Title: Development of subcortical volumes across adolescence in males and females: A multisample study of longitudinal changes

Authors: Megan M. Herting[1], Cory Johnson[1], Kathryn L. Mills[2], Nandita Vijayakumar[2], Meg Dennison[3], Chang Liu[1], Anne-Lise Goddings[5], Ronald E. Dahl[6], Elizabeth R. Sowell[7], Sarah Whittle[4], Nicholas B. Allen[2], Christian K. Tamnes[8]

[1] Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA

[2] Department of Psychology, University of Oregon, Eugene, OR, USA

[3] Phoenix Australia: Centre for Posttraumatic Mental Health, Department of Psychiatry, The University of Melbourne, Melbourne, AU

[4] Melbourne Neuropsychiatry Centre, Department of Psychiatry, The University of Melbourne and Melbourne Health, Melbourne, AU

[5] Institute of Child Health, University College London, London, UK

[6] Institute of Human Development, University of California Berkeley, Berkeley, CA, USA

[7] Department of Pediatrics, Keck School of Medicine, University of Southern California, and Children's Hospital Los Angeles, Los Angeles, CA, USA,

[8] Department of Psychology, University of Oslo, Oslo, Norway

Corresponding author: Megan M. Herting, Department of Preventive Medicine, University of Southern California, 2001 N Soto, Los Angeles, CA, 90032, USA; Email: herting@usc.edu

Abstract

The developmental patterns of subcortical brain volumes in males and females observed in previous studies have been inconsistent. To help resolve these discrepancies, we examined developmental trajectories in three independent longitudinal samples of participants with the age-span of 8-22 years (total 216 participants and 467 scans). These datasets, including *Pittsburgh* (PIT; University of Pittsburgh, USA), *NeuroCognitive Development* (NCD; University of Oslo, Norway), and *Orygen Adolescent Development Study* (OADS; The University of Melbourne, Australia), span three countries and were analyzed in parallel using mixed-effects modeling with both generalized additive models and general linear models. The results showed that absolute volume estimates largely overlapped across samples for the thalamus, caudate, putamen, hippocampus, and amygdala in each sex, whereas volumes were less consistent across samples for the pallidum and nucleus accumbens. In males, the thalamus showed increasing volume until mid-adolescence, whereas females showed no change in two samples and an inverse-U trajectory in NCD. For the caudate and putamen, decreases were seen for both sexes and all samples, except for males in PIT. The trajectories for the pallidum varied by sample, with males and females in OADS having smaller volumes in childhood, but catching up with the other samples by adulthood. Although absolute volumes were different for the nucleus accumbens across samples, sex differences in volume changes were similar with linear decreases observed for females and stability in males. Slightly curvilinear patterns of hippocampal growth were seen in all samples. For the amygdala, females showed early increases followed by little change, whereas males displayed steady increases in volume. Our results help to clarify sex specific patterns in subcortical brain changes across adolescence, and highlight region-specific variations in congruency of developmental trajectories.

Keywords: neurodevelopment; subcortical; adolescence; sex differences; replication; longitudinal

1. Introduction

Developmental patterns of brain morphology, and sex differences in this structural variation, exist due to both global and local maturational changes (Erus et al., 2015; Giedd et al., 2015; Narvacan et al., 2017; Sowell et al., 2004; Tamnes et al., 2013). Determining when and how sex differences emerge in the developing brain is essential to understanding differential risk for disease, especially psychopathology (Kessler et al., 2005; Kessler et al., 1993), as well as life-long sex differences in various cognitive and behavioral traits (Choudhury et al., 2006; Gur and Gur, 2016; Roalf et al., 2014; Rose and Rudolph, 2006). A number of sex differences have been reported in structural MRI growth trajectories of subcortical structures. However, to date, developmental patterns observed in these structures have been inconsistent across studies, and there has yet to be a consensus as to how these patterns differ between sexes (Dennison et al., 2013; Lenroot et al., 2007; Narvacan et al., 2017; Ostby et al., 2009; Sowell et al., 2002; Wierenga et al., 2014).

To date, different studies have reported discrepant findings including growth versus reduction of the thalamus and basal ganglia beginning in late childhood, as well as stability versus continuing growth of the amygdala and hippocampus across adolescence (Giedd et al., 1996; Koolschijn and Crone, 2013; Ostby et al., 2009; Wierenga et al., 2014). Moreover, while some studies have found larger absolute volumes of subcortical regions in males and/or similar trajectories of development between the sexes (Narvacan et al., 2017; Wierenga et al., 2014), other studies suggest that the thalamus may be larger in females as compared to males (Ruigrok et al., 2014) and that sex differences may exist in volumetric change over time in regions including the striatum, thalamus, amygdala, hippocampus, and pallidum (Dennison et al., 2013; Goddings et al., 2014; Koolschijn and Crone, 2013; Lenroot et al., 2007; Raznahan et al., 2014; Sowell et al., 2002). These discrepant observations in studies of patterns of subcortical volume development between sexes may be due to a number of factors, including cohort effects inherent to the sample, variation in study design, image preprocessing, and/or statistical modeling approaches. From a study design perspective, cross-sectional findings that

focus on mean group differences may differ from longitudinal study designs that are able to account for individual differences over time (Crone and Elzinga, 2015). In terms of image processing, dissimilarities have also been reported in the absolute volume estimates as well as in the reliability of subcortical brain structures across different freely available automated segmentation software (Morey et al., 2010). Beyond software, differences in quality control procedures utilized across studies may also impact the results (Ducharme et al., 2016).

Moreover, analytic techniques can vary widely by study (Vijayakumar et al., Accepted). Age-related changes in subcortical volumes have largely been examined using general linear regression for cross-sectional studies. Longitudinal studies have reported difference scores, or more recently, used multi-level modeling to describe age-related changes, but the model terms are diverse. For example, studies have differed in their modeling approach, including use of polynomial terms (e.g. quadratic or cubic), model selection strategy (e.g. top-down or likelihood indices), testing males and females separately and/or including sex as an interaction term, as well as regarding the inclusion of other confounding factors (Ruigrok et al., 2014).

