Received: 13 March 2020 | Revised: 29 June 2020 | Accepted: 7 July 2020

DOI: 10.1002/hbm.25145

RESEARCH ARTICLE

WILEY

Adjusting for allometric scaling in ABIDE I challenges subcortical volume differences in autism spectrum disorder

Camille Michèle Williams[1] | Hugo Peyre[1,2,3] | Roberto Toro[4,5] | Anita Beggiato[3,5] | Franck Ramus[1]

[1]Laboratoire de Sciences Cognitives et Psycholinguistique, Département d'Etudes Cognitives, École Normale Supérieure, EHESS, CNRS, PSL University, Paris, France

[2]INSERM UMR 1141, Paris Diderot University, Paris, France

[3]Department of Child and Adolescent Psychiatry, Robert Debré Hospital, APHP, Paris, France

[4]U1284, Center for Research and Interdisciplinarity (CRI), INSERM, Paris, France

[5]Unité Mixte de Recherche 3571, Human Genetics and Cognitive Functions, Centre National de la Recherche Scientifique, Institut Pasteur, Paris, France

Correspondence
Camille Michèle Williams, LSCP, Département d'Etudes Cognitives, École Normale Supérieure, 29 rue d'Ulm, 75005 Paris, France.
Email: williams.m.camille@gmail.com

Funding information
ANR, Grant/Award Numbers: ANR-10-IDEX-0001-02 PSL, ANR-17-EURE-0017

Abstract

Inconsistencies across studies investigating subcortical correlates of autism spectrum disorder (ASD) may stem from small sample size, sample heterogeneity, and omitting or linearly adjusting for total brain volume (TBV). To properly adjust for TBV, brain allometry—the nonlinear scaling relationship between regional volumes and TBV—was considered when examining subcortical volumetric differences between typically developing (TD) and ASD individuals. Autism Brain Imaging Data Exchange I (ABIDE I; *N* = 654) data was analyzed with two methodological approaches: univariate linear mixed effects models and multivariate multiple group confirmatory factor analyses. Analyses were conducted on the entire sample and in subsamples based on age, sex, and full scale intelligence quotient (FSIQ). A similar ABIDE I study was replicated and the impact of different TBV adjustments on neuroanatomical group differences was investigated. No robust subcortical allometric or volumetric group differences were observed in the entire sample across methods. Exploratory analyses suggested that allometric scaling and volume group differences may exist in certain subgroups defined by age, sex, and/or FSIQ. The type of TBV adjustment influenced some reported volumetric and scaling group differences. This study supports the absence of robust volumetric differences between ASD and TD individuals in the investigated volumes when adjusting for brain allometry, expands the literature by finding no group difference in allometric scaling, and further suggests that differing TBV adjustments contribute to the variability of reported neuroanatomical differences in ASD.

KEYWORDS

allometry, autism spectrum disorder, subcortical volumes, total brain volume

1 | INTRODUCTION

Autism spectrum disorder (ASD) is a neurodevelopmental disorder characterized by early persistent deficits in social communication and interactions and restricted, repetitive patterns of behavior, interests, or activities. These symptoms impair social or occupational functioning and are not restricted to a developmental delay or intellectual deficiencies (American Psychiatric Association, 2013). Although prevalence estimates appear to vary by country and methods of assessments (Adak & Halder, 2017; Elsabbagh et al., 2012; Kim et al., 2011, 2014), ASD prevalence corresponds to 1 child in 59 in the United States (Christensen, 2018), an estimate which is consistent with the number

of diagnoses reported by parents on national surveys (Kogan et al., 2018). ASD is additionally 3–4 times more prevalent in boys than girls (Fombonne, 2009) and is accompanied by an intellectual disability (intelligence quotient, IQ < 70) in one third of patients, while 25% are in the borderline IQ range (from 70 to 85; Christensen et al., 2016). Although diverse genetic (Ramaswami & Geschwind, 2018) and environmental factors (Karimi, Kamali, Mousavi, & Karahmadi, 2017; Modabbernia, Velthorst, & Reichenberg, 2017), as well as their interactions (Abbott, Gumusoglu, Bittle, Beversdorf, & Stevens, 2018; Rijlaarsdam et al., 2017), are thought to contribute to the complex etiology of ASD, ASD's etiology remains poorly understood due to the indirect and small effects of known genetic and environmental factors (Crespi, 2016; Varcin, Alvares, Uljarević, & Whitehouse, 2017). Considering that neuroanatomical markers within the brain are more closely associated to symptoms of a condition, the present study investigated neuroanatomical differences in the Autism Brain Imaging Data Exchange I (ABIDE I, Di Martino et al., 2014; *N* = 1,112) between ASD and typically developing (TD) individuals in terms of their regional (i.e., subcortical and cortical) volumes and the scaling relationship between their regional volumes and total brain volume (TBV; sum of total gray matter (GM) and white matter (WM)).

While toddlers with ASD typically show early brain overgrowth and a larger head circumference (Courchesne, 2002; Hazlett et al., 2005), discrepancies in TBV between ASD and TD individuals after early childhood appear to be relatively subtle—1–2% greater for ASD (Haar, Berman, Behrmann, & Dinstein, 2016; Riddle, Cascio, & Woodward, 2017)—and to depend on age, intelligence, and sex (Redcay & Courchesne, 2005; Sacco, Gabriele, & Persico, 2015; Stanfield et al., 2008; Sussman et al., 2015). As only 20% of autistic individuals experience early brain overgrowth (Zwaigenbaum et al., 2014), global neuroanatomical variation in ASD may reflect a bias in the population norm rather than a trait of ASD (Raznahan et al., 2013). Researchers in turn propose that TBV differences be examined in light of a population's interindividual diversity (Lefebvre, Beggiato, Bourgeron, & Toro, 2015; Raznahan et al., 2014) and that regional volumes may be better proximal factor candidates underlying ASD (Ecker, 2017).

Numerous magnetic resonance imaging (MRI) studies report neuroanatomical differences between individuals with and without ASD in distributed subcortical and cortical regions thought to contribute to the development of ASD (Ha, Sohn, Kim, Sim, & Cheon, 2015; D. Yang, Beam, Pelphrey, Abdullahi, & Jou, 2016). For instance, the reduction in GM volume in the hippocampi of children with ASD may feed their episodic memory and social communication impairments (Duerden, Mak-Fan, Taylor, & Roberts, 2012; Gokcen, Bora, Erermis, Kesikci, & Aydin, 2009), and the decrease in GM volume in the superior temporal sulcus and middle temporal gyrus may reflect ASD patients' social-cognitive deficits (Greimel et al., 2013; Hyde, Samson, Evans, & Mottron, 2010; Wallace, Dankner, Kenworthy, Giedd, & Martin, 2010). However, reported neuroanatomical group differences in this literature are largely inconsistent and difficult to replicate (Lai, Lombardo, Chakrabarti, & Baron-Cohen, 2013; Lenroot & Yeung, 2013; Riddle et al., 2017; Zhang et al., 2018). For instance, van Rooij et al. (2017) reported that ASD subjects between 2 and 64 years old in the ENIGMA cohort (N_{ASD} = 1,571) had smaller amygdala, putamen, pallidum, and nucleus accumbens volumes—regions involved in sociomotivational and cognitive and motor systems (Shafritz, Bregman, Ikuta, & Szeszko, 2015). Yet, Bellani, Calderoni, Muratori, and Brambilla (2013) found that ASD toddlers and young children had larger amygdala volumes in their review of the role of the amygdala in autism and Haar et al. (2016) did not report any subcortical group differences in 9.5–24.9 years old subjects in the ABIDE I (N_{ASD} = 453).

Inconsistencies in regional volumetric differences between ASD and healthy individuals are thought to stem from small sample size and heterogeneity, specifically in age (Lin, Ni, Lai, Tseng, & Gau, 2015; Riddle et al., 2017; Zhang et al., 2018), sex (Lai et al., 2017; Lai, Lombardo, Auyeung, Chakrabarti, & Baron-Cohen, 2015; Mottron et al., 2015; Schaer, Kochalka, Padmanabhan, Supekar, & Menon, 2015; Zhang et al., 2018), and intelligence quotient (IQ) (Stanfield et al., 2008; Zhang et al., 2018). To address these limitations, meta-analyses and cohorts such as the ABIDE I are used to investigate the influence of sex, age, IQ, and TBV on brain volumes in ASD. But the conclusions of these studies tend to vary. For example, a meta-analysis examining total and regional brain volume variations across ages in ASD found that the size of the amygdala decreased with age compared to controls (Stanfield et al., 2008), while a recent ABIDE I study did not replicate this effect and instead reported a smaller putamen in ASD females from 17 to 27 years old (Zhang et al., 2018). Although differences in segmentation algorithms (Katuwal et al., 2016), correction for multiple comparisons, and age range selection may contribute to these discrepancies, studies examining regional neuroanatomical differences in sex (Fish et al., 2017; Jäncke, Mérillat, Liem, & Hänggi, 2015; Mankiw et al., 2017; Reardon et al., 2016, 2018; Sanchis-Segura et al., 2019) and ASD (Lefebvre et al., 2015) report that different methods of adjustment for individual differences in TBV yield varying regional volumetric group differences.

Classical methods of adjustment for TBV (e.g., proportion method [regional volume/TBV], covariate approach) can lead to over and/or underestimating volumetric group differences (Reardon et al., 2016; Sanchis-Segura et al., 2019) for two reasons. First, they omit the potential group variation in the relationship between a regional volume and TBV. Second, they assume that the relationship between TBV and each regional volume is linear when the relationship can be allometric—or nonlinear. If the relationship between TBV and a regional volume was linear, the exponent (α) of the power equation: $y = bx^{\alpha}$ would be equal to 1, indicating isometry. However, the exponent tends to be either hyperallometric (α > 1) or hypoallometric (α < 1) depending on the regional volume (Finlay, Darlington, & Nicastro, 2001; Mankiw et al., 2017; Reardon et al., 2016, 2018). When a region has a hypoallometric coefficient, the regional volume increases less than TBV as TBV increases and when the coefficient is hyperallometric, the regional volume increases more than TBV as TBV increases (e.g., Liu, Johnson, Long, Magnotta, & Paulsen, 2014; Mankiw et al., 2017).

Adjusting for differences in TBV with allometric scaling has two major implications for neuroanatomical research in ASD. First, if the allometric coefficient (α) differs between individuals with and without ASD, the relationship between regional and total volumes may serve as an

additional cerebral marker to differentiate between groups. Second, allometric scaling group differences aside, adjusting for the allometric relationship of each subcortical and cortical volume with total volume yields a more precise estimate of each regional volume, and in turn, provides a more accurate evaluation of volumetric group differences.

To this day, brain allometry in ASD has only been considered in two studies that examined corpus callosum and cerebellar differences between ASD and control individuals (Lefebvre et al., 2015; Traut et al., 2018, respectively). Thus, the primary goal of this study was to investigate allometric scaling and volumetric differences between ASD and control individuals in subcortical volumes while taking into account brain allometry. The second aim was to identify whether neuroanatomical group differences depend on sex, age and/or full scale intelligence quotient (FSIQ), variables previously reported to influence group differences in brain volumes in studies where brain allometry was omitted (Stanfield et al., 2008; Zhang et al., 2018). As the first study to investigate and adjust for allometric scaling differences in regional volumes between TD and ASD individuals, no a priori hypotheses were postulated.

Subcortical allometric and volumetric group differences were investigated in the ABIDE I, a cohort which consists of 539 individuals with ASD and 573 age and sex matched controls (Di Martino et al., 2014). A multiple group confirmatory factor analysis (MGCFA) - a multivariate statistical approach which advantageously tests for global group differences in brain allometry and considers the mutual relationship between regional brain structures (de Jong et al., 2017; Toro et al., 2009) - was conducted on the entire sample and subsamples based on age, sex, and FSIQ. Considering the recency of the MGCFA to examine volumetric group differences (de Mooij, Henson, Waldorp, & Kievit, 2018; Peyre et al., 2020), results from the MGCFA were compared to those obtained from linear mixed effects models (LMEMs). The present study additionally attempted to replicate the age and sex subcortical differences Zhang et al. (2018) found in the ABIDE I without adjusting for brain allometry and examined how different TBV adjustment techniques influence the replicated results.

2 | METHODS

2.1 | Participants

2.1.1 | Participant recruitment

Data was obtained from ABIDE I: a consortium with 1,112 existing resting-state functional MRI datasets with corresponding structural MRI and phenotypic information on 539 ASD patients and 573 age-matched controls between 6 to 64 years old from 17 different scanner sites (http://fcon_1000. projects.nitrc.org/indi/abide; Di Martino et al., 2014). ASD individuals were diagnosed by (a) clinical judgment only, or (b) using the Autism Diagnostic Observation Schedule (ADOS) and/or Autism Diagnostic Interview—Revised, or by (c) combining clinical judgment and the diagnostic instruments only. Di Martino et al. (2014) reported that 94% of the 17 sites using the ADOS and/or Autism Diagnostic Interview-Revised obtained research-reliable administrations and scorings. Data was anonymized and collected by studies approved by the regional Institutional Review Boards. Further details on participant recruitment and phenotypic and imaging data analyses are provided by Di Martino et al. (2014).

