Adolescent to young adult longitudinal development of subcortical volumes in two European sites with four waves

Lea L. Backhausen [1,2], Juliane H. Fröhner[1], Hervé Lemaître*[3, 4], Eric Artiges*[5,6], Marie-Laure Palillère Martinot*[5,7], Megan M. Herting[8], Fabio Sticca[9], Tobias Banaschewski[10], Gareth J. Barker[11], Arun L.W. Bokde[12], Sylvane Desrivières[13], Herta Flor[14,15], Antoine Grigis[3], Hugh Garavan[16], Penny Gowland[17], Andreas Heinz[18], Rüdiger Brühl[19], Frauke Nees[10,14,20], Dimitri Papadopoulos-Orfanos[3], Luise Poustka[21], Sarah Hohmann[10], Lauren Robinson[22], Henrik Walter[18], Jeanne Winterer[18,23], Robert Whelan[24], Gunter Schumann[13,25], Jean-Luc Martinot#[5], Michael N. Smolka#[1], Nora C. Vetter[1,2], and the IMAGEN Consortium

[1] Department of Psychiatry and Neuroimaging Center, Technische Universität Dresden, Germany;

[2] Department of Child and Adolescent Psychiatry, Faculty of Medicine of the Technische Universität Dresden, Germany;

[3] NeuroSpin, CEA, Université Paris-Saclay, Gif-sur-Yvette, France ;

[4] Institut des Maladies Neurodégénératives, UMR 5293, CNRS, CEA, Université de Bordeaux, Bordeaux, France;

[5] Institut National de la Santé et de la Recherche Médicale, INSERM U1299 "Trajectoires développementales en psychiatrie"; Université Paris-Saclay, Ecole Normale supérieure Paris-Saclay, CNRS, Centre Borelli; Gif-sur-Yvette, France;

[6] Psychiatry Department, EPS Barthélémy Durand, Etampes, France;

[7] Department of Child and Adolescent Psychiatry, Pitié-Salpêtrière Hospital, Paris, France;

[8] Department of Preventive Medicine, University of Southern California, Los Angeles, CA 90089, USA;

[9] Marie Meierhofer Children's Institute, Associated Institute of the University of Zurich, Zurich, Switzerland;

[10] Department of Child and Adolescent Psychiatry and Psychotherapy, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Square J5, Mannheim, Germany;

[11] Department of Neuroimaging, Institute of Psychiatry, Psychology & Neuroscience, King's College London, United Kingdom;

[12] Discipline of Psychiatry, School of Medicine and Trinity College Institute of Neuroscience, Trinity College Dublin, Dublin, Ireland;

[13] Centre for Population Neuroscience and Precision Medicine (PONS), Institute of Psychiatry, Psychology & Neuroscience, SGDP Centre, King's College London, United Kingdom;

[14] Institute of Cognitive and Clinical Neuroscience, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Square J5, Mannheim, Germany;

[15] Department of Psychology, School of Social Sciences, University of Mannheim, Mannheim, Germany;

[16] Departments of Psychiatry and Psychology, University of Vermont, Burlington, Vermont, USA;

[17] Sir Peter Mansfield Imaging Centre School of Physics and Astronomy, University of Nottingham, University Park, Nottingham, United Kingdom;

[18] Department of Psychiatry and Psychotherapy CCM, Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Berlin, Germany;

[19] Physikalisch-Technische Bundesanstalt (PTB), Braunschweig and Berlin, Germany;

[20] Institute of Medical Psychology and Medical Sociology, University Medical Center Schleswig Holstein, Kiel University, Kiel, Germany;

[21] Department of Child and Adolescent Psychiatry and Psychotherapy, University Medical Centre Göttingen, Göttingen, Germany;

[22] Department of Psychological Medicine, Section for Eating Disorders, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, SE5 8AF, UK;

[23] Department of Education and Psychology, Freie Universität Berlin, Berlin, Germany.

[24] School of Psychology and Global Brain Health Institute, Trinity College Dublin, Ireland;

[25] PONS Research Group, Dept of Psychiatry and Psychotherapy, Campus Charite Mitte, Humboldt University, Berlin and Leibniz Institute for Neurobiology, Magdeburg, Germany and Institute for Science and Technology of Brain-inspired Intelligence (ISTBI), Fudan University, Shanghai, P.R. China

*# Authors contributed equally

Correspondence to

Lea L. Backhausen or Nora C. Vetter
Universitätsklinikum Carl Gustav Carus
Technische Universität Dresden
Fetscherstraße 74
01307 Dresden, Germany
Email: Lea.backhausen@tu-dresden.de or Nora.vetter@tu-dresden.de
Tel. +49 (0)351 458-7166
Fax +49 (0)351 458-5754

Abstract

Adolescent subcortical structural brain development might underlie psychopathological symptoms, which often emerge in adolescence. At the same time, sex differences exist in psychopathology, which might be mirrored in underlying sex differences in structural development. However, previous studies showed inconsistencies in subcortical trajectories and potential sex differences. Therefore, we aimed to investigate the subcortical structural trajectories and their sex differences across adolescence using for the first time a single cohort design, the same quality control procedure, software and a general additive mixed modeling approach. We investigated two large European sites from ages 14 to 24 with 503 participants and 1408 total scans from France and Germany as part of the IMAGEN project including four waves of data acquisition. We found significantly larger volumes in males versus females in both sites and across all seven subcortical regions. Sex differences in age-related trajectories were observed across all regions in both sites. Our findings provide further evidence of sex differences in longitudinal adolescent brain development of subcortical regions and thus might eventually support the relationship of underlying brain development and different adolescent psychopathology in boys and girls.

6 Keywords: Neurodevelopment, Subcortical, Adolescence, Sex differences, Longitudinal, Structural magnetic resonance imaging (sMRI)

1 Introduction

Adolescence is a core incidence phase for the development of mental disorders (Paus et al., 2008). Mental disorders occur in around 20% of adolescents with important sex differences, i.e. higher rates of internalizing disorders in girls and externalizing disorders in boys (Kessler et al., 2005; Zahn-Waxler et al., 2008). These sex differences might be mirrored in underlying sex differences in subcortical structural development (Gogtay & Thompson, 2010; Paus et al., 2008; Shaw et al., 2010) since alterations in subcortical structures have been associated with psychopathological disorders (Hoogman et al., 2017; Noordermeer et al., 2017; Rogers & Brito, 2016). Hence, there is a need to characterize typical subcortical structural development and the related sex differences across adolescence to better understand adolescent psychopathology.

Cross-sectional (Boedhoe et al., 2020; Brain Development Cooperative Group, 2012; Østby et al., 2009) and longitudinal studies (Dennison et al., 2013; Goddings et al., 2014; Herting et al., 2018; Narvacan et al., 2017; Tamnes et al., 2013; Wierenga et al., 2014) began to investigate the typical development of subcortical structures across adolescence with some studies also analyzing sex effects (Goddings et al., 2014; Herting et al., 2018; Narvacan et al., 2017; Wierenga, Bos, et al., 2018; Wierenga et al., 2014). Differential trajectories across adolescence have been revealed for some subcortical structures. Importantly, we review here the trajectories from the age span 14 to 24, which will also be covered by our study.

The basal ganglia (mainly caudate nucleus, putamen, globus pallidus, and nucleus accumbens) showed volume decreases across adolescence in cross-sectional studies (Brain Development Cooperative Group, 2012; Østby et al., 2009), including a life span mega-analysis (Dima et al., 2021). A longitudinal study with 85 participants from 8-22 years and two waves (170 scans) corroborated these findings (Tamnes et al., 2013). Other longitudinal studies showed similar decreases across adolescence for some basal ganglia structures (Dennison et al., 2013: 120 scans, two waves, n=60, 11-17 years; Goddings et al., 2014: 711 scans, ≥ two waves, n= 275, 7-19 years; Narvacan et al., 2017: 168 scans, two waves, n=84, 5-27 years; Wierenga et al., 2014: 223 scans, up to three waves, n=147, 7–24 years; Raznahan et al., 2014: decrease for the globus pallidus only: 1172 scans, n=618, 5–25 years, about half having ≥ two waves; Wierenga et al., 2018: decrease for the accumbens only: 680 scans, n= 271, 8–28 years, up to three scans).

However, diverging findings from the overall decreasing pattern exist. The caudate nucleus remained stable across adolescence in three studies: for both sexes (Narvacan et al., 2017; Wierenga, Bos, et al., 2018) or for males only (Raznahan et al., 2014). For the globus pallidus three studies found volume increase (Dennison et al., 2013; Wierenga, Bos, et al., 2018; Wierenga et al., 2014) with Wierenga et al. (2014) showing a peak at age 17 with a slight decrease thereafter. A multi-site study found a descriptive, albeit non-significant increase (Herting et al., 2018: three sites, n=216 participants and 467 scans, 10-22 years, two to three waves). For the nucleus accumbens a stable trajectory was found for both sexes (Dennison et al., 2013) or for males only (Herting et al., 2018). For the putamen a non-linear increase was found (Wierenga, Bos, et al., 2018).

For the thalamus, previous results converged in indicating volume decreases from about age 13 (Tamnes et al., 2013) or age 15 to 19 on (Herting et al., 2018 for males only; Raznahan et al., 2014; Wierenga, Bos, et al., 2018) while Narvacan et al. (2017) only showed decreases in their cross-sectional cohort and no changes in their longitudinal cohort.

Diverging findings also exist regarding the trajectories of the amygdala and hippocampus across adolescence (Koolschijn & Crone, 2013; Østby et al., 2009; Sowell et al., 2002; Wierenga et al., 2014). While some cross-sectional (Durston et al., 2001; Giedd et al., 1996) and longitudinal studies (Dennison et al., 2013; Herting et al., 2018) found an increase of the amygdala or hippocampus during adolescence, other longitudinal studies showed decreases (Tamnes et al., 2013) or little to no change (Dennison et al., 2013; Wierenga, Bos, et al., 2018). For both the amygdala and hippocampus a slight decrease for girls at about 15 years versus an increase for boys has been shown in another study (Goddings et al., 2014). Wierenga et al. (2014) reported an increase for both the amygdala and hippocampus with a peak at about age 17/ 18 followed by a slight decrease.