When examining subcortical volumes, one common covariate is a 'global' metric, to account for *between subject* differences in body size or weight (Sanfilipo et al., 2004). That is, because males have been repeatedly found to have larger volumes for total cranial, total gray, total white, and a number of subcortical regions of interest (Ruigrok et al., 2014), a normalization method (proportional or residual) has often been implemented when exploring sex difference across development. However, this reasoning perpetuates a common misunderstanding of the use of covariates when attempting to study preexisting groups (i.e. sexes) that systematically differ on more than one covariate (Miller and Chapman, 2001). That is, the covariate in this case will leave those systematic differences in place (e.g. create a specification bias) and potentially obfuscating an accurate description of sex differences in developmental trajectory. Further, as knowledge about developmental changes in global brain volumes has increased, a number of additional reasons to oppose normalization in longitudinal studies have emerged. Sex differences have been reported in the variability of anatomical

volumes, with males showing larger variance expressed at both upper and lower extremities of the distributions (Wierenga et al., 2017); suggesting that normalization distributions cannot be assumed to be equal between the sexes. Moreover, global metrics (e.g. intracranial or whole brain volume) follow non-linear and different developmental patterns from childhood to adulthood (Mills et al., 2016). Thus, applying a normalization correction to longitudinal data, when aiming to determine if and how regional volumes change over time, generates a fundamentally distinct question of how regional volume changes are coordinated with another volume change within subjects.

For these reasons, previous cross-sectional and longitudinal sex difference studies are also complicated by inconsistent descriptions of either 'raw' or 'corrected' volume estimates, which are often interpreted as being equivalent, but require substantially different interpretations. Because of discrepancies between previous studies in either analysis or interpretation, there remains an important gap in our knowledge regarding the description of raw subcortical neurodevelopmental trajectories in males and females.

The goal of the current study was to utilize identical image processing and analysis methods in three independent longitudinal neuroimaging samples to describe the development of raw subcortical volumes for males and females from late childhood into young adulthood. The study is part of an international collaboration project intended to improve the reliability and efficiency of neurodevelopmental research by simultaneously analyzing multiple existing neuroimaging datasets (Mills et al., 2016; Tamnes et al., 2017). By controlling for software, quality control procedures, and statistical methods across samples, we could assess and interpret the potential impact of sample and acquisition differences on brain development. Thus, in the current study, each independent dataset was analyzed separately to examine the consistency and reproducibility of neurodevelopmental change for subcortical gray matter regions, including the thalamus, caudate, putamen, pallidum, hippocampus, amygdala, and nucleus accumbens in males and females.

2. Materials and Methods

2. 1 Participants

This study analyzed data from typically developing youth from three separate cohorts collected utilizing cohort-sequential longitudinal designs at three separate sites in independent research projects: *Pittsburgh* (PIT; University of Pittsburgh, USA), *NeuroCognitive Development* (NCD; University of Oslo, Norway), and *Orygen Adolescent Development Study* (OADS; The University of Melbourne, Australia). Each project was approved by their respective local review board and informed consent/assent was obtained from parents and children prior to data collection. In order to best account for within-subject variance, only participants with ≥2 scans from each cohort were included in analyses. For the PIT dataset, 126 participants were recruited and scanned at baseline, with 20 not completing their follow-up visit, and 33 excluded due to poor image quality of the MRI (see description of quality checking (QC) procedures in section 2.2). For NCD, 111 participants were recruited and scanned at baseline; 26 were unable to complete their follow-up visit, and 9 were excluded due to poor image quality. For OADS, 177 participants completed baseline and follow-up scans, of which 23 did not complete any additional follow-up visits, 76 were excluded due to psychiatric history and medical illness and 11 were excluded due to poor image quality. The final samples thus included 73 participants from PIT, 76 from NCD, and 67 from OADS. In total, the present study included 216 participants (110 females) and 467 scans covering the age range of 8 to 22 years (see Table 1). Details regarding participant recruitment in each project have been previously described (Herting et al., 2014; Tamnes et al., 2013; Yap et al., 2011).

Table 1. Sample demographics of typically developing youth from three separate cohorts collected utilizing cohort-sequential longitudinal designs at three separate sites in independent research projects. PIT: *Pittsburgh*, USA; NCD: *NeuroCognitive Development*, Norway; OADS: *Orygen Adolescent Development Study*; Australia. Values represent Means and (standard deviations) unless otherwise noted.

	PIT			**NCD**			**OADS**		
	All	Female	Male	All	Female	Male	All	Female	Male
N	73	41	32	76	37	39	67	32	49
Age (years)	12.3 (.9)	11.9 (.7)	12.9 (.7)*	15.2 (3.6)	15.1 (3.5)	15.4 (3.7)	16.2 (2.6)	16.3 (2.5)	16.1 (2.6)
Age Range (years)	10.1-14.0	10.1-13.5	11.4-14.0	8.2-21.9	8.4-21.8	8.2-21.9	11.5-20.3	12.1-19.7	11.5-20.3
Total Scans (N)	146	82	64	152	74	78	169	81	88
Participants with 2 Scans (N)	73	41	32	76	37	39	32	15	17
Participants with 3 Scans (N)	--	--	--	--	--	--	35	17	18
Was	2.2 (.4)	2.2 (.4)	2.1 (.4)	2.6 (.2)	2.7 (.2)	2.6 (.2)	3.09 (.9)	3.09 (.9)	3.19 (.8)

* Age difference between sexes by design (see Supplementary Material for details).

2.2 Image Acquisition and Analysis

T1-weighted anatomical scans were obtained at the three sites using different MRI scanners and sequences (see Supplementary Material). At each site, a radiologist reviewed all scans for incidental findings of gross abnormalities. Image processing, including whole brain segmentation with automated labeling of different neuroanatomical structures, was performed using the longitudinal pipeline of FreeSurfer 5.3 (http://surfer.nmr.mgh.harvard.edu; (Fischl et al., 2002; Reuter et al., 2012). The longitudinal pipeline includes creating an unbiased within-subject template space and image using inverse consistent registration. Skull stripping, Talariach transform and atlas registration, and parcellations are initialized in the common within-subject template, which increases reliability and statistical power. Similar standard quality control (QC) procedures were carried out between sites. QC details are as follows: 1) all raw images were visually inspected for motion prior to processing, 2) post-processed images were visually inspected by trained operators for accuracy of subcortical segmentation by the longitudinal pipeline for each scan per participant, 3) images with inaccurate segmentation were excluded (number of participants excluded during QC is outlined above in section 2.1). No manual intervention (i.e. subcortical editing) was performed. Regions of interest for the present study included the thalamus, caudate, putamen, pallidum, amygdala, hippocampus, and nucleus accumbens for each hemisphere.