2.1.2 | Exclusion/inclusion criteria

As in Zhang et al.'s (2018) study that we aimed to replicate, individuals over 27 years old when scanned were excluded from the analyses since the age distribution was skewed to the left and subjects over 27 years old had a broad age distribution. Moreover, participants with an FSIQ or linearly estimated FSIQ by Lefebvre et al. (2015; details in their Supporting Information Intelligence Score) smaller than 70 and greater than 130 were excluded from the analyses to create a more homogenous sample.

Finally, participants were included based on the visual quality checks that were performed on Freesurfer v.5.1 segmentations (http://surfer.nmr.mgh.harvard.edu/). Considering that segmentation errors can yield large volume estimation errors, we decided to use the stringent image and segmentation quality criteria applied by Lefebvre et al. (2015) at the cost of a reduction in sample size. Since the same segmentation and quality check standard was not available for ABIDE II (Di Martino et al., 2017) or for cortical regions, cortical ABIDE I data and ABIDE II data were not included in this study.

Given that 3 controls and 33 ASD individuals exhibited differing comorbidities (e.g., Attention Hyperactivity Deficit Disorder, Obsessive Compulsive Disorder, phobias [e.g., spiders, darkness]) varying in severity, all individuals with comorbidities were maintained in the main analyses and were removed from the post hoc analyses performed without outlier values to consider their impact on reported group differences.

2.1.3 | Entire sample's descriptive statistics

The entire sample consisted of 654 participants (302 ASD and 352 controls) following the Freesurfer v.5.1 segmentation quality checks. The 302 ASD and 352 TD individuals differed in terms of sex ratio and FSIQ but not in handedness or age (Table 1). There were 218 ASD participants with a total ADOS score (M = 11.85, SD = 3.76).

2.1.4 | Subsamples' descriptive statistics

In addition to the analyses on the entire sample, we ran exploratory MGCFAs and LMEMs on four sufficiently powered subsamples (Mundfrom, Shaw, & Ke, 2005) to investigate age, sex, and FSIQ interactions which cannot be simultaneously investigated with the MGCFA. Girls (N_{ASD} = 37, $N_{Control}$ = 69) and adults from 20–27 years old (N_{ASD} = 46, $N_{Control}$ = 54) could not be examined in the subsample analyses due to the insufficient number of participants (N < 50; Mundfrom et al., 2005).

TABLE 1 Descriptive statistics of the entire sample in sex, age, handedness, ADOS, and FSIQ

	ASD (*N* = 302)	TD (*N* = 352)	Statistics	
Sex ratio (M/F)	265/37	283/69	χ^2 (1) = 6.47	*p* = .012
Age in years (*SD*)				
Mean	14.54 (4.47)	14.54 (4.55)	χ^2 (1) = 0.02	*p* = .877
Min	7.00	6.47		
Max	26.95	26.85		
Handedness (right/other)	170/30	223/24	χ^2 (1) = 2.43	*p* = .119
FSIQ				
Mean (*SD*)	102.18 (14.37)	109.75 (11.05)	χ^2 (1) = 48.58	*p* < .001
Min	71.00	73.00		
Max	129.11	129.00		
ADOS total				
Mean (*SD*)	11.85 (3.76)			
Min	2			
Max	21			
ADOS communication				
Mean (*SD*)	3.72 (1.49)			
Min	0			
Max	7			
ADOS social interactions				
Mean (*SD*)	8.15 (2.72)			
Min	2			
Max	14			

Note: *SD* in parentheses. Other: left, ambidextrous, or mixed. FSIQ: full scale intelligence quotient. ASD: autism spectrum disorder. TD: typically developing. M: male. F: female. Handedness was only provided for a subset of individuals. χ^2 from the Kruskal–Wallis test for FSIQ and Age. ADOS (autism diagnostic observation schedule) total corresponds to the sum of the ADOS communication and ADOS social interactions scores.

Subgroups were defined based on previous studies reporting age effects in ASD (e.g., Lin et al., 2015; Stanfield et al., 2008; Zhang et al., 2018): boys from 6 to under 12 years old (N_{ASD} = 87, $N_{Control}$ = 97) and boys from 12 to under 20 years old (N_{ASD} = 138, $N_{Control}$ = 141). Age did not differ between ASD and TD individuals in each group.

In light of the group differences in FSIQ and the association between FSIQ and brain volume (Maier et al., 2015; McDaniel, 2005), subsamples were additionally created based on the boys' median FSIQ, yielding boys with an FSIQ ≤ 107.8 (N_{ASD} = 165, $N_{Control}$ = 109) and boys with an FSIQ > 107.8 (N_{ASD} = 100, $N_{Control}$ = 174). The FSIQ of ASD boys with an FSIQ ≤ 107.8 (M = 93.40, SE = 0.77) was lower (than their control counterparts (*M* = 98.96, *SE* = 0.72; *ß* = −0.58, *SE* = 0.12, $p = 1.4 \times 10^{-06}$). FSIQ did not differ across boys with an FSIQ >107.8 (*ß* = −0.01, *SE* = 0.13, *p* = .952).

Further descriptive statistics on brain volumes, age, FSIQ score by group and sex are available for the entire sample and descriptive statistics on brain volumes, age, FSIQ score by group are reported for each subsample (Tables S1–S4) with the distribution of all brain volumes, age, and FSIQ of ASD and control participants in Figures S1–13 to compare to those from Zhang et al.'s (2018) study. Finally, since the sample size was predefined, power analyses were run a posteriori on significant LMEM main effects and interactions with the simr package (Green & MacLeod, 2016; Supporting Information 2: Power Analyses).

2.2 | Analyses

Analyses performed on R (R Core Team, 2019) were preregistered on OSF (https://osf.io/wun7s), except where indicated. The data and scripts that support the findings and figures of this study are openly available in "Subcortical-Allometry-in-ASD" at http://doi.org/10.5281/zenodo.3592884.

Since previous research either did not examine the scaling coefficients of some of the presently investigated volumes or potential hemispheric differences (de Jong et al., 2017; Liu et al., 2014; Reardon et al., 2016), we analyzed the scaling relationship between left and right regional volumes and TBV in ASD and TD individuals separately. Although not preregistered, we reported scaling coefficients with the 95% confidence interval and tested whether the scaling coefficients of each regional region with TBV differed from 1 with the car R

package (Fox & Weisberg, 2019). Analyses were conducted with and without age, sex, and age by sex interactions to examine the extent to which these additional variables influence the scaling coefficients. Additional analyses were also conducted without outliers, without individuals with comorbidities, and with medication use (medication vs. no medication) as a covariate to assess whether scaling coefficients were robust to these factors.

MGCFAs and LMEMs were conducted to address the study's primary goal to investigate allometric scaling and volumetric group differences and the study's secondary goal to examine whether allometric scaling and volumetric group differences depend on age, sex and or FSIQ. Briefly, a MGCFA is a multivariate approach that involves simultaneous confirmatory factor analyses (CFA) in two or more groups and tests measurement invariance across groups (i.e., that the same model of equations measures the same latent construct). In a CFA, observed variables (brain volumes) are used to measure an unobserved or latent construct (TBV). A CFA in turn corresponds to a system of equations that describes the relationship the observed variables and the latent construct they measure (TBV). MGCFAs advantageously measure group (i.e., ASD vs. Control) differences across all regional volumes simultaneously (i.e., global test) and in each regional volume (i.e., regional test), while adjusting for the mutual relationships between regional brain volumes. MGCFAs were run with the lavaan R package (Rosseel, 2012). We additionally conducted LMEMs, which measure group differences in each regional volume separately, with the lmerTest R package (Kuznetsova, Brockhoff, & Christensen, 2017) to (a) evaluate the consistency between MGCFA and LMEMs results; (b) adjust for variables that could not be included in MGCFAs; and (c) facilitate result comparisons with previous studies examining neuroanatomical differences in ASD that conducted LMEMs.

2.2.1 | Equations in the MGCFA

The observed variables estimating the latent construct (TBV) were the following 22 regional volumes (Table 2). All brain volumes were log10 transformed in order to take into account the power relationship between each regional volume and TBV within the general linear model framework. This yielded the linear allometric scaling Equation (1) where *i* corresponds to the investigated regional volume, α to the exponent of the power relationship (the allometric coefficient), and group to ASD or Control:

$$\begin{aligned}\text{Log10}\left(\text{Regional Volume}\right)_i^{(group)} &= \text{Intercept}_i^{(group)} \\ &+ \alpha_i^{(group)} \log 10\left(\text{TBV}\right)_i^{(group)} \\ &+ \text{Error}_i^{(group)} \qquad (1)\end{aligned}$$

2.2.2 | Testing for Allometric and Volumetric Group Differences: MGCFA Global and Regional Tests

First, TBV differences between groups identified by regressing TBV on group in the MGCFA models were adjusted for in the configural models of each sample. Second, configural invariance—whether the same observed variables explain the same latent construct across groups—was tested by establishing a configural model with correlated residuals between regional volumes that similarly fits both groups when the intercept and slope values of the allometric equations for each regional volume differs between ASD and Controls. Good model fit was determined using commonly used fit indices: the Tucker Lewis Index (TLI), the Comparative Fit Index (CFI), and the Root Mean Square Error of Approximation (RMSEA) with a TLI and CFI > .95 and a RMSEA ≤ .06 indicating good fit (Hu & Bentler, 1999). The TLI, CFI, and RMSEA robust fit indices were used to correct for non-normality and were obtained from the maximum likelihood robust estimator from the lavaan package (Rosseel, 2012). Although we preregistered that we would additionally use the standardized root mean square residual (SRMR), the SRMR was not used since the lavaan package (Rosseel, 2012) does not provide a robust SRMR.

Third, allometric scaling group differences were identified by testing for metric invariance (equality of slopes, or α_i coefficients from Equation 1) between groups. Fourth, volumetric group differences adjusted for allometric scaling were identified by testing for scalar invariance (equality of intercepts, or Intercepts from Equation 1) between groups.

Metric and scalar invariance were tested with a global test followed by a regional test in each volume if the global test was significant. In a global metric invariance test, regional volumes are simultaneously tested for allometric scaling (slope) group differences by comparing the configural model where the intercept and slope values differ between groups to a model where the slope values are constrained (the same) across groups. In a global scalar invariance test, regional volumes are simultaneously tested for volumetric (intercept) group differences by comparing the configural model where the intercept values differ between groups to a model where intercept values (and slope values, if global metric invariance is rejected) are constrained across groups. If the metric and/or scalar global invariance test is significant (χ^2 difference test *p*-value < .05) and robust TLI, CFI, and RMSEA indicate better model fit for configural model

TABLE 2 Investigated regional volumes

Total cerebral white matter	
Brain-stem	
Right ventral diencephalon	Left ventral diencephalon
Right cerebellum cortex	Left cerebellum cortex
Right accumbens	Left accumbens
Right amygdala	Left amygdala
Right caudate	Left caudate
Right hippocampus	Left hippocampus
Right pallidum	Left pallidum
Right putamen	Left putamen
Right thalamus proper	Left thalamus proper
Right hemisphere cortex	Left hemisphere cortex

(Chen, Curran, Bollen, Kirby, & Paxton, 2008; Chen, 2007; Hu & Bentler, 1999), groups respectively differ in allometric scaling (slopes) and/or volumes (intercept) in one or more of the regional volumes.

Regional volumes that differ in terms of allometric scaling and/or volume between groups are then identified by conducting a regional invariance test on each volume. In a regional invariance test, a model where the parameter (e.g., intercept, slope) values are constrained across groups is compared to a model where all but one of the parameter values of a regional volume are constrained across groups. We initially preregistered the following criteria for significant group differences in parameters in regional invariance tests based on the CFA literature. Groups would differ in parameter value if the χ^2 difference test was significant, if the p-value <0.05/22 (Bonferroni correction for the number of regional volumes in the MGCFA), and if the group difference in model fit exceeded Chen's (2007) cutoffs ($|\Delta CFI| > .005$ and robust $|\Delta RMSEA| \geq .010$) for unequal samples with under 300 participants. These more conservative cutoffs were chosen since our entire sample ASD group ($N = 302$) neared the 300 participants mark. However, these criteria lead to a mismatch between the MGCFA and LMEMs.

As a result, since there is currently no rule of thumb for the cutoff values of fit indices that should be employed in varying conditions (e.g., number of observed variables and factors), leaving researchers to choose fit criteria (Putnick & Bornstein, 2016), we considered there to be a neuroanatomical group difference even if Chen's (2007) criteria indicated invariance, if (a) the χ^2 difference test was significant, (b) the MGCFA effect size was greater than 0.2, (c) the corresponding LMEM (see next section) effect size was similar to the MGCFA effect size, and (d) the corresponding LMEM with False Discovery Rate correction (FDR) was significant with and without outliers.