1.1 Sex differences

Regarding sex differences, cross-sectional and longitudinal studies consistently demonstrate larger subcortical volumes in boys versus girls across adolescence when not correcting for whole brain size (Dennison et al., 2013; Goddings et al., 2014; Herting et al., 2018; Narvacan et al., 2017; Raznahan et al.,

2014; Tamnes et al., 2013; Wierenga, Bos, et al., 2018; Wierenga et al., 2014) ranging from 9% for the putamen to 15% for the amygdala (Narvacan et al., 2017).

There has also been note of sex differences in age-related trajectories of subcortical structures; however, previous studies are rather inconsistent. Overall differences in trajectories between sexes were found in some studies (Dennison et al., 2013; Goddings et al., 2014; Herting et al., 2018) and others reported later peak volumes in males (Lenroot et al., 2007; Raznahan et al., 2014). The thalamus declined for females (Dennison et al., 2013) or males only (Herting et al., 2018) when tested separately for the sexes. A slight decrease for the amygdala and hippocampus was found in girls versus a continued increase for boys (Goddings et al., 2014). Significantly differing nonlinear trajectories across sexes were also found for the amygdala and hippocampus (Herting et al., 2018). Further, no sex differences in trajectories have also been reported (Narvacan et al., 2017; Wierenga, Bos, et al., 2018; Wierenga et al., 2014).

Importantly, in most studies the sex differences in trajectories can only be inferred indirectly when descriptively comparing shapes of trajectories. Only very few studies tested for statistical differences in trajectories between the sexes (Herting et al., 2018; Wierenga, Bos, et al., 2018). Most of the studies further had no or a very low amount of participants with more than two waves (except Goddings et al., 2014; Raznahan et al., 2014; Wierenga, Sexton, et al., 2018), limiting their conclusions on trajectories and their ability to model more complex trajectories. Moreover, some studies only covered few subcortical regions (four or five: Goddings et al., 2014; Raznahan et al., 2014), which limits inferences about trajectories of all subcortical structures.

Taken together, previous studies showed inconsistent results in subcortical age trajectories and their potential sex differences. This might be related to differences in sample characteristics, study design or analysis regarding sampling type, age and sex, sociodemography, inter-scan interval, number of total scans, image acquisition, quality control, segmentation software, and analysis models. Regarding sampling type, most studies used accelerated longitudinal designs (Herting et al., 2018; Narvacan et al., 2017; Raznahan et al., 2014; Wierenga, Bos, et al., 2018). These designs are limited because age effects are confounded with interactions between a given starting age and when the measurement happened and participants only

contribute data to parts of the time span investigated. Some studies included differing sex ratios across ages (Herting et al., 2018) or did not report detailed sociodemographic sample description like IQ or socioeconomic status (Herting et al., 2018; Tamnes 2013) limiting understanding of the generalizability of findings. Image acquisition differed in field strengths or scanner types with most studies using 1.5 Tesla (except Wierenga et al. (2018) using 3 Tesla) and some studies using different scanners within studies (Dennison et al., 2013, using two different 3 Tesla scanners and Herting et al. (2018) with two sub-sites using two different 3 Tesla scanners; see Liu et al., 2020 for the influence of scanner type differences). Differences in quality control procedures may have also impacted the results (Backhausen et al., 2016; Ducharme et al., 2016). The type of segmentation software can also influence volume estimates (Makowski et al., 2018; Morey et al., 2010), even across different versions of the same software (e.g. differences in putamen using FreeSurfer 5.3 versus 6.0, see release notes). Only in some analysis software packages a longitudinal processing approach is included (see Vijayakumar et al., 2018 for a software overview) and thus has not always been used. Previous studies further employed different analysis models, e.g. linear mixed effects modelling (Goddings et al., 2014; Narvacan et al., 2017) or hierarchical regression (Dennison et al. 2013). These different analysis approaches may impact results (Vijayakumar et al., 2018). Parametric models use the a priori assumption of either linear, quadratic, or cubic age-related change and assume similar patterns of age trajectories between groups. In contrast, generalized additive mixed models (GAMMs) are more flexible since they fit curves using 'smooth function' terms and thus have recently become more widely applied (e.g. Herting et al., 2018; Wierenga, Bos, et al., 2018; Wierenga et al., 2014). GAMMs describe the best relationship between predictor and outcome variables of interest without a priori knowledge of the inherent form of the data. Therefore, GAMMs allow for different nonlinear shapes of age trajectories for categories of a variable such as sex with the categories males and females.

Taken together, the methodical differences between studies still limit our ability to draw conclusions about the subcortical structural trajectories in the mid-adolescent to young adult age span, specifically due to the lower amount of studies with more than two waves including a meaningful amount of participants in this age span. Hence, there is still the need to test for possible sex differences in subcortical age trajectories across adolescence. Therefore, we aimed to investigate the subcortical structural trajectories

including the thalamus, globus pallidus, caudate nucleus, putamen, nucleus accumbens, hippocampus, and amygdala as regions of interest (ROIs) and their sex differences across adolescence using the same quality control procedure, segmentation software, i.e. FreeSurfer 6.0.0, and a GAMM approach in a unique study of adolescents from two European sites with large samples. Importantly, differing from most previous accelerated studies, our design was a single cohort approach across two sites where all participants started at the same age and were followed across the entire age-range of interest. Specifically, as part of the Imaging Genetics (IMAGEN) project (Schumann et al., 2010) data stemmed from a site in Paris, France (N=255) and a site in Dresden, Germany (N=248) including four waves of data acquisition at the ages of M=14.47 years (range 13.33 – 15.72), M=16.48 years (range 15.65 – 17.88), M=19.24 years (range 17.26 – 22.54) and M=22.35 years (range 20.1 – 24.76) across sites, with a total of 1408 scans. The sites were largely homogeneous in terms of age range, sex distribution, inter-scan interval, number of total scans and sociodemographic characteristics such as socioeconomic status and ethnicity. In examining two similar sites, we can extend previous research that showed sex differences in trajectories dependent on specific sites (Herting et al., 2018) and thus explore age-constant and time-varying site differences closely focused on the mid-adolescent to young adult age span with more waves and scans within this particular age-range enabling more precise estimation of trajectories.

2 Methods

Access to the IMAGEN dataset is available with an accepted proposal from the IMAGEN executive committee (https://imagen-europe.com/resources/imagen-project-proposal/). All code is posted to Open Science Framework and can be assessed at https://osf.io/5tfh4/.

2.1 Participants and study design

The study is part of the larger European IMAGEN project assessing magnetic resonance imaging (MRI) data from typically developing adolescents at the age of 14, 18, and 22. For a detailed description of recruitment and assessment procedures in the IMAGEN study, please refer to (Schumann et al., 2010). At the sites Dresden and Paris MRI data was assessed additionally at age 16, resulting in four waves (see Table

1 for sample characteristics and Figure 1 for age and sex distributions per site), therefore these two sites were chosen. Participants provided written informed consent and assent (in case of participants <18 years also their legal guardians). The study had been approved by the local ethics committees (Technische Universität Dresden and University of Paris) and was performed in accordance with the Declaration of Helsinki. Exclusion criteria included existing bipolar disorder, schizophrenia and major neuro-developmental disorders such as autism, as well as a premature birth, head trauma and history of several neurological or medical disorders. In this manuscript, female and male is defined as "assigned female at birth" and "assigned male at birth", respectively. For Dresden, 260 participants took part in the overall IMAGEN study of whom 256 were scanned for at least one wave and 248 participants were included in the analyses after quality control. For Paris, 274 participants took part in the study of whom 258 were scanned for at least one wave and 255 participants were included in the analyses after quality control (see below and supplementary Table S1 for details). In total, the present study included 503 participants (250 females) and 1408 scans covering the age range of 13.33–24.76 years, including 447 scans at wave 1 (Dresden: 201, Paris: 246), 336 scans at wave 2 (Dresden: 207, Paris: 129), 372 scans at wave 3 (Dresden: 172, Paris: 200), and 253 scans at wave 4 (Dresden: 130, Paris: 123). Of the 503 participants included in the study 94 % were White (97% Dresden, 90% Paris) and the majority stemmed from rather well-educated households (as a proxy for socioeconomic status), with around 60% of the parents having obtained a university or college (university of applied sciences) degree. For more information about the parental education assessment in the IMAGEN study see Table 1 in Schumann et al. (2010).

Table 1 Sample characteristics.

