bioRxiv preprint doi: https://doi.org/10.1101/2021.02.11.430883; this version posted February 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Tracing Human Amygdala across School Age

Quan Zhou[a,b], Siman Liu[a,b], Chao Jiang[a,b], Ye He[c], Xi-Nian Zuo[a,b,d,e,f,*]

[a] *Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Science, Beijing, 100101, China*
[b] *Department of Psychology, University of Chinese Academy of Sciences, Beijing, 100049, China*
[c] *School of Artificial Intelligence, Beijing University of Posts and Telecommunications, Beijing, 100876, China*
[d] *State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing, 100875, China*
[e] *National Basic Public Science Data Center, Beijing, 100190, China*
[f] *IDG/McGovern Institute for Brain Research, Beijing Normal University, Beijing, 100875, China*

Abstract

The developmental patterns of the amygdala in children and adolescences have been inconsistent in previous studies. This discrepancy may be partly due to methodological differences in segmentation by tracing the human amygdala. To investigate the impact of tracing methods on amygdala volume, we compared *FreeSurfer* and *volBrain* segmentation measurements with those obtained by manual tracing. The manual tracing method, as the 'Gold Standard', exhibited almost perfect intra- and inter-rater reliability. We observed systematic differences in amygdala volumes between automatic and manual methods. Specifically, compared with the manual tracing, *FreeSurfer* estimated larger amygdalae while *volBrain* produced smaller amygdalae. This tracing bias was larger for smaller amygdalae. We further modeled amygdalar growth curves using accelerated longitudinal cohort data from the Chinese Color Nest Project (total 427 magnetic resonance imaging scans from 198 participants aged 6-17 years at baseline). Trajectory modeling and statistical assessments of the manually traced amygdalae revealed linearly increasing and parallel developmental patterns for both girls and boys, although the amygdalae of boys were larger than those of

[*] Corresponding author
Email address: `xinian.zuo@bnu.edu.cn; zuoxn@psych.ac.cn` (Xi-Nian Zuo)

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.11.430883; this version posted February 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

girls. Comparing these trajectories, the shapes of developmental trajectories were similar when using the *volBrain* derived volumes while *FreeSurfer* led to more nonlinear and flattened, but statistically non-significant, growth patterns. The use of amygdala volumes adjusted for total gray-matter volumes, but not intracranial volumes, resolved the shape discrepancies and led to reproducible growth curves across the three methods. Our findings revealed steady growth of the human amygdala, mirroring the functional development across the school age. We argue that methodological improvements are warranted for current automatic tools to achieve more accurate tracing of the amygdala at school age.

Keywords: amygdala, brain development, growth chart, MRI, reliability

1. Introduction

Childhood and adolescence are key periods for socioemotional development, which correlate strongly with the development of risk factors for diverse neuropsychiatric disorders (Paus et al., 2008). Together with enhanced efforts to prevent such disorders, many large-scale studies have been undertaken to explore behavioral and biological development of children and adolescents (Ortiz and Raine, 2004; Silk et al., 2007; Connor, 2004). Rapid progress in in-vivo brain imaging technologies has accelerated the use of structural magnetic resonance imaging (MRI) to quantify volumes of different brain structures. These morphological features have been demonstrated by MRI to be sensitive for developmental brain changes (Tamnes et al., 2013; Albaugh et al., 2017; Wierenga et al., 2018). The accurate developmental trajectories of brain structures using MRI is thus an important requirement for understanding the neurodevelopment mechanism of these disorders occurring during childhood and adolescence.

The amygdala is an almond-shaped brain structure of the limbic system and is highly connected with other brain regions (Schumann and Amaral, 2005). It plays important roles in emotional and cognitive processes, especially fear and threat processing (LeDoux, 1998; Cardinal et al., 2002; Pessoa, 2010) and exhibits network-level connectivity changes across the human lifespan (He et al.,

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.11.430883; this version posted February 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

2016). Abnormal amygdalar structure in children and adolescents has been related to a plethora of neurodevelopmental abnormalities (Scherf et al., 2013; Schumann et al., 2011), including autism (Mosconi et al., 2009; Schumann et al., 2004), anxiety disorder (De Bellis et al., 2000; Redlich et al., 2015) and schizophrenia (Ganzola et al., 2014). Meanwhile, many studies have explored age-related changes of the amygdala in pediatric and adolescent samples (Uematsu et al., 2012; Gilmore et al., 2012; Wierenga et al., 2014; Barnea-Goraly et al., 2014; Herting et al., 2018), indicating the promise of using normal growth patterns for monitoring abnormal development. Growth charts are expected to aid risk evaluation, early diagnosis and educational monitoring by delineating typical development standards. In several recent studies, researchers have tracked the age-related increases of amygdala volume from childhood through adolescence (Herting et al., 2018; Goddings et al., 2014; Albaugh et al., 2017). However, a study including 271 individuals aged 8-29 years reported no significant changes in amygdala volume (Wierenga et al., 2018). This was similar to the observation from a sample of 85 individuals scanned twice across 8-22 years (Tamnes et al., 2013). Thus, there are mixed findings in the literature related to age-related differences or changes in amygdala volume. The anatomical complexity can limit the accurate measurement of amygdalar volume, leading to a large variation in findings obtained using different amygdala segmentation methods (Lyden et al., 2016), which may explain this inconsistency and less reproducibility (Mills and Tamnes, 2014; Lyden et al., 2016).

Manual tracing is commonly considered the ’gold standard’ for amygdala segmentation (Morey et al., 2009). It enables flexible quantification guided by prior anatomical knowledge, without the need to make any of the assumptions built into algorithms. Experienced human tracers can correctly label ambiguous borders by adjusting for variation caused by complex or atypical anatomy and image artifacts. To increase reliability and reduce potential biases associated with manual tracing, multiple protocols have been generated and described in the literature (Schumann et al., 2004; Pruessner et al., 2000; Watson et al., 1992). These protocols significantly increase intra- and inter-rater agree-

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.11.430883; this version posted February 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

ment (Pruessner et al., 2000). However, manual tracing is time-consuming and requires the operator to have sufficient anatomical expertise. For large MRI datasets, the labor cost of manual tracing is prohibitive (Akudjedu et al., 2018; Schmidt et al., 2018). There is also subtly drift in tracing criteria of manual raters during the course of a long study. Accordingly, it is critical to develop automatic techniques that can accurately segment amygdala structures from large and growing datasets while providing consistent results and minimizing the human effort necessary for manual tracing.

Several tools have been developed to achieve automatic segmentation in a time-efficient manner including *FreeSurfer* and *volBrain*, which are both freely available, ease to use, nearly fully automated, and very accurate (Fischl et al., 2002; Manjón and Coupé, 2016; Morey et al., 2009; Akudjedu et al., 2018; Schmidt et al., 2018; Næss-Schmidt et al., 2016). Although automated segmentation has been shown to be comparable to manual tracing for adult populations (Fischl et al., 2002; Manjón and Coupé, 2016; Morey et al., 2009; Grimm et al., 2015), its performance for child and adolescent samples, in which head size and shape as well as the pace of structural growth differ, has not been validated adequately (Herten et al., 2019). In addition, the effects of any differences in the accuracy of automatic and manual amygdala segmentation on the subsequent examination of amygdala development in school-age children and adolescents remain incompletely understood. To fully characterize similarities and discrepancies among techniques, we compared amygdala volumes obtained manually to those extracted by *FreeSurfer* and *volBrain* using 427 longitudinal structural MRI scans from 198 healthy children and adolescents (baseline age: 6-17 years). To answer the aforementioned question, we examined how different tracing methods lead to trajectory differences in amygdala development across school age. Based upon previous reports (Morey et al., 2009; Schoemaker et al., 2016), we expected to observe systematic differences in amygdala segmentation performance among the three tracing methods. We hypothesized that such differences would affect the modeling of human amygdala growth.

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.11.430883; this version posted February 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

2. Materials and Methods

2.1. Participants

The sample described in this study was part of a five-year accelerated longitudinal data of the Chinese Color Nest Project (CCNP) (Liu et al., 2020). It was part of the developmental component of CCNP (devCCNP), and collected at Southwest University (CCNP-SWU), Chongqing, China. The devCCNP was designed to delineate normative trajectories of brain development in the Chinese population across the school-aged years. The participants had no neurological or mental health problem and did not use psychotropic medication; their estimated intelligence quotients were $\geqslant 80$. The CCNP-SWU samples included data from 201 typically developing controls (TDCs) aged 6-17 years who were invited to participate in three consecutive waves of data collection at intervals of approximately 1.25 years (Dong et al., 2020). T1-weighted MRI examinations were performed at these time points, and the images were visually inspected to exclude those with substantial head-motion artifacts and those with structural abnormalities. After this initial quality control, the final sample included 427 scans from 198 participants (105 females; 93 males; **Table 1**). Scans from three time points, two time points, and one time point were available for 79, 71, and 48 participants, respectively. The mean number of scans per participant was 2.16 (standard deviation = 0.79). The current study was approved by review committees of the participating institutions (the Institute of Psychology, Chinese Academy of Sciences, and Southwest University).