2.2.3 | Testing for Allometric and Volumetric Group Differences: LMEMs Regional Tests

Corresponding LMEMs were run on each regional volume in the entire sample an in the exploratory subsamples, on specific volumes that significantly differed between ASD and control participants in terms of allometry (as indicated by the regional metric invariance test) and/or volume (as indicated by the regional scalar invariance test). TBV was calculated as the sum of gray and white matter. The same equation used in the MGCFA was entered in the LMEMs except that scanner site was also included in the Equation (2) as random intercept (slopes were not written out in Equation (2) for clarity).

$$\begin{aligned}\text{Log 10 (Regional Volume)}_i^{(group)} = {} & \text{Intercept}_i^{(group)} + \log 10\,(\text{TBV})_i^{(group)} \\ & + \text{Scanner Site}_i^{(group)} \\ & + \text{Error}_i^{(group)} \end{aligned} \quad (2)$$

An interaction of log10(TBV) by group indicated a significant difference in allometric scaling between groups while a significant group effect suggested a significant volumetric group difference. Additional sensitivity analyses were run excluding outliers and individuals with comorbidities, and with medication use as a covariate to ensure that findings were robust.

2.2.4 | Testing the dependence of allometric and volumetric group differences on age, sex, and FSIQ effects : LMEMs

To address the study's secondary goal, we ran LMEMs in the entire sample with Equation (3) where scanner site was included as a random intercept (slopes were not written out in Equation (3) for clarity).

$$\begin{aligned}\text{Log10 (Regional Volume)}_i^{(group)} = {} & \text{Intercept}_i^{(group)} + \log 10\,(\text{TBV})_i^{(group)} \\ & \times \text{Age}^{(group)} \times \text{Sex}^{(group)} \times \text{FSIQ}^{(group)} \\ & + \text{Scanner Site}_i^{(group)} + \text{Error}_i^{(group)} \end{aligned} \quad (3)$$

2.2.5 | Testing the dependence of allometric and volumetric group differences onage, sex, and FSIQ effects: Exploratory Subsample Analyses

We ran several exploratory MGCFAs and LMEMs on four boy subsamples to further compare MGCFA and LMEM results. In each boy subsample, we applied Equation (1) in the exploratory MGCFAs and Equation (2) in the exploratory LMEMs for each regional volume. Scanner site was always included as a random effect in the LMEMs. If groups in a subsample differed in terms of age and/or FSIQ, we ran LMEMs in the subsamples with age and/or FSIQ as interactive fixed effects (i.e., if groups differed in terms of age, the group × log10 (TBV) × age interaction was included). Total GM volume was also investigated in the latter LMEMs. All possible interactions were maintained in all LMEMs and only significant main effects and interactions were reported. LMEMs revealing significant neuroanatomical group differences were also conducted without outliers, individuals with comorbidities, and with medication use as a covariate to ensure that the findings were robust.

2.2.6 | Testing the relationship of ASD severity with neuroanatomical group differences

We additionally ran post hoc analyses, which were not preregistered, on brain regions exhibiting neuroanatomical group differences to examine whether the total ADOS score in ASD individuals was a significant predictor of the investigated volume and the allometric scaling relationship between that volume and TBV. These LMEMs were run on ASD individuals with the same fixed and random effects as the LMEMs revealing neuroanatomical group differences, except that the group fixed effect was replaced by the total ADOS score. LMEMs were additionally conducted without outliers and individuals with comorbidities and medication status was added as a covariate. Other

ASD scores available in ABIDE I were not employed due to the small number of individuals in each category (Supporting Information 3: MGCFA & LMEMs Assumptions).

2.2.7 | Testing the influence of TBV adjustment techniques on reported neuroanatomical differences

The additional LMEMs, which were conducted to contribute to the literature suggesting that neuroanatomical group differences vary depending on the applied TBV adjustment technique, were not preregistered. We examined the influence of four types of TBV adjustment techniques by comparing results from LMEMs (a) without TBV adjustment (e.g., Zhang et al., 2018), (b) with a linear adjustment considering TBV as a covariate (most common; Prigge et al., 2013; van Rooij et al., 2017; Zhang et al., 2018), (c) with linear adjustment while considering the interaction of TBV by Group (e.g., Lefebvre et al., 2015), and (d) with an allometric scaling adjustment by considering the interaction of log10(TBV) by Group (e.g., Lefebvre et al., 2015; Mankiw et al., 2017; Sanchis-Segura et al., 2019). In the no adjustment and linear adjustment LMEMs, all volumes were standardized raw volumes.

2.2.8 | Testing the influence of TBV adjustment techniques on our replication of Zhang et al.'s (2018) study

We sought to replicate the study by Zhang et al. (2018), who similarly examined the subcortical correlates of ASD with ABIDE I, to assess the reliability of their findings and examine the influence of different adjustment techniques on the findings that we successfully replicated.

Dependent variables in the LMEMs were Cortical WM Volume, Total GM Volume, the caudate, the amygdala, the hippocampus, the thalamus, the pallidum, the putamen, and the accumbens. Scanner site was always included as a random intercept and subject as a random intercept when hemisphere was included in the LMEMs. Fixed effects differed based on the type of adjustment technique, as described below. Dependent and independent variables were entered in the models as raw values except for age (linear and quadratic), which was centered (i.e., demeaned). Significant group main effects and interactions were reported and compared across LMEMs with varying adjustment techniques and p-values were not adjusted for multiple comparison as in Zhang et al.'s (2018) study.

LMEMs without TBV adjustment

Fixed effects were sex, age (quadratic or linear), hemisphere (except for Cerebral WM and Total GM volumes), and group (ASD and Controls). Two replication strategies were put into place: a "result replication" and a "methodological replication." In the "result replication," models were identified based on the significant interactions reported by Zhang et al. (2018) to compare effect sizes even if group interactions and main effects were not statistically significant in our sample. In the "methodological replication," LMEMs were identified using Zhang et al.'s (2018) technique of maintaining main effects in the model and sequentially removing nonsignificant interactions ($p > .05$) from the model.

LMEMs with linear TBV adjustment

As in Zhang et al.'s (2018) analyses, TBV was added as a covariate to the LMEMs identified with the "result replication" and "methodological replication" techniques. Although the authors commented on whether results were similar after covarying for TBV, they did not provide statistics (i.e., effect sizes, *p* values).

Comparing LMEMs with the lack of and differing TBV adjustment techniques

All brain volumes were log 10 transformed prior to scaling. LMEMs identified with the "result replication" and "methodological replication" techniques were run with the interaction of group by log 10 (TBV).

3 | RESULTS

3.1 | Testing for allometry

When examining the relationship of each regional volume with TBV, we found that cerebral WM was hyperallometric (slope > 1), cortical volume was isometric (slope = 1), and most subcortical regions were hypoallometric (slope < 1). After removing outliers and including medication as a fixed effect, all subcortical regions were hypoallometric except for the right amygdala in controls which remained isometric ($\alpha = .74$, CI low = 0.73, CI high = 1.02, $p = .094$; Tables S6–S12). The same results were found when adjusting for the interaction and effect of sex and age (Tables S10–S13).

Medication use was not significant across regional volumes for ASD and Control individuals.

3.2 | Allometric and volumetric group differences

In the MGCFA, the variance of TBV (the latent factor) was set to one to freely estimate the factor loading of the first regional volume. As a result, all *ß* reported from the MGCFA correspond to standardized effect sizes where the variance of regional volume and TBV are set to 1. Group differences in the MGCFA were estimated by calculating the group difference in standardized slopes and intercepts.

In the LMEMs, standardized estimates, *ß*, were reported by centering and scaling dependent and independent variables. Reported *p*-values are not corrected for multiple comparisons in the MGCFAs and were FDR corrected for the LMEMs. Statistics were reported for the age measure (age or age^2) with the largest effect size estimate. Correlated residuals slightly differed across samples (Table S14) and MGCFA model fit were acceptable (Table S15; Supporting Information 4: MGCFA Results).

Since factor levels were set to 1: Controls and 2: ASD in all LMEMs conducted in this study, a negative effect size in the MGCFA suggests that the slope or intercept is greater for Controls compared to ASD individuals, while a positive effect size suggests that the slope or intercept is smaller for Controls compared to ASD individuals.

3.2.1 | TBV group differences

TBV did not differ between individuals with and without ASD in the entire sample ($ß$ = 0.03, SE = 0.06, p = .431) in the MGCFA or LMEM ($ß$ = −0.01, SE = 0.07, p = .878).

3.2.2 | MGCFA

Global metric invariance was supported in the entire sample ($\Delta\chi 2_{[22]}$ = 17.4, p = 0.7395), suggesting that there was no allometric scaling (slope) difference between ASD and TD individuals.

Scalar invariance was supported in the entire sample ($\Delta\chi 2_{[22]}$ = 26.1, p = .2487), suggesting that there are no regional volumetric differences between ASD and TD individuals when adjusting for individual differences in TBV by taking into account allometric scaling.

3.2.3 | LMEMs

LMEMs were consistent with the MGCFA except for a group effect found in the right pallidum ($ß$ = 0.15, SE = 0.06, p = .028). This group effect was no longer significant ($ß$ = 0.07, SE = 0.06, p = .426) after removing outliers and individuals with comorbidities and controlling for medication use.

3.3 | Dependence of allometric scaling and/or volumetric group differences on age, sex, and/or FSIQ effects

Only significant results are reported (see Supporting Information 4: MGCFA Results).

3.3.1 | TBV group differences

MGCFAs only revealed a group difference in TBV for boys with an FSIQ ≤ median (107.8) where ASD individuals had a greater TBV than controls ($ß$ = 0.13, SE = 0.09, p = .023). Results from the LMEMs were consistent with those of the MGCFA. There was a significant interaction of sex by group by FSIQ in the entire sample ($ß$ = −0.52, SE = 0.21, p = .048), which was due to the greater TBV in ASD boys with an FSIQ ≤ median (M = 1,218.37 cm^3, SE = 0.76 cm^3) compared to their control counterparts (M = 1,181.89 cm^3, SE = 1.09 cm^3; $ß$ = 0.23, SE = 0.11, p = .027). Nonsignificant TBV group differences are provided as Supporting Information 4: MGCFA Results.

3.3.2 | Global allometric scaling group differences across subsamples

Global metric invariance was supported in boys from 6 to under 12 years old ($\Delta\chi 2_{[22]}$ = 22.9, p = .405) and in boys with an FSIQ ≤ 107.8 ($\Delta\chi 2_{[22]}$ = 20.0, p = .586), suggesting that there was no allometric scaling (slope) difference between ASD and TD individuals in these samples. However, global metric invariance was not supported in boys from 12 to under 20 years old ($\Delta\chi 2_{[22]}$ = 38.7, p = .015) and in boys with an FSIQ > 107.8 ($\Delta\chi 2_{[22]}$ = 38.5, p = .016). Thus, a regional metric invariance test was conducted on each regional volume of these subsamples to establish where allometric scaling discrepancies between groups lied.

3.3.3 | Regionalallometric scaling group differences in boys aged 12 to under 20 years old

Regional metric invariance χ^2 difference test indicated that the constrained configural model significantly differed from the constrained configural model with one freed slope, when the slope was freed for the brain stem ($ß$ = −0.06, $\Delta\chi 2_{(1)}$ = 11.7, $p = 6.13 \times 10^{-3}$), the left amygdala ($ß$ = 0.08, $\Delta\chi 2_{(1)}$ = 11.3, $p = 7.87 \times 10^{-4}$), and the right hippocampus ($ß$ = 0.22, $\Delta\chi^2_{(1)}$ = 58.2, $p = 2.34 \times 10^{-14}$). Although the robust CFI and robust RMSEA fit indices were invariant across models according to Chen's (2007) metric invariance cutoffs (|ΔCFI| > .005 and |ΔRMSEA| ≥ .010; Table S16.A), the present study's four step procedure for determining invariance suggested that the allometric scaling relationship between the right hippocampus and TBV differed between groups. ASD boys aged 12 to under 20 years old had a smaller allometric scaling coefficient ($ß$ = 0.52, SE = 0.01, $p = 2.21 \times 10^{-8}$) than their control counterparts ($ß$ = 0.74, SE = 0.01, $p = 2.40 \times 10^{-8}$).

Specifically, the χ^2 difference test indicated a group difference in the right hippocampus. The group difference ($ß$ = 0.22) was greater than 0.2. The allometric scaling group difference was replicated in the corresponding LMEM with (Figure S17) and without (Figure 1) outliers and the effect size of the MGCFA and LMEMs were similar (Table 3a,b).

To examine if the allometric scaling group difference reported the right hippocampus of boys from 12 to under 20 years old depended on FSIQ, we ran a LMEM on the right hippocampus a with TBV by group by FSIQ as fixed effects and scanner site as random intercept. Again, ASD individuals had a smaller allometric scaling coefficient compared to controls before and after outlier and comorbidity removal and medication use inclusion (Table 4a,b; a posteriori Power Analyses Table S17).

Post hoc analyses revealed that the total ADOS score did not significantly predict right hippocampal volume ($ß$ = 10.01, SE = 0.01,

p = .695) or the allometric scaling relationship (ß = 0.01, SE = 0.02, p = .695) of that volume in ASD individuals with an available total ADOS score (N = 81).