2.3 Statistical Analyses

Given previous findings highlighting hemispheric differences (Dennison et al., 2013; Herting et al., 2014), we first examined if patterns of change differed by hemisphere by plotting loess curves to each dataset. Overall, trajectories were similar between hemispheres (see Supplementary Material SFigures 1-7), and therefore left and right hemisphere volume estimates were averaged for all subsequent analyses. Given that one of the primary aims of the study was to examine the distinct developmental trajectories in males and females, and preliminary exploratory loess plots confirmed the shape of developmental trajectories varied

between males and females (see Supplementary Material SFigures 1-7), analyses were conducted for females and males separately.

To more fully understand sex specific subcortical volume changes from late childhood and throughout adolescence, we conducted mixed-effects modeling with both generalized additive models (mgcv package version 1.8-17) and general linear models (R version 3.4.0; nlme package version 3.1-131). First, the data were centered at the mean age of the 3 cohorts (15 years). Next, spaghetti plots were created and longitudinal curve fitting was performed using generalized additive mixed modeling (GAMM) to accurately describe the longitudinal developmental trajectories of subcortical volumes in each sex in each sample separately. Unlike parametric general linear modeling, this method does not require *a priori* knowledge of the functional form for the data and as an extension of general linear mixed effects modeling (LME), GAMM replaces one or more of the linear predictor terms with a spline term. The spline term is a non-linear smooth function that describes the best relationship between the covariate(s) and the outcome variable of interest. For GAMM analyses on subcortical structure volume, the main predictor was age. GAMM can be represented by the following formula:

$$G(y) = X^{*}\alpha + \sum_{j=1}^{p} f_j(x_j) + Zb + \varepsilon$$

where G(y) is a monotonic differentiable link function, α is the vector of regression coefficients for the fixed parameters; X* is the fixed-effects matrix; f_j is the smooth function of the covariate x_j; Z is the random-effects model matrix; b is the vector of random-effects coefficients; and ε is the residual error vector.

Second, LME was performed to formally test different developmental models. Moreover, this method is the most commonly used approach in longitudinal studies assessing brain developmental trajectories (Vijayakumar et al., Accepted), and we could therefore more directly compare our results with those reported in previous studies. This method estimates the fixed effect of measured variables on subcortical volume while including within-person variation as

nested random effects in the regression model. This is done to account for individual subject effects and correlation of the data inherent to longitudinal analysis. LME can be represented by the following formula for linear changes with age both between and within participants:

$$Volume_{ij} = Intercept_{0i} + \alpha_{1i}(age)_{ij} + \varepsilon$$

where $Volume_{ij}$ represents the volume in an ROI at the jth timepoint for the ith participant, the $intercept_{0i}$ represents the grand mean at the centered age (age 15), α_{1i} is the grand mean slope of age (linear); and ε is the residual error and reflects within-person variance. All models also included a random intercept for each participant. The linear model was then built upon to also include quadratic and cubic fixed terms to assess linear versus more complex patterns of change. The linear, quadratic, and cubic models were as follows:

1. $Linear\ model: Volume = Intercept + \alpha(age) + \epsilon$
2. $Quadratic\ model: Volume = Intercept + \alpha(age) + \beta(age^2) + \epsilon$
3. $Cubic\ model: Volume = Intercept + \alpha(age) + \beta(age^2) + \gamma(age^3) + \epsilon$

where α, β, and γ represent the effects of each fixed term. Likelihood ratio tests and Akaike Information Criterion (AIC) were used to compare the models and to determine which had the best fit. All models were tested against a null model that included only the intercept term, but not the fixed effect of age. The model with the lowest AIC that was also significantly different from the less complex model as determined by the likelihood ratio test was chosen as the best fit model (e.g. linear had to have a lower AIC and be significantly different from null; quadratic had to have a lower AIC and be significantly different from both the null and linear model).

3. Results

3.1 Description of developmental trajectories using GAMM

Developmental trajectories for volume of regions of interest were visualized with spaghetti plots fitted with GAMM (Figure 1 and 2) for females and males for each of the three independent samples. Overall, despite different shapes of the trajectories, volume estimates in each sex largely overlapped for all three cohorts across age for all regions, except the pallidum and

nucleus accumbens. Developmental volume trajectories for each separate subcortical structure are described below.

Figure 1. Developmental age trajectories for the thalamus, caudate, putamen, and pallidum for females and males. The bolded colored lines represent the GAMM fitting for each sample with 95% confidence intervals. Individual datapoints are shown, connected for each participant, in the appropriate sample color.