Sample characteristics including demographics of typically developing adolescents from two sites of the IMAGEN study using single cohort designs.

|	Dresden, Germany			Paris, France		
	All	Female	Male	All	Female	Male
Total scans across all waves	710	352	358	698	366	332
Age (years)						
Wave 1, M (SD)	14.54 (.32)	14.58 (.32)	14.49 (.33)	14.42 (.5)	14.43 (.5)	14.41 (.5)
Wave 2, M (SD)	16.6 (.42)	16.64 (.4)	16.56 (.42)	16.80 (.62)	16.84 (.59)	16.76 (.64)
Wave 3, M (SD)	18.68 (.57)	18.72 (.56)	18.64 (.58)	19.73 (.73)	19.76 (.74)	19.69 (.73)
Wave 4, M (SD)	22.1 (.74)	22.04 (.62)	22.12 (.85)	22.64 (.55)	22.62 (.52)	22.67 (.59)
N						
Total	248	121	127	255	129	126
Participants with 1 scan	35	14	21	41	17	24
Participants with 2 scans	47	26	21	51	25	28
Participants with 3 scans	83	38	45	95	49	44
Participants with 4 scans	83	43	40	68	38	30
Demographics - *Data from the first wave is reported**						
Pubertal status[a], M(SD)	3.65 (.66)	4.03 (.41)	3.28 (.64)	3.55 (.78)	3.98 (.42)	3.1 (.82)
Height in cm, M(SD)	169.42 (7.21)	166.55 (6.45)	172.13 (6.85)	165.57 (8.03)	162.67 (5.54)	168.6 (9.07)
Weight in kg, M(SD)	59.17 (11.74)	56.82 (10.23)	61.39 (12.64)	55.3 (9.26)	53.61 (7.9)	57.05 (10.25)
BMI, M(SD)	20.53 (3.29)	20.42 (3.03)	20.62 (3.51)	20.11 (2.61)	20.23 (2.6)	19.98 (2.65)
IQ[b], M(SD)	112.95 (10.72)	113.05 (9.93)	112.86 (11.48)	109.35 (9.43)	108.5 (8.95)	110.25 (9.87)

*Note that data for pubertal status, height, weight, BMI and IQ is reported for all participant in the final data set independent of a high-quality scan at the first wave. BMI = Body mass index (BMI = weight in kg/ height in m^2)**.
[a] Pubertal status ranged from 1 for prepubertal to 5 for postpubertal status, measured with the Pubertal Development Scale (PDS; Petersen et al., 1988)
[b] General cognitive ability estimated with the subtests Similarities, Block Design, Vocabulary, and Matrices from the Wechsler Intelligence Scale For Children (WISC-IV, German adaptation; Petermann & Petermann, 2007)
* $p < .05$, ** $p < .01$, *** $p < .001$.

Figure 1: Age and sex distributions for each site.
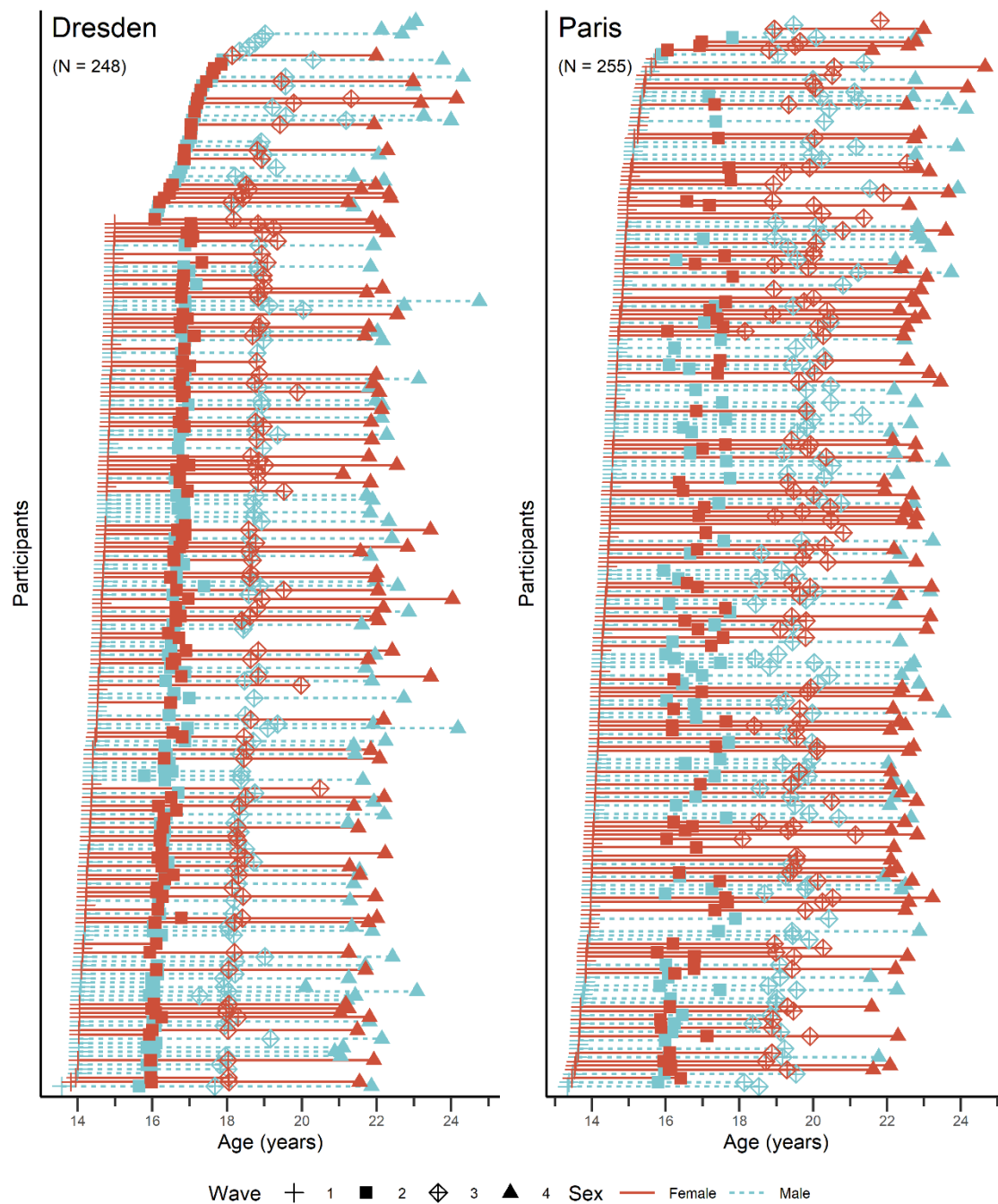
2.2 Structural imaging

2.2.1 Image acquisition

We acquired T1-weighted anatomical scans using the respective local Siemens (Erlangen, Germany) 3T whole-body MR tomograph equipped with a 12-channel head coil in Dresden and a 64-channel head coil in Paris. While in Dresden this was a MAGNETOM Trio for all waves, the MAGNETOM

Trio was updated to a MAGNETOM Prisma during wave 3. For both sites a standardized quality control assessment for the scanner was regularly performed (phantom scan and in-vivo scan, see Schumann et al. (2010)). High-resolution T1-weighted images were collected using a magnetization prepared rapid acquisition gradient-echo (MPRAGE) sequence [Dresden/ Paris: repetition time (TR) = 1900/ 2300 ms, echo time (TE) = 2.26/ 2.93 ms, inversion time (TI) = 900/ 900 ms, voxel size = 0.5 x 0.5 x 1.0/ 1.1 x 1.1 x 1.1 mm, flip angle = 9/ 9°; matrix size = 256 x 256/ 256 x 256 mm; 176/ 160 slices]. The sequence in Dresden varied slightly. This is due to differences in imaging data assessment waves at the each site of the IMAGEN project. Only the sites Dresden and Paris assessed imaging data at age 16 with Dresden using the local MPRAGE sequence only. All images were examined by a clinical neuroradiologist for structural abnormalities.

2.2.2 Image processing

Images (also single time points) were processed using the longitudinal pipeline of FreeSurfer 6.0.0 (http://surfer.nmr.mgh.harvard.edu; Fischl et al., 2002; Reuter et al., 2012). This pipeline creates an unbiased within-subject template space and image using robust, inverse consistent registration (Reuter et al., 2010). Several processing steps, such as skull stripping, Talairach transforms, atlas registration, as well as spherical surface maps and parcellations are then initialized with common information from the within-subject template. This process significantly improves robustness with respect to noise, intensity scaling and outliers, significantly increasing statistical power (Reuter et al., 2012) and scan-rescan reliability as compared to conventional cross-sectional analysis pipelines (Jovicich et al., 2013). We further used the recon-all-flag -3T which alters FreeSurfer's internal N3 bias field correction parameters (Sled et al., 1998) making it more appropriate for 3T MRI (Zheng et al., 2009).

As previous studies demonstrated no evidence of developmental difference between hemispheres in our subcortical ROIs (Brain Development Cooperative Group, 2012; Østby et al., 2009), we report averaged volumes over both hemispheres similarly to previous studies (Herting et al., 2018; Wierenga, Bos, et al., 2018). Correlations between hemispheres across sites at each wave were high for each ROI (r's > .7) except for the nucleus accumbens (r's = .46 to .79) and the amygdala (r's = .6 to .79; see supplementary Table S2 for all correlations). Following the methodology of previous studies (e.g. Herting et al., 2018;

Wierenga, Bos, et al., 2018) we used absolute volume values and did not correct ROI volumes for intra-cranial volume.

2.2.3 Quality control

Before statistical analyses, trained operators performed the following pre- and post-processing quality control: For pre-processing quality control all raw images were visually inspected for motion and technical artifacts according to our lab's three-category (fail, check, pass) rating scheme (Backhausen et al., 2016). We retained "check" and "pass" images and excluded "fail" images from longitudinal pipeline processing to ensure high quality of the within-subject template. In post-processing quality control trained operators inspected longitudinally processed images with questionable raw image quality (rated "check") for accuracy of subcortical segmentation. Subcortical structures were excluded when incorrectly segmented (see supplementary Table S1). No manual adjustments of subcortical volumes were made.

2.2.4 Statistical analyses

For analysis of GAMMs we used R version 3.6.1 with the mgcv package version 1.8-33 (Wood, 2004, 2006) and the packages itsadug version 2.3 and ggplot2 version 3.3.2 for visualization of smooth curves. We used a shrinkage version of penalized cubic regression splines as smooth terms in all models. The k parameters, which specify the number of basis functions to build the curves and thus influence potential over-fitting, were set to 4 after k.check (mgcv package) analyses.