2.2. MRI acquisition

All participants underwent MRI examinations performed with a Siemens Trio$^{\mathrm{TM}}$ 3.0 Tesla MRI scanner. A high-resolution magnetization-prepared rapid gradient-echo (MP-RAGE) T1 sequence (matrix $= 256 \times 256$, FOV $= 256 \times 256$ mm^2, slices thickness = 1mm, repetition time (TR) = 2600 ms, echo time (TE) = 3.02 ms, inversion time (TI) = 900 ms, flip angle = 15°, number of slices = 176) was obtained for each individual.

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.11.430883; this version posted February 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

2.3. Volumetric MRI preprocessing and segmentation

All the images were anonymized by removing all the personal information from the raw MRI data. We removed the facial information by using the **facemasking** tool (Milchenko and Marcus, 2013). The anonymized images were then uploaded to the online image processing system *volBrain* (`http://volbrain.upv.es`) (Manjón and Coupé, 2016), which performed spatially adaptive non-local means of denoising and correction for intensity normalization. All the preprocessed individual brain volumes were in the native space and ready for subsequent manual and automatic tracing procedures.

2.3.1. Manual tracing and reliability assessment

Anatomically trained raters QZ (the first author Quan Zhou) and ZQZ performed manual amygdala segmentation in the native space using the ITK-SNAP software (ver. 3.8.0) (Yushkevich et al., 2006). The anatomical boundaries of amygdala structures were defined and segmented according to the protocol described by Pruessner et al. (2000). This protocol has been demonstrated to achieve almost perfect intra- and inter-rater reliability. The reliability was quantified with intraclass correlation coefficient (ICC), which was interpreted as indicating slight $[0, 0.20)$, fair $[0.20, 0.40)$, moderate $[0.40, 0.60)$, substantial $[0.60, 0.80)$, or almost perfect $[0.80, 1]$ reliability (Landis and Koch, 1977). To assess reliability for the protocol implementation in this study, QZ and ZQZ independently traced the amygdala volumes of 30 scans twice at a two-week interval. They were chosen from 30 subjects at baseline examination balanced for age and sex. The ICCs with a 95% confidence interval (CI) are derived by the following hierarchical linear mixed model on the repeated tracing volumes

$$\begin{aligned} V_{ijk} = \quad & \gamma_{000} + \text{subject}_{i00} + \text{order}_j + \text{rater}_k \\ & + \text{subject} \times \text{order}_{ij} \\ & + \text{subject} \times \text{rater}_{ik} \\ & + \text{order} \times \text{rater}_{jk} \\ & + e_{ijk} \end{aligned} \tag{1}$$

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.11.430883; this version posted February 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

where V_{ijk} represents the amygdalar volume measurement for the $i-$th ($i = 1, 2, \cdots, 30$) participant in the $j-$th ($j = 1, 2$) manual tracing by the $k-$th rater ($k = 1, 2$); γ_{000} is the intercept for a fixed effect of the group average; the following three terms represent random effects for the $i-$th participant, the $j-$th tracing order, the $k-$th rater, respectively; and other three terms denote random interaction effects between the $j-$th tracing and the $i-$th participant, between the $k-$th rater and the $i-$th participant, between the $j-$th tracing and the the $k-$th rater; and r_{ijk} is an error term.

The above-mentioned model assumes that the seven included variables are independent and distributed normally with zero means. The total variances can be decomposed into the variance component:

- among participants $\sigma^2_{\text{subject}}$
- between repeated tracings by the same rater σ^2_{order}
- between raters for the same tracing order σ^2_{rater}
- among participants due to the differences in tracing order $\sigma^2_{\text{subject}\times\text{order}}$
- among participants due to the differences in rater $\sigma^2_{\text{subject}\times\text{rater}}$
- between two raters due to the differences in tracing order $\sigma^2_{\text{order}\times\text{rater}}$
- of the residual σ^2_r.

We define the inter-rater reliability of the human amygdala volumetric measurements by manual tracing as:

$$\text{interICC} = \frac{\sigma^2_{\text{subject}}}{\sigma^2_{\text{subject}} + \sigma^2_{\text{rater}} + \sigma^2_{\text{subject}\times\text{rater}} + \sigma^2_r} \tag{2}$$

and the intra-rater reliability of the human amygdala volumetric measurements by manual tracing as:

$$\text{intraICC} = \frac{\sigma^2_{\text{subject}}}{\sigma^2_{\text{subject}} + \sigma^2_{\text{order}} + \sigma^2_{\text{subject}\times\text{order}} + \sigma^2_r} \tag{3}$$

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.11.430883; this version posted February 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

2.3.2. Automatic tracing and visual inspection

Amygdala volumes were estimated using *volBrain* (`http://volbrain.upv.es`), a fully automated segmentation method that has outperformed other segmentation methods across many brain structures. The operational pipeline has been described and evaluated previously (Manjón and Coupé, 2016). Intracranial volume (ICV) and total gray matter volume (GMV) were also derived using *volBrain*. Automatic segmentation and labeling of the human amygdala were also performed using the "recon-all" pipeline in *FreeSurfer* (ver. 6.0.0; `http://surfer.nmr.mgh.harvard.edu`). The *FreeSurfer* processing stages have been documented in (Fischl et al., 2002). Amygdala volumes provided in **aseg.stats** files were used in the subsequent analysis, and **aseg.mgz** volume files were converted into NIFTI files in native space for visualization. Transformation for segmentation and its inverse transformation to native space for volumetric comparison have been described in (Morey et al., 2009). Both ICV and GMV measurements were also provided by the *FreeSurfer* outputs. Visual inspection of the traced amygdala volumes in a representative subject using manual and automatic methods is illustrated in **Figure 1**.

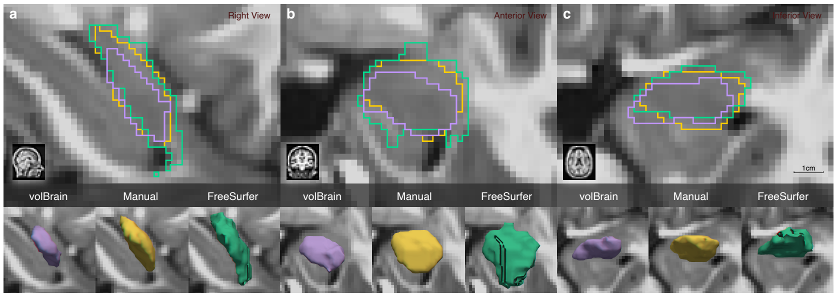

Figure 1: **Tracing left amygdala with the three methods in a sample subject**. Purple: *volBrain*; Yellow: manual; Green: *FreeSurfer*

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.11.430883; this version posted February 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

2.4. Accuracy assessments on automatic segmentation

QZ manually traced all the 427 amygdala of the CCNP-SWU samples, which served as the reference volumes (i.e., gold standard) for the subsequent analyses. We validated the accuracy of automatic segmentation separately for each of the three waves of the samples. For each wave, we performed paired t-tests on traced volumes between the automatic and manual methods. We quantified volume difference between the automatic and manual tracing as the equation 4. A greater volume difference indicates increased discrepancy relative to the manually segmented amygdala volumes. To examine systematic changes of the traced volumes, we tested the Pearson's correlation of traced volumes between the automatic and manual methods across individual subjects. A strong correlation ($R \geq 0.8$) is taken to indicate good consistency on the individual differences in amygdala volumes between the manual and automatic methods. We further calculated the spatially overlapping volumes and the false positive rate to quantitatively measure the degree of correct or incorrect estimation of the automatic methods. These metrics are defined as:

- percentage of volume difference

$$D(V_A, V_M) = \frac{|V_A - V_M|}{V_M} \times 100\% \tag{4}$$

- percentage of spatial overlap

$$P(V_A, V_M) = \frac{V_A \cap V_M}{0.5(V_A + V_M)} \times 100\% \tag{5}$$

- false-positive rate

$$F(V_A, V_M) = \frac{V_A - V_A \cap V_M}{V_A} \times 100\% \tag{6}$$

In these equations, V_A is the volume measured automatically and V_M is that measured manually (the reference, i.e., the gold standard). The maximum $P(V_A, V_M)$ value is 100%, reflecting identical tracing between manual and automatic method while smaller values indicated less perfect spatial overlaps (Morey et al., 2009), implying the worse performance of the automatic tracing. The

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.11.430883; this version posted February 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

minimum $F(V_A, V_M)$ value is 0, reflecting identical tracing between manual and automatic method while larger values indicate higher error rates of automatic segmentation, i.e., the inclusion of larger proportions of non-amygdalar structure(s).

To further investigate how the accuracy of the automatic tracing methods varies with amygdala sizes, we employed a generalized additive mixed model (GAMM) to model the size effect of amygdala on the automatic tracing accuracy. Specifically, we plotted the spatial overlap (the overlap percentage P) between automatic and manual segmentation as a function of the reference (i.e., the manually traced) volumes. Unlike the common parametric linear models (Herting et al., 2018), GAMM does not require a-priori knowledge of the relationship between the response and predictors, which enables more flexible and efficient estimation of changing patterns (Mills and Tamnes, 2014; Wood, 2017). In addition, GAMMs are well suited for the repeated measurements (e.g., our accelerated longitudinal samples from developing brains), as they account for both within-subject dependency and developmental differences among participants at the time of study enrollment (Alexander-Bloch et al., 2014; Harezlak et al., 2005). Such a GAMM was implemented using the following formula in R language with the **mgcv** package:

$$P(V_A, V_M) \sim s(V_M) + (1|\text{subject}) \tag{7}$$

where the $s()$ is a smoothing function with a fixed degree of freedom and cubic B-splines, whose number of knots is set at 5 (determined to be optimal for our data). This was set to be sufficiently large to have adequate degrees of freedom across both spline terms from fits of the model to the amygdala volume, but sufficiently small to maintain reasonable computational efficiency.