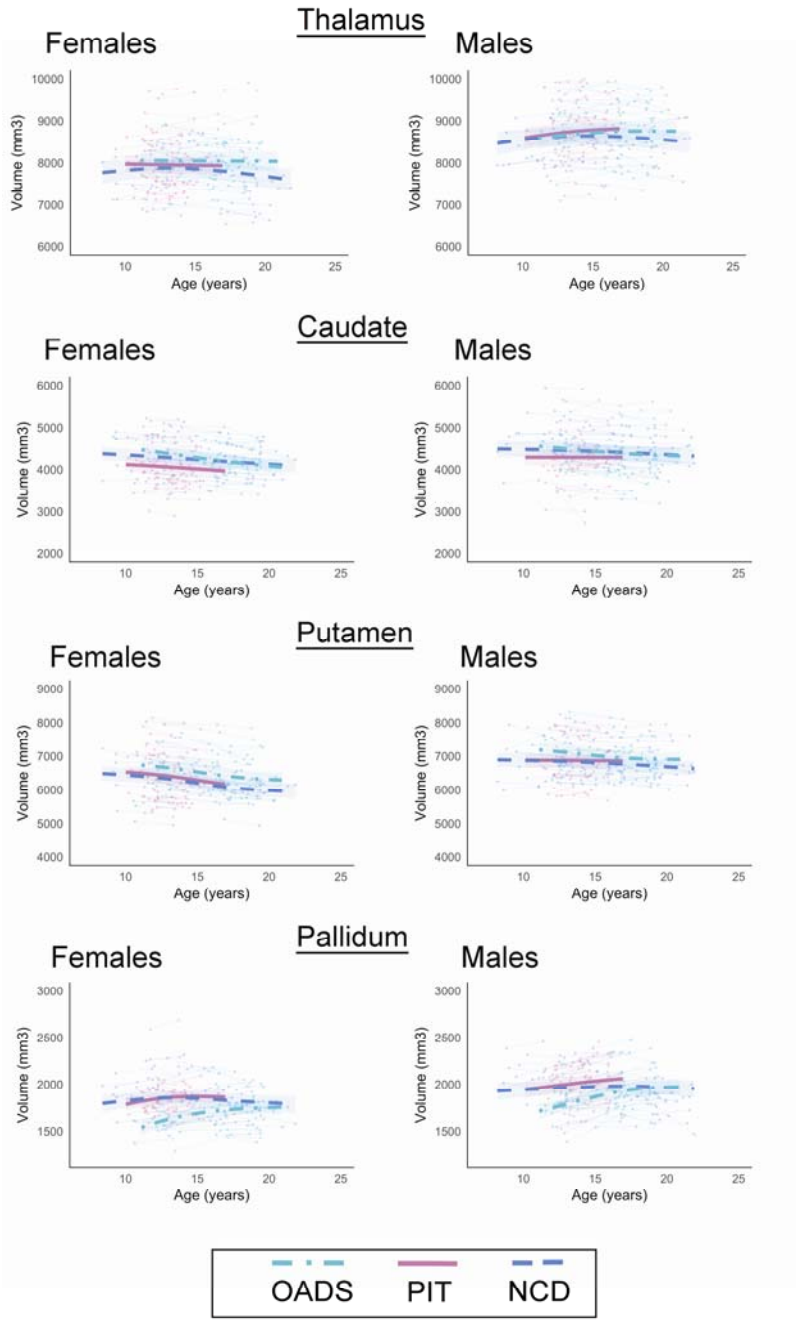
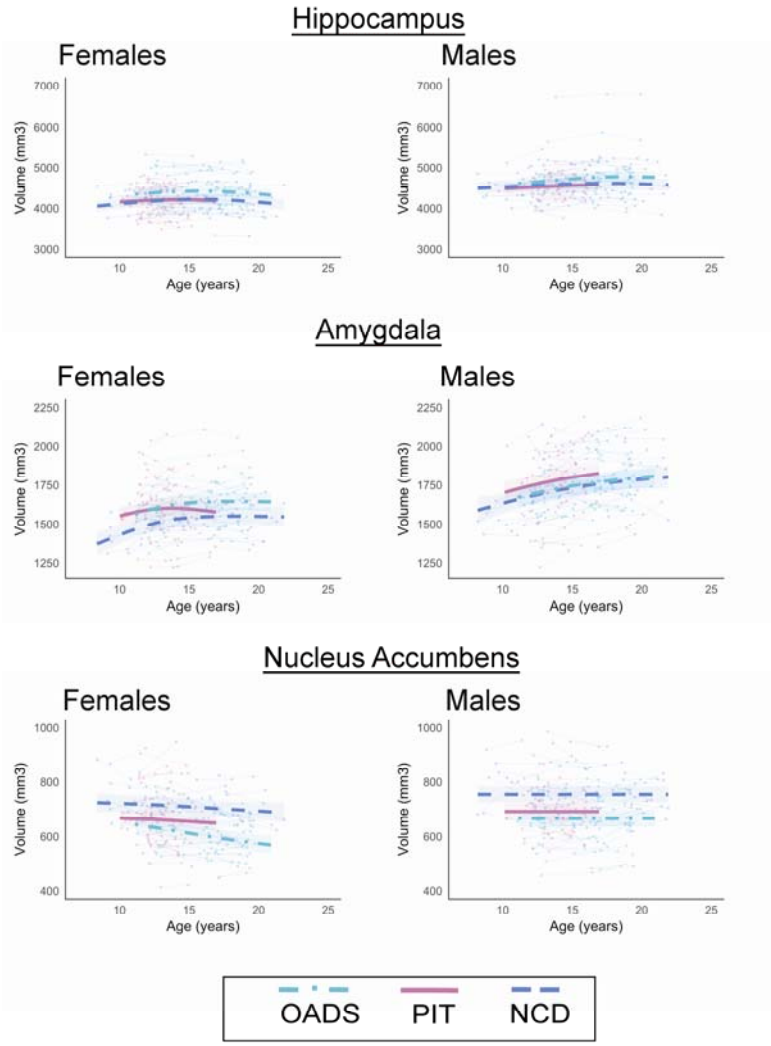


Figure 2. Developmental age trajectories for the hippocampus, amygdala, and nucleus accumbens for females and males. The bolded colored lines represent the GAMM fitting for each sample while the lighter color reflects the confidence intervals (95%). Each individual can also be seen, with colors of points and lines reflecting sample.



3.1.1 Thalamus

In the NCD sample, females showed a curvilinear trajectory between ages 8 and 22 years, with a slight increase in volumes until mid-adolescence, followed by a decrease into young adulthood. In contrast, no changes were seen in thalamus volumes for females in the OADS and PIT samples across adolescence and into young adulthood. In males, the developmental trajectories in all three samples were similar, showing increases in thalamus volumes until mid-adolescence, followed by deceleration in young adulthood.

3.1.2 Caudate

In females, a relatively steady decrease was seen in caudate volume in all samples, with a slightly steeper slope in the OADS sample. In males, a decrease in caudate volumes was also seen for the OADS and NCD, whereas little change was found with age in the PIT sample.

3.1.3 Putamen

In females, the putamen volume decreased across the whole age range in all three samples, with slight acceleration in decline during the adolescent period as compared to late childhood or young adulthood. In males, a decrease in volume was also seen in the NCD and OADS with age, while males in the PIT samples showed no changes in volumes with age. Overall, the reduction in putamen volumes also appeared greater in females as compared to males across the ages of 8 to 22 years.

3.1.4 Pallidum

In both males and females, increases in pallidum volumes were seen in early childhood through mid-adolescence, followed by little change or slight decreases in volumes in young adulthood. Trajectories, however, varied drastically by samples, with both females and males from the OADS showing the largest change over time, with volumes catching up to PIT and NCD estimates by later adolescence and young adulthood.

3.1.5 Hippocampus

All three samples showed relatively small developmental changes in hippocampal volumes. Females showed curvilinear patterns of hippocampal growth, with an increase until mid-adolescence, followed by a decrease into young adulthood. Males showed early volume increases, which decelerated in adolescence.

3.1.6 Amygdala

Largely parallel trajectories for amygdala volumes were found between the three cohorts, with apparent sex differences in the pattern of change. Females showed an increase in amygdala volume during childhood and early adolescence, followed by little changed in late adolescence and young adulthood. In contrast, males showed relatively steady increases in amygdala volume across the included age ranges in all three samples.