We closely followed the approach of Herting et al. (2018) and built GAMM models to examine age-constant and time-varying sex and site differences in trajectories of subcortical volumes. We explored differences in trajectories by examining significance of difference smooths between the trajectories of males versus females in the Dresden and Paris site. GAMM models were implemented to (1) test an age-constant sex difference (main effect of sex) as well as a sex difference in the age trajectory (age*sex interaction), while controlling for site at the level of main effect and site*interaction and (2) test an age-constant site difference (main effect of site) as well as a site difference in the age trajectory (age*site interaction), while controlling for sex at the level of main effect and sex*age interaction. All models included an individual-level random effect intercept per participant. Sex and site were coded as factors (male = 0, female = 1; Dresden = 0, Paris = 1). To directly test sex and site differences, previous GAMM

models were updated in coding sex/ site and sex*age /site*age as contrasting ordered factors. For GAMM

estimates of these smooth terms see Table 2.

Table 2. GAMMS examining sex and age effects across both sites on subcortical regions of

interest.

THALAMUS

Main effect	β	SE	t	p
Sex (F vs M)	-734.91	53.75	-13.67	***
Trajectory	edf	Ref.df	F	p
S(age)	2.85	3	24.44	***
S(age, F vs M)	2.29	3	11.44	***

GLOBUS PALLIDUS

Main effect	β	SE	t	p
Sex (F vs M)	-184.97	17.39	-10.64	***
Trajectory	edf	Ref.df	F	p
S(age)	0.81	3	0.45	**
S(age, F vs M)	2.46	3	56.84	***

CAUDATE NUCLEUS

Main effect	β	SE	t	p
Sex (F vs M)	-302.60	40.35	-7.50	***
Trajectory	edf	Ref.df	F	p
S(age)	1.88	3	1.78	***
S(age, F vs M)	2.62	3	40.02	***

PUTAMEN

Main effect	β	SE	t	p
Sex (F vs M)	-595.96	45.89	-12.99	***
Trajectory	edf	Ref.df	F	p
S(age)	1.76	3	1.53	***
S(age, F vs M)	2.53	3	20.27	***

NUCLEUS ACCUMBENS

Main effect	β	SE	t	p
Sex (F vs M)	-63.17	9.0	-7.02	***
Trajectory	edf	Ref.df	F	p
S(age)	0.00	3	0.00	0.62
S(age, F vs M)	1.80	3	11.52	***

HIPPOCAMPUS

Main effect	β	SE	t	p
Sex (F vs M)	-360.67	30.39	-11.87	***
Trajectory	edf	Ref.df	F	p
S(age)	2.31	3	1.88	0.55
S(age, F vs M)	2.33	3	19.17	***

AMYGDALA

Main effect	β	SE	t	p
Sex (F vs M)	-238.48	15.28	-15.61	***
Trajectory	edf	Ref.df	F	p
S(age)	0.00	3	0.00	0.76
S(age, F vs M)	2.27	3	23.22	***

GAMM estimates for age, sex, and age*sex across both sites. Sex was coded as a factor (male = 0, female =1), allowing for each term to reflect the following: Sex (F vs M) reflected the main difference in females as compared to males; S(age) reflected the global smooth trajectory of age; S(age, F vs M) reflected the difference in trajectory of females compared males. Site and site*age were included as covariates in all models. Estimate, standard error (SE), t-value and associated p-values for each main difference. Smooth function (S(age) / edf), degrees of freedom (Ref.df), F-statistic and associated p-values for each smooth term. * $p < .05$, ** $p < .01$, *** $p < .001$.

To test whether the inclusion of age*sex interaction terms significantly improve model fit and explain variance, we analyzed three models following procedures by Pedersen et al. (2019) and Wierenga et al. (2018; 2019) for each subcortical structure as a smooth function of age of the individual i at wave j, a random intercept per participant effect (u_i) and error ($error_{ij}$), while controlling for site (in main effect and site*age interaction) in each model. We compared these models using Akaike Information Criterion (AIC; Akaike, 1973). The model that had the lowest AIC score and significantly differed from more parsimonious models was selected as the best fit model.

Model 1: age only model including only a global smooth term for age

$$y_{ij} = \beta_0 + \beta_1(site_i) + s_1(age_{ij}) + s_2(age_{ij})site_j + u_i + error_{ij}$$

Model 2: age plus main effect of sex model including a global smooth term for age and a main effect of sex

$$y_{ij} = \beta_0 + \beta_1(site_i) + \beta_2(sex_i) + s_1(age_{ij}) + s_2(age_{ij})site_j + u_i + error_{ij}$$

Model 3: age plus sex*age interaction including a global smooth term for age, a main effect of sex, and group-level smooth terms for males and females with differing complexity

$$y_{ij} = \beta_0 + \beta_1(sample_i) + \beta_2(sex_i) + s_1(age_{ij}) + s_2(age_{ij})site_j + s_3(age_{ij})sex_j + u_i + error_{ij}$$

Here, s_1, s_2, and s_3 denote the arbitrary smooth functions (shrinkage version of cubic regression splines). Further, β_0 stands for the random intercepts per participant, β_1 for parameter estimates of site, and β_2 for parameter estimates of sex.

3 Results

For all seven subcortical structures, models with the interaction age*sex (Model 3) had lower AIC values than the more parsimonious models, and were thus deemed best fitting models (supplementary Table S3). Moreover, for all seven ROIs we found significant sex differences in age trajectories when examining difference smooths between females and males. Hence, we present results for GAMM models including age, sex, as well as an age*sex interaction to investigate age-constant sex differences as well as a sex difference in age trajectories (age*sex interaction), while controlling for site (main effect and age*site interaction). Notably, trajectories for each sex follow the global age trajectory (i.e. across all participants of both sexes) as they are penalized by the model to not differ too much from it (Pedersen et al., 2019). To better understand these differences, we examined age trajectories in GAMMs for each sex separately (independent from each other and the global age trajectory) while controlling for site (see Table 3 and Figure 2).

Table 3. GAMM estimates for age across both sites in females and males separately.

FEMALES					**MALES**				
THALAMUS	edf	Ref.df	*F*	*p*		edf	Ref.df	*F*	*p*
S(age)	2.77	3	54.82	***	S(age)	2.58	3	5.18	***
GLOBUS PALLIDUS									
S(age)	2.06	3	58.88	***	S(age)	2.70	3	80.86	***
CAUDATE NUCLEUS									
S(age)	2.70	3	197.81	***	S(age)	2.75	3	177.71	***
PUTAMEN									
S(age)	2.58	3	127.74	***	S(age)	2.68	3	61.43	***
NUCLEUS ACCUMBENS									
S(age)	1.75	3	10.61	***	S(age)	0.00	3	0.00	0.78
HIPPOCAMPUS									
S(age)	2.45	3	4.22	**	S(age)	1.76	3	1.15	***
AMYGDALA									
S(age)	0.00	3	0.00	0.74	S(age)	2.70	3	18.82	***

Smooth function (S(age) / edf), degrees of freedom (Ref.df), F-statistic and associated p-values for each term. * $p <$ .05, ** $p <$ .01, *** $p <$ .001.

3.1 Sex differences in raw volumes

To investigate volume differences between the sexes for all data sets and in each wave separately, we calculated percentage differences between males and females from raw values, not from GAMM modeling. For all structures, males had larger volumes as compared to females across the entire age range with overall differences ranging from 8.7% for the hippocampus to 13.7% for the amygdala (see Table 4). Percentage differences increased with age in thalamus, globus pallidus, caudate nucleus, hippocampus, and amygdala, which is partly mirrored by increasingly diverging trajectories of females and males over time (see Table 4; Figure 2).

Table 4. Percentage volume differences of subcortical brain structures between sexes, ordered by largest to smallest in the overall difference.

Structure	Difference in volume (male – female, %)				
	All (N = xx)	TP1 (N = xx)	TP2 (N = xx)	TP3 (N = xx)	TP4 (N = xx)
Amygdala	13.7	12.3	12.4	15.4	15.2
Putamen	10.8	9.4	10.6	12.5	10.7
Thalamus	9.8	8.6	9.7	10.2	11.6
Pallidum	9.4	7.2	9.0	11.0	11.8
Nucleus accumbens	9.2	8.0	8.8	11.9	7.9
Hippocampus	8.8	7.7	8.2	9.6	10.2
Caudate nucleus	7.4	5.6	7.5	8.7	8.2

For n of each structure per wave please refer to supplementary Table S1.

3.2 Trajectories for all participants; i.e. not differentiating sex and site

Estimation of smooth terms including all participants indicated overall volume decrease in caudate nucleus and putamen and overall volume increase in globus pallidus. Further, we found a nonlinear pattern of change in the thalamus and no significant volume change in nucleus accumbens, hippocampus and amygdala.

3.3 Sex differences in age trajectories

Thalamus: Males showed a volume increase until about age 17, while females remained rather stable until about age 16. After this, both sexes followed a similar nonlinear trajectory starting with a decrease and then a slight increase.

Globus pallidus: While females showed a nonlinear increase across the entire age range, the curve for males increased steeper until age 16 and flattened afterwards.

Caudate nucleus: For females, volume decreased until about age 22, then decrease flattened. Males showed a flatter decrease until about age 17, followed by a nonlinear decrease.

Putamen: For females, volume decreased until about age 20, then decrease flattened. Males showed a flatter decrease until about age 16, followed by a nonlinear decrease.

Nucleus accumbens: We did not find volume change in males, but a significant volume decrease in females which differed from the global age trajectory.

Hippocampus: While both groups showed significant change over time we did not find obvious decrease or increase over time. Males and females differed significantly from each other showing nonlinear patterns with females showing a volume decrease until age 20, then increasing again to about their initial volume level. Males increased slightly until age 17, followed by a slight increase.

Amygdala: We found no volume change in females but a volume increase in males until about age 16 in males with a following non-linear trajectory. Males also differed significantly from the global trajectory while females did not.

When examining sex differences in each site separately, significant differences between males' and females' trajectories were seen for all subcortical structures in Paris and for all subcortical structures except the nucleus accumbens in Dresden (see Figure 3).