3.3.4 | Regional allometric scaling group differences in boys with an FSIQ > median (107.8)

The constrained configural model with one freed slope significantly differed from the constrained configural model, when the slope was freed for the left hippocampus (ß = 0.11, $\Delta\chi^2_{(1)}$ = 9.1, p = .003), the left caudate (ß = 0.04, $\Delta\chi2_{(1)}$ = 4.84, p = .028), the left accumbens (ß = 0.21, $\Delta\chi2_{(1)}$ = 6.2, p = .013), left pallidum (ß = 0.22, $\Delta\chi2_{(1)}$ = 7.8, p = .005), and the right ventral diencephalon (ß = 0.05, $\Delta\chi2_{(1)}$ = 5.9, p = .015). Since the covariance matrix of the residuals was not positive definite in group 2, we were not able to interpret the cortical white matter freed slope model. Although the robust CFI and RMSEA fit indices were invariant across models according to Chen's (2007) metric invariance cutoffs (|ΔCFI| > .005 & |ΔRMSEA| ≥ .010; Table S16.B, the present study's four step procedure for determining invariance supports that the allometric scaling relationship between the left accumbens and TBV differed between groups. ASD boys with an FSIQ > median had a smaller allometric scaling coefficient (ß = 0.32, SE = 0.01, $p = 2.13 \times 10^{-3}$) than their control counterparts (ß = 0.52, SE = 0.02, $p = 3.76 \times 10^{-3}$). Specifically, the χ^2 difference test indicated a group difference in the left accumbens. The group difference (ß = 0.21) was greater than 0.2. The allometric scaling group difference was replicated in the corresponding LMEM with (Figure S21) and without outliers (Figure 2) and the effect size of the MGCFA and LMEMs were similar (Table 5a,b; a posteriori Power Analyses Table S17).

Although the left pallidum had a group difference over 0.2 (ß > 0.2) in the MGCFA, which was replicated in the corresponding LMEM after FDR correction (ß = −0.26, SE = 0.10, p = .023), the allometric scaling group difference was no longer significant after including medication as a covariate and removing outliers and comorbidities (ß = 0.08, SE = 0.11, p = .607). Although cortical WM was not investigated in the MGCFA due to model convergence issues, allometric scaling did not differ between groups in the LMEM (ß = 0.00, SE = 0.04, p = .864).

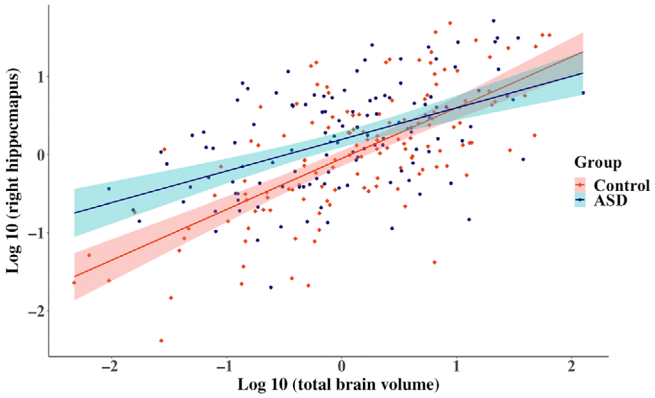

FIGURE 1 Relationship between the right hippocampus and total brain volume across groups after outlier and comorbidity removal ($N_{Control}$ = 137, N_{ASD} = 123) in boys from 12 to under 20 years old. ASD, autism spectrum disorder. 95% confidence region are given by group. Volumes were log transformed and scaled

TABLE 3 Right hippocampus LMEM results (a) and unstandardized allometric coefficients (b) for boys from 12 to 20 years old

(a)	Right hippocampus ~ group × log 10 (TBV)					
				No outliers and comorbidities		
	(N_{ASD} = 138 and N_C = 141)			(N_{ASD} = 123 and N_C = 137)		
	ß	*SE*	pFDR	ß	*SE*	pFDR
Medication				−0.03	0.12	0.855
Log 10(TBV)	0.83	0.07	1.76×10^{-24}	0.65	0.06	4.48×10^{-23}
Group	0.22	0.09	0.020	0.21	0.08	0.010
Group × log 10 (TBV)	−0.33	0.10	0.003	−0.25	0.08	0.010
(b)	**Right hippocampus ~ group × log 10 (TBV)**					
				No outliers and comorbidities		
	(N_{ASD} = 138 and N_C = 141)			(N_{ASD} = 123 and N_C = 137)		
Log 10(TBV)	**α**	***SE***	**pFDR**	**α**	***SE***	**pFDR**
ASD	0.66	0.09	0.000	0.53	0.09	0.000
Control	1.08	0.10	0.000	0.82	0.08	0.000

Note: ß corresponds to standardized beta for all main effects and interactions. α corresponds to the unstandardized allometric scaling coefficient of log 10 (TBV) with log 10 (left accumbens). C corresponds to controls, FDR to false discovery rate correction for multiple comparison, TBV to total brain volume and FSIQ to full scale intelligence quotient.

TABLE 4 Right hippocampus LMEM results (a) and unstandardized allometric coefficients (b) for boys from 12 to 20 years old

(a)	Right hippocampus ~ group × log 10(TBV) × FSIQ					
				No outliers and comorbidities		
	(N_{ASD} = 138 and N_C = 141)			(N_{ASD} = 119 and N_C = 133)		
	ß	*SE*	pFDR	*ß*	*SE*	pFDR
Medication				−0.04	0.11	0.739
Log 10(TBV)	0.86	0.08	9.34×10^{-22}	0.69	0.06	2.71×10^{-23}
Group	0.22	0.10	0.059	0.21	0.09	0.021
FSIQ	−0.09	0.08	0.503	0.03	0.06	0.739
Group × log 10(TBV)	−0.35	0.11	0.004	−0.22	0.09	0.037
Group × FSIQ	0.08	0.10	0.503	−0.08	0.07	0.537
FSIQ × log 10(TBV)	0.06	0.07	0.503	0.03	0.06	0.739
Group × FSIQ × log 10(TBV)	−0.02	0.10	0.816	0.06	0.08	0.739
(b)	**Right hippocampus ~ group × log 10 (TBV) × FSIQ**					
				No outliers and comorbidities		
	(N_{ASD} = 138 and N_C = 141)			(N_{ASD} = 119 and N_C = 133)		
Log 10(TBV)	**α**	***SE***	**pFDR**	**α**	***SE***	**pFDR**
ASD	0.67	0.09	0.000	0.62	0.09	0.000
Control	1.11	0.11	0.000	0.86	0.08	0.000

Note: *ß* corresponds to standardized beta for all main effects and interactions. α corresponds to the unstandardized allometric scaling coefficient of log 10 (TBV) with log 10(left accumbens). C corresponds to controls, FDR to false discovery rate correction for multiple comparison, TBV to Total Brain Volume and FSIQ to Full Scale Intelligence Quotient.

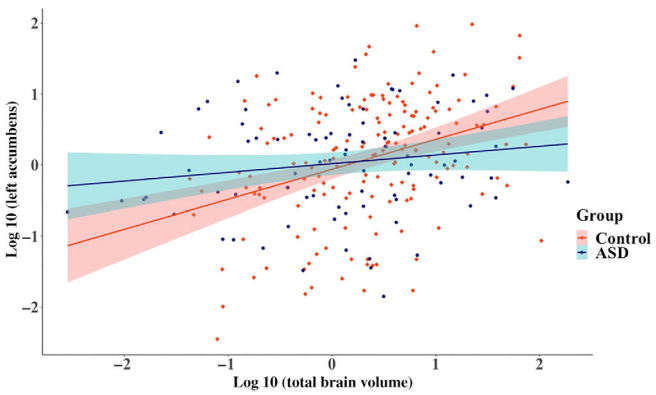

FIGURE 2 Relationship between the left accumbens and total brain volume across groups after outlier and comorbidity removal ($N_{Control}$ = 167, N_{ASD} = 85) in boys with a full scale intelligence quotient < median (107.8). ASD, autism spectrum disorder. 95% confidence region are given by group. Volumes were log transformed and scaled

To examine if the allometric scaling group difference reported in the left accumbens of boys with an FSIQ > median depended on age, we ran a LMEM on the left accumbens with TBV by group by Age (linear or quadratic) as fixed effects and scanner site as random intercept (Table 6a). Again, ASD individuals had a smaller allometric scaling coefficient compared to controls before and after outlier and comorbidity removal and medication use inclusion (Table 6a,b). Linear age and age effects were similar, although the effect sizes were slightly greater in the model with quadratic age (Table S18).

Post hoc analyses revealed that the total ADOS did not significantly predict left accumbens volume (*ß* = −0.01, *SE* = 0.02, *p* = .770) or the allometric scaling relationship (*ß* = −0.02, *SE* = 0.02, *p* = .770) of that volume in ASD individuals with an available total ADOS score (*N* = 59).

3.3.5 | Global volumetric group differences

Scalar invariance was supported in boys from 6 to under 12 years old ($\Delta\chi 2_{[22]}$ = 24.37, *p* = .328), in boys aged 12 to under 20 years old ($\Delta\chi 2_{[22]}$ = 30.6, *p* = .104), boys with an FSIQ ≤107.8 ($\Delta\chi 2_{[22]}$ = 28.0, *p* = .176), and in boys with an FSIQ >107.8 ($\Delta\chi 2_{[22]}$ = 27.5, *p* = .194), suggesting that there are no volumetric differences between ASD and TD individuals in these subsamples. However, unlike the exploratory MGCFA, LMEMs revealed a volumetric group difference in the right hippocampus of boys from 12 to under 20 years old (Table 4). Specifically, ASD individuals (*M* = 4,300.41 mm^3, *SD* = 501.38 mm^3) had a greater volume than their control counterparts (*M* = 4,201.98 mm^3, *SD* = 582.45 mm^3). A volumetric difference in the left caudate was

TABLE 5 Left accumbens LMEM results (a) and unstandardized allometric coefficients (b) for boys with a full scale intelligence quotient > median (107.8)

(a)	Log 10(left accumbens) ~ group × log 10(TBV)					
				Without outliers and comorbidities		
	(N_{ASD} = 100 and N_C = 174)			(N_{ASD} = 85 and N_C = 167)		
	ß	*SE*	pFDR	ß	*SE*	pFDR
Medication				0.11	0.15	0.46
Log 10(TBV)	0.54	0.08	2.90×10^{-11}	0.41	0.07	2.92×10^{-8}
Group	0.12	0.10	0.295	0.02	0.10	0.822
Group × log 10(TBV)	−0.32	0.10	0.003	−0.24	0.10	0.044
(b)	**Log 10(left accumbens) ~ group × log 10(TBV)**					
				Without outliers and comorbidities		
	(N_{ASD} = 100 and N_C = 174)			(N_{ASD} = 85 and N_C = 167)		
Log 10(TBV)	**α**	***SE***	**pFDR**	**α**	***SE***	**pFDR**
ASD (*N* = 100)	0.45	0.15	0.009	0.26	0.18	0.377
Control (*N* = 174)	1.21	0.18	0.000	0.97	0.16	0.000

Note: ß corresponds to standardized beta for all main effects and interactions. α corresponds to the unstandardized allometric scaling coefficient of log 10(TBV) with log 10(left accumbens). C corresponds to controls, FDR to false discovery rate correction for multiple comparison, and TBV to total brain volume.

TABLE 6 Left accumbens LMEM results with age (a) and unstandardized allometric coefficients (b) for boys with a full scale intelligence quotient > median (107.8)

(a)	Log 10(left accumbens) ~ group × log 10(TBV) × age					
				Without outliers and comorbidities		
	(N_{ASD} = 100 and N_C = 174)			(N_{ASD} = 79 and N_C = 162)		
	ß	*SE*	pFDR	ß	*SE*	pFDR
Medication				0.15	0.15	0.551
Log 10(TBV)	0.55	0.07	1.89×10^{-11}	0.48	0.07	6.63×10^{-10}
Group	0.07	0.10	0.618	0.03	0.10	0.804
Age	−0.14	0.07	0.12	−0.12	0.06	0.132
Log 10(TBV) × group	−0.28	0.10	0.024	−0.32	0.10	0.011
Log 10(TBV) × age	0.04	0.07	0.705	0.04	0.07	0.783
Group × age	0.04	0.10	0.705	−0.02	0.09	0.804
Group × log 10(TBV) × age	0.15	0.10	0.311	0.21	0.12	0.176
(b)	**Log 10(left accumbens) ~ group × log 10(TBV) × age**					
				Without outliers and comorbidities		
	(N_{ASD} = 100 and N_C = 174)			(N_{ASD} = 79 and N_C = 162)		
Log 10(TBV)	***α***	***SE***	**pFDR**	***α***	***SE***	**pFDR**
ASD (*N* = 100)	0.54	0.16	0.005	0.27	0.19	0.280
Control (*N* = 174)	1.26	0.18	0.000	1.10	0.16	0.000

Note: ß corresponds to standardized beta for all main effects and interactions. α corresponds to the unstandardized allometric scaling coefficient of log 10 (TBV) with log 10(left accumbens). C corresponds to controls, FDR to false discovery rate correction for multiple comparison, and TBV to total brain volume.

also found for boys with an FSIQ > the median before (ß = 0.30, *SE* = 0.12, *p* = .013) and after including medication as a covariate and removing comorbidities and outliers (ß = 0.30, *SE* = 0.10, *p* = .011; a posteriori Power Analyses Table S17). ASD individuals (*M* = 4,299.23 mm^3, *SD* = 451.55 mm^3) had a greater volume their control counterparts (*M* = 4,227.20 mm^3, *SD* = 646.23 mm^3).