2.5. Modeling growth curves of human amygdala development

To fully model method-related differences in the growth curves of human amygdala volumes, we employed the following GAMM to examine age-related

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.11.430883; this version posted February 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

changes of the human amygdala by including the tracing method and its interaction with age as variables of interests:

$$V \sim s(age) + \text{method} + s(age, by = \text{method}) + (1|\text{subject}) \tag{8}$$

where V represents the amygdala volume and $s()$ is a smoothing function, with a fixed degree of freedom and cubic B-splines (the number of knots = 5). Tracing method was entered as an ordinal factor (manual = 0, automatic = 1). The method term reflects the method differences in the intercept (i.e., the main effect of method). The first smoothing term models the slope of age for manual tracing, and the second smoothing term models the difference in the age-related slope between methods (i.e., age×method interaction). The p value associated with this term is the basis of statistical inference regarding methodological differences in developmental trajectories of bilateral amygdala volume.

To more specifically understand differences in age trajectories between methods, a set of GAMMs (see the equation 8) were proposed to detect age-related changes revealed by each method separately:

$$\begin{aligned} &V \sim s(age) + (1|\text{subject}) \\ &V \sim s(age) + \text{sex} + (1|\text{subject}) \\ &V \sim s(age) + \text{sex} + s(age, by = \text{sex}) + (1|\text{subject}) \end{aligned} \tag{9}$$

The first GAMM models the traced volume as a smoothing function of age. As previous studies have consistently shown larger brain regions in males than in females (Herting et al., 2018), we established the second GAMM model with sex as a fixed term to assess the sex difference in the trajectory intercept as well as the third GAMM model including age×sex interaction to test the sex differences in the trajectory slope. The Akaike Information Criterion (AIC) was used to determine which model had the best fit (the lowest AIC value). These analyses were performed using the **mgcv** (Wood, 2017) and **ggplot2** (Wickham, 2016) packages in R (R Core Team, 2014).

We also tested growth curves of the human amygdala by accounting for global brain features in the GAMMs. The volumes of subcortical structures

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.11.430883; this version posted February 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

are known to be related to brain size (Brown et al., 2014; Brain Development Cooperative Group, 2012; Uematsu et al., 2012). Accordingly, we included ICV as a co-variate for regression control to enable the removal of individual variability that can be explained by brain size (Narvacan et al., 2017; Sawiak et al., 2018; Herting et al., 2014). Researchers have also demonstrated that the size of the amygdala often scales with the GMV (Van Petten, 2004; Rice et al., 2014). We thus accounted for brain size by controlling for the GMV in the GAMMs. We performed the analysis with ICV and GMV measurements obtained by *FreeSurfer* and *volBrain*, respectively.

3. Results

3.1. Measurement reliability of manually traced human amygdala

We reported almost perfect reliability of the human amygdala volumes measured by the manual tracing protocol. Specifically, as in **Table 2**, both intra-rater and inter-rater reliability of the volumes for the manually traced amygdala were achieved. Inter-rater ICCs were around 0.88 with 95%CI$=[0.80, 0.96]$ for the left amygdala, and 0.89 with 95%CI$=[0.83, 0.95]$ for the right amygdala. Intra-rater ICCs were also almost perfect: 0.91 with 95%CI$=[0.82, 0.96]$ for rater ZQZ while 0.95 with 95%CI$=[0.89, 0.97]$ for rater QZ. These results confirmed that the raters' manual tracings could be used as the gold standard or the reference for comparisons with automatic segmentation.

3.2. Measurement accuracy of automatically traced human amygdala

For the first-wave samples, one-way analysis of variance with repeated measures indicated significant differences in volumes of human amygdala across the three segmentation methods (left amygdala: $F = 925.70, p < 0.001$; right amygdala: $F = 725.60, p < 0.001$). Our post-hoc paired comparisons revealed that volumes obtained with *FreeSurfer* were significantly larger than those obtained by manual tracing (left amygdala: $t = 15.45, p < 0.001$; right amygdala: $t = 14.51, p < 0.001$), which in turn were larger than those obtained with

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.11.430883; this version posted February 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

235 *volBrain* segmentation (left amygdala: $t = 53.32, p < 0.001$; right amygdala: $t = 50.09, p < 0.001$). These findings (**Figure 2**) are reproducible for the second and third waves of samples (see Supplementary Figure S1 and S2).

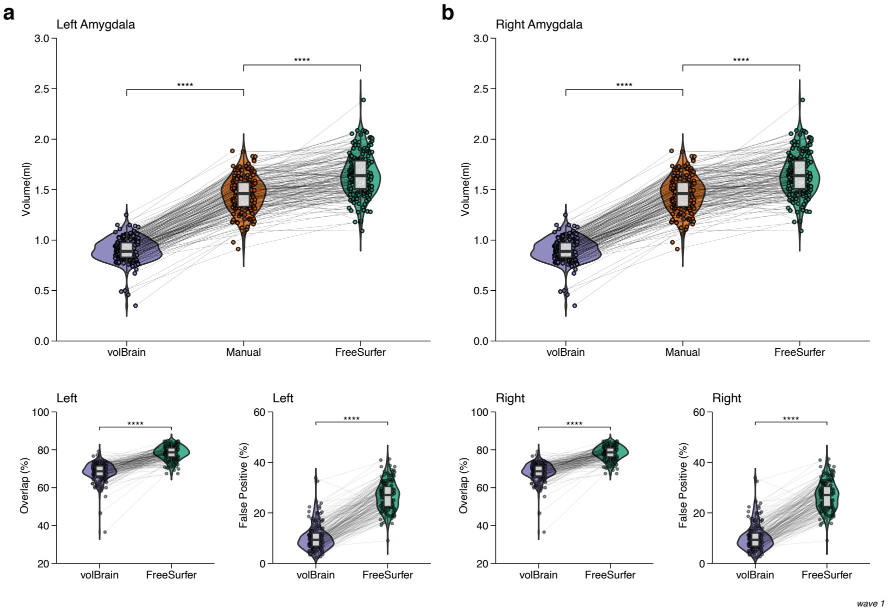

Figure 2: **Human amygdala volumes produced by the three tracing methods in CCNP wave 1 samples**. Brackets indicate differences between manual and automated methods on pairwise comparisons. In the left (a) and right (b) amygdala, spatial overlap and false positive rate for segmentation using *volBrain* and *FreeSurfer* compared to the manual "gold standard". Percentage of volume overlap between *volBrain* segmentation and manual tracing is lower than that of the overlap between *FreeSurfer* and manual tracing for the left and right amygdala. The false-positive rate was significantly lower for *volBrain* than for *FreeSurfer* segmentation for the left and right amygdala. $*p < 0.05$; $**p < 0.01$; $***p < 0.001$

As depicted in **Figure 2**, paired two-sample t-tests revealed that *FreeSurfer* had higher percentages of spatial overlap than *volBrain* with the manual tracing 240 for the first-wave data (left amygdala: $t = 22.16, p < 0.001$; right amygdala: $t = 26.09, p < 0.001$). The false-positive rates were significantly lower for *volBrain* than for *FreeSurfer* segmentation of the left amygdala ($t = 38.12, p < 0.001$) and the right amygdala ($t = 31.78, p < 0.001$). Both the left and right amyg-

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.11.430883; this version posted February 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

dala volumes obtained with the two automatic methods only showed moderate Pearson's correlations with those obtained by the manual tracing although statistically significant ($Rs = 0.58 - 0.62, ps < 0.001$; **Table 3**), but did not exceed 0.8. This indicated that the individual differences measured by the automatic methods were not fully consistent with those measured by the manual tracing method. These findings are reproducible for the second-wave and the third-wave samples (see Supplementary Figures S1 and S2).