Figure 2. Age trajectories for subcortical ROIs across the two sites Dresden and Paris.
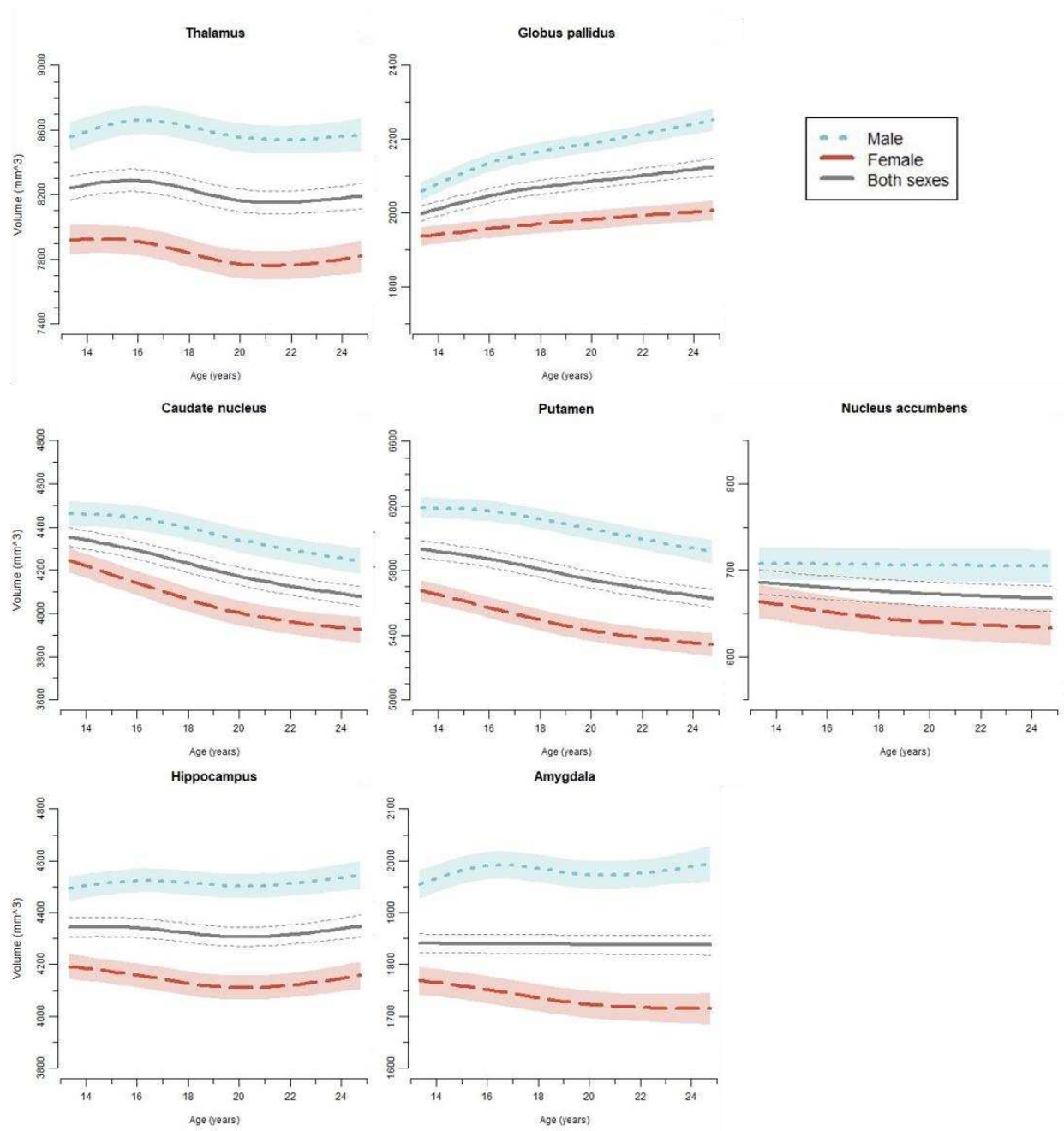
GAMM fitting with 95% confidence intervals plotted separately for females, males, and both sexes across Dresden and Paris.

3.4 Age-constant site differences in volumes

Overall, we found significantly smaller volumes in the Paris site for thalamus, globus pallidus, and hippocampus. Significantly larger volumes in the Paris site were seen for putamen, nucleus accumbens, and amygdala, while the caudate nucleus volume did not differ between the sites.

3.5 Site differences in age trajectories

We followed a similar logic for examining the site differences as for the sex differences. For all subcortical structures we found significant differences in age trajectories between the Dresden and Paris site. Examining the sites separately revealed significant volume change in thalamus, globus pallidus, caudate nucleus, and putamen for both sites, significant change only for Dresden in the nucleus accumbens, and no change in both sites for the amygdala.

Figure 3. Site differences of developmental trajectories for the subcortical structures.
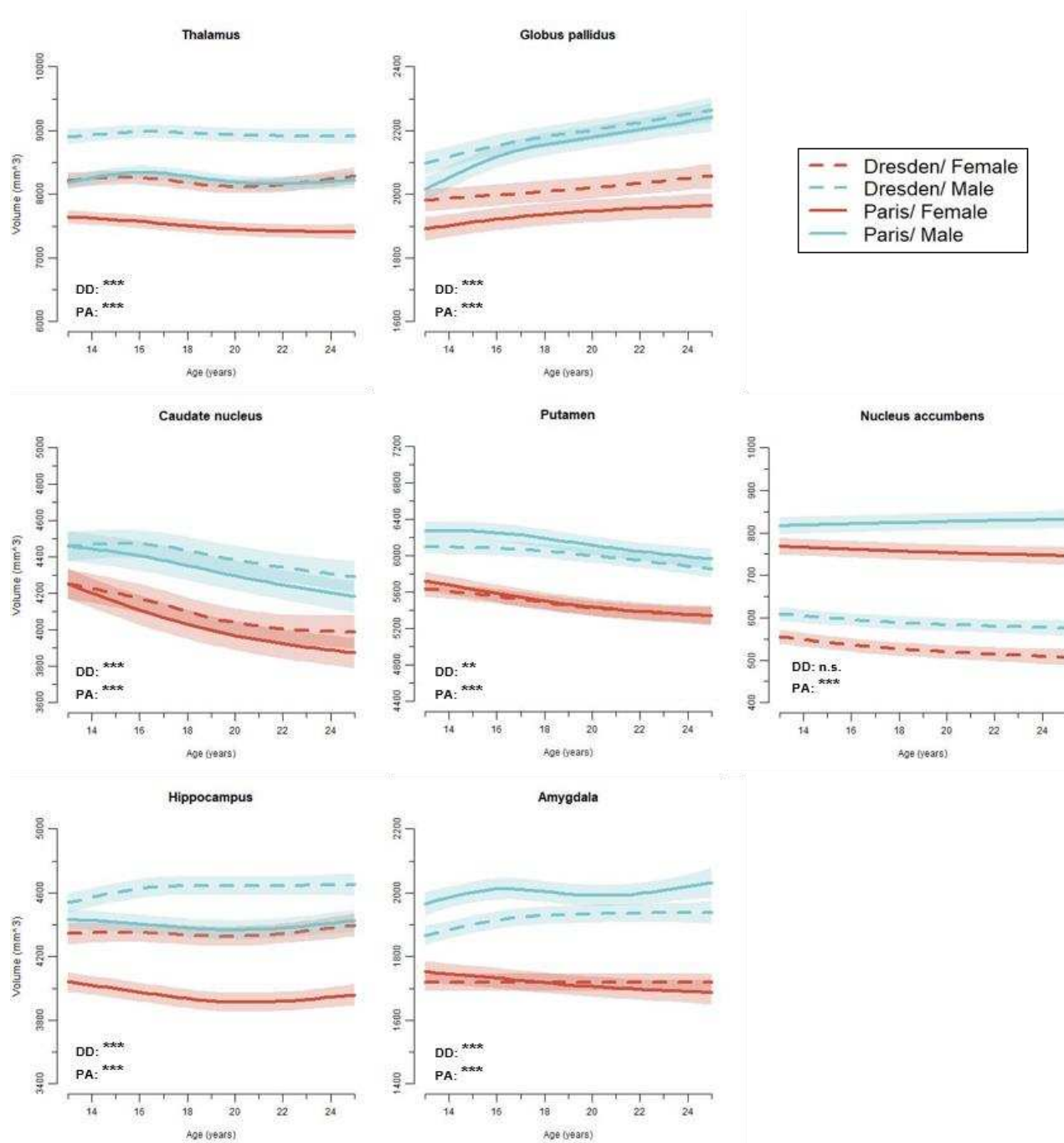
GAMM fitting with 95% confidence intervals plotted separately for females and males in Dresden and Paris. P-values represent sex differences per site; DD = Dresden; PA = Paris; n.s. = not significant; ** p<.01, *** $p < .001$.

4 Discussion

This study investigated age trajectories and their sex differences across ages 14 to 24 of the volumes of seven subcortical structures in two large samples from Germany and France as part of the IMAGEN

project including four waves of data acquisition. This is the first study with a single cohort design to focus on the mid-adolescent to young adult age span providing a continuous data set with 1408 total scans in 503 participants with 36% of participants having three, and 30% of participants having four waves of data acquisition. Our data indicate larger volumes in males compared to females and different age trajectories between the sexes across all subcortical regions. Although the current sample was partly included in the meta-analysis of Dima et al. (2021) we here present all 4 waves from age 14 to 24 processed in the longitudinal stream with FreeSurfer 6.0 (available only for the Dresden and Paris site) versus only the first wave in Dima et al. (2021) was processed using the cross-sectional stream with FreeSurfer 5.3 (available for all IMAGEN sites).

