. https://doi.org/10.1007/s00247-016-3677-9

Andersen, S. L., Tomada, A., Vincow, E. S., Valente, E., Polcari, A., & Teicher, M. H. (2008). Preliminary Evidence for Sensitive Periods in the Effect of Childhood Sexual Abuse on Regional Brain Development. *The Journal of Neuropsychiatry and Clinical Neurosciences*, *20*(3), 292–301. https://doi.org/10.1176/jnp.2008.20.3.292

Avishai-Eliner, S., Yi, S. J., & Baram, T. Z. (1996). Developmental profile of messenger RNA for the corticotropin-releasing hormone receptor in the rat limbic system. *Brain Research. Developmental Brain Research*, *91*(2), 159–163. https://doi.org/0165380695001581 [pii]

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bernard, K., Zwerling, J., & Dozier, M. (2015). Effects of early adversity on young children's diurnal cortisol rhythms and externalizing behavior. *Developmental Psychobiology*, *57*(8), 935–947. https://doi.org/10.1002/dev.21324

Bradbury, M. J. (1994). Roles of type I and II corticosteroid receptors in regulation of basal activity in the hypothalamo-pituitary-adrenal axis during the diurnal trough and the peak: Evidence for a nonadditive effect of combined receptor occupation. *Endocrinology*, *134*(3), 1286–1296. https://doi.org/10.1210/en.134.3.1286

Bremner, J. D., Narayan, M., Anderson, E. R., Staib, L. H., Miller, H. L., & Charney, D. S. (2000). Hippocampal volume reduction in major depression. *The American Journal of Psychiatry*, *157*(1), 115–118. https://doi.org/10.1176/ajp.157.1.115

Burghy, C. A., Stodola, D. E., Ruttle, P. L., Molloy, E. K., Armstrong, J. M., Oler, J. a, Fox, M. E., Hayes, A. S., Kalin, N. H., Essex, M. J., Davidson, R. J., & Birn, R. M. (2012). Developmental pathways to amygdala-prefrontal function and internalizing symptoms in adolescence. *Nature Neuroscience*, *15*(12), 1736–1741. https://doi.org/10.1038/nn.3257 PMCID: PMC3509229

Buss, C., Davis, E. P., Shahbaba, B., Pruessner, J. C., Head, K., & Sandman, C. A. (2012). Maternal cortisol over the course of pregnancy and subsequent child amygdala and hippocampus volumes and affective problems. *Proceedings of the National Academy of Sciences*, *109*(20), E1312–E1319. https://doi.org/10.1073/pnas.1201295109

Butterworth, P., Cherbuin, N., Sachdev, P., & Anstey, K. J. (2012). The association between financial hardship and amygdala and hippocampal volumes: Results from the PATH through life project. *Social Cognitive and Affective Neuroscience*, *7*(5), 548–556. https://doi.org/10.1093/scan/nsr027

Calem, M., Bromis, K., McGuire, P., Morgan, C., & Kempton, M. J. (2017). Meta-analysis of associations between childhood adversity and hippocampus and amygdala volume in non-clinical and general population samples. *NeuroImage: Clinical*, *14*, 471–479. https://doi.org/10.1016/j.nicl.2017.02.016

Castillo-Gómez, E., Pérez-Rando, M., Bellés, M., Gilabert-Juan, J., Llorens, J. V., Carceller, H., Bueno-Fernández, C., García-Mompó, C., Ripoll-Martínez, B., Curto, Y., Sebastiá-Ortega, N., Moltó, M. D., Sanjuan, J., & Nacher, J. (2017). Early Social Isolation Stress

and Perinatal NMDA Receptor Antagonist Treatment Induce Changes in the Structure and Neurochemistry of Inhibitory Neurons of the Adult Amygdala and Prefrontal Cortex. *ENeuro*, *4*(2). https://doi.org/10.1523/ENEURO.0034-17.2017

Dabbs, J. M. (1991). Salivary testosterone measurements: Collecting, storing, and mailing saliva samples. *Physiology & Behavior*, *49*(4), 815–817. https://doi.org/10.1016/0031-9384(91)90323-G

Dahmen, B., Puetz, V. B., Scharke, W., von Polier, G. G., Herpertz-Dahlmann, B., & Konrad, K. (2017). Effects of Early-Life Adversity on Hippocampal Structures and Associated HPA Axis Functions. *Developmental Neuroscience*. https://doi.org/10.1159/000484238

Daniel Lüdecke. (2018). *sjstats: Statistical Functions for Regression Models.* Zenodo. https://doi.org/10.5281/zenodo.1489175

Dannlowski, U., Stuhrmann, A., Beutelmann, V., Zwanzger, P., Lenzen, T., Grotegerd, D., Domschke, K., Hohoff, C., Ohrmann, P., Bauer, J., Lindner, C., Postert, C., Konrad, C., Arolt, V., Heindel, W., Suslow, T., & Kugel, H. (2012). Limbic scars: Long-term consequences of childhood maltreatment revealed by functional and structural magnetic resonance imaging. *Biological Psychiatry*, *71*(4), 286–293. https://doi.org/10.1016/j.biopsych.2011.10.021

Davis, E. G., Humphreys, K. L., McEwen, L. M., Sacchet, M. D., Camacho, M. C., MacIsaac, J. L., Lin, D., Kobor, M. S., & Gotlib, I. H. (2017). Diurnal cortisol, DNA methylation age, and hippocampal volume in adolescent girls. *Nature Publishing Group*, *7*(June), 10–14. https://doi.org/10.1038/tp.2017.188

De Bellis, M. D., Casey, B. J., Dahl, R. E., Birmaher, B., Williamson, D. E., Thomas, K. M., Axelson, D. A., Frustaci, K., Boring, A. M., Hall, J., & Ryan, N. D. (2000). A pilot study

of amygdala volumes in pediatric generalized anxiety disorder. *Biological Psychiatry*, *48*(1), 51–57. https://doi.org/10.1016/S0006-3223(00)00835-0

DePasquale, C. E., Donzella, B., & Gunnar, M. R. (2019). Pubertal recalibration of cortisol reactivity following early life stress: A cross-sectional analysis. *Journal of Child Psychology and Psychiatry*, *60*(5), 566–575. https://doi.org/10.1111/jcpp.12992

Dozier, M., Manni, M., Gordon, M. K., Peloso, E., Gunnar, M. R., Stovall-Mcclough, K. C., Eldreth, D., & Levine, S. (2006). Foster children's diurnal production of cortisol: An exploratory study. *Child Maltreatment*, *11*(2), 189–197. https://doi.org/10.1177/1077559505285779

Ebesutani, C., Chorpita, B. F., Higa-Mcmillan, C. K., Nakamura, B. J., Regan, J., & Lynch, R. E. (2011). A psychometric analysis of the revised child anxiety and depression scales-parent version in a school sample. *Journal of Abnormal Child Psychology*, *39*, 173–185. https://doi.org/10.1007/s10802-010-9460-8

Ebesutani, C., Tottenham, N., & Chorpita, B. (2015). The Revised Child Anxiety and Depression Scale—Parent Version: Extended Applicability and Validity for Use with Younger Youth and Children with Histories of Early-Life Caregiver Neglect. *Journal of Psychopathology and Behavioral Assessment*, *37*(4), 705–718. https://doi.org/10.1007/s10862-015-9494-x

Edmiston, E. E. (2011). Corticostriatal-Limbic Gray Matter Morphology in Adolescents With Self-reported Exposure to Childhood Maltreatment. *Archives of Pediatrics & Adolescent Medicine*, *165*(12), 1069. https://doi.org/10.1001/archpediatrics.2011.565

Eiland, L., Ramroop, J., Hill, M. N., Manley, J., & McEwen, B. S. (2012). Chronic juvenile stress produces corticolimbic dendritic architectural remodeling and modulates emotional

behavior in male and female rats. *Psychoneuroendocrinology*, *37*(1), 39–47. https://doi.org/10.1016/j.psyneuen.2011.04.015

Ellwood-Lowe, M. E., Humphreys, K. L., Ordaz, S. J., Camacho, M. C., Sacchet, M. D., & Gotlib, I. H. (2018). Time-varying effects of income on hippocampal volume trajectories in adolescent girls. *Developmental Cognitive Neuroscience*, *30*, 41–50. https://doi.org/10.1016/j.dcn.2017.12.005

Fischl, B., Salat, D. H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., van der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., & Dale, A. M. (2002). Whole Brain Segmentation: Automated Labeling of Neuroanatomical Structures in the Human Brain. *Neuron*, *33*(3), 341–355. https://doi.org/10.1016/S0896-6273(02)00569-X

Fisher, P. A., & Stoolmiller, M. (2008). Intervention effects on foster parent stress: Associations with child cortisol levels. *Development and Psychopathology*, *20*(3), 1003–1021. https://doi.org/10.1017/S0954579408000473

Flannery, J. E., Gabard-Durnam, L. J., Shapiro, M., Goff, B., Caldera, C., Louie, J., Gee, D. G., Telzer, E. H., Humphreys, K. L., Lumian, D. S., & Tottenham, N. (2017). Diurnal cortisol after early institutional care—Age matters. *Developmental Cognitive Neuroscience*, *25*(March), 160–166. https://doi.org/10.1016/j.dcn.2017.03.006

Gee, D. G., Gabard-Durnam, L. J., Flannery, J., Goff, B., Humphreys, K. L., Telzer, E. H., Hare, T. a, Bookheimer, S. Y., & Tottenham, N. (2013). Early developmental emergence of human amygdala-prefrontal connectivity after maternal deprivation. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(39), 15638–15643. https://doi.org/10.1073/pnas.1307893110, PMCID: PMC3785723

Gilmore, J. H., Shi, F., Woolson, S. L., Knickmeyer, R. C., Short, S. J., Lin, W., Zhu, H., Hamer, R. M., Styner, M., & Shen, D. (2012). Longitudinal Development of Cortical and Subcortical Gray Matter from Birth to 2 Years. *Cerebral Cortex*, *22*(11), 2478–2485. https://doi.org/10.1093/cercor/bhr327

Goddings, A. L., Mills, K. L., Clasen, L. S., Giedd, J. N., Viner, R. M., & Blakemore, S. J. (2014). The influence of puberty on subcortical brain development. *NeuroImage*, *88*, 242–251. https://doi.org/10.1016/j.neuroimage.2013.09.073

Gorka, A. X., Hanson, J. L., Radtke, S. R., & Hariri, A. R. (2014). Reduced hippocampal and medial prefrontal gray matter mediate the association between reported childhood maltreatment and trait anxiety in adulthood and. *Biology of Mood & …*, 1–10.

Green, J. G., Mclaughlin, K. a, Berglund, P. a, Gruber, M. J., Sampson, N. a, Zaslavsky, A. M., & Kessler, R. C. (2010). Childhood Adversities and Adult Psychiatric Disorders in the National Comorbidity Survey Replication I. *Archives of General Psychiatry*, *67*(2), 113–123. https://doi.org/10.1001/archgenpsychiatry.2009.187, PMCID: PMC2822662

Guadagno, A., Wong, T. P., & Walker, C.-D. (2018). Morphological and functional changes in the preweaning basolateral amygdala induced by early chronic stress associate with anxiety and fear behavior in adult male, but not female rats. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *81*, 25–37. https://doi.org/10.1016/j.pnpbp.2017.09.025

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

Gunnar, M R, Bruce, J., & Grotevant, H. D. (2000). International adoption of institutionally reared children: Research and policy. *Development and Psychopathology*, *12*(4), 677–693. https://doi.org/10.1017/S0954579400004077

Gunnar, M. R., Morison, S. J., Chisholm, K., & Schuder, M. (2001). Salivary cortisol levels in children adopted from romanian orphanages. *Development and Psychopathology*, *13*(3), 611–628.