3.1.7 Nucleus Accumbens

Similar results were found across samples within each sex for the nucleus accumbens. No changes were seen for males in any of the samples across the age range. Conversely, linear decreases were seen for females, with the OADS sample showing a steeper decrease across adolescence as compared to the PIT and NCD samples.

3.2 Testing of developmental models using LME

Linear, quadratic, and cubic LME were used to formally determine best fit models for females and males of each sample. The highest-order polynomial model for each brain region is summarized in Table 2 (for AIC comparisons, see Supplementary Tables 1-7). Overall, LME best fit models per sample were in agreement with the GAMM trajectories. The only exception to this was for the hippocampus in NCD males (quadratic: p=.09), where the best fit trajectory that was similar to GAMM estimates did not reach statistical significance. However, best fits were different between samples in both sexes for most ROIs, except for the caudate and nucleus accumbens. For the caudate, both males and females in each sample showed similar trajectories with age (PIT: linear, NCD: quadratic, and OADS: linear). For the nucleus

accumbens, no significant change was found in all samples, except for OADS females, which showed a linear decrease with age.

Table 2. LME best fit models for each brain region by sex and sample. Highest-order best fit polynomial model summary reported. Test refers to the highest-order model compared to previous sequential model tested (1 = null model, 2 = linear, 3 = quadratic, 4= cubic) and its associated Likelihood Ratio (L.Ratio) and p-value (***bold*** highlights p<.05).

THALAMUS				
FEMALES	**Model Result**	**Test**	**L.Ratio**	**p-value**
PIT	Null Model	1 vs 2	0.75	0.386
NCD	Quadratic Age Model	2 vs 3	15.35	***0.000***
OADS	Null Model	1 vs 2	0.83	0.362
MALES	**Model Result**	**Test**	**L.Ratio**	**p-value**
PIT	Quadratic Age Model	2 vs 3	7.29	***0.007***
NCD	Cubic Age Model	2 vs 4	12.47	***0.002***
OADS	Linear Age Model	1 vs 2	6.91	***0.009***
CAUDATE				
FEMALES	**Model Result**	**Test**	**L.Ratio**	**p-value**
PIT	Linear Age Model	1 vs 2	13.70	***0.000***
NCD	Quadratic Age Model	2 vs 3	5.02	***0.025***
OADS	Linear Age Model	1 vs 2	65.50	***<.0001***
MALES	**Model Result**	**Test**	**L.Ratio**	**p-value**
PIT	Null Model	1 vs 2	3.45	0.063
NCD	Quadratic Age Model	2 vs 3	4.61	***0.032***
OADS	Linear Age Model	1 vs 2	38.91	***<.0001***
PUTAMEN				
FEMALES	**Model Result**	**Test**	**L.Ratio**	**p-value**
PIT	Linear Age Model	1 vs 2	19.52	***<.0001***
NCD	Linear Age Model	1 vs 2	33.31	***<.0001***
OADS	Cubic Age Model	2 vs 4	8.90	***0.012***
MALES	**Model Result**	**Test**	**L.Ratio**	**p-value**
PIT	Null Model	1 vs 2	0.38	0.537
NCD	Quadratic Age Model	2 vs 3	3.96	***0.047***
OADS	Linear Age Model	1 vs 2	25.40	***<.0001***
PALLIDUM				
FEMALES	**Model Result**	**Test**	**L.Ratio**	**p-value**
PIT	Quadratic Age Model	2 vs 3	8.61	***0.003***
NCD	Cubic Age Model	2 vs 4	6.50	***0.039***
OADS	Quadratic Age Model	2 vs 3	5.13	***0.024***
MALES	**Model Result**	**Test**	**L.Ratio**	**p-value**
PIT	Linear Age Model	1 vs 2	4.93	***0.026***
NCD	Null Model	1 vs 2	0.49	0.482
OADS	Quadratic Age Model	2 vs 3	8.14	***0.004***
HIPPOCAMPUS				
FEMALES	**Model Result**	**Test**	**L.Ratio**	**p-value**
PIT	Cubic Age Model	2 vs 3	6.66	***0.036***
NCD	Quadratic Age Model	2 vs 3	14.20	***0.000***
OADS	Quadratic Age Model	2 vs 3	14.46	***0.000***
MALES	**Model Result**	**Test**	**L.Ratio**	**p-value**
PIT	Null Model	1 vs 2	0.57	0.452
NCD	Null Model	1 vs 2	0.49	0.482
OADS	Cubic Age Model	2 vs 3	8.66	***0.013***
AMYGDALA				
FEMALES	**Model Result**	**Test**	**L.Ratio**	**p-value**
PIT	Quadratic Age Model	2 vs 3	4.60	***0.032***
NCD	Quadratic Age Model	2 vs 3	10.92	***0.001***
OADS	Linear Age Model	1 vs 2	6.19	***0.013***
MALES	**Model Result**	**Test**	**L.Ratio**	**p-value**
PIT	Linear Age Model	1 vs 2	4.91	***0.027***
NCD	Cubic Age Model	2 vs 3	8.53	***0.014***
OADS	Linear Age Model	1 vs 2	21.36	***<.0001***
NUCLEUS ACCUMBENS				
FEMALES	**Model Result**	**Test**	**L.Ratio**	**p-value**
PIT	Null Model	1 vs 2	0.43	0.510
NCD	Null Model	1 vs 2	2.70	0.100
OADS	Linear Age Model	1 vs 2	21.82	***<.0001***
MALES	**Model Result**	**Test**	**L.Ratio**	**p-value**
PIT	Null Model	1 vs 2	0.09	0.764
NCD	Null Model	1 vs 2	0.09	0.766
OADS	Null Model	1 vs 2	0.12	0.725

4. Discussion:

This is the first within-study replication of subcortical neurodevelopmental trajectories in males and females using a multisample approach with longitudinal datasets spanning late childhood and adolescence. By implementing mixed-effects modeling with both GAMM and LME in three independent samples, our findings refine and extend knowledge of developmental trajectories of subcortical volumes in males and females. The GAMM models reframe and bring additional clarity in the pattern of development for the thalamus, caudate, putamen, hippocampus, amygdala, and nucleus accumbens, highlighting both similarities and differences in the trajectories in males versus females. Our GAMM results also show that deciphering an accurate pattern of pallidum development may be more challenging. While GAMM predictions and LME best fits were in agreement for males and females per cohort, the best fit model was strikingly different between cohorts within a given sex for most ROIs. As such, the similarities in subcortical trajectories between three independent cohorts using GAMM, but rather striking differences in best-model fits using LME, provides additional 'meta-science' insight on how practices in neurodevelopmental imaging can influence interpretations and conclusions.