4.1 Overall age trajectories across both sexes

We found a decrease in the thalamus from age 16, which was in the range of other studies (15-19 years; Herting et al., 2018; Raznahan et al., 2014; Wierenga, Bos, et al., 2018). Our data show a stabilization thereafter. The slight increase from about age 21 is unexpected and might stem from the few scans at later ages (40 scans after age 23). We found an overall volume increase in the globus pallidus. This is in line with four other studies including a multi-site study (Dennison et al., 2013; Herting et al., 2018; Wierenga, Bos, et al., 2018; Wierenga et al., 2014). Wierenga et al. (2014) had very few scans after age 17, presumably leading to a peak with a following decrease. In contrast, Raznahan et al. (2014), found a decrease. Overall, we found volume decreases in putamen and caudate. This is in line with cross-sectional studies (Brain Development Cooperative Group, 2012; Østby et al., 2009) and a recent mega-analysis (Dima et al., 2021) as well as longitudinal studies (Dennison et al., 2013; Goddings et al., 2014; Tamnes et al., 2013; Wierenga et al., 2014). In contrast, other findings suggest no volume change for the caudate nucleus (Narvacan et al., 2017; Wierenga, Bos, et al., 2018) or an increase for the putamen (Wierenga, Bos, et al., 2018). We demonstrated no volume change in the nucleus accumbens. This is in line with (Dennison et al., 2013), while Wierenga, Bos, et al. (2018) found a decrease. Further, we found a decrease in the hippocampus followed by a slight increase from about age 20. The decreasing pattern is in line with Tamnes et al. (2013) but others found an increase (Dennison et al., 2013; Herting et al., 2018). We demonstrated no volume change in the amygdala which is in line with some previous studies (Dennison et al., 2013; Wierenga, Bos,

et al., 2018) but divergent to others (Goddings et al., 2014; Østby et al., 2009; Wierenga et al., 2014). However, findings seem to be driven by sex differences (see e.g. Goddings et al., 2014 and next section on sex differences).

4.2 Sex differences

4.2.1 Sex differences in raw volumes

We found larger volumes across subcortical structures for females versus males without allometric scaling. This is in line with previous studies (Herting et al., 2018; Wierenga, Bos, et al., 2018; Wierenga et al., 2014; Wierenga, Sexton, et al., 2018) and the volume differences of 7 to 14 % are in a similar range as found by (Narvacan et al., 2017).

4.2.2 Sex differences in age trajectories

While many previous studies only had few data sets per age group our single cohort study should have the computational strength to precisely model age trajectories during adolescence and detect sex differences. We found significant sex differences in age trajectories for all subcortical ROIs across the two sites. This is in line with Herting et al. (2018) and Raznahan et al. (2014) who also found sex differences in age trajectories for all their subcortical ROIs (Raznahan et al., 2014: striatum, globus pallidus, thalamus). We also found sex differences for each site separately, except for the nucleus accumbens in Dresden. This differs from Herting et al. (2018), who did not find sex differences for all regions when examining their three sites separately. The consistent sex differences for each site might be due to the more homogenous samples and more times points, i.e. power in our study. No sex differences in age trajectories were found in three studies (Narvacan et al., 2017; Wierenga et al., 2014; Wierenga, Sexton, et al., 2018). This is most likely an issue of power (studies covering about 1/ 2, 1/ 6, or 1/ 10 of our total scans) or larger age-ranges (from about primary school until mid-to late twenties). Interestingly, Dennison et al. (2013) had similarly low power (around 1/ 11 of our scans) in a longitudinal design with participants in the adolescent age range and still found sex differences in three subcortical regions. Only Herting et al. (2018) and Raznahan et al. (2014) had similarly powered samples but covered larger age-ranges. Single ROIs will be discussed below.

Overall, descriptively, we observed volume changes for males across several structures until age 16/ 17 while females showed different trajectories in this time span (i.e. stabilization/ decrease). Presumably, females reached the peaks and turning points earlier.

While the thalamus has been shown to change for females or males only (Dennison et al., 2013; Herting et al., 2018), we found a significant change in both sexes. In line with Dennison et al. (2013) and Herting et al. (2018) both sexes first decreased from about age 16 (females) or 17 (males) on. Differing from our study, Dennison et al. (2013) did not cover the age beyond 17. Raznahan et al. (2014) found a rather stable trajectory in males with a slight decrease in females after about age 14. In contrast to our study, Raznahan et al. (2014) had an accelerated design and assessed fewer participants in adolescence while covering earlier ages. Our slight increase from about age 21 is not in line with previous studies and might stem from the few scans after age 23.

For the globus pallidus, females showed a nonlinear increase across the entire age range, while the curve for males increased steeper until age 16 and flattened afterwards. Herting et al. (2018) only found a descriptive albeit non-significant increase for each sex. In contrast, Raznahan et al. (2014) demonstrated a similar slight decrease between the sexes across the adolescent age-range, while Dennison et al. (2013) found an increase in the globus pallidus and no sex differences. Taken together, sex differences in volume age trajectories for the globus pallidus were not convergingly found in the previous literature (also including Narvacan et al., 2017; Wierenga et al., 2014; Wierenga, Sexton, et al., 2018) and our trajectories show an increase in both sexes with just a steeper increase in males.

Since caudate nucleus and putamen form the striatum and show similar trajectories in our data, we discuss them together. We found a decrease for both structures: For females, volume decreased until about age 22 for the the caudate nucleus, respectivley age 20 for the putamen 20, then decrease flattened. Males showed a flatter decrease until about age 17 (putamen 16), followed by a nonlinear decrease. This is in line with Dennison et al. (2013) who found a decrease for both sexes in both structures. No decrease was shown by Herting et al. (2018) for males, but a significant decrease for females in both structures. Raznahan et al.

(2014) investigated both structures together as 'the striatum' and found an earlier peak for females at age 12 versus 14 for males with a slight decrease thereafter.

In line with Herting et al. (2018) we found a stable trajectory for the nucleus accumbens in males and a decrease in females while other studies (Dennison et al., 2013; Narvacan et al., 2017; Wierenga et al., 2014; Wierenga, Sexton, et al., 2018) did not find sex differences for this region. Taken together, our study together with Herting et al. (2018) is the only one so far to indicate sex differences in the nucleus accumbens.

We discuss amygdala and hippocampus trajectories together since they have similar trajectories. We can only descriptively compare our data with Herting et al. (2018) since they investigated also younger age groups from 8 years on and therefore sex differences in trajectories from age 14 to 24 were not statistically tested. We found no volume change in females but a volume increase in males until about age 16 with a following non-linear trajectory. Males also differed significantly from the global trajectory while females did not. For the amygdala, in line with Herting et al. (2018) who reported stable volumes from age 14 on for females and a slight increase in males, we did not find significant change in females but a nonlinear increase in males. Like Herting et al. (2018) we found significant nonlinear change in the hippocampus for both males and females while our data indicated slight increase from age 20 which was not shown by Herting et al. (2018). Other studies (Dennison et al., 2013; Narvacan et al., 2017; Wierenga, Bos, et al., 2018; Wierenga et al., 2014) did not find sex differences for these regions.

Notably, a recent study investigated sex and puberty effects on structural maturation subcortical structures in a subsample (waves at age 14 and 16) from Dresden and Paris using a voxel-based morphometry and ROI approach (Bézivin-Frere et al., 2020). The authors reported significant differences in volume trajectories with age and puberty in the amygdala-hippocampal complex, caudate, putamen, and thalamus. Together with another study pointing to independent and interactive influences of puberty and age on subcortical development (Goddings et al., 2014) the authors call attention to the importance of pubertal markers when looking at brain development during adolescence, as these processes are likely influenced by differences in sex hormones between females and males. As we only had information on

pubertal status for waves one and two and assume pubertal maturation complete afterwards, we did not include pubertal effects in our analyses. Still, differences in subcortical brain development with age and puberty, next to cognitive and social-emotional differences, may add insight into the development of sexual dimorphism in mental disorders during adolescence and later in life (Zahn-Waxler et al., 2008). One mechanism might be the influence of amygdala and hippocampus volume change which differently related to the development of self-reported positive characteristics (e.g. how generous, affectionate, and caring a person is) during adolescence between females and males (Bézivin-Frere et al., 2020).

Reviewing results across all subcortical structures, there is still need for further studies. So far, our study is only the seventh study with three studies spanning only low numbers of 60-147 participants and including only two waves or few participants with three time points (Dennison et al., 2013; Narvacan et al., 2017; Wierenga et al., 2014). All except one study had accelerated designs, so there is a need for larger studies with single cohort designs.

4.3 Site effects

Similar to Herting et al. (2018) who also used consistent quality control, longitudinal preprocessing and statistical analyses, we observed site differences and interactions of site by sex depending on the subcortical ROIs. Overall differences in volume between Dresden and Paris were especially visible in thalamus, hippocampus, and nucleus accumbens, and Herting et al. (2018) point to site differences in the thalamus, globus pallidus, caudate, and hippocampus. Interestingly, while Herting et al. (2018) analyzed more heterogeneous sites (differences in number of total scans, age spans, and sex distributions), our two European samples were vastly comparable on sample characteristics such as number of total scans, inter-scan interval, age range, sex distribution and pubertal and socioeconomic status. It is hence unlikely that differences in sample characteristics lead to overall volume differences in Dresden versus Paris. These systematic effects may rather stem from differences in the imaging sequence used on the two sites (higher resolution in Dresden) which may have locally affected the FreeSurfer segmentation algorithm. Similarly, sequence-related differences in subcortical structures have recently been found using CAT-12 (personal communication with J.-L. Martinot). Further, a scanner update was completed during wave three in Paris (Siemens Trio to Prisma) which may have introduced additional bias in results as Medawar et al. (2020)

and Plitman et al. (2020) reported very good ICC but also percent volume differences in several cortical and subcortical regions when comparing Siemens Scanners (Trio versus Prisma / Verio versus Skyra). Since many recent longitudinal studies on brain development included several samples with differences in imaging sequence and scanner type (Dennison et al., 2013; Mills et al., 2021; Tamnes et al., 2017; Wierenga, Sexton, et al., 2018), results should be interpreted with caution. Future studies in this field should consider using tools for multi-site harmonization like the recent algorithm longitudinal ComBat shown control type I error rate even better than the unharmonized data with scanner as a covariate approach (Beer et al., 2020).