Gunnar, Megan R., DePasquale, C. E., Reid, B. M., & Donzella, B. (2019). Pubertal stress recalibration reverses the effects of early life stress in postinstitutionalized children. *Proceedings of the National Academy of Sciences*, 201909699. https://doi.org/10.1073/pnas.1909699116

Gunnar, Megan R, & Quevedo, K. M. (2008). Early care experiences and HPA axis regulation in children: A mechanism for later trauma vulnerability. *Progress in Brain Research*, *167*, 137–149. https://doi.org/10.1016/S0079-6123(07)67010-1

Hanson, J. L., Chandra, A., Wolfe, B. L., & Pollak, S. D. (2011). Association between income and the hippocampus. *PLoS ONE*, *6*(5), 1–8. https://doi.org/10.1371/journal.pone.0018712

Hanson, J. L., Nacewicz, B. M., Sutterer, M. J., Cayo, A. a, Schaefer, S. M., Rudolph, K. D., Shirtcliff, E. a, Pollak, S. D., & Davidson, R. J. (2014). Behavioral Problems After Early Life Stress: Contributions of the Hippocampus and Amygdala. *Biological Psychiatry*, 1–9. https://doi.org/10.1016/j.biopsych.2014.04.020

Hatalski, C. G., Guirguis, C., & Baram, T. Z. (1998). Corticotropin Releasing Factor mRNA Expression in the Hypothalamic Paraventricular Nucleus and the Central Nucleus of the

Amygdala is Modulated by Repeated Acute Stress in the Immature Rat. *Journal of Neuroendocrinology*, *10*(9), 663–669.

Hellerstedt, W. L., Madsen, N. J., Gunnar, M. R., Grotevant, H. D., Lee, R. M., & Johnson, D. E. (2008). The International Adoption Project: Population-based Surveillance of Minnesota Parents Who Adopted Children Internationally. *Maternal and Child Health Journal*, *12*(2), 162–171. https://doi.org/10.1007/s10995-007-0237-9

Herman, J. P., McKlveen, J. M., Solomon, M. B., Carvalho-Netto, E., & Myers, B. (2012). Neural regulation of the stress response: Glucocorticoid feedback mechanisms. *Brazilian Journal of Medical and Biological Research*, *45*(4), 292–298. https://doi.org/10.1590/S0100-879X2012007500041

Herting, M. M., Johnson, C., Mills, K. L., Vijayakumar, N., Dennison, M., Liu, C., Goddings, A.-L., Dahl, R. E., Sowell, E. R., Whittle, S., Allen, N. B., & Tamnes, C. K. (2017). Development of subcortical volumes across adolescence in males and females: A multisample study of longitudinal changes. *BioRxiv*, *172*(August 2017), 171389. https://doi.org/10.1101/171389

Hodel, A. S., Hunt, R. H., Cowell, R. a, Van Den Heuvel, S. E., Gunnar, M. R., & Thomas, K. M. (2014). Duration of early adversity and structural brain development in post-institutionalized adolescents. *NeuroImage*, *105C*, 112–119. https://doi.org/10.1016/j.neuroimage.2014.10.020

Humphreys, K. L., King, L. S., Sacchet, M. D., Camacho, M. C., Colich, N. L., Ordaz, S. J., Ho, T. C., & Gotlib, I. H. (2019). Evidence for a sensitive period in the effects of early life stress on hippocampal volume. *Developmental Science*, *22*(3), e12775. https://doi.org/10.1111/desc.12775

Keding, T. J., & Herringa, R. J. (2015). Abnormal Structure of Fear Circuitry in Pediatric Post-Traumatic Stress Disorder. *Neuropsychopharmacology*, *40*(3), 537–545. https://doi.org/10.1038/npp.2014.239

Kessler, R. C., McLaughlin, K. A., Green, J. G., Gruber, M. J., Sampson, N. A., Zaslavsky, A. M., Aguilar-Gaxiola, S., Alhamzawi, A. O., Alonso, J., Angermeyer, M., Benjet, C., Bromet, E., Chatterji, S., de Girolamo, G., Demyttenaere, K., Fayyad, J., Florescu, S., Gal, G., Gureje, O., … Williams, D. R. (2010). Childhood adversities and adult psychopathology in the WHO World Mental Health Surveys. *The British Journal of Psychiatry*, *197*(5), 378–385. https://doi.org/10.1192/bjp.bp.110.080499 PMCID: PMC2966503

King, L.S., Humphreys, K. L., Camacho, M. C., & Gotlib, I. H. (2018). A person-centered approach to the assessment of early life stress: Associations with the volume of stress-sensitive brain regions in early adolescence. *Development and Psychopathology*, *i*. https://doi.org/10.1017/S0954579418000184

King, Lucy S, Colich, N. L., Lemoult, J., Humphreys, K. L., Ordaz, S. J., Price, A. N., & Gotlib, I. H. (2017). The impact of the severity of early life stress on diurnal cortisol: The role of puberty. *Psychoneuroendocrinology*, *77*, 68–74. https://doi.org/10.1016/j.psyneuen.2016.11.024

Kirschbaum, C., & Hellhammer, D. (2000). Salivary cortisol. *Encycl. Stress*, *3*, 379–383.

Koolschijn, P. C. M. P., van IJzendoorn, M. H., Bakermans-Kranenburg, M. J., & Crone, E. A. (2013). Hippocampal volume and internalizing behavior problems in adolescence. *European Neuropsychopharmacology*, *23*(7), 622–628. https://doi.org/10.1016/j.euroneuro.2012.07.001

Koss, K. J., Hostinar, C. E., Donzella, B., & Gunnar, M. R. (2014). Social deprivation and the HPA axis in early development. *Psychoneuroendocrinology*, *50*, 1–13. https://doi.org/10.1016/j.psyneuen.2014.07.028

Kumsta, R., Schlotz, W., Golm, D., Moser, D., Kennedy, M., Knights, N., Kreppner, J., Maughan, B., Rutter, M., & Sonuga-Barke, E. (2017). HPA axis dysregulation in adult adoptees twenty years after severe institutional deprivation in childhood. *Psychoneuroendocrinology*, *86*, 196–202. https://doi.org/10.1016/j.psyneuen.2017.09.021

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(1), 1–26. https://doi.org/10.18637/jss.v082.i13

Lee, A., Poh, J. S., Wen, D. J., Tan, H. M., Chong, Y.-S., Tan, K. H., Gluckman, P. D., Fortier, M. V., Rifkin-Graboi, A., & Qiu, A. (2019). Maternal care in infancy and the course of limbic development. *Developmental Cognitive Neuroscience*, *40*. https://doi.org/10.1016/j.dcn.2019.100714

Liu, J., Chaplin, T., Wang, F., Sinha, R., Mayes, L. C., & Blumberg, H. P. (2012). Stress Reactivity and Corticolimbic Response to Emotional Faces in Adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry*, *51*(3), 304–312. https://doi.org/10.1016/j.jaac.2011.12.014

Luby, J. (2013). The Effects of Poverty on Childhood Brain Development. *JAMA Pediatrics*, *1093*(12), 1–8. https://doi.org/10.1001/jamapediatrics.2013.3139

Luby, J. L., Tillman, R., & Barch, D. M. (2019). Association of Timing of Adverse Childhood Experiences and Caregiver Support With Regionally Specific Brain Development in

Adolescents. *JAMA Network Open*, *2*(9), e1911426–e1911426. https://doi.org/10.1001/jamanetworkopen.2019.11426

Lupien, S. J., Parent, S., Evans, A. C., Tremblay, R. E., David, P., & Corbo, V. (2011). Larger amygdala but no change in hippocampal volume in 10-year-old children exposed to maternal depressive symptomatology since birth. *PNAS*, *108*(34), 14324–14329. https://doi.org/10.1073/pnas.1105371108/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1105371108

Lyden, H., Gimbel, S. I., Del Piero, L., Tsai, A. B., Sachs, M. E., Kaplan, J. T., Margolin, G., & Saxbe, D. (2016). Associations between Family Adversity and Brain Volume in Adolescence: Manual vs. Automated Brain Segmentation Yields Different Results. *Frontiers in Neuroscience*, *10*(September), 398. https://doi.org/10.3389/fnins.2016.00398

Maas, C. J. M., & Hox, J. J. (2004). Robustness issues in multilevel regression analysis. *Statistica Neerlandica*, *58*(2), 127–137. https://doi.org/10.1046/j.0039-0402.2003.00252.x

Madhyastha, T., Peverill, M., Koh, N., McCabe, C., Flournoy, J., Mills, K., King, K., Pfeifer, J., & McLaughlin, K. A. (2018). Current methods and limitations for longitudinal fMRI analysis across development. *Developmental Cognitive Neuroscience*, *33*, 118–128. https://doi.org/10.1016/j.dcn.2017.11.006

Magarin˜os, A. M., & McEwen, B. S. (1995). Stress-induced atrophy of apical dendrites of hippocampal CA3c neurons: Involvement of glucocorticoid secretion and excitatory amino acid receptors. *Neuroscience*, *69*(1), 89–98. https://doi.org/10.1016/0306-4522(95)00259-L

McEwen, B. S., Nasca, C., & Gray, J. D. (2016). Stress Effects on Neuronal Structure: Hippocampus, Amygdala, and Prefrontal Cortex. *Neuropsychopharmacology*, *41*(1), 3–23. https://doi.org/10.1038/npp.2015.171

Meaney, M. J., Diorio, J., Francis, D., Widdowson, J., LaPlante, P., Caldji, C., Sharma, S., Seckl, J. R., & Plotsky, P. M. (1996). Early environmental regulation of forebrain glucocorticoid receptor gene expression: Implications for adrenocortical responses to stress. *Developmental Neuroscience*, *18*(1–2), 49–72. https://doi.org/10.1159/000111395

Mehta, M. a, Golembo, N. I., Nosarti, C., Colvert, E., Mota, A., Williams, S. C. R., Rutter, M., & Sonuga-Barke, E. J. S. (2009). Amygdala, hippocampal and corpus callosum size following severe early institutional deprivation: The English and Romanian Adoptees study pilot. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *50*(8), 943–951. https://doi.org/10.1111/j.1469-7610.2009.02084.x

Merz, E. C., Tottenham, N., & Noble, K. G. (2017). Socioeconomic Status, Amygdala Volume, and Internalizing Symptoms in Children and Adolescents. *Journal of Clinical Child & Adolescent Psychology*, *00*(00), 1–12. https://doi.org/10.1080/15374416.2017.1326122

Miller, G. E., Chen, E., & Zhou, E. S. (2007). If it goes up, must it come down? Chronic stress and the hypothalamic-pituitary-adrenocortical axis in humans. *Psychological Bulletin*, *133*(1), 25–45. https://doi.org/10.1037/0033-2909.133.1.25

Myers, B., McKlveen, J. M., & Herman, J. P. (2012). Neural regulation of the stress response: The many faces of feedback. *Cellular and Molecular Neurobiology*. https://doi.org/10.1007/s10571-012-9801-y