3.4 | Comparing TBV adjustment techniques

3.4.1 | Present study

For the right hippocampus in the sample of boys from 12 to under 20 years old, the linear covariate and allometric scaling TBV adjustment technique revealed volumetric group differences that were absent when omitting TBV and adjusting for the linear interaction (Table 7a). However, TBV adjustment techniques yielded similar results for the left accumbens in the sample of boys with and FSIQ > median (107.8; Table 7b). Overall, these results suggest that the extent to which the type of adjustment technique influences reported volumetric and scaling group differences varies across GM volumes.

3.4.2 | Replication of Zhang et al. (2018)

In the LMEMs without TBV adjustment, we replicated the significant interaction of group by linear age by sex in the hippocampus. We were unable to replicate the remaining group differences reported by Zhang et al. (2018); Table 8). Although Zhang et al. (2018) reported that the interaction of group by linear age by sex in the hippocampus was no longer significant when covarying for TBV (no statistics were provided), the interaction remained minimally significant in our sample (Table 8).

When comparing results from LMEMs across all brain volumes with varying TBV adjustment techniques (Table 8 and Tables S19–S27), we found that the effect size of TBV was smaller when considering allometric scaling across all volumes. Although generally consistent, there were some differences in effect size and significance across TBV adjustment techniques. For instance, the interaction of group by linear age by sex in the hippocampus previously reported in LMEMs without TBV and with linear TBV adjustment was no longer significant when adjusting for TBV with allometric scaling (Table 8). Instead, the interaction of group by log10 (TBV) by sex was significant (*ß* = −0.40, *SE* = 0.20, *p* = .041, *d* = −0.08) when linear age was included in the model (Table S19). The interaction was no longer significant following FDR correction for multiple comparisons and was not significant when linear age was included in the model.

4 | DISCUSSION

The primary aim of this study was to investigate subcortical allometric scaling and volumetric differences between TD and ASD individuals from the ABIDE I, while adjusting for individual differences in TBV by taking into account brain allometry. The secondary goal of

TABLE 7 Variations in exploratory neuroanatomical group differences across TBV adjustment techniques without outliers and comorbidities

(a)				
Right hippocampus	**Effect**	***B***	***SE***	**pFDR**
No adjustment	Group	0.25	0.13	0.083
Group × FSIQ + medication				
Linear covariate adjustment	Group	0.27	0.10	0.027*
TBV + group × FSIQ + medication				
Linear interactive adjustment	Group	0.26	0.10	0.054
TBV × group × FSIQ + medication	Group by TBV	−0.21	0.10	0.103
Allometric interactive adjustment	Group	0.21	0.07	0.021*
Log 10(TBV) × group × FSIQ + medication	Group by log 10(TBV)	−0.22	0.09	0.037*
(b)				
Left accumbens	**Effect**	***B***	***SE***	**pFDR**
No adjustment	Group	−0.08	0.13	0.948
Group × age + medication				
Linear covariate adjustment	Group	−0.06	0.12	0.859
TBV + group × age + medication				
Linear interactive adjustment	Group	−0.11	0.12	0.546
TBV × group × age + medication	Group by TBV	−0.29	0.10	0.022*
Allometric interactive adjustment	Group	0.03	0.10	0.804
Log 10(TBV) × group × age + medication	Group by log10(TBV)	−0.32	0.10	0.011*

Note: *ß* corresponds to standardized beta, TBV to total brain volume, FSIQ to full scale intelligence quotient, and FDR to false discovery rate correction for multiple comparisons (*: significance at 0.05 after FDR correction).

TABLE 8 Replication of the significant group effects reported by Zhang et al. (2018) and TBV adjustment comparison

	Zhang model		Unadjusted model				TBV + model				Log 10(TBV) * group + model			
Volume, model, effect	***d***	***p***	***B***	***SE***	***d***	***p***	***B***	***SE***	***d***	***p***	***B***	***SE***	***d***	***p***
Accumbens														
Group + age + sex + hemi	0.14	0.040*	−2.48	9.97	−0.01	0.803	0.47	8.74	0.00	0.957	0.24	0.73	0.01	0.738
Hippocampus														
Group × age × sex + hemi	0.19	0.006*	58.09	22.34	0.10	0.010*	40.55	17.62	0.09	0.022*	0.25	0.28	0.04	0.361
Caudate														
Group × Age^2 × hemi + sex	0.16	0.022*	0.13	0.13	0.04	0.321	2.31	1.20	0.08	0.050	0.00	0.00	0.07	0.092
Putamen														
Group × age × hemi × sex	0.15	0.026*	19.01	18.98	0.04	0.317	19.01	18.98	0.04	0.317	0.13	0.18	0.03	0.479
Group × age × sex × hemi	0.20	0.008*	73.73	37.57	0.08	0.050	48.51	31.99	0.06	0.130	−0.07	0.30	−0.01	0.807
Group × Age × sex × hemi	0.14	0.038*	−9.89	20.42	−0.02	0.628	−9.89	20.42	−0.02	0.628	0.16	0.20	0.03	0.418

Note: Statistics are reported for the underlined effects of the model with or without adjustment for total brain volume (TBV). Zhang model corresponds to Zhang et al.'s (2018) linear mixed effects models. Hemi corresponds to hemisphere and d to Cohen's *d*. Groups (1: controls, 2: ASD). Hemi (1: left, 2: right). *P* values were not corrected for multiple comparisons (*$p < .05$). *B* are unstandardized estimates.

this article was to identify if subcortical allometric scaling and volumetric group differences depend on sex, age, and/or FSIQ. We compared the results of two statistical methods: MGCFAs, which advantageously test global and regional cerebral group differences while considering the mutual relationships between volumes, and LMEMs, to evaluate result consistency across methods and facilitate result comparison with the literature on volumetric differences in ASD. MGCFAs and LMEMs were generally consistent. While no robust neuroanatomical group differences were reported in the entire sample, exploratory MGCFAs and LMEMs revealed group differences in allometry for the right hippocampus in boys aged 12 to under 20 years old and the left accumbens in boys with an FSIQ > median. Our findings additionally further support that the type of adjustment techniques for TBV can influence reported volumetric and scaling group differences and suggest that allometric scaling should be considered to reduce the risk of reporting biased neuroanatomical group differences.

4.1 | Allometric scaling in ABIDE I

In line with previous studies (Liu et al., 2014; Reardon et al., 2016), the right and left cortex were isometric ($\alpha = 1$), cerebral white matter was hyperallometric ($\alpha > 1$), and subcortical volumes in TD and ASD individuals were hypoallometric ($\alpha < 1$). Yet, following outlier removal, the scaling coefficient of the right amygdala in controls were also isometric when sex and age effects were considered. While our findings could suggest that allometry is not a characteristic of all brain regions, allometry may still be present in subcortical subregions. A recent study examining surface area scaling coefficients reported different scaling coefficients within brain regions (e.g., both, negative and positive scaling in the amygdala (Reardon et al., 2018)). Brain allometry should in turn be investigated in cortical and subcortical subregions (not examined in the present study) since allometric scaling across these regions may serve as cerebral markers of ASD.

4.2 | Absence of general group differences in TBV

TBV only differed between ASD and TD individuals in the sample of boys with an FSIQ ≤107.8 and TBV was greater for individuals with ASD compared to their control counterparts. However, this difference in TBV between groups may be artifactual considering that IQ and brain size are differently correlated between ASD subjects ($r = 0.08$) and controls ($r = 0.31$). The study that provided the ABIDE I data simulated the impact of matching patient and control subjects by FSIQ and reported that FSIQ matching can bias TBV group differences by increasing the number of patient with a large TBV (Lefebvre et al., 2015). This biasing effect of IQ matching on TBV differences may also explain why one ABIDE I study reported a subtle TBV group differences (1–2%) after controlling for IQ in the matched but not the entire cohort (Riddle et al., 2017).

The lack of a general TBV difference is consistent with past ABIDE I studies examining volumetric group differences (Haar et al., 2016;Riddle et al., 2017; Zhang et al., 2018). While previous studies reported neuroanatomical differences between ASD and TD individuals across stages of development (Duerden et al., 2012; Stanfield et al., 2008), no group differences in TBV were found in children and adolescent boys in the present study. Since the studies that report a greater TBV in children with ASD suggest that TBV group differences are greater in early childhood and disappear in 10 year old children (Courchesne, Campbell, & Solso, 2011; Lange et al., 2015), children in the present sample may be too old to exhibit TBV group differences (First Quartile Age = 9.3 years old). As for adolescents, the majority of studies were either underpowered (Freitag et al., 2009; Hazlett et al., 2005) or grouped adolescent and children (Duerden et al., 2012), suggesting that their findings may be unreliable or biased by the younger children in their sample. The present study provides further evidence that enlarged TBV may not serve as a reliable biomarker of ASD after young childhood and may instead represent a bias in population norm (Raznahan et al., 2013).

4.3 | No regional group differences in the entire sample

ASD and TD individuals did not differ in terms volume or allometric scaling across presently investigated cortical and subcortical volumes. Although consistent with recent large-scale studies (Riddle et al., 2017; Zhang et al., 2018), this finding contrasts with the largest study to our knowledge (N_{ASD} = 1, 571 and $N_{Controls}$ = 1, 651; van Rooij et al., 2017) examining cortical and subcortical differences in ASD. The authors linearly adjusted for TBV (covariate approach) and reported volumetric group differences in the pallidum, putamen, amygdala, and nucleus accumbens (Cohen's d = −0.08 to −0.13). While the absence of such small volumetric group differences may stem from our smaller sample size, the covariate approach for TBV adjustment has also been shown to yield a higher rate of false positives (Liu et al., 2014; Sanchis-Segura et al., 2019), suggesting that these results should be replicated with an allometric scaling adjustment for TBV to be judged robust.

Volumetric group differences may lie in other cortical areas and WM volumes that make up the large-scale neurocognitive systems assumed to mediate ASD symptoms. Reported group differences in cortical regions (e.g., the insula; and prefrontal cortex (Duerden et al., 2012) thought to be involved in social cognition (Blakemore, 2008)) and in WM volumes (e.g., corpus callosum assumed to enable the integration of multiple sources of stimulation; Just, Cherkassky, Keller, Kana, & Minshew, 2007) must nonetheless be replicated in sufficiently powered studies (Di & Biswal, 2016; Haar et al., 2016; Lefebvre et al., 2015) that appropriately adjust for TBV (Liu et al., 2014; Sanchis-Segura et al., 2019) to be judged as robust neuroanatomical markers of ASD.

4.4 | No regional group differences depending on age, sex, and FSIQ in the entire sample

When considering age and sex effects and their interactions, we did not find group differences in allometric scaling or volume. This contrasts with several cross-sectional studies and meta-analyses on the neuroanatomical variations of ASD (Duerden et al., 2012; Greimel et al., 2013; D. Yang, Beam, et al., 2016; X. Yang et al., 2016) and the ABIDE I study we aimed to replicate (Zhang et al., 2018), which reported that ASD male adolescents and adults had smaller hippocampal volumes and that ASD female adolescents and adults had a smaller right putamen compared to their control counterparts.

These discrepancies with the literature may stem from (a) limited statistical power, (b) publication bias in favor of positive results, and (c) from the lack of correction for multiple comparison across a majority of studies, which increases the risk of false positives. Consistent with our entire sample analyses, the largest-scale ASD study to date addressing these limitations did not report age by sex or age by diagnostic effects when the linear effects of age were considered (van Rooij et al., 2017). However, based on previous findings that omitting brain allometry can lead to underestimating group differences (Mankiw et al., 2017; Reardon et al., 2016), we cannot rule out the presence of small age by sex or age by diagnostic effects on the investigated regional volumes since they would not be detectable with our current sample size.

Unlike the largest study to date on cerebral markers of ASD, which linearly corrected for TBV (covariate approach) and found volumetric sex differences in the thalamus, caudate, putamen, amygdala, and nucleus (van Rooij et al., 2017), no sex effects were found in our study. Although the absence of sex effects may be due to the few females (N = 106) in our sample, some significant sex effects may be false positives considering that the covariate TBV adjustment tends to overestimate volumetric sex differences (Reardon et al., 2016; Sanchis-Segura et al., 2019). In light of the numerous methodological discrepancies in the studies on the neuroanatomical group differences in ASD, more large-scale studies with an allometric scaling adjustment for TBV will be necessary to unbiasedly estimate cerebral differences in ASD across sexes.