4.4 Limitations

All findings of the present study must be considered in light of the following limitations. Some subcortical structures might suffer from low reliability, i.e. the nucleus accumbens (Mills et al., 2021) putamen or globus pallidus (Wonderlick et al., 2009). The globus pallidus is less distinct from its surrounding white matter as compared to other subcortical regions such as the thalamus or caudate nucleus (Fischl et al., 2002; Wonderlick et al., 2009). Further, the putamen seems to be systematically overestimated in Freesurfer 5.3, which has been improved in the Freesurfer 6.0 (see release notes). This becomes obvious when e.g. comparing putamen volumes at around age 14 in the samples of Herting et al. (2018) of around 6500 mm^3 for female and 7000 mm^3 for male with our values (5600 mm^3 for female 6200 mm^3 for male).

Concerning GAMM statistics, when global age smooth term is included additionally to group-level smooth term (as done here following Herting et al., 2018) the group-level smooth terms are penalized to not deviate too much from the global smooth terms. This inherently prevents female and male age trajectories from differing too much when fitted in these types of GAMM compared to when fitted separately. Further, concurvity (nonparametric equivalent of multicollinearity) is present when fitting models with separate global and group-level smooth terms as the global term could be approximated by a combination of female and male group-level smooth terms. This might lead to estimation issues in the model and inflated type 1 errors.

Moreover, current results are naturally dependent on power effects: when classes of sexes/ sites are modeled separately the number of data sets is divided in half and significance of effects might be changed (see discussion in Herting et al., 2018). Results further depend on differences in site composition over time as in our study more females than males returned in later waves which might have affected estimation of sex trajectories. Our data also provides more power for the first two waves. Thus, we can be more confident in describing trajectories during mid to late adolescence (until about age 17) while conclusions become less reliable after this time due to attrition of the site (especially after age 23).

While the strength of some previous studies was to show the development in-and-out of adolescence, we focused on the mid-adolescent to young adult age span and provide 1408 total scans within this age span. However, we cannot draw any conclusions about a potential turning point of trajectories before age 14. Additionally, in line with most other studies, the young participants were ethnically homogeneous (i.e. mostly White participants) and stemmed from a well-educated background which restricts generalizability. Future studies could try to recruit participants with a more diverse ethnical and socioeconomic background.

5 Conclusions

Our detailed descriptions of regional subcortical trajectories provide further evidence of sex differences in longitudinal adolescent brain development from mid-adolescence to young adulthood. This might eventually add to understanding of the relationship between underlying brain development and different adolescent psychopathology in boys and girls. Additionally, we illuminate pitfalls in large multi-site longitudinal brain imaging studies in an attempt to improve replicability of typical subcortical trajectories over the lifespan.

Individual contributions

Acknowledgments:

Disclosures

Dr Banaschewski served in an advisory or consultancy role for ADHS digital, Infectopharm, Lundbeck, Medice, Neurim Pharmaceuticals, Oberberg GmbH, Roche, and Takeda. He received conference support or speaker's fee by Medice and Takeda. He received royalties from Hogrefe, Kohlhammer, CIP Medien, Oxford University Press. The present work is unrelated to the above grants and relationships. Dr Barker has received honoraria from General Electric Healthcare for teaching on scanner programming courses. Dr Poustka served in an advisory or consultancy role for Roche and Viforpharm and received speaker's fee by Shire. She received royalties from Hogrefe, Kohlhammer and Schattauer. The present work is unrelated to the above grants and relationships. The other authors report no biomedical financial interests or potential conflicts of interest.

References

Akaike, H. (1973). *Information Theory and and Extension of the Maximum Likelihood Principle*. 267–281.

Backhausen, L. L., Herting, M. M., Buse, J., Roessner, V., Smolka, M. N., & Vetter, N. C. (2016). Quality Control of Structural MRI Images Applied Using FreeSurfer—A Hands-On Workflow to Rate Motion Artifacts. *Frontiers in Neuroscience*, *10*. https://doi.org/10.3389/fnins.2016.00558

Beer, J. C., Tustison, N. J., Cook, P. A., Davatzikos, C., Sheline, Y. I., Shinohara, R. T., Linn, K. A., & Initiative, for the A. D. N. (2020). Longitudinal ComBat: A Method for Harmonizing Longitudinal Multi-scanner Imaging Data. *BioRxiv*, 868810. https://doi.org/10.1101/868810

Bézivin-Frere, P., Vetter, N. C., Artiges, E., Filippi, I., Miranda, R., Vulser, H., Paillère-Martinot, M.-L., Ziesch, V., Conrod, P., Cattrell, A., Walter, H., Gallinat, J., Bromberg, U., Jurk, S., Menningen, E., Frouin, V., Papadopoulos Orfanos, D., Stringaris, A., Penttilä, J., … Lemaître, H. (2020). Sex effects on structural maturation of the limbic system and outcomes on emotional regulation during adolescence. *NeuroImage*, *210*, 116441. https://doi.org/10.1016/j.neuroimage.2019.116441

Boedhoe, P. S. W., van Rooij, D., Hoogman, M., Twisk, J. W. R., Schmaal, L., Abe, Y., Alonso, P., Ameis, S. H., Anikin, A., Anticevic, A., Arango, C., Arnold, P. D., Asherson, P., Assogna, F., Auzias, G., Banaschewski, T., Baranov, A., Batistuzzo, M. C., Baumeister, S., … van den Heuvel, O. A. (2020). Subcortical Brain Volume, Regional Cortical Thickness, and Cortical Surface Area Across Disorders: Findings From the ENIGMA ADHD, ASD, and OCD Working Groups. *American Journal of Psychiatry*, *177*(9), 834–843. https://doi.org/10.1176/appi.ajp.2020.19030331

Brain Development Cooperative Group. (2012). Total and regional brain volumes in a population-based normative sample from 4 to 18 years: The NIH MRI Study of Normal Brain Development. *Cerebral Cortex (New York, N.Y.: 1991)*, *22*(1), 1–12. https://doi.org/10.1093/cercor/bhr018

Dennison, M., Whittle, S., Yücel, 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. *Developmental Science*, *16*(5), 772–791. https://doi.org/10.1111/desc.12057

Dima, D., Modabbernia, A., Papachristou, E., Doucet, G. E., Agartz, I., Aghajani, M., Akudjedu, T. N., Albajes-Eizagirre, A., Alnæs, D., Alpert, K. I., Andersson, M., Andreasen, N. C., Andreassen, O. A., Asherson, P., Banaschewski, T., Bargallo, N., Baumeister, S., Baur-Streubel, R., Bertolino, A., … Frangou, S. (2021). Subcortical volumes across the lifespan: Data from 18,605 healthy individuals aged 3–90 years. *Human Brain Mapping*, *n/a*(n/a). https://doi.org/10.1002/hbm.25320

Ducharme, S., Albaugh, M. D., Nguyen, T.-V., Hudziak, J. J., Mateos-Pérez, J. M., Labbe, A., Evans, A. C., & Karama, S. (2016). Trajectories of cortical thickness maturation in normal brain development—The importance of quality control procedures. *NeuroImage*, *125*, 267–279. https://doi.org/10.1016/j.neuroimage.2015.10.010

Durston, S., Hulshoff Pol, H. E., Casey, B. J., Giedd, J. N., Buitelaar, J. K., & Van Engeland, H. (2001). Anatomical MRI of the Developing Human Brain: What Have We Learned?

Journal of the American Academy of Child & Adolescent Psychiatry, *40*(9), 1012–1020. https://doi.org/10.1097/00004583-200109000-00009

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

Giedd, J. N., Snell, J. W., Lange, N., Rajapakse, J. C., Casey, B. J., Kozuch, P. L., Vaituzis, A. C., Vauss, Y. C., Hamburger, S. D., Kaysen, D., & Rapoport, J. L. (1996). Quantitative Magnetic Resonance Imaging of Human Brain Development: Ages 4–18. *Cerebral Cortex*, *6*(4), 551–559. https://doi.org/10.1093/cercor/6.4.551

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

Gogtay, N., & Thompson, P. M. (2010). Mapping gray matter development: Implications for typical development and vulnerability to psychopathology. *Brain and Cognition*, *72*(1), 6–15. https://doi.org/10.1016/j.bandc.2009.08.009

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. (2018). Development of subcortical volumes across adolescence in males and females: A multisample study of longitudinal changes. *NeuroImage*, *172*, 194–205. https://doi.org/10.1016/j.neuroimage.2018.01.020

Hoogman, M., Bralten, J., Hibar, D. P., Mennes, M., Zwiers, M. P., Schweren, L. S. J., van Hulzen, K. J. E., Medland, S. E., Shumskaya, E., Jahanshad, N., Zeeuw, P. de, Szekely, E., Sudre, G., Wolfers, T., Onnink, A. M. H., Dammers, J. T., Mostert, J. C., Vives-Gilabert, Y., Kohls, G., … Franke, B. (2017). Subcortical brain volume differences in participants with attention deficit hyperactivity disorder in children and adults: A cross-sectional mega-analysis. *The Lancet Psychiatry*, *4*(4), 310–319. https://doi.org/10.1016/S2215-0366(17)30049-4

Jovicich, J., Marizzoni, M., Sala-Llonch, R., Bosch, B., Bartrés-Faz, D., Arnold, J., Benninghoff, J., Wiltfang, J., Roccatagliata, L., Nobili, F., Hensch, T., Tränkner, A., Schönknecht, P., Leroy, M., Lopes, R., Bordet, R., Chanoine, V., Ranjeva, J.-P., Didic, M., … Frisoni, G. B. (2013). Brain morphometry reproducibility in multi-center 3T MRI studies: A comparison of cross-sectional and longitudinal segmentations. *NeuroImage*, *83*, 472–484. https://doi.org/10.1016/j.neuroimage.2013.05.007