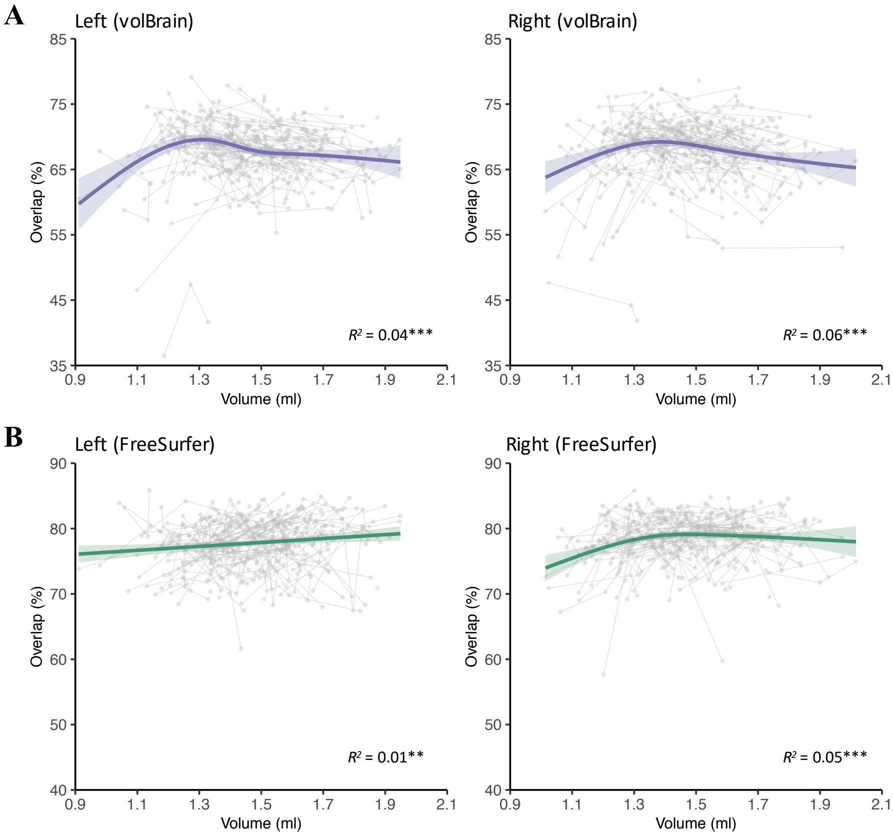

Figure 3: **Percentage of spatial overlap of automatic methods as function of the amygdala volume**. A: *volBrain*; B: *FreeSurfer*; $*p < 0.05$; $**p < 0.01$; $***p < 0.001$

The GAMM-based regression showed that the accuracy of *volBrain* segmentation (i.e., the percentage of spatial overlap with manual tracing) increased

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.11.430883; this version posted February 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

with the amygdala size before reaching a stable accuracy with a larger volume of the amygdala (**Figure 3A**). For *FreeSurfer*, as in **Figure 3B**, the segmentation accuracy displayed a linearly significant increase pattern with the left amygdala size while a two-stage (first increase and then remain stable) pattern with the right amygdala size. In all cases of the automatic segmentation methods, a smaller amygdala structure is associated with the worse segmentation accuracy, especially for those small amygdalae. These results indicated that neuroanatomical features can possibly affect the accuracy of automatic segmentation in a systematic way.

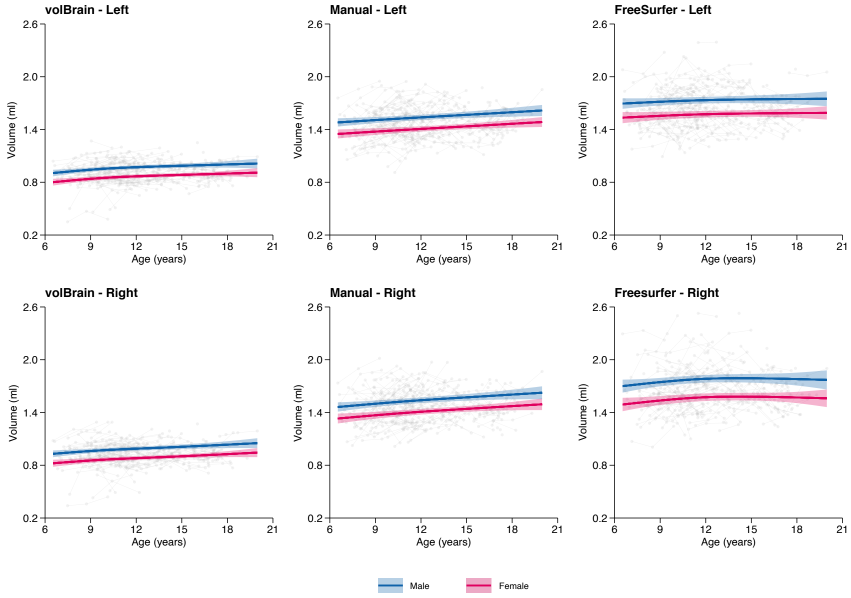

Figure 4: **Longitudinal developmental trajectories of volume for human amygdala traced by *volBrain*, manual and *FreeSurfer*.** The blue color indicates trajectories for boys while the red color for girl's trajectories. The trajectories are surrounded by shaded 95% confidence intervals. Note that boys and girls showed very similar developmental trajectories with no significant age-by-sex interactions, although boys had significantly larger amygdala volumes across the school ages (all $ps < 0.001$)

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.11.430883; this version posted February 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

3.3. Growth curves of human amygdala volume

The unified GAMM method, which includes age and interactions terms indicated that the age effects on the human amygdala were not consistent across the automatic tracing methods (**Table 4**). Specifically, these models reproduced the results of measurement accuracy for both *volBrain* and *FreeSurfer* reported in the previous section. The *volBrain* produced amygdala's age-related changes highly similar to that of manual tracing, i.e., no age×method interactions (all $ps > 0.05$). In contrast, the age-related amygdala changes showed discrepancies between *FreeSurfer* and the manual tracing. This led to a much lower explained variance using the GAMMs with *FreeSurfer* compared to that by the GAMMs with *volBrain* (left amygdala: 16% versus 77%; right amygdala: 16% versus 74%, respectively). Specifically, the age×method (*FreeSurfer* versus manual tracing) interaction was detectable (with a nearly marginal significance: $p = 0.193$) for the right amygdala but not for the left amygdala (**Table 4**). This indicated a trend toward a significant difference in the growth rate of the right amygdala volume between the *FreeSurfer* and manual segmentation.

The post-hoc method-wise GAMMs further revealed the growth patterns of the human amygdala as well as their sex differences. For all methods, the best models were determined by AIC as the second model, which included sex as a fixed effect (**Table 5**), indicating no need for an interaction between age and sex. This model revealed bigger amygdalae in boys than in girls, but their growth rates did not differ by sex. Specifically, as shown in **Figure 4**, the growth curve patterns were parallel in girls and boys for both manual and automatic tracing methods although boys demonstrated larger volumes of their amygdalae than girls across the entire school age range (6-18 years old). As the reference standard, the manual tracing method revealed that the human amygdala (both left and right) exhibited linear growth during the school-age years in both boys and girls. The *volBrain* tracing method yielded growth curves very similar to those established by the manual tracing method. *FreeSurfer* tracing method produced less linear and flatter curves and not statistically significant, except for a marginal significant growth curve in the right amygdala ($p = 0.066$). This

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.11.430883; this version posted February 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

growth curve had an inverted U shape: increasing during childhood and early adolescence, and then decreasing in late adolescence (the peak age around 14.18 years old).

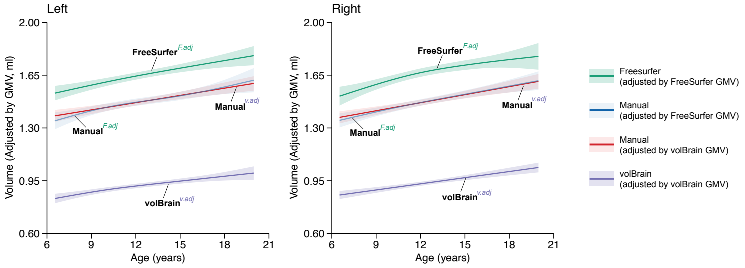

Figure 5: **Longitudinal developmental trajectories of volume for human amygdala adjusted by gray matter volume (GMV).** Amygdala volumes are adjusted by GMV in different ways: F.adj, adjusted by *FreeSurfer* produced GMV; v.adj, adjusted by *volBrain* produced GMV. The trajectories are surrounded by shaded 95% confidence intervals.

Correction for GMV abolished the significant sex differences across the entire age range (see details on the parameters for best-fitting models in Supplementary **Table S1**). This correction highly increased the reproducibility of the human amygdala growth curves across the three tracing methods (**Figure 5**). The growth patterns derived by the manual tracing method after controlling for either *volBrain*-estimated GMV or *FreeSurfer*-estimated GMV remain consistent with those without the GMV corrections. After the GMV-based correction, the growth patterns derived by the two automatic tracing methods showed almost identical shapes to those obtained using the manual tracing method (**Figure 5**). In contrast, correction for ICV reduced the reproducibility of the human amygdala growth curves across the three tracing methods (see Supplementary Figure S3). The growth patterns derived by the manual tracing method remained consistent with those without ICV corrections, but with much less statistical power: controlling for *volBrain*-estimated ICV led to much less significant age-related changes while controlling for *FreeSurfer*-estimated ICV led to no significant age-related changes and even sex-related differences. The significant positive linear

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.11.430883; this version posted February 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

association with age remained for *volBrain* traced amygdala with less statistical power, even after controlling for the ICV. However, correction for ICV changed the *FreeSurfer*-derived growth curves of the amygdala volume from nonlinear (not significant) to linear decrease (significant) patterns (Figure S3).