Nelson, C. A. (2007). A Neurobiological Perspective on Early Human Deprivation. *Child Development Perspectives*, *1*(1), 13–18. https://doi.org/10.1111/j.1750-8606.2007.00004.x

Noble, K. G., Houston, S. M., Brito, N. H., Bartsch, H., Kan, E., Kuperman, J. M., Akshoomoff, N., Amaral, D. G., Bloss, C. S., Libiger, O., Schork, N. J., Murray, S. S., Casey, B. J., Chang, L., Ernst, T. M., Frazier, J. a, Gruen, J. R., Kennedy, D. N., Van Zijl, P., … Sowell, E. R. (2015). Family income, parental education and brain structure in children and adolescents. *Nature Neuroscience*, *18*(5). https://doi.org/10.1038/nn.3983

Noble, K. G., Houston, S. M., Kan, E., & Sowell, E. R. (2012). Neural correlates of socioeconomic status in the developing human brain. *Developmental Science*, *15*(4), 516–527. https://doi.org/10.1111/j.1467-7687.2012.01147.x

Ostby, Y., Tamnes, C. K., Fjell, A. M., Westlye, L. T., Due-Tonnessen, P., & Walhovd, K. B. (2009). Heterogeneity in Subcortical Brain Development: A Structural Magnetic Resonance Imaging Study of Brain Maturation from 8 to 30 Years. *Journal of Neuroscience*, *29*(38), 11772–11782. https://doi.org/10.1523/JNEUROSCI.1242-09.2009

Pagliaccio, D., Luby, J. L., Bogdan, R., Agrawal, A., Gaffrey, M. S., Belden, A. C., Botteron, K. N., Harms, M. P., & Barch, D. M. (2013). Stress-System Genes and Life Stress Predict Cortisol Levels and Amygdala and Hippocampal Volumes in Children. *Neuropsychopharmacology*, *39*(5), 1245–1253. https://doi.org/10.1038/npp.2013.327

Pagliaccio, D., Luby, J. L., Bogdan, R., Agrawal, A., Gaffrey, M. S., Belden, A. C., Botteron, K. N., Harms, M. P., & Barch, D. M. (2015). Amygdala Functional Connectivity, HPA Axis Genetic Variation, and Life Stress in Children and Relations to Anxiety and Emotion

Regulation. *J Abnorm Psychol*, *124*(4), 817–833. https://doi.org/DOI: 10.1037/abn0000094, PMCID: PMC4662045

Pagliaccio, D., Luby, J. L., Luking, K. R., Belden, A. C., & Barch, D. M. (2014). Brain–behavior relationships in the experience and regulation of negative emotion in healthy children: Implications for risk for childhood depression. *Development and Psychopathology*, *26*, 1289–1303. https://doi.org/doi:10.1017/S0954579414001035

Payne, C., Machado, C. J., Bliwise, N. G., & Bachevalier, J. (2010). Maturation of the hippocampal formation and amygdala in Macaca mulatta: A volumetric magnetic resonance imaging study. *Hippocampus*, *20*(8), 922–935. https://doi.org/10.1002/hipo.20688

Piccolo, L. R., & Noble, K. G. (2018). Perceived stress is associated with smaller hippocampal volume in adolescence. *Psychophysiology*, *55*(5), 1–10. https://doi.org/10.1111/psyp.13025

Pitula, C. E., DePasquale, C. E., Mliner, S. B., & Gunnar, M. R. (2019). Peer Problems Among Postinstitutionalized, Internationally Adopted Children: Relations to Hypocortisolism, Parenting Quality, and ADHD Symptoms. *Child Development*, *90*(3), e339–e355. https://doi.org/10.1111/cdev.12986

Plotsky, P. M., Thrivikraman, K. V., Nemeroff, C. B., Caldji, C., Sharma, S., & Meaney, M. J. (2005). Long-Term Consequences of Neonatal Rearing on Central Corticotropin-Releasing Factor Systems in Adult Male Rat Offspring. *Neuropsychopharmacology*, *30*(12), 2192–2204. https://doi.org/10.1038/sj.npp.1300769

Preacher, K. J., Curran, P. J., & Bauer, D. J. (2006). Computational Tools for Probing Interactions in Multiple Linear Regression, Multilevel Modeling, and Latent Curve

Analysis. *Journal of Educational and Behavioral Statistics*, *31*(4), 437–448. https://doi.org/10.3102/10769986031004437

Quevedo, K., Johnson, A., Loman, M., Lafavor, T., & Gunnar, M. (2012). The confluence of adverse early experience and puberty on the cortisol awakening response. *International Journal of Behavioral Development*, *36*(1), 19–28. https://doi.org/10.1177/0165025411406860.The

Raineki, C., Opendak, M., Sarro, E., Showler, A., Bui, K., McEwen, B. S., Wilson, D. A., & Sullivan, R. M. (2019). During infant maltreatment, stress targets hippocampus, but stress with mother present targets amygdala and social behavior. *Proceedings of the National Academy of Sciences*, 201907170. https://doi.org/10.1073/pnas.1907170116

Rao, U., Chen, L. A., Bidesi, A. S., Shad, M. U., Thomas, M. A., & Hammen, C. L. (2010). Hippocampal Changes Associated with Early-Life Adversity and Vulnerability to Depression. *Biological Psychiatry*, *67*(4), 357–364. https://doi.org/10.1016/j.biopsych.2009.10.017

Reul, J. M., & de Kloet, E. R. (1985). Two receptor systems for corticosterone in rat brain: Microdistribution and differential occupation. *Endocrinology*, *117*(6), 2505–2511. https://doi.org/10.1210/endo-117-6-2505

Reuter, M. (2016). *Re: Registration Errors in Longitudinal Analysis*. Freesurfer Email List. https://www.mail-archive.com/freesurfer@nmr.mgh.harvard.edu/msg44963.html

Riem, M. M. E., Alink, L. R. A., Out, D., Van Ijzendoorn, M. H., & Bakermans-Kranenburg, M. J. (2015). Beating the brain about abuse: Empirical and meta-analytic studies of the association between maltreatment and hippocampal volume across childhood and

adolescence. *Development and Psychopathology*, *27*(02), 507–520. https://doi.org/10.1017/S0954579415000127

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, *48*(1), 1–36. https://doi.org/10.18637/jss.v048.i02

Roth, M. C., Humphreys, K. L., King, L. S., & Gotlib, I. H. (2018). Self-reported neglect , amygdala volume , and symptoms of anxiety in adolescent boys. *Child Abuse & Neglect*, *80*(March), 80–89. https://doi.org/10.1016/j.chiabu.2018.03.016

Santiago, A., Aoki, C., & Sullivan, R. M. (2017). From attachment to independence: Stress hormone control of ecologically relevant emergence of infants' responses to threat. *Current Opinion in Behavioral Sciences*, *14*, 78–85. https://doi.org/10.1016/j.cobeha.2016.12.010

Satterthwaite, F. E. (1946). An Approximate Distribution of Estimates of Variance Components. *Biometrics Bulletin*, *2*(6), 110–114. JSTOR. https://doi.org/10.2307/3002019

Sheline, Y. I., Wang, P. W., Gado, M. H., Csernansky, J. G., & Vannier, M. W. (1996). Hippocampal atrophy in recurrent major depression. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(9), 3908–3913. https://doi.org/10.1073/pnas.93.9.3908

Sheridan, M. a, Fox, N. a, Zeanah, C. H., McLaughlin, K. a, & Nelson, C. a. (2012). Variation in neural development as a result of exposure to institutionalization early in childhood. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(32), 12927–12932. https://doi.org/10.1073/pnas.1200041109

Silvers, J. A., Insel, C., Powers, A., Franz, P., Helion, C., Martin, R. E., Weber, J., Mischel, W., Casey, B. J., & Ochsner, K. N. (2016). VlPFC – vmPFC – Amygdala Interactions

Underlie Age-Related Differences in Cognitive Regulation of Emotion. *Cerebral Cortex*, *27*(7), 3502–3514. https://doi.org/10.1093/cercor/bhw073, PMID: 27341851

Teicher, M. H., Anderson, C. M., & Polcari, A. (2012). Childhood maltreatment is associated with reduced volume in the hippocampal subfields CA3, dentate gyrus, and subiculum. *Proceedings of the National Academy of Sciences*, *109*(9), E563–E572. https://doi.org/10.1073/pnas.1115396109

Teicher, M.H., Samson, J. A., Anderson, C. M., & Ohashi, K. (2016). The effects of childhood maltreatment on brain structure, function and connectivity. *Nature Reviews Neuroscience*, *17*(10), 652–666. https://doi.org/10.1038/nrn.2016.111

ter Heegde, F., De Rijk, R. H., & Vinkers, C. H. (2015). The brain mineralocorticoid receptor and stress resilience. *Psychoneuroendocrinology*, *52*, 92–110. https://doi.org/10.1016/j.psyneuen.2014.10.022

Tottenham, N. (2012). Risk and developmental heterogeneity in previously institutionalized children. *The Journal of Adolescent Health : Official Publication of the Society for Adolescent Medicine*, *51*(2 Suppl), S29-33. https://doi.org/10.1016/j.jadohealth.2012.04.004

Tottenham, N., Hare, T. a, Quinn, B. T., McCarry, T. W., Nurse, M., Gilhooly, T., Millner, A., Galvan, A., Davidson, M. C., Eigsti, I.-M., Thomas, K. M., Freed, P. J., Booma, E. S., Gunnar, M. R., Altemus, M., Aronson, J., & Casey, B. J. (2010). Prolonged institutional rearing is associated with atypically large amygdala volume and difficulties in emotion regulation. *Developmental Science*, *13*(1), 46–61. https://doi.org/10.1111/j.1467-7687.2009.00852.x

Tottenham, N., & Sheridan, M. A. (2009). A review of adversity, the amygdala and the hippocampus: A consideration of developmental timing. *Frontiers in Human Neuroscience*, *3*, 68. https://doi.org/10.3389/neuro.09.068.2009, PMCID: PMC2813726

Uematsu, A., Matsui, M., Tanaka, C., Takahashi, T., Noguchi, K., Suzuki, M., & Nishijo, H. (2012). Developmental trajectories of amygdala and hippocampus from infancy to early adulthood in healthy individuals. *PloS One*, *7*(10), e46970. https://doi.org/10.1371/journal.pone.0046970

van Velzen, L. S., Schmaal, L., Jansen, R., Milaneschi, Y., Opmeer, E. M., Elzinga, B. M., Wee, N. J. A., Veltman, D. J., & Penninx, B. W. J. H. (2016). Effect of childhood maltreatment and brain-derived neurotrophic factor on brain morphology. *Social Cognitive and Affective Neuroscience*, *11*(11), 1841–1852. https://doi.org/10.1093/scan/nsw086

Vázquez, D. M., Neal, C. R., Patel, P. D., Kaciroti, N., & López, J. F. (2012). Regulation of corticoid and serotonin receptor brain system following early life exposure of glucocorticoids: Long term implications for the neurobiology of mood. *Psychoneuroendocrinology*, *37*(3), 421–437. https://doi.org/10.1016/j.psyneuen.2011.07.012

Vyas, A., Mitra, R., Rao, B. S. S., & Chattarji, S. (2002). Chronic Stress Induces Contrasting Patterns of Dendritic Remodeling in Hippocampal and Amygdaloid Neurons. *Journal of Neuroscience*, *22*(15), 6810–6818.