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. *Archives of General Psychiatry*, *62*(6), 593. https://doi.org/10.1001/archpsyc.62.6.593

Koolschijn, P. C. M. P., & Crone, E. A. (2013). Sex differences and structural brain maturation from childhood to early adulthood. *Developmental Cognitive Neuroscience*, *5*, 106–118. https://doi.org/10.1016/j.dcn.2013.02.003

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*(4), 1065–1073. https://doi.org/10.1016/j.neuroimage.2007.03.053

Liu, S., Hou, B., Zhang, Y., Lin, T., Fan, X., You, H., & Feng, F. (2020). Inter-scanner reproducibility of brain volumetry: Influence of automated brain segmentation software. *BMC Neuroscience*, *21*(1), 35. https://doi.org/10.1186/s12868-020-00585-1

Makowski, C., Béland, S., Kostopoulos, P., Bhagwat, N., Devenyi, G. A., Malla, A. K., Joober, R., Lepage, M., & Chakravarty, M. M. (2018). Evaluating accuracy of striatal, pallidal, and thalamic segmentation methods: Comparing automated approaches to manual delineation. *NeuroImage*, *170*, 182–198. https://doi.org/10.1016/j.neuroimage.2017.02.069

Medawar, E., Thieleking, R., Manuilova, I., Paerisch, M., Villringer, A., Witte, A. V., & Beyer, F. (2020). Estimating the effect of a scanner upgrade on measures of grey matter structure for longitudinal designs. *BioRxiv*, 2020.08.28.271296. https://doi.org/10.1101/2020.08.28.271296

Mills, K. L., Siegmund, K. D., Tamnes, C. K., Ferschmann, L., Wierenga, L. M., Bos, M. G. N., Luna, B., & Herting, M. M. (2021). Individual variability in structural brain development from late childhood to young adulthood. *BioRxiv*, 2021.02.04.429671. https://doi.org/10.1101/2021.02.04.429671

Morey, R. A., Selgrade, E. S., Wagner II., H. R., Huettel, S. A., Wang, L., & McCarthy, G. (2010). Scan-rescan reliability of subcortical brain volumes derived from automated segmentation. *Human Brain Mapping*, *31*(11), 1751–1762. Scopus. https://doi.org/10.1002/hbm.20973

Narvacan, K., Treit, S., Camicioli, R., Martin, W., & Beaulieu, C. (2017). Evolution of deep gray matter volume across the human lifespan. *Human Brain Mapping*, *38*(8), 3771–3790. https://doi.org/10.1002/hbm.23604

Noordermeer, S. D., Luman, M., Greven, C. U., Veroude, K., Faraone, S. V., Hartman, C. A., Hoekstra, P. J., Franke, B., Buitelaar, J. K., Heslenfeld, D. J., & Oosterlaan, J. (2017). Structural Brain Abnormalities of Attention-Deficit/Hyperactivity Disorder With Oppositional Defiant Disorder. *Biological Psychiatry*, *82*(9), 642–650. https://doi.org/10.1016/j.biopsych.2017.07.008

Østby, Y., Tamnes, C. K., Fjell, A. M., Westlye, L. T., Due-Tønnessen, 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

Paus, T., Keshavan, M., & Giedd, J. N. (2008). Why do many psychiatric disorders emerge during adolescence? *Nature Reviews Neuroscience*, *9*(12), 947–957. https://doi.org/10.1038/nrn2513

Pedersen, E. J., Miller, D. L., Simpson, G. L., & Ross, N. (2019). Hierarchical generalized additive models in ecology: An introduction with mgcv. *PeerJ*, *7*, e6876. https://doi.org/10.7717/peerj.6876

Petermann, F., & Petermann, U. (2007). *Hamburg Wechsler Intelligenztest für Kinder –IV (HAWIK-IV)*. Huber.

Petersen, A. C., Crockett, L., Richards, M., & Boxer, A. (1988). A self-report measure of pubertal status: Reliability, validity, and initial norms. *Journal of Youth and Adolescence*, *17*(2), 117–133. https://doi.org/10.1007/BF01537962

Plitman, E., Bussy, A., Valiquette, V., Salaciak, A., Patel, R., Béland, M.-L., Tullo, S., Tardif, C., Rajah, M. N., Near, J., Devenyi, G. A., & Chakravarty, M. M. (2020). The Impact of the Siemens Trio to Prisma Upgrade and Volumetric Navigators on MRI Indices: A Reliability Study with Implications for Longitudinal Study Designs. *BioRxiv*, 2020.06.19.162420. https://doi.org/10.1101/2020.06.19.162420

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. *Proceedings of the National Academy of Sciences*, *111*(4), 1592–1597. https://doi.org/10.1073/pnas.1316911111

Reuter, M., Rosas, H. D., & Fischl, B. (2010). Highly accurate inverse consistent registration: A robust approach. *NeuroImage*, *53*(4), 1181–1196. https://doi.org/10.1016/j.neuroimage.2010.07.020

Reuter, M., Schmansky, N. J., Rosas, H. D., & Fischl, B. (2012). Within-subject template estimation for unbiased longitudinal image analysis. *NeuroImage*, *61*(4), 1402–1418. https://doi.org/10.1016/j.neuroimage.2012.02.084

Rogers, J. C., & Brito, S. A. D. (2016). Cortical and Subcortical Gray Matter Volume in Youths With Conduct Problems: A Meta-analysis. *JAMA Psychiatry*, *73*(1), 64–72. https://doi.org/10.1001/jamapsychiatry.2015.2423

Schumann, G., Loth, E., Banaschewski, T., Barbot, A., Barker, G., Büchel, C., Conrod, P. J., Dalley, J. W., Flor, H., Gallinat, J., Garavan, H., Heinz, A., Itterman, B., Lathrop, M., Mallik, C., Mann, K., Martinot, J.-L., Paus, T., Poline, J.-B., … Struve, M. (2010). The IMAGEN study: Reinforcement-related behaviour in normal brain function and psychopathology. *Molecular Psychiatry*, *15*(12), 1128–1139. https://doi.org/10.1038/mp.2010.4

Shaw, P., Gogtay, N., & Rapoport, J. (2010). Childhood psychiatric disorders as anomalies in neurodevelopmental trajectories. *Human Brain Mapping*, *31*(6), 917–925. https://doi.org/10.1002/hbm.21028

Sled, J. G., Zijdenbos, A. P., & Evans, A. C. (1998). A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Transactions on Medical Imaging*, *17*(1), 87–97. https://doi.org/10.1109/42.668698

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 & Child Neurology*, *44*(1), 4–16. https://doi.org/10.1111/j.1469-8749.2002.tb00253.x

Tamnes, C. K., Herting, M. M., Goddings, A.-L., Meuwese, R., Blakemore, S.-J., Dahl, R. E., Güroğlu, 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. *Journal of Neuroscience*, *37*(12), 3402–3412. https://doi.org/10.1523/JNEUROSCI.3302-16.2017

Tamnes, C. K., Walhovd, K. B., Dale, A. M., Østby, Y., Grydeland, H., Richardson, G., Westlye, L. T., Roddey, J. C., Hagler, D. J., Due-Tønnessen, P., Holland, D., & Fjell, A. M. (2013). Brain development and aging: Overlapping and unique patterns of change. *NeuroImage*, *68*, 63–74. https://doi.org/10.1016/j.neuroimage.2012.11.039

Vijayakumar, N., Mills, K. L., Alexander-Bloch, A., Tamnes, C. K., & Whittle, S. (2018). Structural brain development: A review of methodological approaches and best practices.

Developmental Cognitive Neuroscience, *33*, 129–148. https://doi.org/10.1016/j.dcn.2017.11.008

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*, 105–114. https://doi.org/10.1016/j.psyneuen.2018.02.034

Wierenga, L. M., Bos, M. G. N., van Rossenberg, F., & Crone, E. A. (2019). Sex Effects on Development of Brain Structure and Executive Functions: Greater Variance than Mean Effects. *Journal of Cognitive Neuroscience*, *31*(5), 730–753. https://doi.org/10.1162/jocn_a_01375

Wierenga, L. M., 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. https://doi.org/10.1016/j.neuroimage.2014.03.072

Wierenga, L. M., Sexton, J. A., Laake, P., Giedd, J. N., Tamnes, C. K., & Pediatric Imaging, Neurocognition, and Genetics Study. (2018). A Key Characteristic of Sex Differences in the Developing Brain: Greater Variability in Brain Structure of Boys than Girls. *Cerebral Cortex (New York, N.Y.: 1991)*, *28*(8), 2741–2751. https://doi.org/10.1093/cercor/bhx154

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*(4), 1324–1333. https://doi.org/10.1016/j.neuroimage.2008.10.037

Wood, S. N. (2004). Stable and Efficient Multiple Smoothing Parameter Estimation for Generalized Additive Models. *Journal of the American Statistical Association*, *99*(467), 673–686. https://doi.org/10.1198/016214504000000980

Wood, S. N. (2006). *Generalized Additive Models: An Introduction with R*. Chapman and Hall/CRC. https://doi.org/10.1201/9781420010404

Zahn-Waxler, C., Shirtcliff, E. A., & Marceau, K. (2008). Disorders of Childhood and Adolescence: Gender and Psychopathology. *Annual Review of Clinical Psychology*, *4*(1), 275–303. https://doi.org/10.1146/annurev.clinpsy.3.022806.091358

Zheng, W., Chee, M. W. L., & Zagorodnov, V. (2009). Improvement of brain segmentation accuracy by optimizing non-uniformity correction using N3. *NeuroImage*, *48*(1), 73–83. https://doi.org/10.1016/j.neuroimage.2009.06.039