4. Discussion

This study evaluated the performance of segmentation of the amygdala using either the automatic software *volBrain* and *Freesurfer* compared to manual tracing in a longitudinal developmental sample. Importantly, we also explored how the segmentation differences could impact the growth curve modeling of the amygdala development. The findings indicated systematic differences in tracing performance across the three methods. *FreeSurfer* overestimated the volumes with more spatial overlapping with the manual tracing method, but had higher false-positive rates. In contrast, *volBrain* tended to underestimate the volumes with less spatial overlap with the manual tracing method, but had lower false-positive rates. We noted that the tracing accuracy of automatic methods was worse for smaller amygdalae. Furthermore, the growth curves of the amygdala volume estimated by different methods were inconsistent. These discrepancies indicated the importance to evaluate the segmentation performance across methods, especially in a developmental sample. To our knowledge, this study performed manual tracing of the amygdalae in the largest longitudinal sample to date and presented a systematic investigation of the method-wise variability of the growth curves of the human amygdala across school age. This variability of growth patterns could be normalized by adjusting for the total gray matter volume, but not adjusting for intracranial volume. The manual tracing method revealed linear growth of the amygdala in both boys and girls throughout the school-aged years, which is valuable to provide a growth norm for pediatric studies in the future.

The measurement accuracy of the amygdala volume varied across the automatic methods. *FreeSurfer* overestimated amygdala volumes (13–17%), and this

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.11.430883; this version posted February 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

overestimation has been observed in previous studies of the amygdala volume measurement by *Freesurfer* (Morey et al., 2009; Schoemaker et al., 2016). It is likely due to the greater variability in the definition of the amygdala boundary and liberal inclusion of voxels near this boundary (Morey et al., 2009; Schoemaker et al., 2016). The degree of overestimation observed here was greater than that reported for adults (7−9%) (Morey et al., 2009), but less than that reported for children aged 6–11 years (93–100%) (Schoemaker et al., 2016). Schoemaker et al.(2016) suggested it might be caused by using a standard brain template derived from 39 adults (mean age 38 ± 10 years). This may introduce greater bias when applied to a pediatric sample, in which amygdala sizes and shapes differ from adults. Another possible reason for this overestimation might be artifacts caused by more movements in children during imaging, causing a less precise differentiation and classification of amygdala structures by *FreeSurfer*. In contrast, *volBrain* underestimated the amygdala volume (35–37%) compared to the manual tracing. The underestimation may reflect the stringent inclusion of the amygdala during the segmentation by *volBrain*. This underestimation has been also observed previously, but is greater in children than for adults (3.38%) (Manjón and Coupé, 2016). *volBrain* segmentation uses manually labeled brain templates from 50 individuals with ages from 2 years old and 24-80 years old (Manjón and Coupé, 2016), which have no overlap with the age range of the current study (6-19 years old). The opposite directions of the estimation differences from the two automatic methods imply, other than using unmatched templates, the potential opposition in tracing algorithms between the two methods may exist. Further studies are clearly warranted to explore whether the use of age-matched templates could improve the accuracy of automatic amygdala segmentation (Dong et al., 2020). Given the systematic differences in the amygdala volume between automatic and manual segmentation, it calls for caution on interpreting the results of the absolute amygdala volumes obtained by using the automatic methods in children and adolescents.

FreeSurfer exhibited more spatial volume overlap than *volBrain* with the manual tracing method. The spatial overlap (76 − 79%) observed between

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.11.430883; this version posted February 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

FreeSurfer and the manual segmentation is consistent with the results reported by Morey et al. (2009). A higher overlap of *volBrain* was reported in a previous study (Manjón and Coupé, 2016), which is inconsistent with the observation in the present work. This could be related to the excessive underestimation of volume caused by the age-mismatched brain templates used by *volBrain* when segmenting amygdala for children and adolescents. In terms of spatial overlap, *FreeSurfer* outperformed *volBrain* for human pediatric amygdala segmentation. However, in terms of false-positive rate, *FreeSurfer* performed less than *volBrain*. The high false-positive rate of *FreeSurfer* could be an indication of its overestimation of the volume. A previous study suggested that it was due to excessive segmentation of brain structures in *FreeSurfer* by including structures and areas not part of the target structure (Næss-Schmidt et al., 2016). Although few studies have explored the performance of *volBrain* on human amygdala segmentation in terms of the false-positive rates, similar performance results have been shown for the automatic segmentation of the hippocampus and thalamic volume (Næss-Schmidt et al., 2016). According to inter-individual differences in segmented amygdala volumes, the two automatic methods only demonstrated moderate correlation with the manual segmentation, which are consistent with previous work (Morey et al., 2009; Grimm et al., 2015), implying its potential challenge for reliable measurements of their growth curves. Overall, the two automatic tracing methods have advantages and disadvantages for the assessment of amygdala volume. The complex amygdala structure adds difficulty to reliably and validly estimate its volume. It's a trade-off to choose which method should be used, requiring careful evaluation, and also demonstrates which facet of the automatic methods should be further improved in the future.

In this study, we found that the automatic segmentation performed worse in smaller amygdalae in developmental neuroimaging studies of school age children. The segmentation accuracy increased with amygdala volume, and then remained stable when the amygdala has reached a large enough size. Previous studies have found that smaller brain structures were associated with greater automatic segmentation errors (Schoemaker et al., 2016; Biffen et al., 2020;

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.11.430883; this version posted February 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Sánchez-Benavides et al., 2010). Our results are consistent with that neuroanatomical and geometric features could systematically influence the accuracy of their automatic segmentation. This bias is likely less problematic in adults, whose structures are commonly larger than in children. The human amygdala has been widely investigated in pediatric studies and associated with many developmental disorders such as autism (Mosconi et al., 2009; Schumann et al., 2009, 2004) and anxiety disorder (De Bellis et al., 2000; Hill et al., 2010; Milham et al., 2005). Our findings further highlighted the importance of improving the measurement accuracy of automatic segmentation for developing individuals. We argue that, in the current stage, manual tracing should be given priority for amygdala volume estimation in pediatric research. In the future, the technique development to eliminate the bias in automatic segmentation methods will be of great importance.

Although the statistical models indicated that the systematic differences in amygdala volume exhibited moderately marginal effects on growth curve modeling between the automatic and manual segmentation, our post-hoc growth chart analyses demonstrated remarkable discrepancies in age-related changes of the human amygdala across development. As the 'gold standard', manually traced amygdala volumes exhibited linear growth patterns without sex differences in growth rate. This is completely consistent with the patterns validated by the manual tracing method for amygdala growth from youth to adulthood in the macaque monkey (Schumann et al., 2019). Most previous studies of amygdala development in children and adolescents have been based on automatic segmentations (Wierenga et al., 2014; Goddings et al., 2014; Herting et al., 2018; Uematsu et al., 2012) while manual segmentation has been used in only two studies (Giedd et al., 1996; Merke et al., 2003). We noted that the developmental patterns of the amygdala have been inconsistent across these studies between automatic and manual methods. The growth patterns we detected by manual tracing were generally consistent with that by Giedd et al. (1996) and Merke et al. (2003) although they observed volume increases only in boys, but not in girls. In our study, the amygdala volumes grew in both boys and girls

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.11.430883; this version posted February 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

along highly similar trajectories. Such distinction may be an indication of the difference in scanning field strengths (3T versus 1.5T). Higher-resolution MRI enabled us to detect subtle changes in the human amygdala volume. Regarding automatic segmentation, previous studies generated amygdala growth curves with inverted U shapes from childhood to adolescence with peaks around 12–15 years old (Wierenga et al., 2014; Goddings et al., 2014; Herting et al., 2018; Uematsu et al., 2012). These were similar to our findings based on *FreeSurfer* segmentation, which showed a nonlinear trend of growth, especially for the right amygdala, with an inverted U-shaped trajectory (the volume peak at 14.18 years old). *volBrain* segmentation yielded growth curves very similar to that obtained by the manual tracing for the amygdala development. *volBrain* seems to have less error modeling growth curves than *FreeSurfer*. However, given limited studies using *volBrain* to investigate amygdala development in children and adolescents, it is hard to compare our results with others directly. It is interesting that the growth curves show similar shapes between the automatic and manual tracing methods when we adjusted amygdala volumes by the total gray matter volume rather than the intracranial volume. This may reflect the reduction in the bias related to the amygdala size in automatic segmentation as mentioned above correcting the amygdala volume. It works only for the gray matter volume and probably because that the measurement bias of the amygdala volume is more highly associated with the gray matter than the whole brain. These results suggest that controlling for the gray matter volume improved the accuracy of curve-fitting on the automatic segmentation of amygdala from childhood to adolescence.

Accurate delineation of the development of the human amygdala is fundamentally important to providing a reference to develop neuroimaging markers for various developmental disorders (DiMartino et al., 2014; Zuo, 2020; Holla et al., 2020). Our findings present a challenge for charting the growth of the human amygdala across school-age development considering that the growth curve modeling was highly dependent on the segmentation method. The methodological differences may contribute to the inconsistencies among previous findings

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.11.430883; this version posted February 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

regarding the patterns of amygdala development during childhood and adolescence (Wierenga et al., 2018; Uematsu et al., 2012; Albaugh et al., 2017; Herting et al., 2018). Given the inconsistency, we suggest that researchers working on the amygdala of children and adolescents should, 1) manually trace the amygdala if possible; 2) check and correct the automatic segmentation of the amygdala by a trained professional to improve the accuracy and save the effort if the manual segmentation not feasible; 3) use age-matched brain templates for automatic segmentation (Dong et al., 2020); 4) use high-resolution MRI protocol on scanning the amygdala; 5) adjust the amygdala volume by total gray matter volume when conducting statistical analysis; 6) cautiously compare and interpret previous findings using different segmentation methods than the study proposed. To facilitate the use of the growth curves we developed for human amygdala development at school age, we generated their charts and made them publicly open to the community (LINK TO BE ADDED after a final publication).