4.5 | Exploratory regional group differences depending on age, sex, and FSIQ

Based on the LMEMs in the entire sample, allometric scaling and volumetric group differences did not depend on sex, age, and/or FSIQ. Exploratory analyses were nonetheless run on previously examined ASD subsamples (e.g., Lin et al., 2015; Maier et al., 2015) to compare our findings with previous studies and to further examine result consistency between MGCFAs and LMEMs. Exploratory MGCFAs and LMEMs revealed that allometric scaling coefficients were smaller for ASD individuals in the right hippocampus for boys aged 12 to under 20 years old and in the left accumbens for boys with an FSIQ < median.

This finding suggests that although both groups had hypoallometric scaling coefficients, indicating that these regional volumes grow at a slower rate than TBV, the regional volume increased less with TBV in ASD individuals compared to controls.

Hypoallometry (exponent < 1) in the right hippocampus and left accumbens regions of ASD boy subsamples did not covary with ASD severity, although previous studies suggest that the neuroanatomy of ASD is heterogeneous and varies with ASD severity (Bedford et al., 2020; H. Chen et al., 2019). One possibility is that the size of the present sample is not sufficient to detect a link between the allometric scaling coefficient and ASD severity. Another is that the severity of ASD may not correlate with allometry in the investigated subcortical structures.

While allometric scaling group differences were consistent across methods, LMEMs revealed a greater right hippocampal volume in boys from 12 to under 20 years old, which was not present in the MGCFA. Discrepancies in how parameter values are estimated in LMEMs and MGCFAs may explain inconsistencies across methods. For instance, unlike LMEMs, the MGCFA considers all regional volumes when predicting allometric scaling and volumetric group differences and takes into account correlated residuals when estimating parameter values. Yet, in light of the absence of allometric and volumetric group differences when examining the entire sample and the exploratory nature of these results, these results must be replicated in a larger sample to be judged as robust.

4.6 | MGCFAs and LMEMs: Methodology

Although MGCFAs and LMEMs generally provided similar results, MGCFAs may not be optimal to investigate neuroanatomical differences between groups in future studies for several reasons. First, although the MGCFA can simultaneously conduct global and regional tests, the MGCFA cannot simultaneously examine FSIQ, age, and sex effects, factors thought to influence brain anatomy (Duerden et al., 2012; Mankiw et al., 2017; Reardon et al., 2016; Sacco et al., 2015; van Rooij et al., 2017; Zhang et al., 2018). The present use of the MGCFA was nonetheless appropriate considering that the primary goal was to examine neuroanatomical group differences regardless of age, sex, and FSIQ. Second, the latent construct in the MGCFA cannot be equated with log10(TBV) which is typically employed to examine allometric scaling (Finlay et al., 2001), as in LMEMs. Instead, the latent construct reflects the shared variance between the observed variables: the log-transformed regional volumes. Third, numerous correlated residuals (overlap in variance between volumes that measure something else than TBV) were included in each MGCFA to reach appropriate fit and these correlated residuals slightly differed in the entire sample and each subsample. Since brain regions across and within hemispheres are highly interconnected, the measurement error of one volume correlates with the measurement error of another volume. However, it is unclear to what extent the correlated residuals established in the present model reflect general relationships between brain regions, and to what extent they reflect idiosyncratic properties of the present sample. Only a comparison with another large dataset would allow one to assess how generalizable this model is. Nonetheless, we emphasize that the model fit of all MGCFAs were similar across groups and the results between LMEMs and MGCFAs were overall consistent.

Fourth, while the number of participants included in each subsample was sufficient to provide a MGCFA factor solution in agreement with the population structure from which the sample was taken (Mundfrom et al., 2005), more MGCFA simulation studies and the development of packages to estimate MGCFA power are needed to establish the number of participants required to observe a specific group difference in parameter (slope or intercept) at 80% power. Finally, additional simulation studies are required to ensure that the current MGCFA thresholds employed in the literature reflect "real" rather than mathematical differences (Putnick & Bornstein, 2016). In the present study, Chen's (2007) cutoff values for fit indices to determine regional metric invariance between groups were too conservative to detect the small neuroanatomical group differences reported by the χ^2 difference tests and the LMEMs. One possibility is that Chen's (2007) cutoff values for fit indices may be appropriate for testing invariance between groups on medium effect sizes but not for testing the small differences in parameter values in the current article. Yet, this interpretation requires validation from future simulations studies conducted to identify appropriate fit indices cutoff values to detect small group differences in models with a varying number of factors and observed variables.

4.7 | Replication of Zhang et al. (2018)

Although the present article used similar inclusion/exclusion criteria and analyzed data from the same cohort with the same statistical method, the only robust reproducible result from the latest ABIDE I study was the significant interaction of group by age by sex in the hippocampus without TBV adjustment. Discrepancies between our findings and Zhang et al.'s (2018) can be explained by several factors. First, the small effect size and borderline *p*-values of the interactions reported by Zhang et al. (2018), suggest that these interactions without correcting for multiple comparisons were weak and perhaps not reliable. Second, while we selected similar age and FSIQ inclusion criteria, segmentation and quality checks differed between studies. While Zhang et al. (2018) used the FMRIB's Automated Segmentation Tool (FAST) from the FMRIB's Software Library (FSL), the present study used FreeSurfer. As a result, the mean of the investigated regional volumes and the distribution of participants across scanner sites for each volume somewhat differed between studies. Third, the current study's smaller sample size (N = 654) following segmentation and quality checks may explain why fewer significant interactions were found compared to Zhang et al. (2018; N = 859). Inconsistencies between the present and replicated study provide further evidence for the fragility of many reported results and emphasize the need to reexamine results and identify the reasons for failures in replication to improve future research (Button et al., 2013).

4.8 | Comparing TBV adjustment techniques

In line with previous findings (Barnes et al., 2010; Mankiw et al., 2017; Sanchis-Segura et al., 2019), neuroanatomical group differences depended on the techniques used to adjust for individual differences in TBV. Analyses from the replication revealed that the volumetric group differences in the hippocampus identified without TBV and with linear TBV adjustment were no longer significant when adjusting for TBV with allometric scaling (effect size was halved). The change in effect size suggests that omitting brain allometry can overestimate volumetric group differences even in the absence of TBV group differences. We additionally compared TBV adjustment techniques in the right hippocampus in boys from 12 to under 20 years old and in left accumbens for boys with an FSIQ over the median. Consistent with our findings from the replication, the type of adjustment technique and number of predictors included in the exploratory models influenced reported neuroanatomical differences in some volumes (i.e., in the right hippocampus and not the left accumbens). In light of our results and the literature reporting an effect of the TBV adjustment technique on reported neuroanatomical group differences (Liu et al., 2014; Sanchis-Segura et al., 2019), future studies should consider brain allometry to provide unbiased estimates of the cerebral markers of ASD.

4.9 | Limitations

The current article is limited in its capacity to study sex, age, and FSIQ effects on allometric scaling and volumetric group differences due to the insufficient number of girls, adults aged over 20, and individuals with an FSIQ < 70 in the ABIDE I sample. Further research on these populations is necessary to better understand ASD's etiology for numerous reasons. For instance, while some females exhibit symptoms similar to males at an early age, high functioning females are thought to have more efficient coping strategies than males, specifically in the social domain (Dworzynski, Ronald, Bolton, & Happé, 2012; Lai et al., 2015, 2017), which mask the severity of their ASD until later in adolescence or adulthood (Lai et al., 2015). By examining such individuals, who vary in ASD symptomatology, future studies may shed a light on the neuroanatomical markers related to specific ASD traits. In light of the cognitive changes associated with age-related brain volume alterations in the adult population (Scahill et al., 2003; Takao, Hayashi, & Ohtomo, 2012; Vinke et al., 2018), more adults in the young adult and older adult age ranges must be scanned and studied to accurately depict how age influences neuroanatomical differences reported in ASD. Finally, given that 1/3 of ASD individuals have an FSIQ < 70 (Christensen et al., 2016) and that they have a high within-group variability at the genomic level (Srivastava & Schwartz, 2014), neuroanatomical variations in these individuals likely depend on specific genetic components, warranting the investigation of cerebral differences with an imaging genetics approach in this population (Jack & Pelphrey, 2017).

Considering the heterogeneity of symptoms experienced by autistic individuals (Jack & Pelphrey, 2017; McIntyre et al., 2017; Rao, Beidel, & Murray, 2008) and the diverse genetic contributions to ASD (Ramaswami & Geschwind, 2018), multimodal approaches (e.g., imaging genetics) should be applied by future studies across ASD individuals to better characterize ASD's heterogeneity within etiologically dissimilar samples. Despite the need to inspect diverse ASD samples to fully understand the heterogeneity of ASD, investigating neuroanatomical group differences in ABIDE I is a primordial step to identifying robust allometric scaling and volumetric group differences in high functioning (FSIQ > 70) children and adolescents.

4.10 | Implications of studying allometry

Correcting for TBV with allometric scaling provides more accurate estimates of group differences in cerebral volumes and investigates whether allometric scaling could serve as a neuroanatomical marker for group differences in behavior and cognition. However, while numerous studies have proposed functional correlates for regional volume changes, the influence of allometric scaling on behavior and cognition remains unknown. For instance, while a reduced hippocampal volume has previously been linked to impaired episodic memory (Salmond et al., 2005; Williams, Goldstein, & Minshew, 2006) and a decrease in the left putamen volume to greater repetitive and stereotyped behavior in ASD (Cheung et al., 2010; Estes et al., 2011), atypical allometric scaling relationships may or may not translate to such cognitive and behavioral symptoms. Yet, prior to linking cerebral markers to variations in cognition and behavior, robust neuroanatomical markers that consider additional factors thought to influence cerebral diversity in the TD (e.g., sex, age) and in the ASD (e.g., minimally verbal subtype, IQ) population must be established.

There are numerous efforts aimed at identifying cerebral markers of ASD with brain imaging techniques for diagnosis purposes (e.g., Alvarez-Jimenez, Múnera-Garzón, Zuluaga, Velasco, & Romero, 2020; Kong et al., 2019; Nielsen et al., 2013). However, our study along with the increasing literature reporting the absence of (Haar et al., 2016; Lefebvre et al., 2015) or very subtle (van Rooij et al., 2017) volumetric group differences, suggest that previous group differences in subcortical volumes are potentially false positives or that individual regions may not constitute useful cerebral markers to employ for the diagnosis of ASD. This is consistent with the emerging literature that focuses on training classification algorithms with numerous brain regions and various methods, such as resting state functional MRI, to generate a more accurate diagnostic tool for ASD (Heinsfeld, Franco, Craddock, Buchweitz, & Meneguzzi, 2018; Plitt, Barnes, & Martin, 2015). Thus, from a clinical standpoint, our findings further support that the cerebral markers of ASD, which could be used for diagnosis, should not be restricted to a specific region in the brain.

Once robust cerebral markers that covary with cognitive abilities and disease severity are identified, mediation models can be conducted by future studies to uncover the diverse causal links of ASD that integrate genetic, environmental, cognitive, and behavioral information (Lai et al., 2013). These advances may enable the creation of more accurate ASD subgroups, offer more accurate diagnostic criteria, which are

increasingly being used to automate diagnosis (H. Chen et al., 2019; Nielsen et al., 2013), as well as facilitate person-centered treatment by providing insights on ASD's complex etiology.

5 | CONCLUSION

The primary goal of this study was to identify allometric scaling and volumetric differences between TD and ASD individuals when taking into account brain allometry. The second goal was to examine whether cerebral group differences depended on age, sex, and/or FSIQ. We analyzed data from ABIDE I using a common univariate approach, LMEMs, and a multivariate approach part of structural equation modeling, MGCFA. No robust allometric and volumetric group differences were observed in the entire sample, although exploratory analyses on subsamples based on age, sex, and FSIQ suggested that allometric scaling and volume may depend on age, sex, and/or FSIQ. While the LMEMs and the MGCFA were generally consistent, we propose that LMEMs may be more efficient to examine neuroanatomical group differences in light of the encountered methodological MGCFA constraints (e.g., no interaction effects, correlated residuals inclusion). Additional LMEM analyses with different TBV adjustment techniques revealed that the effect sizes and significance of cerebral differences between TD and ASD individuals differed across TBV adjustment techniques.

In addition to being the first study to examine allometric scaling and volumetric differences between ASD and TD individuals in the presently investigated volumes, the study adds to the literature by offering reference scaling coefficients for future studies in both ASD and TD individuals and by comparing two statistical methods: the MGCFA and LMEMs. Finally, in its difficulty to replicate a recent similar study, the article contributes to the literature on the replication crisis and, through its comparison of TBV adjustment techniques, supports the consideration of brain allometry to reduce reporting biased estimates of neuroanatomical group differences.

ACKNOWLEDGMENTS

This work received support under the program “Investissements d'Avenir” launched by the French Government and implemented by ANR with the references ANR-17-EURE-0017 and ANR-10-IDEX-0001-02 PSL.