LME is one of the most commonly used statistical approaches to determining both between and within-person changes in longitudinal neuroimaging studies (Vijayakumar et al., Accepted). At the outset of the study, it was assumed that using LME might therefore allow for a more direct comparison of our results between the three independent cohorts as well as results from previous studies. However, despite similarities in GAMM trajectories between cohorts in males and females for each subcortical region (except the pallidum), LME results were more variable. This highlights that age-range and statistical power may greatly influence best fits, illustrating a potential limitation of this common approach. Because studies often differ in their age-ranges, scan intervals, and sample size, the conclusion that a particular region shows a linear, quadratic, or even cubic developmental pattern will not necessarily generalize to another study. For example, although all 3 samples of both sexes showed similar patterns using GAMM, the current LME results show both linear and quadratic best fits for caudate volumes in males

and females, and in contrast the putamen showed linear and cubic fits in females, but linear and quadratic for males. Similarly, previous single-sample longitudinal studies have also reported linear decreases in caudate and putamen volumes in both sexes (7-24 years; n=223 scans from 147 individuals; (Wierenga et al., 2014); but also no change for caudate and quadratic for putamen volumes from 5 to 27 years (n=175 scans from 84 individuals; (Narvacan et al., 2017) or quadratic for females and cubic for males from 3 to 26 years (n=829 scans from 387 individuals (Lenroot et al., 2007). This poses a challenge for the field, especially given that we are rarely interested in exact ages, but rather periods of development across the life span. Thus, replication efforts such as the current study and the use of GAMM are not only important, but also provide greater precision in the description of volumetric change by maximizing our ability to reliably detect developmental changes across study populations.

By examining each sex separately, differences in patterns of subcortical maturation were revealed for the amygdala and nucleus accumbens. Subcortical longitudinal studies usually include sex in the LME age model, either as a covariate or as both a main effect and interaction term. However, in the case where the shape of the trajectory may differ between groups, putting both males and females in the same model may incorrectly assume similar shapes in growth in both sexes. For example, the amygdala and nucleus accumbens in the current study showed striking differences in trajectories between males and females. Despite both trajectories being curvilinear, amygdala growth in females decreased during mid-adolescence and into young adulthood, while males continued to show steeper rates of growth across this period. In addition, females displayed decreases in nucleus accumbens volumes whereas males showed no change with age from 8 to 22 years. Previous studies suggest that within-subject changes in puberty (both physical and hormonal) are important factors for amygdala growth in male and female adolescents (Goddings et al., 2014; Herting et al., 2014); with very similar curvilinear amygdala growth patterns seen as reported here when raw volumes were estimated based on Tanner stage in males and females separately (Goddings et al., 2014). Puberty has also been found to relate to nucleus accumbens volumes (Goddings et al., 2014), although the trajectories do not mirror the patterns of volumetric change identified in the current study. Unfortunately,

pubertal metrics were not consistent across the three cohorts; making it impossible for us to investigate how puberty may contribute to differences in amygdala and nucleus accumbens trajectories in males and females in the current study. Thus, future research is warranted to examine the contributions of hormones to sex differences in the neurodevelopmental trajectories of the amygdala and nucleus accumbens.

Of the regions examined, rather discordant trajectories were found for the pallidum in both males and females between the three cohorts, with the OADS showing a unique profile of growth as compared to PIT and NCD. While it is unclear as to the reason behind this discrepancy, reliability estimates for each of the subcortical volumes have been previously examined in the OADS sample and pallidum showed relatively low reliability in this sample (Dennison et al., 2013). Others have also reported the pallidum to be the least reliable out of subcortical volume measurements using FreeSurfer, with volume estimates for this region found to be impacted by MP-RAGE acquisition parameters, such as isotropic versus anisotropic voxel size (Wonderlick et al., 2009). Low reliability of pallidum volumes using automated segmentation software have been attributed to the T1-weighted contrast profile of the pallidum which is less distinct from its surrounding white matter as compared to other subcortical regions such as the thalamus or caudate (Fischl et al., 2002; Wonderlick et al., 2009). Not surprisingly, our findings suggest that region-specific variations exist in the synchronization of developmental trajectories when studying multiple cohorts, with greatest congruency in trajectories of growth seen for the more reliable FreeSurfer subcortical structures.

Overall, our findings indicate a strong consensus across three independent samples when using GAMM, providing much needed clarification of subcortical neurodevelopment during childhood and adolescence. Our findings provide a within-study approach of replication that are in agreement with previous research that sensory, motor, and cognitive related subcortical regions, such as the dorsal striatum (caudate and putamen) and the thalamus, undergo reduction into young adulthood in both males and females. We also show within-study replication that the hippocampus tends to follow a shallow inverted-U shape growth in both

sexes, whereas the amygdala and nucleus accumbens show deviating trajectories between males and females. By replicating these trajectories in typical development, we now have the ability to further focus our inquiry on the factors influencing sex differences and individual differences in subcortical growth; including genetic, and/or environmental effects that may contribute to the observed differences at the group and individual-level. Given the inherent differences in study design and data obtained from the three independent samples, the current study was limited in its ability to explore these potential factors. However, because a number of subcortical abnormalities are implicated in disorders typically associated with child and adolescent onset, such as ADHD (Greven et al., 2015), schizophrenia (van Erp et al., 2016), and mood disorders (Schmaal et al., 2016), expanding on the 'age residual' of within-subject changes in deep gray matter structures will help us further understand risk and resilience for psychopathology during development.

5. Conclusions:

Using three independent samples, mixed-effects modeling with GAMM revealed consistent patterns of volumetric development across adolescence for the thalamus, caudate, putamen, hippocampus, amygdala, and nucleus accumbens, whereas the trajectory for pallidum volume was not reproducible across the samples. The more common linear mixed-effects modeling approach led to conspicuous differences in best fit models between samples for any given region; a cautionary limitation of this method when comparing samples with different age ranges. Using a multisite approach with consistent preprocessing (software and quality checking) and statistical analyses, reproducible patterns in subcortical brain changes can be found across adolescence, with region-specific variations in both sex differences (amygdala and nucleus accumbens) and congruency (pallidum).