Our study has some limitations that should be noted. First, the age span of our sample might not be sufficient for examining the full range of development of the human amygdala from childhood, adolescence and into young adulthood. While the previous work in the macaque monkey revealed the linear pattern of amygdala growth from youth to adulthood (Schumann et al., 2019), further work would benefit from the extension of the age span into adulthood for direct growth assessments in human in future. Second, we did not investigate the measurement reliability across different versions of automatic segmentation tools, which has been shown remarkable influences on the brain segmentation (Gronenschild et al., 2012). This factor should be carefully evaluated by using different versions of these tools to model amygdala growth. Third, we only examined the overall volume measurement of the human amygdala. In the future, we will employ more local and shape measurements (Li et al., 2012; Roshchupkin et al., 2016) for investigating more details of human amygdala growth. To provide more efficient and accurate tracing of the pediatric amygdala, we also plan to develop an automatic algorithm based upon the manually traced samples using more advanced methods such as deep learning (Ataloglou et al., 2019).

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.11.430883; this version posted February 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

5. Conclusion

By manually tracing a large-sample pediatric MRI dataset from the accelerated longitudinal cohort, we charted the growth of human amygdala across school age. We identified measurement biases for the automatic amygdala segmentation methods and their impacts on modeling growth curves of the amygdala volumes from childhood to adolescence. There is considerable room for the methodological improvement of automatic tools to achieve more accurately tracing of the human amygdala during development. Our work provides not only a practical guideline for future studies on amygdala in children and adolescents but also its growth standard resources for translational and educational applications.

Acknowledgments

We thank all families participating in the Chinese Color Nest Program and all the support from schools and communities. This work was supported in part by the Key-Area Research and Development Program of Guangdong Province (2019B030335001), the Startup Funds for Leading Talents at Beijing Normal University, the Beijing Municipal Science and Tech Commission (Z161100002616023, Z171100000117012, Z181100001518003), the National Natural Science Foundation of China (81220108014), the China - Netherlands CAS-NWO Programme (153111KYSB20160020), the Guangxi BaGui Scholarship (201621) and National Basic Research (973) Program (2015CB351702). Finally, we would like to thank Tonya White, MD, PhD for her comments, suggestions, and for proofing the article for English language.

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.11.430883; this version posted February 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References

Akudjedu, T.N., Nabulsi, L., Makelyte, M., Scanlon, C., Hehir, S., Casey, H., Ambati, S., Kenney, J., O'Donoghue, S., McDermott, E., et al., 2018. A comparative study of segmentation techniques for the quantification of brain subcortical volume. Brain Imaging and Behavior 12, 1678–1695. doi:10.1007/s11682-018-9835-y.

Albaugh, M.D., Nguyen, T.V., Ducharme, S., Collins, D.L., Botteron, K.N., D'Alberto, N., Evans, A.C., Karama, S., Hudziak, J.J., Group, B.D.C., et al., 2017. Age-related volumetric change of limbic structures and subclinical anxious/depressed symptomatology in typically developing children and adolescents. Biological Psychology 124, 133–140. doi:10.1016/j.biopsycho.2017.02.002.

Alexander-Bloch, A.F., Reiss, P.T., Rapoport, J., McAdams, H., Giedd, J.N., Bullmore, E.T., Gogtay, N., 2014. Abnormal cortical growth in schizophrenia targets normative modules of synchronized development. Biological Psychiatry 76, 438–446. doi:10.1016/j.biopsych.2014.02.010.

Ataloglou, D., Dimou, A., Zarpalas, D., Daras, P., 2019. Fast and precise hippocampus segmentation through deep convolutional neural network ensembles and transfer learning. Neuroinformatics 17, 563–582. doi:10.1007/s12021-019-09417-y.

Barnea-Goraly, N., Frazier, T.W., Piacenza, L., Minshew, N.J., Keshavan, M.S., Reiss, A.L., Hardan, A.Y., 2014. A preliminary longitudinal volumetric mri study of amygdala and hippocampal volumes in autism. Progress in Neuro-Psychopharmacology and Biological Psychiatry 48, 124–128. doi:10.1016/j.pnpbp.2013.09.010.

Biffen, S.C., Warton, C.M., Dodge, N.C., Molteno, C.D., Jacobson, J.L., Jacobson, S.W., Meintjes, E.M., 2020. Validity of automated freesurfer segmentation compared to manual tracing in detecting prenatal alcohol exposure-

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.11.430883; this version posted February 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

related subcortical and corpus callosal alterations in 9-to 11-year-old children. NeuroImage: Clinical 28, 102368. doi:10.1016/j.nicl.2020.102368.

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 22, 1–12. doi:10.1093/cercor/bhr018.

Brown, C.J., Miller, S.P., Booth, B.G., Andrews, S., Chau, V., Poskitt, K.J., Hamarneh, G., 2014. Structural network analysis of brain development in young preterm neonates. Neuroimage 101, 667–680. doi:10.1016/j.neuroimage.2014.07.030.

Cardinal, R.N., Parkinson, J.A., Hall, J., Everitt, B.J., 2002. Emotion and motivation: the role of the amygdala, ventral striatum, and prefrontal cortex. Neuroscience Biobehavioral Reviews 26, 321–352. doi:10.1016/s0149-7634(02)00007-6.

Connor, D.F., 2004. Aggression and antisocial behavior in children and adolescents: Research and treatment. Guilford Press.

De Bellis, M.D., Casey, B., Dahl, R.E., Birmaher, B., Williamson, D.E., Thomas, K.M., Axelson, D.A., Frustaci, K., Boring, A.M., Hall, J., et al., 2000. A pilot study of amygdala volumes in pediatric generalized anxiety disorder. Biological Psychiatry 48, 51–57. doi:10.1016/s0006-3223(00)00835-0.

DiMartino, A., Fair, D., Kelly, C., Satterthwaite, T., Castellanos, F., Thomason, M., Craddock, R., Luna, B., Leventhal, B., Zuo, X.N., Milham, M., 2014. Unraveling the miswired connectome: A developmental perspective. Neuron 83, 1335–1353. doi:10.1016/j.neuron.2014.08.050.

Dong, H.M., Castellanos, F.X., Yang, N., Zhang, Z., Zhou, Q., He, Y., Zhang, L., Xu, T., Holmes, A., Thomas Yeo, B., et al., 2020. Charting brain growth

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.11.430883; this version posted February 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

in tandem with brain templates for schoolchildren. Science Bulletin 65, 1924–1934. doi:10.1016/j.scib.2020.07.027.

Fischl, B., Salat, D.H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., Van Der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., et al., 2002. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron 33, 341–355. doi:10.1016/s0896-6273(02)00569-x.

Ganzola, R., Maziade, M., Duchesne, S., 2014. Hippocampus and amygdala volumes in children and young adults at high-risk of schizophrenia: research synthesis. Schizophrenia Research 156, 76–86. doi:10.1016/j.schres.2014.03.030.

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. Journal of Comparative Neurology 366, 223–230. doi:10.1002/(SICI)1096-9861(19960304)366:2<223::AID-CNE3>3.0.CO;2-7.

Gilmore, J.H., Shi, F., Woolson, S.L., Knickmeyer, R.C., Short, S.J., Lin, W., Zhu, H., Hamer, R.M., Styner, M., Shen, D., 2012. Longitudinal development of cortical and subcortical gray matter from birth to 2 years. Cerebral Cortex 22, 2478–2485. doi:10.1093/cercor/bhr327.

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

Grimm, O., Pohlack, S., Cacciaglia, R., Winkelmann, T., Plichta, M.M., Demirakca, T., Flor, H., 2015. Amygdalar and hippocampal volume: a comparison between manual segmentation, freesurfer and vbm. Journal of Neuroscience Methods 253, 254–261. doi:10.1016/j.jneumeth.2015.05.024.

Gronenschild, E., Habets, P., Jacobs, H., Mengelers, R., Rozendaal, N., van Os, J., Marcelis, M., 2012. The effects of freesurfer version, workstation type,

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.11.430883; this version posted February 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

and macintosh operating system version on anatomical volume and cortical thickness measurements. PLoS One 7, e38234. doi:10.1371/journal.pone.0038234.

Harezlak, J., Ryan, L.M., Giedd, J.N., Lange, N., 2005. Individual and population penalized regression splines for accelerated longitudinal designs. Biometrics 61, 1037–1048. doi:10.1111/j.1541-0420.2005.00376.x.

He, Y., Xu, T., Zhang, W., Zuo, X.N., 2016. Lifespan anxiety is reflected in human amygdala cortical connectivity. Human Brain Mapping 37, 1178–1193. doi:10.1002/hbm.23094.

Herten, A., Konrad, K., Krinzinger, H., Seitz, J., von Polier, G.G., 2019. Accuracy and bias of automatic hippocampal segmentation in children and adolescents. Brain Structure and Function 224, 795–810. doi:10.1007/s00429-018-1802-2.