Vythilingam, M., Heim, C., Ph, D., Newport, J., Miller, A. H., Anderson, E., Bronen, R., Brummer, M., Staib, L., Vermetten, E., Charney, D. S., Nemeroff, C. B., & Bremner, J. D. (2002). Volume in Women With Major Depression. *American Journal of Psychiatry*, *159*(December), 2072–2080. https://doi.org/10.1176/appi.ajp.159.12.2072

Warmingham, J. M., Handley, E. D., Rogosch, F. A., Manly, J. T., & Cicchetti, D. (2019). Identifying maltreatment subgroups with patterns of maltreatment subtype and chronicity: A latent class analysis approach. *Child Abuse & Neglect*, *87*, 28–39. https://doi.org/10.1016/j.chiabu.2018.08.013

Weems, C. F., & Carrión, V. G. (2009). Brief report: Diurnal salivary cortisol in youth—Clarifying the nature of posttraumatic stress dysregulation. *Journal of Pediatric Psychology*, *34*(4), 389–395. https://doi.org/10.1093/jpepsy/jsn087

Wierenga, L. M., Bos, M. G. N., Schreuders, E., vd Kamp, F., Peper, J. S., Tamnes, C. K., & Crone, E. A. (2018). Unraveling age, puberty and testosterone effects on subcortical brain development across adolescence. *Psychoneuroendocrinology*, *91*(September 2017), 105–114. https://doi.org/10.1016/j.psyneuen.2018.02.034

Zalewski, M., Lengua, L. J., Thompson, S. F., & Kiff, C. J. (2016). Income, cumulative risk, and longitudinal profiles of hypothalamic-pituitary-adrenal axis activity in preschool-age children. *Development and Psychopathology*, *28*(2), 341–353. https://doi.org/10.1017/S0954579415000474

Appendix A: Supplemental Tables

Table S1. Descriptive information for diurnal cortisol data.

Variable	*M*	*SD*	1	2	3
1. Age (years)	10.54	4.13			
2. Morning cortisol (nmol/L)	20.14	10.09	.39** [.30, .47]		
3. Evening cortisol (nmol/L)	4.97	3.97	.25** [.15, .34]	.32** [.23, .41]	
4. Saliva sample batch	3.60	1.98	.22** [.12, .31]	.42** [.34, .50]	.17** [.07, .27]

Note. M and *SD* are used to represent mean and standard deviation, respectively. Values in square brackets indicate the 95% confidence interval for each correlation. The confidence interval is a plausible range of population correlations that could have caused the sample correlation (Cumming, 2014). * indicates $p < .05$. ** indicates $p < .01$.

Table S2. Descriptive information for structural MRI data.

Variable	*M*	*SD*	1	2	3	4	5
1. Age (years)	11.76	3.78					
2. Motion Rating	1.67	0.68	-.30** [-.39, -.19]				
3. Segmentation Rating	1.27	0.49	-.15** [-.26, -.04]	.14* [.03, .24]			
4. Intracranial Volume (ICV)	1,572,533	175,496	-.29** [-.39, -.19]	.11 [-.01, .22]	.06 [-.05, .17]		
5. Amygdala Volume	1608.84	186.84	.29** [.18, .39]	-.07 [-.18, .04]	.01 [-.10, .12]	.40** [.30, .49]	
6. Hippocampal Volume	3978.83	412.70	.19** [.08, .29]	-.03 [-.14, .09]	.00 [-.11, .11]	.47** [.38, .56]	.70** [.63, .75]

Note. *M* and *SD* are used to represent mean and standard deviation, respectively. Values in square brackets indicate the 95% confidence interval for each correlation. The confidence interval is a plausible range of population correlations that could have caused the sample correlation (Cumming, 2014). * indicates $p < .05$. ** indicates $p < .01$.

Table S3. Group differences in relevant covariates for diurnal cortisol analyses. Results are depicted for 2-sampled t-tests (PI - COMP). Ages and batch are averaged across waves for subjects with more than one wave of data included in analysis.

Variable	t value	df	p value
Sex	2.44	193.92	0.015
Age	2.29	208.51	0.023
Batch	-1.02	180.07	0.309
Medication	5.20	116.22	< 0.001

Table S4. Group differences in relevant covariates for structural analyses. Results are depicted for 2-sampled t-tests (PI – COMP). Ages, ICV, motion and segmentation ratings are averaged across waves for subjects with more than one wave of data included in analysis.

Variable	t value	df	p value
Sex	1.15	142.72	0.251
Age	1.11	152.30	0.269
Motion Rating	-1.67	147.22	0.097
Segmentation Rating	-0.31	145.22	0.755
ICV	-3.98	143.36	< 0.001

Table S5. Demographic information for sub-sample in cross-lagged SEM models. Path 1 represents longitudinal path from morning cortisol at T1 to subcortical brain volume at T2. Path 2 represents longitudinal path from subcortical brain volume at T1 to morning cortisol at T2. PI = Previously Institutionalized, COMP = comparisons.

				Age (years)			Age (change)	
GROUP	Path	N	M/F	Mean	SD	Range	Mean	SD
	T1	66	32/34	10.36	3.70	4.25-17.58		
COMP	Path1	41	19/22	10.65	3.85	4.25-17.58	2.59	0.86
	Path2	45	21/24	10.23	3.83	4.25-17.58	2.63	0.9
	T2	41	19/22	13.24	3.85	6.67-20.33		
	T1	49	19/30	10.62	3.11	4.58-17.5		
PI	Path1	31	12/19	10.55	3.15	4.58-17.5	2.95	1.11
	Path2	40	15/25	10.71	3.23	4.58-17.5	2.73	1.26
	T2	31	12/19	13.50	3.15	7.08-18.83		

Table S6. Full model results for diurnal cortisol model with piecewise Age X Group X Time of day (morning, evening) effects. Age1 represents the effect of age before the piecewise inflection point (13.1 years) and Age2 represents the effect of age after 13.1 years old.

Coefficient	Estimate	SE	df	t value	p value
(Intercept)	22.03	0.79	651	28.00	< 0.001
Age1	0.68	0.14	705	5.03	< 0.001
Age1 X Time of Day	-0.42	0.17	2,207	-2.52	0.012
Age2	0.48	0.30	647	1.60	0.11
Batch	0.90	0.14	805	6.22	< 0.001
Group (PI - COMP)	-3.90	1.20	649	-3.25	0.001
Group X Age1	-0.40	0.24	654	-1.69	0.091
Group x Age1 X Time of Day	-0.16	0.30	2,198	-0.52	0.602
Group X Age2	2.17	0.51	668	4.29	< 0.001
Group X Time of Day	1.56	1.44	2,193	1.08	0.279
Group X Age2 X Time of Day	-2.26	0.59	2,184	-3.85	< 0.001
Day	0.09	0.19	2,292	0.44	0.657
Wave	0.23	0.27	108	0.86	0.389
Medications	0.38	0.83	1,933	0.46	0.644
Sex (M - F)	1.10	0.52	226	2.12	0.035
Time of Day	-16.04	0.93	2,189	-17.32	< 0.001
Age2 x Time of Day	-0.21	0.35	2,183	-0.60	0.545

Table S7. Full model results for morning cortisol model with piecewise Age X Group effects. Age1 represents the effect of age before the piecewise inflection point (13.1 years) and Age2 represents the effect of age after 13.1 years.

Coefficient	Estimate	SE	DF	t value	p value
(Intercept)	22.43	1.08	320.91	20.82	< 0.001
Age1	0.73	0.18	320.84	4.04	< 0.001
Age2	0.34	0.41	313.01	0.84	0.404
Batch	1.42	0.24	713.91	6.01	< 0.001
Group (PI - COMP)	-4.61	1.63	318.66	-2.83	0.005
Group X Age1	-0.68	0.32	300.66	-2.12	0.034
Group X Age2	2.03	0.70	348.87	2.89	0.004
Day	0.36	0.31	1,056.74	1.14	0.255
Wave	0.54	0.48	107.15	1.12	0.263
Medications	-0.33	1.32	1,175.06	-0.25	0.804
Sex (M - F)	1.77	0.84	236.00	2.10	0.037

Table S8. Full model results for evening cortisol with piecewise Age X Group effects. Age1 represents the effect of age before the piecewise inflection point (13.1 years) and Age2 represents the effect of age after 13.1 years.

Coefficient	Estimate	SE	DF	t value	p value
(Intercept)	5.61	0.53	334.62	10.53	< 0.001
Age1	0.21	0.10	377.51	2.21	0.028
Age2	0.51	0.20	391.14	2.55	0.011
Batch	0.25	0.13	929.59	1.95	0.051
Group (PI - COMP)	-1.43	0.84	359.85	-1.71	0.088
Group X Age1	-0.24	0.17	415.38	-1.41	0.159
Group X Age2	-0.12	0.32	519.60	-0.39	0.698
Day	-0.10	0.17	1,014.97	-0.62	0.533
Wave	-0.08	0.19	1,156.16	-0.41	0.679
Medications	1.38	0.74	1,167.64	1.87	0.062
Sex (M - F)	0.48	0.44	219.68	1.08	0.282

Table S9. Full model results for amygdala volume with piecewise Age X Group effects. Age1 represents the effect of age before the piecewise inflection point (9.4 years) and Age2 represents the effect of age after 9.4 years.

Coefficient	Estimate	SE	df	t value	p value
(Intercept)	17.36	0.24	276.66	71.51	< 0.001
Age1	0.65	0.08	250.99	8.26	< 0.001
Age2	-0.01	0.04	295.65	-0.17	0.868
Group (PI - COMP)	-0.54	0.32	250.79	-1.70	0.091
Group X Age1	-0.45	0.15	210.58	-3.04	0.003
Group X Age2	-0.02	0.06	283.57	-0.38	0.7
ICV	0.28	0.06	280.68	5.17	< 0.001
Motion Rating	-0.27	0.09	190.64	-2.87	0.005
Scanner	-0.92	0.13	220.27	-7.15	< 0.001
Sex (F - M)	-0.52	0.24	131.81	-2.16	0.032

Table S10. Full model results for hippocampal volume.

Coefficient	Estimate	SE	df	t value	p value
(Intercept)	37.84	0.75	266.04	50.38	< 0.001
Age	0.23	0.06	290.15	4.03	< 0.001
Group (PI - COMP)	-1.78	0.55	129.82	-3.24	0.002
ICV	0.46	0.10	225.98	4.45	< 0.001
Motion Rating	0.06	0.16	156.32	0.38	0.703
Scanner	-1.63	0.23	218.33	-7.01	< 0.001
Sex (F - M)	-1.38	0.55	127.25	-2.52	0.013

Table S11. Cross-lag model results for amygdala volume and morning cortisol with standardized coefficients and 95% confidence intervals.