Acknowledgments: The authors thank all participating individuals and families in these longitudinal studies.

Funding: This study was supported by the National Institutes of Health [R01 DA018910 (RED), R01 HD053893 and R01 MH087563 (ERS), F32 HD0708084 (MMH), K01 MH1087610 (MMH), R01 MH107418 (NV)]; the Research Council of Norway and the University of Oslo [FRIMEDBIO 230345 to CKT]; the Colonial Foundation, the National Health and Medical Research Council [NHMRC; Australia; Program Grant 350241; Career Development Fellowship 1007716 (SW)], the Australian Research Council [ARC; Discovery Grant DP0878136].

References:

Choudhury, S., Blakemore, S.J., Charman, T., 2006. Social cognitive development during adolescence. Soc Cogn Affect Neurosci 1, 165-174.

Crone, E.A., Elzinga, B.M., 2015. Changing brains: how longitudinal functional magnetic resonance imaging studies can inform us about cognitive and social-affective growth trajectories. Wiley Interdiscip Rev Cogn Sci 6, 53-63.

Dennison, M., Whittle, S., Yucel, M., Vijayakumar, N., Kline, A., Simmons, J., Allen, N.B., 2013. Mapping subcortical brain maturation during adolescence: evidence of hemisphere- and sex-specific longitudinal changes. Dev Sci 16, 772-791.

Ducharme, S., Albaugh, M.D., Nguyen, T.V., Hudziak, J.J., Mateos-Perez, J.M., Labbe, A., Evans, A.C., Karama, S., Brain Development Cooperative, G., 2016. Trajectories of cortical thickness maturation in normal brain development--The importance of quality control procedures. Neuroimage 125, 267-279.

Erus, G., Battapady, H., Satterthwaite, T.D., Hakonarson, H., Gur, R.E., Davatzikos, C., Gur, R.C., 2015. Imaging patterns of brain development and their relationship to cognition. Cereb Cortex 25, 1676-1684.

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, 341-355.

Giedd, J.N., Raznahan, A., Alexander-Bloch, A., Schmitt, E., Gogtay, N., Rapoport, J.L., 2015. Child psychiatry branch of the National Institute of Mental Health longitudinal structural magnetic resonance imaging study of human brain development. Neuropsychopharmacology 40, 43-49.

Giedd, J.N., Vaituzis, A.C., Hamburger, S.D., Lange, N., Rajapakse, J.C., Kaysen, D., Vauss, Y.C., Rapoport, J.L., 1996. Quantitative MRI of the temporal lobe, amygdala, and hippocampus in normal human development: ages 4-18 years. J Comp Neurol 366, 223-230.

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.

Greven, C.U., Bralten, J., Mennes, M., O'Dwyer, L., van Hulzen, K.J., Rommelse, N., Schweren, L.J., Hoekstra, P.J., Hartman, C.A., Heslenfeld, D., Oosterlaan, J., Faraone, S.V., Franke, B., Zwiers, M.P., Arias-Vasquez, A., Buitelaar, J.K., 2015. Developmentally stable whole-brain volume reductions and developmentally sensitive caudate and putamen volume alterations in those with attention-deficit/hyperactivity disorder and their unaffected siblings. JAMA Psychiatry 72, 490-499.

Gur, R.E., Gur, R.C., 2016. Sex differences in brain and behavior in adolescence: Findings from the Philadelphia Neurodevelopmental Cohort. Neurosci Biobehav Rev 70, 159-170.

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. Hum Brain Mapp 35, 5633-5645.

Kessler, R.C., Berglund, P., Demler, O., Jin, R., Merikangas, K.R., Walters, E.E., 2005. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. Arch Gen Psychiatry 62, 593-602.

Kessler, R.C., McGonagle, K.A., Swartz, M., Blazer, D.G., Nelson, C.B., 1993. Sex and depression in the National Comorbidity Survey. I: Lifetime prevalence, chronicity and recurrence. J Affect Disord 29, 85-96.

Koolschijn, P.C., Crone, E.A., 2013. Sex differences and structural brain maturation from childhood to early adulthood. Dev Cogn Neurosci 5, 106-118.

Lenroot, R.K., Gogtay, N., Greenstein, D.K., Wells, E.M., Wallace, G.L., Clasen, L.S., Blumenthal, J.D., Lerch, J., Zijdenbos, A.P., Evans, A.C., Thompson, P.M., Giedd, J.N., 2007. Sexual dimorphism of brain developmental trajectories during childhood and adolescence. Neuroimage 36, 1065-1073.

Miller, G.A., Chapman, J.P., 2001. Misunderstanding analysis of covariance. J Abnorm Psychol 110, 40-48.

Mills, K.L., Goddings, A.L., Herting, M.M., Meuwese, R., Blakemore, S.J., Crone, E.A., Dahl, R.E., Guroglu, B., Raznahan, A., Sowell, E.R., Tamnes, C.K., 2016. Structural brain development between childhood and adulthood: Convergence across four longitudinal samples. Neuroimage 141, 273-281.

Morey, R.A., Selgrade, E.S., Wagner, H.R., 2nd, Huettel, S.A., Wang, L., McCarthy, G., 2010. Scan-rescan reliability of subcortical brain volumes derived from automated segmentation. Hum Brain Mapp 31, 1751-1762.

Narvacan, K., Treit, S., Camicioli, R., Martin, W., Beaulieu, C., 2017. Evolution of deep gray matter volume across the human lifespan. Hum Brain Mapp 38, 3771-3790.

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. J Neurosci 29, 11772-11782.

Raznahan, A., Shaw, P.W., Lerch, J.P., Clasen, L.S., Greenstein, D., Berman, R., Pipitone, J., Chakravarty, M.M., Giedd, J.N., 2014. Longitudinal four-dimensional mapping of subcortical anatomy in human development. Proc Natl Acad Sci U S A 111, 1592-1597.

Reuter, M., Schmansky, N.J., Rosas, H.D., Fischl, B., 2012. Within-subject template estimation for unbiased longitudinal image analysis. Neuroimage 61, 1402-1418.

Roalf, D.R., Gur, R.E., Ruparel, K., Calkins, M.E., Satterthwaite, T.D., Bilker, W.B., Hakonarson, H., Harris, L.J., Gur, R.C., 2014. Within-individual variability in neurocognitive performance: age- and sex-related differences in children and youths from ages 8 to 21. Neuropsychology 28, 506-518.