Herting, M.M., Gautam, P., Spielberg, J.M., Kan, E., Dahl, R.E., Sowell, E.R., 2014. The role of testosterone and estradiol in brain volume changes across adolescence: a longitudinal structural mri study. Human Brain Mapping 35, 5633–5645. doi:10.1002/hbm.22575.

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

Hill, S.Y., Tessner, K., Wang, S., Carter, H., McDermott, M., 2010. Temperament at 5 years of age predicts amygdala and orbitofrontal volume in the right hemisphere in adolescence. Psychiatry Research 182, 14–21. doi:10.1016/j.pscychresns.2009.11.006.

Holla, B., Seidlitz, J., Bethlehem, R., Schumann, G., 2020. Population norma-

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.11.430883; this version posted February 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

tive models of human brain growth across development. Science Bulletin 65, 1872–1873. doi:10.1016/j.scib.2020.08.040.

Landis, J.R., Koch, G.G., 1977. The measurement of observer agreement for categorical data. Biometrics 33, 159–174.

LeDoux, J., 1998. The emotional brain: The mysterious underpinnings of emotional life. Simon and Schuster.

Li, S., Wang, Y., Xu, P., Pu, F., Li, D., Fan, Y., Gong, G., Luo, Y., 2012. Surface morphology of amygdala is associated with trait anxiety. PLoS One 7, e47817. doi:10.1371/journal.pone.0047817.

Liu, S., Zhang, Z., Yang, N., Zhang, Q., Zhou, Q., Zuo, X.N., 2020. Cohort profile: Chinese color nest project. PsyArXiv doi:10.31234/osf.io/d8kpx.

Lyden, H., Gimbel, S.I., Del Piero, L., Tsai, A.B., Sachs, M.E., Kaplan, J.T., Margolin, G., Saxbe, D., 2016. Associations between family adversity and brain volume in adolescence: Manual vs. automated brain segmentation yields different results. Frontiers in Neuroscience 10, 398. doi:10.3389/fnins.2016.00398.

Manjón, J.V., Coupé, P., 2016. volbrain: an online mri brain volumetry system. Frontiers in Neuroinformatics 10, 30. doi:10.3389/fninf.2016.00030.

Merke, D.P., Fields, J.D., Keil, M.F., Vaituzis, A.C., Chrousos, G.P., Giedd, J.N., 2003. Children with classic congenital adrenal hyperplasia have decreased amygdala volume: potential prenatal and postnatal hormonal effects. The Journal of Clinical Endocrinology & Metabolism 88, 1760–1765. doi:10.1210/jc.2002-021730.

Milchenko, M., Marcus, D., 2013. Obscuring surface anatomy in volumetric imaging data. Neuroinformatics 11, 65–75. doi:10.1007/s12021-012-9160-3.

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.11.430883; this version posted February 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Milham, M.P., Nugent, A.C., Drevets, W.C., Dickstein, D.S., Leibenluft, E., Ernst, M., Charney, D., Pine, D.S., 2005. Selective reduction in amygdala volume in pediatric anxiety disorders: a voxel-based morphometry investigation. Biological Psychiatry 57, 961–966. doi:10.1016/j.biopsych.2005.01.038.

Mills, K.L., Tamnes, C.K., 2014. Methods and considerations for longitudinal structural brain imaging analysis across development. Developmental Cognitive Neuroscience 9, 172–190. doi:10.1016/j.dcn.2014.04.004.

Morey, R.A., Petty, C.M., Xu, Y., Hayes, J.P., Wagner II, H.R., Lewis, D.V., LaBar, K.S., Styner, M., McCarthy, G., 2009. A comparison of automated segmentation and manual tracing for quantifying hippocampal and amygdala volumes. NeuroImage 45, 855–866. doi:10.1016/j.neuroimage.2008.12.033.

Mosconi, M.W., Cody-Hazlett, H., Poe, M.D., Gerig, G., Gimpel-Smith, R., Piven, J., 2009. Longitudinal study of amygdala volume and joint attention in 2-to 4-year-old children with autism. Archives of General Psychiatry 66, 509–516. doi:10.1001/archgenpsychiatry.2009.19.

Næss-Schmidt, E., Tietze, A., Blicher, J.U., Petersen, M., Mikkelsen, I.K., Coupé, P., Manjón, J.V., Eskildsen, S.F., 2016. Automatic thalamus and hippocampus segmentation from mp2rage: comparison of publicly available methods and implications for dti quantification. International Journal of Computer Assisted Radiology and Surgery 11, 1979–1991. doi:10.1007/s11548-016-1433-0.

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, 3771–3790. doi:10.1002/hbm.23604.

Ortiz, J., Raine, A., 2004. Heart rate level and antisocial behavior in children and adolescents: A meta-analysis. Journal of the American Academy of Child & Adolescent Psychiatry 43, 154–162. doi:10.1097/00004583-200402000-00010.

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.11.430883; this version posted February 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Paus, T., Keshavan, M., Giedd, J., 2008. Why do many psychiatric disorders emerge during adolescence? Nature Reviews Neuroscience 9, 947–957. doi:10.1038/nrn2513.

Pessoa, L., 2010. Emotion and cognition and the amygdala: from "what is it?" to "what's to be done?". Neuropsychologia 48, 3416–3429. doi:10.1016/j.neuropsychologia.2010.06.038.

Pruessner, J.C., Li, L.M., Serles, W., Pruessner, M., Collins, D.L., Kabani, N., Lupien, S., Evans, A.C., 2000. Volumetry of hippocampus and amygdala with high-resolution mri and three-dimensional analysis software: minimizing the discrepancies between laboratories. Cerebral Cortex 10, 433–442. doi:10.1093/cercor/10.4.433.

R Core Team, 2014. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. URL: http://www.R-project.org/.

Redlich, R., Grotegerd, D., Opel, N., Kaufmann, C., Zwitserlood, P., Kugel, H., Heindel, W., Donges, U.S., Suslow, T., Arolt, V., et al., 2015. Are you gonna leave me? separation anxiety is associated with increased amygdala responsiveness and volume. Social Cognitive and Affective Neuroscience 10, 278–284. doi:10.1093/scan/nsu055.

Rice, K., Viscomi, B., Riggins, T., Redcay, E., 2014. Amygdala volume linked to individual differences in mental state inference in early childhood and adulthood. Developmental Cognitive Neuroscience 8, 153–163. doi:10.1016/j.dcn.2013.09.003.

Roshchupkin, G., Gutman, B., Vernooij, M., Jahanshad, N., Martin, N., Hofman, A., McMahon, K., Van Der Lee, S., Van Duijn, C., De Zubicaray, G., Uitterlinden, A., Wright, M., Niessen, W., Thompson, P., Ikram, M., Adams, H., 2016. Heritability of the shape of subcortical brain structures in the general population. Nature Communications 7, 13738. doi:10.1038/ncomms13738.

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.11.430883; this version posted February 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Sánchez-Benavides, G., Gómez-Ansón, B., Sainz, A., Vives, Y., Delfino, M., Peña-Casanova, J., 2010. Manual validation of freesurfer's automated hippocampal segmentation in normal aging, mild cognitive impairment, and alzheimer disease subjects. Psychiatry Research: Neuroimaging 181, 219–225. doi:10.1016/j.pscychresns.2009.10.011.

Sawiak, S., Shiba, Y., Oikonomidis, L., Windle, C., Santangelo, A.M., Grydeland, H., Cockcroft, G., Bullmore, E., Roberts, A., 2018. Trajectories and milestones of cortical and subcortical development of the marmoset brain from infancy to adulthood. Cerebral Cortex 28, 4440–4453. doi:10.1093/cercor/bhy256.

Scherf, K.S., Smyth, J.M., Delgado, M.R., 2013. The amygdala: an agent of change in adolescent neural networks. Hormones and Behavior 64, 298–313. doi:10.1016/j.yhbeh.2013.05.011.

Schmidt, M.F., Storrs, J.M., Freeman, K.B., Jack Jr, C.R., Turner, S.T., Griswold, M.E., Mosley Jr, T.H., 2018. A comparison of manual tracing and freesurfer for estimating hippocampal volume over the adult lifespan. Human Brain Mapping 39, 2500–2513. doi:10.1002/hbm.24017.

Schoemaker, D., Buss, C., Head, K., Sandman, C.A., Davis, E.P., Chakravarty, M.M., Gauthier, S., Pruessner, J.C., 2016. Hippocampus and amygdala volumes from magnetic resonance images in children: Assessing accuracy of freesurfer and fsl against manual segmentation. NeuroImage 129, 1–14. doi:10.1016/j.neuroimage.2016.01.038.

Schumann, C., Scott, J., Lee, A., Bauman, M., Amaral, D., 2019. Amygdala growth from youth to adulthood in the macaque monkey. Journal of Comparative Neurology 527, 3034–3045. doi:10.1002/cne.24728.

Schumann, C.M., Amaral, D.G., 2005. Stereological estimation of the number of neurons in the human amygdaloid complex. The Journal of Comparative Neurology 491, 320–329. doi:10.1002/cne.20704.