		Coefficient	Estimate	SE	Z	p value	lower	upper
Amygdala Volume T2	~	Amygdala Volume T1	0.72	0.06	11.55	< 0.001	0.60	0.84
		Morning Cortisol T1	0.04	0.06	0.72	0.47	-0.07	0.15
		Age T2	-0.09	0.09	-0.95	0.342	-0.27	0.09
		Group (PI-COMP)	-0.02	0.06	-0.40	0.691	-0.13	0.09
		Sex	0.04	0.08	0.46	0.645	-0.12	0.20
		ICV T2	0.28	0.15	1.89	0.059	-0.01	0.56
		Scanner	0.32	0.06	5.70	< 0.001	0.21	0.43
Morning Cortisol T2	**~**	**Amygdala Volume T1**	**0.32**	**0.12**	**2.68**	**< 0.01**	**0.09**	**0.55**
		Morning Cortisol T1	0.07	0.09	0.76	0.447	-0.11	0.26
		Age T2	0.24	0.17	1.37	0.17	-0.10	0.58
		Group (PI-COMP)	-0.04	0.10	-0.45	0.653	-0.23	0.15
		Sex	0.36	0.09	3.84	< 0.001	0.17	0.54
Amygdala Volume T1	~	Age T1	0.64	0.13	5.01	< 0.001	0.39	0.89
		Group (PI-COMP)	0.04	0.12	0.36	0.717	-0.18	0.27
		Sex	-0.12	0.10	-1.21	0.226	-0.32	0.08
		ICV T1	0.75	0.28	2.68	< 0.01	0.20	1.30
Morning Cortisol T1	~	Age T1	0.17	0.08	2.28	0.023	0.02	0.32
		Group (PI-COMP)	-0.06	0.07	-0.92	0.356	-0.20	0.07
		Sex	0.08	0.06	1.32	0.187	-0.04	0.21
Amygdala Volume T1	~~	Morning Cortisol T1	-0.29	0.08	-3.37	< 0.01	-0.45	-0.12
Amygdala Volume T2	~~	Morning Cortisol T2	-0.14	0.12	-1.25	0.213	-0.37	0.08

Table S12. Cross-lag model results for hippocampal volume and morning cortisol with standardized coefficients and 95% confidence intervals.

		Coefficient	Estimate	SE	Z	p value	lower	upper
Hippocampus Volume T2	~	Hippocampus Volume T1	1.01	0.07	13.68	< 0.001	0.86	1.15
		Morning Cortisol T1	0.04	0.04	0.88	0.378	-0.05	0.13
		Age T2	-0.14	0.06	-2.23	0.026	-0.26	-0.02
		Group (PI-COMP)	-0.04	0.05	-0.85	0.398	-0.13	0.05
		Sex	-0.13	0.05	-2.88	< 0.01	-0.22	-0.04
		ICV T2	-0.17	0.09	-1.94	0.052	-0.34	0.00
		Scanner	0.10	0.05	2.17	0.03	0.01	0.20
Morning Cortisol T2	**~**	**Hippocampus Volume T1**	**0.41**	**0.17**	**2.36**	**0.018**	**0.07**	**0.75**
		Morning Cortisol T1	-0.02	0.11	-0.17	0.864	-0.23	0.19
		Age T2	0.44	0.13	3.27	< 0.01	0.17	0.70
		Group (PI-COMP)	0.00	0.11	-0.04	0.964	-0.22	0.21
		Sex	0.41	0.10	4.11	< 0.001	0.21	0.60
Hippocampus Volume T1	~	Age T1	0.75	0.27	2.78	< 0.01	0.22	1.28
		Group (PI-COMP)	-0.09	0.10	-0.88	0.379	-0.28	0.10
		Sex	-0.16	0.07	-2.11	0.035	-0.30	-0.01
		ICV T1	0.45	0.15	3.03	< 0.01	0.16	0.74
Morning Cortisol T1	~	Age T1	0.18	0.09	2.02	0.043	0.01	0.36
		Group (PI-COMP)	-0.07	0.06	-1.10	0.27	-0.20	0.06
		Sex	0.09	0.06	1.34	0.179	-0.04	0.21
Hippocampus Volume T1	~~	Morning Cortisol T1	-0.04	0.11	-0.42	0.678	-0.25	0.16
Hippocampus Volume T2	~~	Morning Cortisol T2	-0.01	0.13	-0.05	0.957	-0.26	0.25

Table S13. Cross-lag model results for amygdala volume and morning cortisol with group interactions, showing standardized coefficients and 95% confidence intervals.

		Coefficient	Estimate	SE	Z	p value	lower	upper
Amygdala Volume T2	~	Morning Cortisol T1 X Group	-0.07	0.07	-1.12	0.263	-0.20	0.06
		Amygdala Volume T1	0.76	0.06	12.30	< 0.001	0.64	0.88
		Morning Cortisol T1	0.08	0.06	1.25	0.21	-0.04	0.20
		Age T2	-0.04	0.09	-0.40	0.691	-0.22	0.14
		Group (PI-COMP)	-0.03	0.05	-0.60	0.552	-0.13	0.07
		ICV T2	0.30	0.08	3.93	< 0.001	0.15	0.45
		Sex	0.03	0.06	0.61	0.541	-0.07	0.14
		Scanner	0.33	0.06	6.00	< 0.001	0.22	0.44
Morning Cortisol T2	~	Amygdala Volume T1 X Group	0.14	0.12	1.12	0.263	-0.10	0.38
		Amygdala Volume T1	0.23	0.14	1.61	0.108	-0.05	0.51
		Morning Cortisol T1	0.07	0.09	0.79	0.431	-0.11	0.26
		Age T2	0.17	0.11	1.61	0.108	-0.04	0.38
		Group (PI-COMP)	-0.04	0.10	-0.40	0.692	-0.22	0.15
		Sex	0.34	0.10	3.51	< 0.001	0.15	0.52
Amygdala Volume T1	~	Age T1	0.58	0.14	4.29	< 0.001	0.32	0.85
		Group (PI-COMP)	0.03	0.07	0.43	0.666	-0.10	0.16
		ICV T1	0.49	0.08	5.90	< 0.001	0.33	0.65
		Sex	-0.15	0.07	-2.17	0.03	-0.28	-0.01
Morning Cortisol T1	~	Age T1	0.16	0.07	2.29	0.022	0.02	0.30
		Group (PI-COMP)	-0.08	0.06	-1.16	0.245	-0.20	0.05
		Sex	0.09	0.06	1.36	0.172	-0.04	0.21
Amygdala Volume T1	~~	Morning Cortisol T1	-0.25	0.09	-2.84	< 0.01	-0.42	-0.08
Amygdala Volume T2	~~	Morning Cortisol T2	-0.16	0.12	-1.43	0.153	-0.39	0.06

Table S14. Cross-lag model for hippocampal volume and morning cortisol with group interactions, showing standardized coefficients and 95% confidence intervals.

		Coefficient	Estimate	SE	Z	p value	lower	upper
Hippocampus Volume T2	~	Morning Cortisol T1 X Group	-0.10	0.07	-1.51	0.132	-0.23	0.03
		Hippocampus Volume T1	1.24	0.08	15.00	< 0.001	1.08	1.41
		Morning Cortisol T1	0.08	0.07	1.23	0.219	-0.05	0.22
		Age T2	-0.28	0.08	-3.63	< 0.001	-0.43	-0.13
		Group (PI-COMP)	0.00	0.05	0.07	0.941	-0.09	0.10
		ICV T2	-0.25	0.09	-2.71	< 0.01	-0.44	-0.07
		Sex	-0.10	0.05	-2.02	0.044	-0.20	0.00
		Scanner	0.06	0.05	1.17	0.244	-0.04	0.16
Morning Cortisol T2	~	Hippocampus Volume T1 X Group	-0.04	0.18	-0.20	0.844	-0.39	0.32
		Hippocampus Volume T1	0.28	0.21	1.32	0.186	-0.14	0.70
		Morning Cortisol T1	0.02	0.10	0.21	0.836	-0.17	0.21
		Age T2	0.23	0.11	2.00	0.046	0.00	0.45
		Group (PI-COMP)	0.03	0.11	0.23	0.815	-0.19	0.24
		Sex	0.36	0.10	3.73	< 0.001	0.17	0.55
Hippocampus Volume T1	~	Age T1	-4.32	3.68	-1.18	0.24	-11.53	2.88
		Group (PI-COMP)	0.01	0.34	0.04	0.966	-0.65	0.68
		ICV T1	-1.05	1.29	-0.81	0.416	-3.57	1.48
		Sex	-0.23	0.32	-0.71	0.478	-0.86	0.40
Morning Cortisol T1	~	Age T1	-0.43	0.29	-1.46	0.145	-1.01	0.15
		Group (PI-COMP)	-0.01	0.08	-0.16	0.869	-0.18	0.15
		Sex	0.10	0.08	1.35	0.178	-0.05	0.26
Hippocampus Volume T1	~~	Morning Cortisol T1	0.60	0.17	3.58	< 0.001	0.27	0.92
Hippocampus Volume T2	~~	Morning Cortisol T2	-0.06	0.14	-0.40	0.691	-0.34	0.22

Table S15. Full model results showing main effects of group, amygdala and hippocampal volume on total internalizing scores.

Coefficient	Estimate	SE	df	t value	p value
(Intercept)	43.34	2.15	204	20.19	< 0.001
Age	0.08	0.21	190	0.41	0.68
Amygdala Volume	1.00	0.50	209	2.01	0.046
Group (PI-COMP)	9.62	1.56	132	6.16	< 0.001
Hippocampus Volume	-0.57	0.23	197	-2.43	0.016
ICV	1.01	0.47	190	2.18	0.031
Medications	3.14	1.74	207	1.80	0.073
Scanner	-1.44	1.17	161	-1.23	0.22
Sex (F-M)	2.89	1.51	128	1.91	0.058
Morning Cortisol	0.00	0.05	166	-0.06	0.951

Table S16. Full model results for the 3-way Group X Age X Morning cortisol interaction on internalizing scores.

Coefficient	Estimate	SE	df	t value	p value
(Intercept)	42.69	2.45	206.00	17.42	< 0.001
Age	-0.04	0.32	202.00	-0.14	0.891
Amygdala Volume	1.11	0.51	203.00	2.18	0.03
Group	8.64	2.74	203.00	3.15	0.002
Group x Age	0.99	0.61	195.00	1.62	0.106
Group x Morning Cortisol	0.10	0.12	157.00	0.85	0.396
Group x Morning Cortisol X Age	-0.07	0.02	162.00	-2.72	0.007
Hippocampal Volume	-0.59	0.23	192.00	-2.56	0.011
ICV	0.94	0.46	186.00	2.03	0.044
Medications	3.22	1.76	201.00	1.83	0.069
Scanner	-1.46	1.17	160.00	-1.24	0.215
Sex (F-M)	3.24	1.53	129.00	2.12	0.036
Morning Cortisol	0.04	0.06	155.00	0.65	0.519
Morning Cortisol x Age	0.00	0.01	165.00	0.20	0.842

Table S17. Post-hoc analyses for the Group X Age X Morning cortisol effect on internalizing scores. Simple slopes of the 3-way interaction were calculated to estimate the effect of morning cortisol on internalizing scores for each group at three different ages (mean -1SD, mean age, mean + 1SD). Results show no direct relationship between morning cortisol and internalizing scores at any level tested. Estimates are provided with 95% confidence intervals. PI = Previously Institutionalized, COMP = comparisons.

GROUP	Age (years)	Estimate	t value	p value	lower	upper
	7.67	0.03	0.31	0.76	-0.15	0.20
COMP	11.52	0.04	0.65	0.52	-0.08	0.15
	15.37	0.05	0.83	0.40	-0.06	0.16
	7.67	0.24	1.47	0.14	-0.08	0.56
PI	11.52	0.10	0.95	0.34	-0.10	0.30
	15.37	-0.05	-0.58	0.56	-0.20	0.11

Appendix B: Supplemental Analyses

Supplementary Analyses for Morning Cortisol

1. Retention

The current study was by design not fully longitudinal, with a subset of participants asked to completed follow-up visits 2 and/or 4 years later. Control analyses tested for systematic differences between participants who did and did not complete follow-up assessments for diurnal cortisol. When comparing groups of participants who completed 1, 2, or 3 time points of cortisol data collection, there were no significant differences in sex (*Chisq* (2) = 2.916, $p = 0.233$), age (F (1, 237) = 2.021, $p = 0.135$), or group (*Chisq* (2) = 4.938, $p = 0.085$).