CONFLICT OF INTERESTS

On behalf of all authors, the corresponding author states that there is no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in “Subcortical-Allometry-in-Autism” at http://doi.org/10.5281/zenodo.3592884

ORCID

Camille Michèle Williams https://orcid.org/0000-0002-1471-6566
Hugo Peyre https://orcid.org/0000-0001-8757-0783

REFERENCES

Abbott, P. W., Gumusoglu, S. B., Bittle, J., Beversdorf, D. Q., & Stevens, H. E. (2018). Prenatal stress and genetic risk: How prenatal stress interacts with genetics to alter risk for psychiatric illness. *Psychoneuroendocrinology*, *90*, 9–21. https://doi.org/10.1016/j.psyneuen.2018.01.019

Adak, B., & Halder, S. (2017). Systematic review on prevalence for autism spectrum disorder with respect to gender and socio-economic status. *Journal of Mental Disorders and Treatment*, *3*(1), 1–9. https://doi.org/10.4172/2471-271X.1000133

Alvarez-Jimenez, C., Múnera-Garzón, N., Zuluaga, M. A., Velasco, N. F., & Romero, E. (2020). Autism spectrum disorder characterization in children by capturing local-regional brain changes in MRI. *Medical Physics*, *47*(1), 119–131. https://doi.org/10.1002/mp.13901

American Psychiatric Association. (2013). Neurodevelopmental disorders. *Diagnostic and statistical manual of mental disorders (DSM-5)*, 1–0, 5th ed., Arlington, VA: American Psychiatric Association. https://doi.org/10.1176/appi.books.9780890425596.dsm01.

Barnes, J., Ridgway, G. R., Bartlett, J., Henley, S. M. D., Lehmann, M., Hobbs, N., … Fox, N. C. (2010). Head size, age and gender adjustment in MRI studies: A necessary nuisance? *NeuroImage*, *53*(4), 1244–1255. https://doi.org/10.1016/j.neuroimage.2010.06.025

Bedford, S. A., Park, M. T. M., Devenyi, G. A., Tullo, S., Germann, J., Patel, R., … Chakravarty, M. M. (2020). Large-scale analyses of the relationship between sex, age and intelligence quotient heterogeneity and cortical morphometry in autism spectrum disorder. *Molecular Psychiatry*, *25*(3), 614–628. https://doi.org/10.1038/s41380-019-0420-6

Bellani, M., Calderoni, S., Muratori, F., & Brambilla, P. (2013). Brain anatomy of autism spectrum disorders II. Focus on amygdala. *Epidemiology and Psychiatric Sciences*, *22*(4), 309–312. https://doi.org/10.1017/S2045796013000346

Blakemore, S.-J. (2008). The social brain in adolescence. *Nature Reviews Neuroscience*, *9*(4), 267–277. https://doi.org/10.1038/nrn2353

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*(5), 365–376. https://doi.org/10.1038/nrn3475

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, *14*(3), 464–504. https://doi.org/10.1080/10705510701301834.

Chen, F., Curran, P. J., Bollen, K. A., Kirby, J., & Paxton, P. (2008). An empirical evaluation of the use of fixed cutoff points in RMSEA test statistic in structural equation models. *Sociological Methods & Research*, *36*(4), 462–494. https://doi.org/10.1177/0049124108314720

Chen, H., Uddin, L. Q., Guo, X., Wang, J., Wang, R., Wang, X., … Chen, H. (2019). Parsing brain structural heterogeneity in males with autism spectrum disorder reveals distinct clinical subtypes. *Human Brain Mapping*, *40*(2), 628–637. https://doi.org/10.1002/hbm.24400

Cheung, C., Yu, K., Fung, G., Leung, M., Wong, C., Li, Q., … McAlonan, G. (2010). Autistic disorders and schizophrenia: Related or remote? An anatomical likelihood estimation. *PLoS One*, *5*(8), e12233. https://doi.org/10.1371/journal.pone.0012233

Christensen, D. L. (2018). Prevalence and characteristics of autism spectrum disorder among children aged 8 years—Autism and developmental disabilities monitoring network, 11 sites, United States, 2012. *MMWR. Surveillance Summaries*, *65*, 1–23. https://doi.org/10.15585/mmwr.ss6513a1

Christensen, D. L., Baio, J., Van Naarden Braun, K., Bilder, D., Charles, J., Constantino, J. N., … Centers for Disease Control and Prevention (CDC). (2016). Prevalence and characteristics of autism Spectrum disorder among children aged 8 years—Autism and developmental disabilities monitoring network, 11 sites, United States, 2012. *Morbidity and Mortality Weekly Report. Surveillance Summaries (Washington, D.C.: 2002)*, *65*(3), 1–23. https://doi.org/10.15585/mmwr.ss6503a1

Courchesne, E. (2002). Abnormal early brain development in autism. *Molecular Psychiatry*, *7*(S2), S21–S23. https://doi.org/10.1038/sj.mp.4001169

Courchesne, E., Campbell, K., & Solso, S. (2011). Brain growth across the life span in autism: Age-specific changes in anatomical pathology. *Brain Research*, *1380*, 138–145. https://doi.org/10.1016/j.brainres.2010.09.101

Crespi, B. J. (2016). Autism as a disorder of high intelligence. *Frontiers in Neuroscience*, *10*, 1–17. https://doi.org/10.3389/fnins.2016.00300

de Jong, L. W., Vidal, J.-S., Forsberg, L. E., Zijdenbos, A. P., Haight, T., Sigurdsson, S., … Launer, L. J. (2017). Allometric scaling of brain regions to intra-cranial volume: An epidemiological MRI study. *Human Brain Mapping*, *38*(1), 151–164. https://doi.org/10.1002/hbm.23351

de Mooij, S. M. M., Henson, R. N. A., Waldorp, L. J., & Kievit, R. A. (2018). Age differentiation within gray matter, white matter, and between memory and white matter in an adult life span cohort. *The Journal of Neuroscience*, *38*(25), 5826–5836. https://doi.org/10.1523/JNEUROSCI.1627-17.2018

Di Martino, A., Yan, C.-G., Li, Q., Denio, E., Castellanos, F. X., Alaerts, K., … Milham, M. P. (2014). The autism brain imaging data exchange: Towards a large-scale evaluation of the intrinsic brain architecture in autism. *Molecular Psychiatry*, *19*(6), 659–667. https://doi.org/10.1038/mp.2013.78

Di Martino, A., O'Connor, D., Chen, B., Alaerts, K., Anderson, J. S., Assaf, M., … Milham, M. P. (2017). Enhancing studies of the connectome in autism using the autism brain imaging data exchange II. *Scientific Data*, *4*, 170010. https://doi.org/10.1038/sdata.2017.10

Di, X., & Biswal, B. B. (2016). Similarly expanded bilateral temporal lobe volumes in female and male children with autism spectrum disorder. *Biological Psychiatry. Cognitive Neuroscience and Neuroimaging*, *1*(2), 178–185. https://doi.org/10.1016/j.bpsc.2015.11.006

Duerden, E. G., Mak-Fan, K. M., Taylor, M. J., & Roberts, S. W. (2012). Regional differences in grey and white matter in children and adults with autism spectrum disorders: An activation likelihood estimate (ALE) meta-analysis. *Autism Research*, *5*(1), 49–66. https://doi.org/10.1002/aur.235

Dworzynski, K., Ronald, A., Bolton, P., & Happé, F. (2012). How different are girls and boys above and below the diagnostic threshold for autism spectrum disorders? *Journal of the American Academy of Child & Adolescent Psychiatry*, *51*(8), 788–797. https://doi.org/10.1016/j.jaac.2012.05.018

Ecker, C. (2017). The neuroanatomy of autism spectrum disorder: An overview of structural neuroimaging findings and their translatability to the clinical setting. *Autism*, *21*(1), 18–28. https://doi.org/10.1177/1362361315627136

Elsabbagh, M., Divan, G., Koh, Y.-J., Kim, Y. S., Kauchali, S., Marcín, C., … Fombonne, E. (2012). Global prevalence of autism and other pervasive developmental disorders. *Autism Research*, *5*(3), 160–179. https://doi.org/10.1002/aur.239

Estes, A., Shaw, D. W. W., Sparks, B. F., Friedman, S., Giedd, J. N., Dawson, G., … Dager, S. R. (2011). Basal ganglia morphometry and repetitive behavior in young children with autism spectrum disorder. *Autism Research*, *4*(3), 212–220. https://doi.org/10.1002/aur.193

Finlay, B. L., Darlington, R. B., & Nicastro, N. (2001). Developmental structure in brain evolution. *The Behavioral and Brain Sciences*, *24*(2), 263–278; discussion 278–308. https://doi.org/10.1017/S0140525X01003958

Fish, A. M., Cachia, A., Fischer, C., Mankiw, C., Reardon, P. K., Clasen, L. S., … Raznahan, A. (2017). Influences of brain size, sex, and sex chromosome complement on the architecture of human cortical folding. *Cerebral Cortex (New York, N.Y.: 1991)*, *27*(12), 5557–5567. https://doi.org/10.1093/cercor/bhw323

Fombonne, E. (2009). Epidemiology of pervasive developmental disorders. *Pediatric Research*, *65*(6), 591–598. https://doi.org/10.1203/PDR.0b013e31819e7203

Fox, J., & Weisberg, S. (2019). *An {R} companion to applied regression* (3rd ed.). Thousand Oaks, CA: Sage. Retrieved from https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Freitag, C. M., Luders, E., Hulst, H. E., Narr, K. L., Thompson, P. M., Toga, A. W., … Konrad, C. (2009). Total brain volume and corpus callosum size in medication-Naïve adolescents and Young adults with autism spectrum disorder. *Biological Psychiatry*, *66*(4), 316–319. https://doi.org/10.1016/j.biopsych.2009.03.011

Gokcen, S., Bora, E., Erermis, S., Kesikci, H., & Aydin, C. (2009). Theory of mind and verbal working memory deficits in parents of autistic children. *Psychiatry Research*, *166*(1), 46–53. https://doi.org/10.1016/j.psychres.2007.11.016

Green, P., & MacLeod, C. J. (2016). Simr: An R package for power analysis of generalised linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*(4), 493–498. https://doi.org/10.1111/2041-210X.12504

Greimel, E., Nehrkorn, B., Schulte-Rüther, M., Fink, G. R., Nickl-Jockschat, T., Herpertz-Dahlmann, B., … Eickhoff, S. B. (2013). Changes in grey matter development in autism spectrum disorder. *Brain Structure and Function*, *218*(4), 929–942. https://doi.org/10.1007/s00429-012-0439-9

Ha, S., Sohn, I.-J., Kim, N., Sim, H. J., & Cheon, K.-A. (2015). Characteristics of brains in autism spectrum disorder: Structure, function and connectivity across the lifespan. *Experimental Neurobiology*, *24*(4), 273–284. https://doi.org/10.5607/en.2015.24.4.273

Haar, S., Berman, S., Behrmann, M., & Dinstein, I. (2016). Anatomical abnormalities in autism? *Cerebral Cortex*, *26*(4), 1440–1452. https://doi.org/10.1093/cercor/bhu242

Hazlett, H. C., Poe, M., Gerig, G., Smith, R. G., Provenzale, J., Ross, A., … Piven, J. (2005). Magnetic resonance imaging and head circumference study of brain size in autism: Birth through age 2 years. *Archives of General Psychiatry*, *62*(12), 1366–1376. https://doi.org/10.1001/archpsyc.62.12.1366

Heinsfeld, A. S., Franco, A. R., Craddock, R. C., Buchweitz, A., & Meneguzzi, F. (2018). Identification of autism spectrum disorder using deep learning and the ABIDE dataset. *NeuroImage: Clinical*, *17*, 16–23. https://doi.org/10.1016/j.nicl.2017.08.017

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Hyde, K. L., Samson, F., Evans, A. C., & Mottron, L. (2010). Neuroanatomical differences in brain areas implicated in perceptual and other core features of autism revealed by cortical thickness analysis and voxel-based morphometry. *Human Brain Mapping*, *31*(4), 556–566. https://doi.org/10.1002/hbm.20887

Jack, A., & Pelphrey, K. A. (2017). Annual research review: Understudied populations within the autism spectrum—Current trends and future directions in neuroimaging research. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *58*(4), 411–435. https://doi.org/10.1111/jcpp.12687

Jäncke, L., Mérillat, S., Liem, F., & Hänggi, J. (2015). Brain size, sex, and the aging brain. *Human Brain Mapping*, *36*(1), 150–169. https://doi.org/10.1002/hbm.22619

Just, M. A., Cherkassky, V. L., Keller, T. A., Kana, R. K., & Minshew, N. J. (2007). Functional and anatomical cortical underconnectivity in autism: Evidence from an fMRI study of an executive function task and corpus callosum morphometry. *Cerebral Cortex*, *17*(4), 951–961. https://doi.org/10.1093/cercor/bhl006

Karimi, P., Kamali, E., Mousavi, S. M., & Karahmadi, M. (2017). Environmental factors influencing the risk of autism. *Journal of Research in Medical Sciences: The Official Journal of Isfahan University of Medical Sciences*, *22*, 27. https://doi.org/10.4103/1735-1995.200272