Rose, A.J., Rudolph, K.D., 2006. A review of sex differences in peer relationship processes: potential trade-offs for the emotional and behavioral development of girls and boys. Psychol Bull 132, 98-131.

Ruigrok, A.N., Salimi-Khorshidi, G., Lai, M.C., Baron-Cohen, S., Lombardo, M.V., Tait, R.J., Suckling, J., 2014. A meta-analysis of sex differences in human brain structure. Neurosci Biobehav Rev 39, 34-50.

Sanfilipo, M.P., Benedict, R.H., Zivadinov, R., Bakshi, R., 2004. Correction for intracranial volume in analysis of whole brain atrophy in multiple sclerosis: the proportion vs. residual method. Neuroimage 22, 1732-1743.

Schmaal, L., Veltman, D.J., van Erp, T.G., Samann, P.G., Frodl, T., Jahanshad, N., Loehrer, E., Tiemeier, H., Hofman, A., Niessen, W.J., Vernooij, M.W., Ikram, M.A., Wittfeld, K., Grabe, H.J., Block, A., Hegenscheid, K., Volzke, H., Hoehn, D., Czisch, M., Lagopoulos, J., Hatton, S.N., Hickie, I.B., Goya-Maldonado, R., Kramer, B., Gruber, O., Couvy-Duchesne, B., Renteria, M.E., Strike, L.T., Mills, N.T., de Zubicaray, G.I., McMahon, K.L., Medland, S.E., Martin, N.G., Gillespie, N.A., Wright, M.J., Hall, G.B., MacQueen, G.M., Frey, E.M., Carballedo, A., van Velzen, L.S., van Tol, M.J., van der Wee, N.J., Veer, I.M., Walter, H., Schnell, K., Schramm, E., Normann, C., Schoepf, D., Konrad, C., Zurowski, B., Nickson, T., McIntosh, A.M., Papmeyer, M., Whalley, H.C., Sussmann, J.E., Godlewska, B.R., Cowen, P.J., Fischer, F.H., Rose, M., Penninx, B.W., Thompson, P.M., Hibar, D.P., 2016. Subcortical brain alterations in major depressive disorder: findings from the ENIGMA Major Depressive Disorder working group. Mol Psychiatry 21, 806-812.

Sowell, E.R., Thompson, P.M., Toga, A.W., 2004. Mapping changes in the human cortex throughout the span of life. Neuroscientist 10, 372-392.

Sowell, E.R., Trauner, D.A., Gamst, A., Jernigan, T.L., 2002. Development of cortical and subcortical brain structures in childhood and adolescence: a structural MRI study. Developmental Medicine and Child Neurology 44, 4-16.

Tamnes, C.K., Herting, M.M., Goddings, A.L., Meuwese, R., Blakemore, S.J., Dahl, R.E., Guroglu, B., Raznahan, A., Sowell, E.R., Crone, E.A., Mills, K.L., 2017. Development of the Cerebral Cortex across Adolescence: A Multisample Study of Inter-Related Longitudinal Changes in Cortical Volume, Surface Area, and Thickness. J Neurosci 37, 3402-3412.

Tamnes, C.K., Walhovd, K.B., Dale, A.M., Ostby, Y., Grydeland, H., Richardson, G., Westlye, L.T., Roddey, J.C., Hagler, D.J., Jr., Due-Tonnessen, P., Holland, D., Fjell, A.M., Alzheimer's

Disease Neuroimaging, I., 2013. Brain development and aging: overlapping and unique patterns of change. Neuroimage 68, 63-74.

van Erp, T.G., Hibar, D.P., Rasmussen, J.M., Glahn, D.C., Pearlson, G.D., Andreassen, O.A., Agartz, I., Westlye, L.T., Haukvik, U.K., Dale, A.M., Melle, I., Hartberg, C.B., Gruber, O., Kraemer, B., Zilles, D., Donohoe, G., Kelly, S., McDonald, C., Morris, D.W., Cannon, D.M., Corvin, A., Machielsen, M.W., Koenders, L., de Haan, L., Veltman, D.J., Satterthwaite, T.D., Wolf, D.H., Gur, R.C., Gur, R.E., Potkin, S.G., Mathalon, D.H., Mueller, B.A., Preda, A., Macciardi, F., Ehrlich, S., Walton, E., Hass, J., Calhoun, V.D., Bockholt, H.J., Sponheim, S.R., Shoemaker, J.M., van Haren, N.E., Hulshoff Pol, H.E., Ophoff, R.A., Kahn, R.S., Roiz-Santianez, R., Crespo-Facorro, B., Wang, L., Alpert, K.I., Jonsson, E.G., Dimitrova, R., Bois, C., Whalley, H.C., McIntosh, A.M., Lawrie, S.M., Hashimoto, R., Thompson, P.M., Turner, J.A., 2016. Subcortical brain volume abnormalities in 2028 individuals with schizophrenia and 2540 healthy controls via the ENIGMA consortium. Mol Psychiatry 21, 547-553.

Vijayakumar, N., Mills, K.L., Alexander-Bloch, A., Tamnes, C.K., Whittle, S., Accepted. Structural brain development: a review of methodological approaches and best practices.

Wierenga, L., Langen, M., Ambrosino, S., van Dijk, S., Oranje, B., Durston, S., 2014. Typical development of basal ganglia, hippocampus, amygdala and cerebellum from age 7 to 24. Neuroimage 96, 67-72.

Wierenga, L.M., Sexton, J.A., Laake, P., Giedd, J.N., Tamnes, C.K., the Pediatric Imaging, N., and Genetics Study, 2017. A Key Characteristic of Sex Differences in the Developing Brain: Greater Variability in Brain Structure of Boys than Girls. Cerebral Cortex 1.

Wonderlick, J.S., Ziegler, D.A., Hosseini-Varnamkhasti, P., Locascio, J.J., Bakkour, A., van der Kouwe, A., Triantafyllou, C., Corkin, S., Dickerson, B.C., 2009. Reliability of MRI-derived cortical and subcortical morphometric measures: effects of pulse sequence, voxel geometry, and parallel imaging. Neuroimage 44, 1324-1333.

Yap, M.B., Allen, N.B., O'Shea, M., di Parsia, P., Simmons, J.G., Sheeber, L., 2011. Early adolescents' temperament, emotion regulation during mother-child interactions, and depressive symptomatology. Dev Psychopathol 23, 267-282.