2. Secondary analyses

2.1 Age of adoption

Prior research suggests that the timing of ECA exposure may influence the development of the HPA axis. For example, later age of adoption or foster care placement has previously been associated with blunted morning cortisol (Kumsta et al., 2017) and blunted cortisol reactivity to a stressor (McLaughlin et al., 2015). In order to assess timing effects of early institutional care on morning cortisol, we tested the effects of age of adoption on morning cortisol in the PI group only (N = 84 with known adoption timing). Although the median age of adoption was 16.5 months, some participants were adopted at much later ages (e.g. 5-10 years old). Because of this non-normal distribution, (Shapiro Wilk = 0.568, $p < 0.0001$), age of adoption was log-transformed prior to analysis. As shown in Figure S1, age of adoption was negatively associated with morning cortisol levels ($b = -1.648$, t (64.497) = -2.302, $p = 0.025$, CI = -3.028, -0.269), controlling for age, wave, cortisol batch, medications, and sex. This effect remained significant

when we also controlled for (log) age of placement in the institution (b = -2.034, t (61.505) = -2.169, p = 0.034, CI = -3.827, -0.243) and when using non-transformed age of adoption (b = -0.106, t (54.912) = -2.95, p < 0.01, CI = -0.175, -0.037). These results suggest that later age of adoption is associated with more blunted morning cortisol, regardless of age effects on these phenotypes. These findings correspond with prior research showing a dose-dependent effect of early caregiving adversity on cortisol phenotypes (Kumsta et al., 2017; McLaughlin et al., 2015).

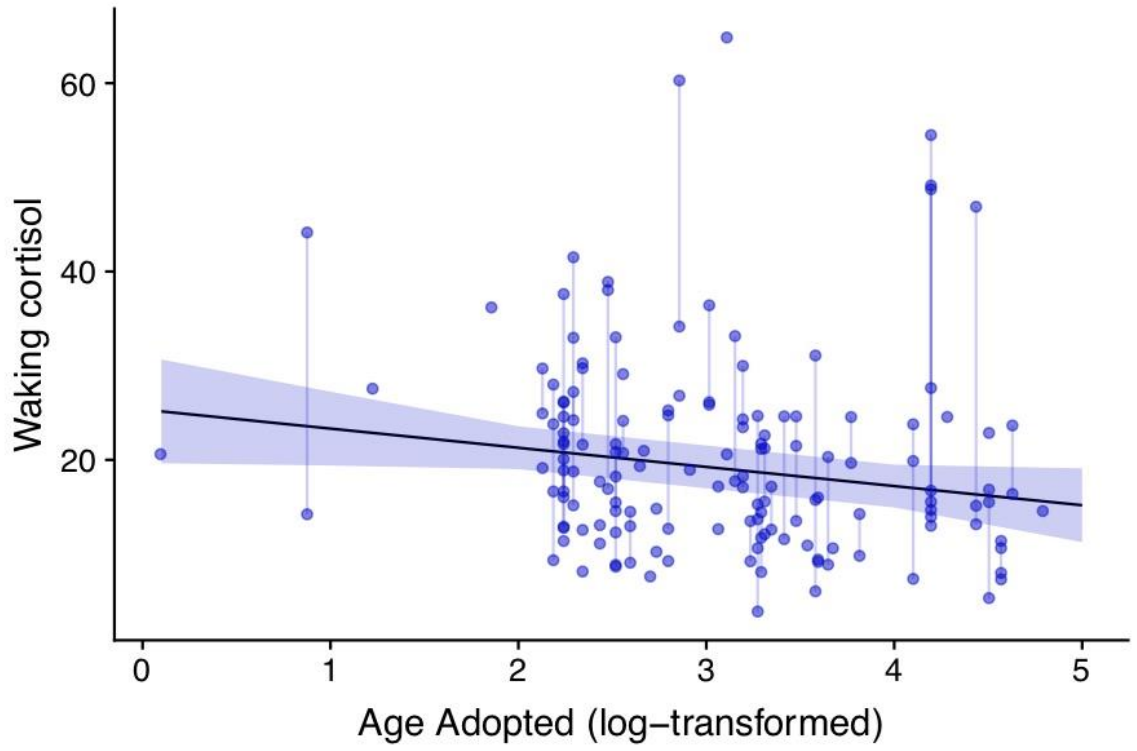

Figure S1: Age of adoption is negatively associated with morning cortisol in PI youth. Specifically, later age of adoption (log-transformed) is associated with lower morning cortisol. Ribbons indicate 95% confidence interval of the effect, with raw data points overlaid and lines connecting within-subject observations (i.e., waves of data from the same subject).

2.2 Sex effects

As shown in table S8, a main effect of sex was observed on morning cortisol, such that females showed higher cortisol concentration levels than males. Follow-up analyses tested whether sex moderated group or age effects on morning cortisol. Sex significantly interacted with age (b = 0.55, t (297.001) = 2.719, $p < 0.01$, CI = 0.155, 0.943), such that females had steeper age-related increases in morning cortisol relative to males. Simple slopes showed that females had higher levels of cortisol than males at age 15 years old (1SD above the mean; t (230.57) = 3.6, $p < 0.001$, b = 4.53) but not at 6 years old (1SD below the mean; t (230.57) = 0.03, p = 0.974, b = 0.04). Given the small number of cortisol observations from PI males older than 13 years old (N = 2), we did not examine these sex effects in piecewise age models. However, sex did not moderate group effects on morning cortisol (b = -0.346, t (230.822) = -0.187, p = 0.851, CI = -3.935, 3.242), suggesting that although females overall had higher morning cortisol, the group differences observed in the current sample were not driven by females.

2.3 Pubertal hormones

Several studies have implicated pubertal maturation as the driver of cortisol changes following early adversity (Flannery et al., 2017; King et al., 2017; Gunnar et al., 2019). In follow-up analyses, we assessed the role of the pubertal hormone testosterone on morning cortisol level in the PI group. Due to the high collinearity between testosterone and age (r = 0.609, p < 0.001), we tested the effects of testosterone with and without controlling for age.

Testosterone was assayed from the same saliva samples used for cortisol analyses. Because testosterone varied significantly by time of day (b = 0.228, t (1917.83) = 6.712, $p <$

0.001, CI = 0.161, 0.295), with morning values higher than evening values, we used the average testosterone values from the two morning samples only. Implausible testosterone values (e.g. > 500) were excluded, and then values were log-transformed and outliers > 3SD were omitted. Second, to account for sex differences in testosterone, values were rank-normed within sex by dividing the difference between the maximum and minimum values within males and females to obtain a proportion of the maximum possibility score for each participant (e.g. 0 to 1; Cohen 1999, Fareri et al., 2015). Testosterone values were obtained from the majority of participants included in the original cortisol analyses (missing 9 PI samples and 14 COMP samples from wave 1).

We observed a significant main effect of sex-normed testosterone on morning cortisol (b = 12.756, t (342.248) = 5.733, $p < 0.001$, CI = 8.342, 17.158). When including age in the model, testosterone effects remained significant (b = 7.918, t (372.55) = 3.154, $p < 0.01$, CI = 2.963, 12.833), as did age (b = 0.454, t (258.091) = 3.787, $p < 0.001$, CI = 0.221, 0.688), suggesting that testosterone and age have independent and significant effects on morning cortisol development. However, testosterone did not moderate the effects of PI status on morning cortisol (with age as covariate: b = 0.87, t (392.538) = 0.207, p = 0.836, CI = -7.288, 9.089]; without age as covariate: b = -1.512, t (401.652) = -0.355, p = 0.723, CI = -9.8, 6.821) .These findings suggest that while the main effects of testosterone on morning cortisol changes are robust, the pubertal hormone is not driving the developmental changes in morning cortisol that we observed in the current sample of PI youth.

Supplemental analyses for Amygdala and Hippocampal volume

1. Control Analyses

1.1. Retention

Control analyses tested for systematic differences between participants who did and did not complete follow-up scans. When comparing groups of participants who completed 1, 2, or 3 time points of structural MRI data, there were no significant differences in sex (*Chisq* (2) = 0.956, $p = 0.62$) age (F (1, 153) = 1.119, $p = 0.329$) or group status (*Chisq* (2) = 1.333, $p = 0.513$).

1.2 Limiting age range from 6 to 19

For the structural MRI data, there was an imbalance in distribution of age between groups, with more comparison scans at the extreme ages (Figure S2). Main analyses included all available data from ages 4 to 20 to examine age effects for the widest possible age-range and mirror analyses conducted on cortisol data, which had more coverage at the extreme ages. Secondary analyses of amygdala and hippocampal volume were conducted for ages 6 to 19 to validate findings with balanced age distributions per group using a parallel analytic strategy.

When testing the piecewise model for amygdala volume, we detected similar results to the main analyses. Specifically, we detected a significant Group x Age interaction in children younger than 9.5 years old ($b = -0.449$, t (179.563) = -2.531, $p = 0.012$, CI = -0.794, -0.106), with significant positive effect of age in the comparison group (t (207) = 6.349, $p < 0.001$), but not the PI group (t (176) = 1.615, $p = 0.108$). After 9.5 years old, there was no Group X Age interaction ($b = -0.054$, t (262.364) = -0.843, $p = 0.4$, CI = -0.18, 0.072), and no significant age-

related change in amygdala volume was observed in comparisons (t (277) = 0.702, p = 0.483) or PIs (t (274) = -0.443, p = 0.658).

For hippocampal volume, a significant group main effect was detected in the sub-sample of ages 6 to 19, such that hippocampal volume was smaller in PI group relative to comparisons (b = -1.824, t (128.481) = -3.232, p < 0.01, CI = -2.934, -0.728). A significant main effect of age was also detected (b = 0.206, t (279.686) = 3.421, p < 0.01, CI = 0.089, 0.323). In summary, amygdala and hippocampal volume analyses conducted in the sub-sample of PI and comparisons from ages 6 to 19 showed the same pattern of results to the main analyses using the entire data set available.