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.11.430883; this version posted February 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Schumann, C.M., Barnes, C.C., Lord, C., Courchesne, E., 2009. Amygdala enlargement in toddlers with autism related to severity of social and communication impairments. Biological Psychiatry 66, 942–949. doi:10.1016/j.biopsych.2009.07.007.

Schumann, C.M., Bauman, M.D., Amaral, D.G., 2011. Abnormal structure or function of the amygdala is a common component of neurodevelopmental disorders. Neuropsychologia 49, 745–759. doi:10.1016/j.neuropsychologia.2010.09.028.

Schumann, C.M., Hamstra, J., Goodlin-Jones, B.L., Lotspeich, L.J., Kwon, H., Buonocore, M.H., Lammers, C.R., Reiss, A.L., Amaral, D.G., 2004. The amygdala is enlarged in children but not adolescents with autism; the hippocampus is enlarged at all ages. The Journal of Neuroscience 24, 6392–6401. doi:10.1523/JNEUROSCI.1297-04.2004.

Silk, J.S., Vanderbilt-Adriance, E., Shaw, D.S., Forbes, E.E., Whalen, D.J., Ryan, N.D., Dahl, R.E., 2007. Resilience among children and adolescents at risk for depression: Mediation and moderation across social and neurobiological contexts. Development and Psychopathology 19, 841–865. doi:10.1017/S0954579407000417.

Tamnes, C.K., Walhovd, K.B., Dale, A.M., Østby, Y., Grydeland, H., Richardson, G., Westlye, L.T., Roddey, J.C., Hagler Jr, D.J., Due-Tønnessen, P., et al., 2013. Brain development and aging: overlapping and unique patterns of change. NeuroImage 68, 63–74. doi:10.1016/j.neuroimage.2012.11.039.

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

Van Petten, C., 2004. Relationship between hippocampal volume and memory ability in healthy individuals across the lifespan: review and meta-analysis.

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.11.430883; this version posted February 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Neuropsychologia 42, 1394–1413. doi:10.1016/j.neuropsychologia.2004.04.006.

Watson, C., Andermann, F., Gloor, P., Jones-Gotman, M., Peters, T., Evans, A., Olivier, A., Melanson, D., Leroux, G., 1992. Anatomic basis of amygdaloid and hippocampal volume measurement by magnetic resonance imaging. Neurology 42, 1743–1743. doi:10.1212/wnl.42.9.1743.

Wickham, H., 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. URL: https://ggplot2.tidyverse.org.

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. doi:10.1016/j.neuroimage.2014.03.072.

Wierenga, L.M., Bos, M.G., 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. doi:10.1016/j.psyneuen.2018.02.034.

Wood, S., 2017. Generalized Additive Models: An Introduction with R. 2 ed., Chapman and Hall/CRC.

Yushkevich, P.A., Piven, J., Hazlett, H.C., Smith, R.G., Ho, S., Gee, J.C., Gerig, G., 2006. User-guided 3d active contour segmentation of anatomical structures: significantly improved efficiency and reliability. NeuroImage 31, 1116–1128. doi:10.1016/j.neuroimage.2006.01.015.

Zuo, X.N., 2020. Editorial: Mapping the miswired connectome in autism spectrum disorder. Journal of the American Academy of Child and Adolescent Psychiatry 59, 348–349. doi:10.1016/j.jaac.2020.01.001.

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.11.430883; this version posted February 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 1 Sample characteristics for each wave

	Wave 1	Wave 2	Wave 3
n	183	149	95
n females/males	100/83	75/74	48/47
Age, mean (SD)	11.82 (3.14)	12.33 (2.87)	12.77 (2.61)
Age, range	6-17	7-18	9-19

Table 2 Intra- and inter-rater reliability for manual tracing human amygdala

Reliability Type	Rater	Hemisphere	ICC	95% Confidence Interval	
				Lower Bound	Upper Bound
Intra-rater reliability	Rater QZ	Left	0.95	0.89	0.97
		Right	0.94	0.86	0.97
	Rater ZQZ	Left	0.91	0.82	0.96
		Right	0.91	0.82	0.96
Inter-rater reliability	Between rater QZ and ZQZ	Left	0.88	0.80	0.96
		Right	0.89	0.83	0.95

Table 3 Comparison of segmented amygdala volumes between methods

Wave	Technique	Structure volume (mean cm^3 ±SD)		Comparison of techniques to manual tracing							
				%Volume difference ±SD		%Volume overlap ±SD		%False positive ±SD		Correlation	
		Left	Right	Left	Right	Left	Right	Left	Right	Left	Right
Wave 1	Manual	1.46±0.18	1.45±0.18								
	volBrain	0.90±0.12	0.92±0.13	37.61±7.70	35.98±7.78	68.16±4.92	68.21±4.86	10.28±5.17	11.65±5.95	0.61***	0.62***
	FreeSurfer	1.65±0.21	1.66±0.25	14.99±11.22	16.67±13.63	78.32±3.65	78.78±3.33	26.76±5.61	26.69±6.13	0.62***	0.59***
Wave 2	Manual	1.48±0.17	1.48±0.18								
	volBrain	0.92±0.11	0.93±0.12	37.56±6.42	36.73±6.91	67.78±4.06	68.23±4.67	11.17±4.81	11.05±5.27	0.63***	0.61***
	FreeSurfer	1.65±0.22	1.63±0.23	14.81±10.29	12.61±10.15	77.69±3.46	79.11±3.38	26.78±4.95	25.42±7.94	0.54***	0.66***
Wave 3	Manual	1.49±0.21	1.51±0.21								
	volBrain	0.92±0.15	0.94±0.16	37.84±8.05	37.44±8.09	67.54±4.87	67.34±5.33	10.81±5.13	11.59±5.13	0.67***	0.70***
	FreeSurfer	1.63±0.23	1.68±0.25	12.81±10.11	14.32±12.88	76.90±4.88	77.25±4.15	26.91±5.70	27.06±6.44	0.66***	0.59***

Note: Summary of automated segmentation performance, percent volume difference, percent volume overlap, percent false positive and Pearson's correlations between automated and manual segmentations. *$p < 0.05$; **$p < 0.01$;***$p < 0.001$

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.11.430883; this version posted February 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 4 GAMM statistical tests on developmental trajectories of bilateral amygdala volumes

Left Amygdala					**Right Amygdala**				
Manual vs. volBrain									
Intercept	**Estimate**	**SE**	***t***	***p* - value**	**Intercept**	**Estimate**	**SE**	***t***	***p* - value**
Method: volBrain	-0.40	0.00	-81.77	**< .001**	Method: volBrain	-0.38	0.00	-77.22	**< .001**
Slope	**edf**	**Ref.df**	**F**	**p - value**	**Slope**	**edf**	**Ref.df**	**F**	**p - value**
s (age)	3.57	3.57	4.00	**0.006**	s (age)	3.12	3.12	5.15	**0.001**
s (age): volBrain	1.00	1.00	0.80	0.371	s (age): volBrain	1.59	1.59	0.52	0.627
R^2 = 0.77					R^2 = 0.74				
Manual vs. FreeSurfer									
Intercept	**Estimate**	**SE**	***t***	***p* - value**	**Intercept**	**Estimate**	**SE**	***t***	***p* - value**
Method:FreeSurfer	0.12	0.01	19.85	**< .001**	Method: FreeSurfer	0.13	0.01	19.79	**< .001**
Slope	**edf**	**Ref.df**	**F**	**p - value**	**Slope**	**edf**	**Ref.df**	**F**	**p - value**
s (age)	3.43	3.43	1.89	0.142	s (age)	2.57	2.57	3.59	**0.026**
s (age): FreeSurfer	1.00	1.00	0.14	0.710	s (age): FreeSurfer	1.73	1.73	1.17	0.193
R^2 = 0.16					R^2 = 0.16				

Note: Smooth function (edf) as well as degrees of freedom (Ref.df) and *F*-statistic and associated *p*-value for age (***bold*** highlights *p*<.05).

Table 5

GAMM statistical tests on developmental trajectories of bilateral amygdala volumes for manual tracing, FreeSurfer and volBrain segmentation, respectively

	Hemisphere	**Best model fit**	**R^2(adjusted)**	**Sex**				**Age spline**			
				Estimate	**SE**	***t***	***p*-value**	**edf**	**Red.df**	***F***	***p*-value**
Manual	Left amygdala	age + sex	0.14	0.13	0.02	6.07	**0.000**	1.12	1.12	8.32	**0.003**
	Right amygdala	age + sex	0.12	0.12	0.02	5.86	**0.000**	1.38	1.38	8.47	**0.001**
volBrain	Left amygdala	age + sex	0.19	0.10	0.01	7.15	**0.000**	2.17	2.17	8.07	**0.000**
	Right amygdala	age + sex	0.17	0.11	0.02	6.84	**0.000**	1.83	1.83	7.34	**0.001**
FreeSurfer	Left amygdala	age + sex	0.15	0.16	0.03	6.08	**0.000**	1.48	1.48	0.69	0.317
	Right amygdala	age + sex	0.17	0.21	0.03	7.15	**0.000**	2.05	2.05	2.77	0.066

Note: Smooth function (edf) as well as degrees of freedom (Ref.df) and *F*-statistic and associated *p*-value (***bold*** highlights *p* < .05) for age.