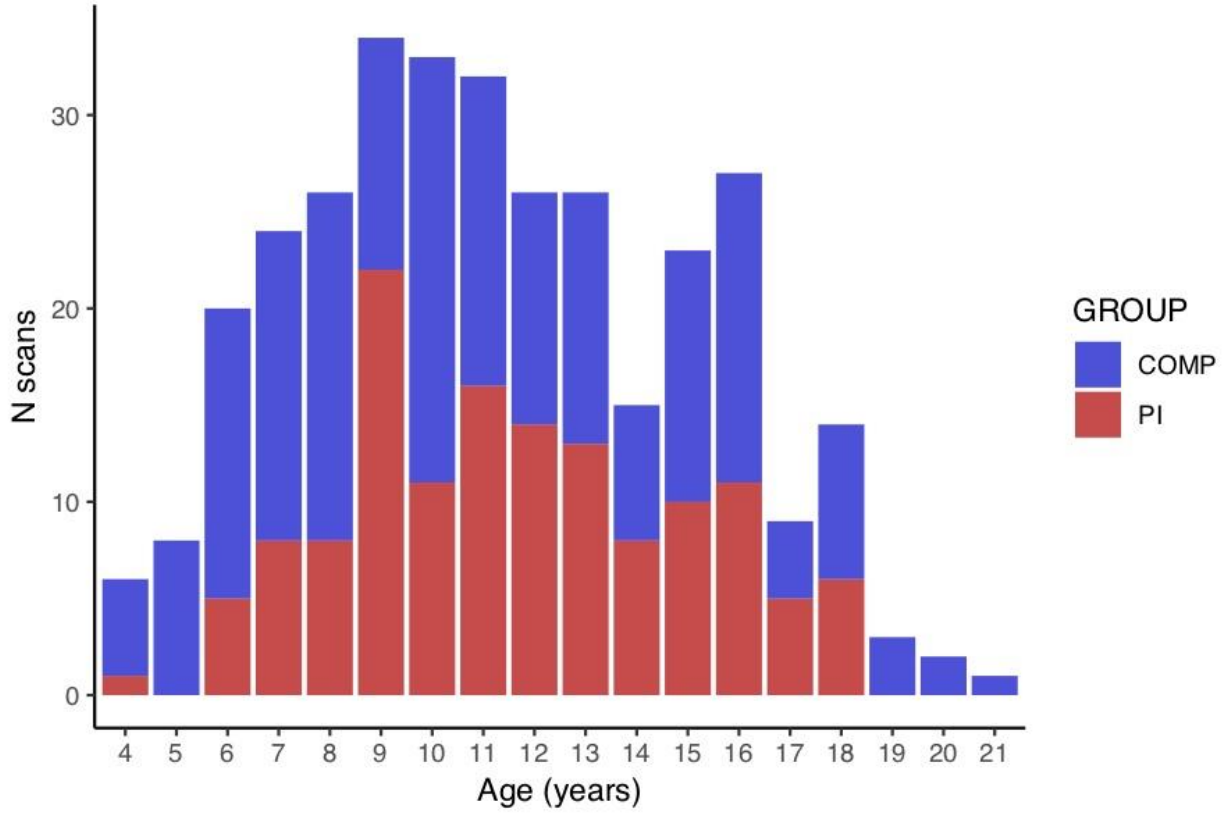

Figure S2: Age distribution for usable MRI scans by group.

1.3 Scanner change

In order to assess bias due to scanner change in wave 3, we conducted sensitivity analyses for amygdala and hippocampal volume that only included waves 1 and 2 (conducted on same scanner, with same protocol). For both amygdala and hippocampal models, results were similar to the main analyses. Specifically, the amygdala volume model with piecewise age effects (9.5 breakpoint) showed a significant Group X Age effect during childhood (younger than 9.5 years old; b = -0.527, t (156.408) = -2.91, $p < 0.01$, CI = -0.884, -0.177), but no Group X Age effects after 9.5 years old (b = 0.02, t (215.814) = 0.248, p = 0.805, CI = -0.138, 0.18). For hippocampal volume, we detected a significant main effect of group (b = -1.479, t (119.403) = -2.689, $p < 0.01$, CI = -2.572, -0.407) and age (b = 0.288, t (209.991) = 4.946, $p < 0.001$, CI = 0.175, 0.401) but no group x age interaction (b = -0.059, t (190.094) = -0.488, p = 0.626, CI = -0.293, 0.175). These results suggest that the change in scanner in wave 3 did not bias results of the main analyses.

1.4 Right and Left Hemispheres

Secondary analyses were performed on left and right hemispheres separately using the same covariates. When testing piecewise age effects with a breakpoint of 9.5 years old, we detected a significant Group X Age effect during childhood (younger than 9.5 years old) in both left and right hemispheres (Right: b = -0.51, t (249.932) = -2.733, $p < 0.01$, CI = -0.877, -0.149; Left: b = -0.48, t (245.005) = -2.444, p = 0.015, CI = -0.861, -0.1), but no Group X Age effects after 9.5 years old (Right: b = -0.005, t (293.228) = -0.064, p = 0.949, CI = -0.144, 0.137; Left b = 0.03, t (295.586) = 0.4, p = 0.689, CI = -0.114, 0.175). These effects parallel the results obtained using bilateral amygdala volume.

A significant main effect of group on hippocampal volume was detected for both the right and left hemispheres (Right: b = -1.53, t (130.445) = -2.647, p < 0.01, CI = -2.668, -0.406]; Left: b = -1.996, t (133.555) = -3.645, p < 0.001, CI = -3.074, -0.931). Similar to main analyses with bilateral hippocampus, there were also significant age effects (Right: b = 0.23, t (279.053) = 3.639, p < 0.001, CI = 0.107, 0.352]; Left: b = 0.237, t (277.766) = 3.971, p < 0.001, CI = 0.121, 0.354) but group did not moderate these effects for either right (b = -0.107, t (275.973) = -1.013, p = 0.312, CI = -0.312, 0.1) or left hemispheres (b = -0.009, t (277.376) = -0.092, p = 0.927, CI = -0.205, 0.186). Together, results of left and right hemispheres tested separately show the same pattern of results that were obtained in the main analyses with bilateral amygdala and hippocampus.

2. Secondary analyses

2.1 Age of Adoption

Prior studies have identified timing effects of early adversity on limbic brain development. For example, one influential study found that later age of adoption is associated with larger (i.e. more atypical) amygdala volume in PI children (Tottenham et al., 2010). Supplemental analyses examined whether age of adoption influenced subcortical brain development within the sample of PI youth. These analyses were performed with a sub-sample of PIs (N = 64) with known adoption timing. Age of adoption was non-normally distributed (W = 0.779, p < 0.001) so this variable was log-transformed. Mixed effects modeling was used, controlling for age effects, ICV, motion, scanner, and sex. No effect of age of adoption was detected for amygdala volume (b = 0.041, t (53.022) = 0.18, p = 0.858, CI = -0.399, 0.481) or hippocampal volume (b = -0.444, t (48.89) = -0.935, p = 0.355, CI = -1.361, 0.471). Similar

results were found when also controlling for age of placement in the institution. These findings suggest that the effects of ECA on limbic brain development are not strongly influenced by age of adoption in the current sample.

2.3 Sex effects

Main analyses showed a main effect of sex on amygdala volume, such that males have larger volumes relative to females (b = -0.52, t (131.809) = -2.162, p = 0.032). Follow-up analyses were performed using linear models to test whether sex moderated the observed group or age effects on amygdala volume, controlling for ICV, motion, and scanner. A significant Group X Sex interaction was detected (b = 1.57, t (132.888) = 3.237, p < 0.01, CI = 0.63, 2.513), shown in Figure S3. Post-hoc tests showed group differences for males (t (139) = -2.88, p < 0.01, b = -1.09), such that PI males had smaller volumes relative to comparison males, but there were no group differences for females (t (139) = 1.58, p = 0.117, b = 0.48). Likewise, while comparison males had significantly larger amygdala volumes relative to comparison females (t (135) = -3.91, p < 0.001, b = -1.22), no sex differences were observed within the PI group (t (135) = 0.95, p = 0.343, b = 0.35).

However, sex did not moderate the main effect of age (b = 0.06, t (280.281) = 1.173, p = 0.242, CI = -0.039, 0.158) and there was no Sex X Group X Age interactions (b = -0.108, t (290.075) = -1.013, p = 0.312, CI = -0.314, 0.101). These results suggest that although PI males have smaller amygdala volumes relative to comparison males, they did not drive the group differences in amygdala growth (Group X Age effects) observed in the entire sample.

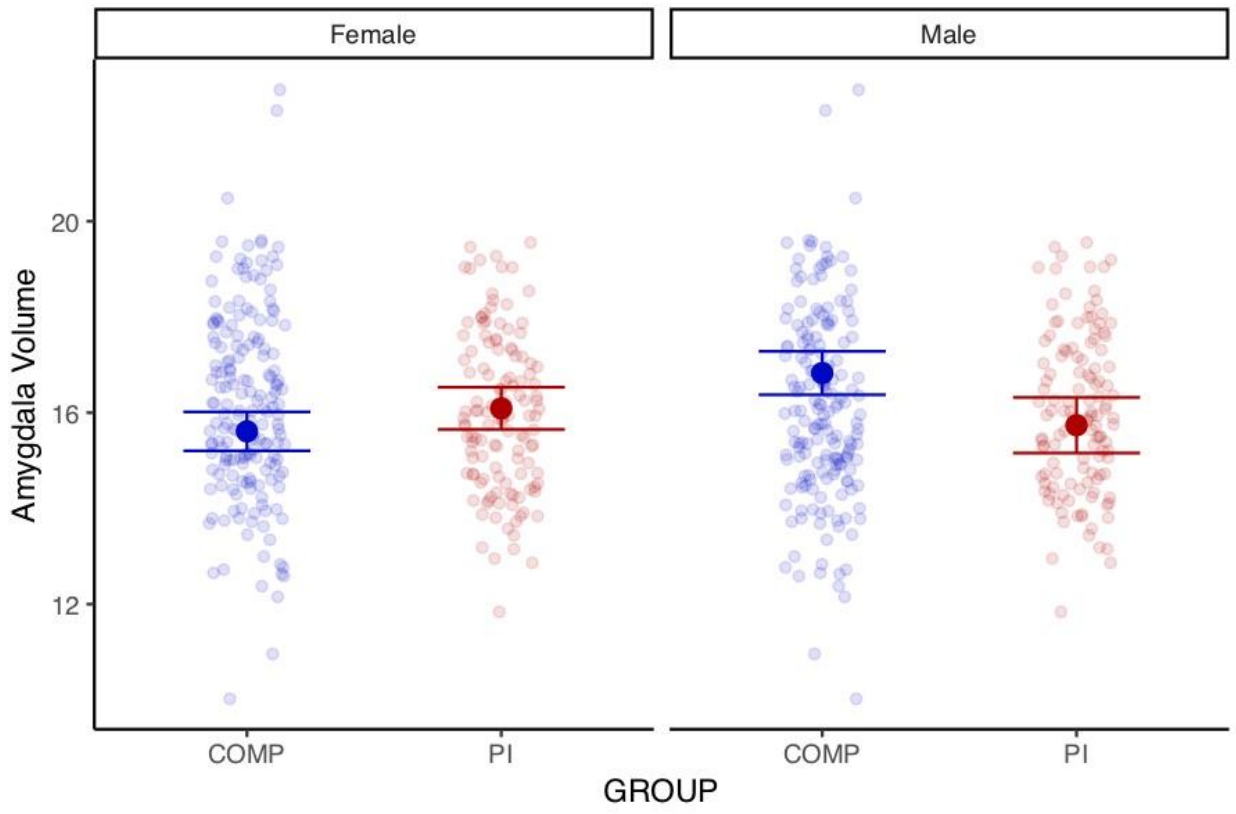

Figure S3: The effect of PI status on amygdala volume is moderated by sex, with PI males showing significantly smaller amygdala volume relative to comparison males. Means and 95% confidence intervals are shown with raw data points overlaid.

Similar to amygdala volume, we detected a main effect of sex on hippocampal volume, such that females have smaller hippocampal volumes relative to males (b = -1.379, t (127.251) = -2.523, p = 0.013). Follow-up analyses were performed to test whether sex moderated the observed group or age effects on hippocampal volume, controlling for ICV, motion, and scanner. As shown in Figure S4, a significant Group X Sex interaction was detected (b = 2.894, t (130.199) = 2.609, p = 0.01, CI = 0.744, 5.049). Post-hoc tests showed group differences for males (t (133) = -4.06, p < 0.001, b = -3.49), such that PI males had smaller volumes relative to comparison males, but there were no group differences for females (t (133) = -0.85, p = 0.4, b = -0.59). Likewise, while comparison males had significantly larger hippocampus volumes relative

to comparison females (t (132) = -3.65, p < 0.001, b = -2.59), no sex differences were observed within the PI group (t (132) = 0.36, p = 0.719, b = 0.3).

However, sex did not moderate the main effect of age on hippocampal volume (b = -0.104, t (269.675) = -1.119, p = 0.264, CI = -0.283, 0.079) and there was no Sex X Group X Age interactions (b = -0.011, t (260.656) = -0.057, p = 0.955, CI = -0.388, 0.364). In summary, these results suggest that the main effects of ECA on hippocampal volume in the current sample are driven by smaller volumes in PI males relative to comparison males.

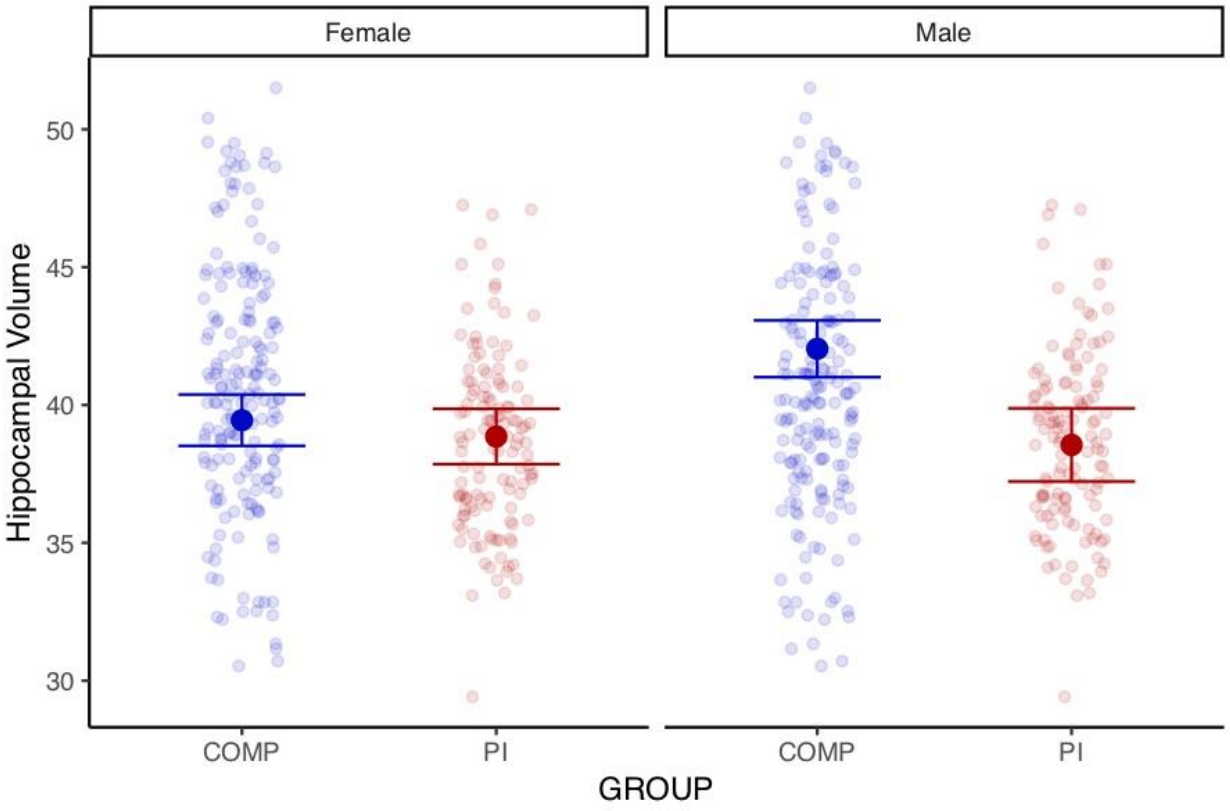

Figure S4. Sex moderates the effect of PI status on hippocampal volume, with PI males showing significantly smaller hippocampal volume relative to comparison males. Means and 95% confidence intervals are shown with raw data points overlaid.

2.3 Pubertal hormones

Recent research has suggested that subcortical brain development may be driven by changes in pubertal hormones (e.g. testosterone), as opposed to age (Herting et al., 2014). For these reasons, we conducted follow-up analyses to assess the role of testosterone on group differences observed on amygdala and hippocampal volume. As was done for the cortisol analyses, testosterone was rank-normed within sex to account for sex differences. We used linear mixed effects modeling to assess the interaction between sex-normed testosterone and PI status. Because testosterone and age are highly collinear ($r = 0.609$, $p < 0.001$) each model was conducted two ways: (1) with testosterone alone, and (2) with testosterone and age in the same model. These analyses were conducted in a sub-sample of participants with available testosterone obtained from saliva samples (Table S18) and controlled for batch of saliva sample processing, sex, scanner, ICV and motion.

Table S18. Demographic information for sub-sample of subjects with usable structural MRI scans and testosterone.

			Age (years)			Testosterone (normed)		
GROUP	Wave	N	Mean	SD	Range	Mean	SD	Range
COMP	1	56	11.31	3.70	4.33-18.58	0.50	0.20	0-0.92
	2	47	11.77	4.37	4.83-20.33	0.57	0.23	0.01-1
	3	24	12.15	3.92	6.67-19.08	0.57	0.24	0.15-0.94
PI	1	38	10.58	3.00	4.58-16.58	0.47	0.18	0.07-0.94
	2	32	12.20	3.38	6.75-18.25	0.61	0.17	0.21-0.97
	3	28	14.00	3.40	7.08-18.83	0.61	0.20	0-0.94

A significant main effect of testosterone was detected such that amygdala volume increased as a function of increasing testosterone (b = 1.362, t (211.125) = 2.752, $p < 0.01$, CI = 0.399, 2.326). When including age in the model, both testosterone (b = 0.608, t (191.665) = 1.083, p = 0.28, CI = -0.487, 1.693), and age effects were significant (b = 0.11, t (168.812) = 2.901, $p < 0.01$, CI = 0.035, 0.185). Testosterone also moderated group effects on amygdala volume (b = -1.729, t (180.072) = -2.145, p = 0.033, CI = -3.318, -0.16). These interaction effects are similar to the main findings with age: we detected a significant increase in amygdala volume with increasing testosterone in comparisons (t (195) = 2.73, $p < 0.01$, b = 1.52) and a negative, non-significant effect in PIs (t (195) = -0.33, p = 0.744, b = -0.21). When including age in the model, we detected a similar interaction effect (b = -1.602, t (184.748) = -1.957, p = 0.052, CI = -3.198, -0.022). However, simple slopes of this interaction showed that the relationship between testosterone on amygdala volume was not significant in comparisons (t (195) = 2.73, p = 0.226, b = 0.81) or the PI group (t (195) = -0.33, p = 0.252, b = -0.79). In addition, the effect of age on amygdala volume was still significant in this model (b = 0.086, t (188) = 2.294, p = 0.032). These results do not provide strong evidence that testosterone is driving the observed developmental effects of ECA exposure on amygdala volume. Instead, these results suggest that relative to comparisons, the PI group shows reduced growth of the amygdala relative to comparisons, regardless of whether 'growth' is measured by age or pubertal hormones.

For hippocampal volume, we detected a significant main effect of testosterone, such that hippocampal volume increased with increasing testosterone levels (b = 2.54, t (176.09) = 2.651, $p < 0.01$, CI = 0.672, 4.414). However, when controlling for significant age effects in the model (b = 0.176, t (165.67) = 2.128, p = 0.035, CI = 0.013, 0.335), testosterone effects were not significant (b = 1.624, t (139.508) = 1.531, p = 0.128, CI = -0.425, 3.669) suggesting shared

variance between age and testosterone on hippocampal growth. Testosterone did not moderate group effects on hippocampal volume (without age: b = -1.239, t (135.3) = -0.829, p = 0.409, CI = -4.177, 1.664]; with age: (b = -0.975, t (130.975) = -0.65, p = 0.517, CI = -3.908, 1.925). Together, these results provide further evidence that although early institutional care is associated with reduced hippocampal size, it does not influence hippocampal growth rates relating to pubertal hormones or age.

References for Appendix B Supplemental Analyses:

Cohen, P., Cohen, J., Aiken, L. S., & West, S. G. (1999). The problem of units and the circumstance for POMP. *Multivariate Behavioral Research*, *34*(3), 315–346. https://doi.org/10.1207/S15327906MBR3403_2

Fareri, D. S., Gabard-Durnam, L., Goff, B., Flannery, J., Gee, D. G., Lumian, D. S., … Tottenham, N. (2015). Normative development of ventral striatal resting state connectivity in humans. *NeuroImage*, *118*, 422–437. https://doi.org/10.1016/j.neuroimage.2015.06.022, PMCID: PMC5553607

Flannery, J. E., Gabard-Durnam, L. J., Shapiro, M., Goff, B., Caldera, C., Louie, J., … Tottenham, N. (2017). Diurnal cortisol after early institutional care—Age matters. *Developmental Cognitive Neuroscience*, *25*(March), 160–166. https://doi.org/10.1016/j.dcn.2017.03.006

Gunnar, Megan R., DePasquale, C. E., Reid, B. M., & Donzella, B. (2019). Pubertal stress recalibration reverses the effects of early life stress in postinstitutionalized children. *Proceedings of the National Academy of Sciences*, 201909699. https://doi.org/10.1073/pnas.1909699116

Herting, M. M., Gautam, P., Spielberg, J. M., Kan, E., Dahl, R. E., & Sowell, E. R. (2014). The role of testosterone and estradiol in brain volume changes across adolescence: A longitudinal structural MRI study. *Human Brain Mapping*, *In Press*(February), In Press. https://doi.org/10.1002/hbm.22575

King, L. S, Colich, N. L., Lemoult, J., Humphreys, K. L., Ordaz, S. J., Price, A. N., & Gotlib, I. H. (2017). The impact of the severity of early life stress on diurnal cortisol: The role of

puberty. *Psychoneuroendocrinology*, *77*, 68–74. https://doi.org/10.1016/j.psyneuen.2016.11.024

Kumsta, R., Schlotz, W., Golm, D., Moser, D., Kennedy, M., Knights, N., … Sonuga-Barke, E. (2017). HPA axis dysregulation in adult adoptees twenty years after severe institutional deprivation in childhood. *Psychoneuroendocrinology*, *86*, 196–202. https://doi.org/10.1016/j.psyneuen.2017.09.021

Mclaughlin, K. A., Sheridan, M. A., Tibu, F., Fox, N. A., Zeanah, C. H., & Nelson, C. A. (2015). Causal effects of the early caregiving environment on development of stress response systems in children. *Proceedings of the National Academy of Sciences*, *112*(18). https://doi.org/10.1073/pnas.1423363112

Tottenham, N., Hare, T. a, Quinn, B. T., McCarry, T. W., Nurse, M., Gilhooly, T., … Casey, B. J. (2010). Prolonged institutional rearing is associated with atypically large amygdala volume and difficulties in emotion regulation. *Developmental Science*, *13*(1), 46–61. https://doi.org/10.1111/j.1467-7687.2009.00852.x