Katuwal, G., Baum, S., Cahill, N., Dougherty, C., Evans, E., Evans, D., … Michael, A. (2016). Inter-method discrepancies in brain volume estimation may drive inconsistent findings in autism. *Frontiers in Neuroscience*, *10*(439), 1–16. https://doi.org/10.3389/fnins.2016.00439

Kim, Y. S., Fombonne, E., Koh, Y.-J., Kim, S.-J., Cheon, K.-A., & Leventhal, B. L. (2014). A comparison of DSM-IV pervasive developmental disorder and DSM-5 autism spectrum disorder prevalence in an epidemiologic sample. *Journal of the American Academy of Child &*

Adolescent Psychiatry, *53*(5), 500–508. https://doi.org/10.1016/j.jaac.2013.12.021

Kim, Y. S., Leventhal, B. L., Koh, Y.-J., Fombonne, E., Laska, E., Lim, E.-C., … Grinker, R. R. (2011). Prevalence of autism spectrum disorders in a total population sample. *The American Journal of Psychiatry*, *168*(9), 904–912. https://doi.org/10.1176/appi.ajp.2011.10101532

Kogan, M. D., Vladutiu, C. J., Schieve, L. A., Ghandour, R. M., Blumberg, S. J., Zablotsky, B., … Lu, M. C. (2018). The prevalence of parent-reported autism spectrum disorder among US children. *Pediatrics*, *142*(6), e20174161. https://doi.org/10.1542/peds.2017-4161

Kong, Y., Gao, J., Xu, Y., Pan, Y., Wang, J., & Liu, J. (2019). Classification of autism spectrum disorder by combining brain connectivity and deep neural network classifier. *Neurocomputing*, *324*, 63–68. https://doi.org/10.1016/j.neucom.2018.04.080

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Lai, M.-C., Lerch, J. P., Floris, D. L., Ruigrok, A. N. V., Pohl, A., Lombardo, M. V., & Baron-Cohen, S. (2017). Imaging sex/gender and autism in the brain: Etiological implications. *Journal of Neuroscience Research*, *95*(1–2), 380–397. https://doi.org/10.1002/jnr.23948

Lai, M.-C., Lombardo, M. V., Auyeung, B., Chakrabarti, B., & Baron-Cohen, S. (2015). Sex/gender differences and autism: Setting the scene for future research. *Journal of the American Academy of Child & Adolescent Psychiatry*, *54*(1), 11–24. https://doi.org/10.1016/j.jaac.2014.10.003

Lai, M.-C., Lombardo, M. V., Chakrabarti, B., & Baron-Cohen, S. (2013). Subgrouping the autism "spectrum": Reflections on DSM-5. *PLoS Biology*, *11*(4), e1001544. https://doi.org/10.1371/journal.pbio.1001544

Lange, N., Travers, B. G., Bigler, E. D., Prigge, M. B. D., Froehlich, A. L., Nielsen, J. A., … Lainhart, J. E. (2015). Longitudinal volumetric brain changes in autism Spectrum disorder ages 6–35 years. *Autism Research*, *8*(1), 82–93. https://doi.org/10.1002/aur.1427

Lefebvre, A., Beggiato, A., Bourgeron, T., & Toro, R. (2015). Neuroanatomical diversity of corpus callosum and brain volume in autism: Meta-analysis, analysis of the autism brain imaging data exchange project, and simulation. *Biological Psychiatry*, *78*(2), 126–134. https://doi.org/10.1016/j.biopsych.2015.02.010

Lenroot, R. K., & Yeung, P. K. (2013). Heterogeneity within autism spectrum disorders: What have we learned from neuroimaging studies? *Frontiers in Human Neuroscience*, *7*(733), 1–16. https://doi.org/10.3389/fnhum.2013.00733

Lin, H.-Y., Ni, H.-C., Lai, M.-C., Tseng, W.-Y. I., & Gau, S. S.-F. (2015). Regional brain volume differences between males with and without autism spectrum disorder are highly age-dependent. *Molecular Autism*, *6*, 29. https://doi.org/10.1186/s13229-015-0022-3

Liu, D., Johnson, H. J., Long, J. D., Magnotta, V. A., & Paulsen, J. S. (2014). The power-proportion method for intracranial volume correction in volumetric imaging analysis. *Frontiers in Neuroscience*, *8*(356), 1–16. https://doi.org/10.3389/fnins.2014.00356

Maier, S., Tebartz van Elst, L., Beier, D., Ebert, D., Fangmeier, T., Radtke, M., … Riedel, A. (2015). Increased hippocampal volumes in adults with high functioning autism spectrum disorder and an IQ > 100: A manual morphometric study. *Psychiatry Research: Neuroimaging*, *234*(1), 152–155. https://doi.org/10.1016/j.pscychresns.2015.08.002

Mankiw, C., Park, M. T. M., Reardon, P. K., Fish, A. M., Clasen, L. S., Greenstein, D., … Raznahan, A. (2017). Allometric analysis detects brain size-independent effects of sex and sex chromosome complement on human cerebellar organization. *The Journal of Neuroscience*, *37*(21), 5221–5231. https://doi.org/10.1523/JNEUROSCI.2158-16.2017

McDaniel, M. A. (2005). Big-brained people are smarter: A meta-analysis of the relationship between in vivo brain volume and intelligence. *Intelligence*, *33*(4), 337–346. https://doi.org/10.1016/j.intell.2004.11.005

McIntyre, N. S., Solari, E. J., Grimm, R. P., Lerro, E., Gonzales, E., & Mundy, P. C. (2017). A comprehensive examination of reading heterogeneity in students with high functioning autism: Distinct reading profiles and their relation to autism symptom severity. *Journal of Autism and Developmental Disorders*, *47*(4), 1086–1101. https://doi.org/10.1007/s10803-017-3029-0

Modabbernia, A., Velthorst, E., & Reichenberg, A. (2017). Environmental risk factors for autism: An evidence-based review of systematic reviews and meta-analyses. *Molecular Autism*, *8*(1), 13. https://doi.org/10.1186/s13229-017-0121-4

Mottron, L., Duret, P., Mueller, S., Moore, R. D., Forgeot d'Arc, B., Jacquemont, S., & Xiong, L. (2015). Sex differences in brain plasticity: A new hypothesis for sex ratio bias in autism. *Molecular Autism*, *6*(1), 33. https://doi.org/10.1186/s13229-015-0024-1

Mundfrom, D. J., Shaw, D. G., & Ke, T. L. (2005). Minimum sample size recommendations for conducting factor analyses. *International Journal of Testing*, *5*(2), 159–168. https://doi.org/10.1207/s15327574ijt0502_4

Nielsen, J. A., Zielinski, B. A., Fletcher, P. T., Alexander, A. L., Lange, N., Bigler, E. D., … Anderson, J. S. (2013). Multisite functional connectivity MRI classification of autism: ABIDE results. *Frontiers in Human Neuroscience*, *7*(599), 1–12. https://doi.org/10.3389/fnhum.2013.00599

Peyre, H., Mohanpuria, N., Jednoróg, K., Heim, S., Grande, M., van Ermingen-Marbach, M., … Ramus, F. (2020). Neuroanatomy of dyslexia: An allometric approach. *European Journal of Neuroscience*, 1–15. https://doi.org/10.1111/ejn.14690

Plitt, M., Barnes, K. A., & Martin, A. (2015). Functional connectivity classification of autism identifies highly predictive brain features but falls short of biomarker standards. *NeuroImage: Clinical*, *7*, 359–366. https://doi.org/10.1016/j.nicl.2014.12.013

Prigge, M. B. D., Lange, N., Bigler, E. D., Merkley, T. L., Neeley, E. S., Abildskov, T. J., … Lainhart, J. E. (2013). Corpus callosum area in children and adults with autism. *Research in Autism Spectrum Disorders*, *7*(2), 221–234. https://doi.org/10.1016/j.rasd.2012.09.007

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review: DR*, *41*, 71–90. https://doi.org/10.1016/j.dr.2016.06.004

R Core Team. (2019). *R: A language and environment for statistical computing*, Vienna, Austria: R Foundation for Statistical Computing Retrieved from https://www.R-project.org/

Ramaswami, G., & Geschwind, D. H. (2018). Chapter 21—Genetics of autism spectrum disorder. In D. H. Geschwind, H. L. Paulson, & C. Klein (Eds.), *Handbook of clinical neurology* (Vol. 147, pp. 321–329). Amsterdam, The Netherlands: Elsevier. https://doi.org/10.1016/B978-0-444-63233-3.00021-X

Rao, P. A., Beidel, D. C., & Murray, M. J. (2008). Social skills interventions for children with Asperger's syndrome or high-functioning autism: A review and recommendations. *Journal of Autism and Developmental Disorders*, *38*(2), 353–361. https://doi.org/10.1007/s10803-007-0402-4

Raznahan, A., Lee, N. R., Greenstein, D., Wallace, G. L., Blumenthal, J. D., Clasen, L. S., & Giedd, J. N. (2014). Globally divergent but locally convergent X-and Y-chromosome influences on cortical development. *Cerebral Cortex*, *26*(1), 70–79.

Raznahan, A., Wallace, G. L., Antezana, L., Greenstein, D., Lenroot, R., Thurm, A., … Giedd, J. N. (2013). Compared to what? Early brain overgrowth in autism and the perils of population norms. *Biological Psychiatry*, *74*(8), 563–575. https://doi.org/10.1016/j.biopsych.2013.03.022

Reardon, P. K., Clasen, L., Giedd, J. N., Blumenthal, J., Lerch, J. P., Chakravarty, M. M., & Raznahan, A. (2016). An allometric analysis of sex and sex chromosome dosage effects on subcortical anatomy in humans. *The Journal of Neuroscience: The Official Journal of the Society*

for Neuroscience, *36*(8), 2438–2448. https://doi.org/10.1523/JNEUROSCI.3195-15.2016

Reardon, P. K., Seidlitz, J., Vandekar, S., Liu, S., Patel, R., Park, M. T. M., … Raznahan, A. (2018). Normative brain size variation and brain shape diversity in humans. *Science*, *360*(6394), 1222–1227. https://doi.org/10.1126/science.aar2578

Redcay, E., & Courchesne, E. (2005). When is the brain enlarged in autism? A meta-analysis of all brain size reports. *Biological Psychiatry*, *58*(1), 1–9. https://doi.org/10.1016/j.biopsych.2005.03.026

Riddle, K., Cascio, C. J., & Woodward, N. D. (2017). Brain structure in autism: A voxel-based morphometry analysis of the autism brain imaging database exchange (ABIDE). *Brain Imaging and Behavior*, *11*(2), 541–551. https://doi.org/10.1007/s11682-016-9534-5

Rijlaarsdam, J., van IJzendoorn, M. H., Verhulst, F. C., Jaddoe, V. W. V., Felix, J. F., Tiemeier, H., & Bakermans-Kranenburg, M. J. (2017). Prenatal stress exposure, oxytocin receptor gene (OXTR) methylation, and child autistic traits: The moderating role of OXTR rs53576 genotype. *Autism Research*, *10*(3), 430–438. https://doi.org/10.1002/aur.1681

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36.

Sacco, R., Gabriele, S., & Persico, A. M. (2015). Head circumference and brain size in autism spectrum disorder: A systematic review and meta-analysis. *Psychiatry Research: Neuroimaging*, *234*(2), 239–251. https://doi.org/10.1016/j.pscychresns.2015.08.016

Salmond, C. H., Ashburner, J., Connelly, A., Friston, K. J., Gadian, D. G., & Vargha-Khadem, F. (2005). The role of the medial temporal lobe in autistic spectrum disorders. *The European Journal of Neuroscience*, *22*(3), 764–772. https://doi.org/10.1111/j.1460-9568.2005.04217.x

Sanchis-Segura, C., Ibañez-Gual, M. V., Adrián-Ventura, J., Aguirre, N., Gómez-Cruz, Á. J., Avila, C., & Forn, C. (2019). Sex differences in gray matter volume: How many and how large are they really? *Biology of Sex Differences*, *10*(1), 32. https://doi.org/10.1186/s13293-019-0245-7

Scahill, R. I., Frost, C., Jenkins, R., Whitwell, J. L., Rossor, M. N., & Fox, N. C. (2003). A longitudinal study of brain volume changes in normal aging using serial registered magnetic resonance imaging. *Archives of Neurology*, *60*(7), 989–994. https://doi.org/10.1001/archneur.60.7.989

Schaer, M., Kochalka, J., Padmanabhan, A., Supekar, K., & Menon, V. (2015). Sex differences in cortical volume and gyrification in autism. *Molecular Autism*, *6*(1), 42. https://doi.org/10.1186/s13229-015-0035-y

Shafritz, K. M., Bregman, J. D., Ikuta, T., & Szeszko, P. R. (2015). Neural systems mediating decision-making and response inhibition for social and nonsocial stimuli in autism. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *60*, 112–120. https://doi.org/10.1016/j.pnpbp.2015.03.001

Srivastava, A. K., & Schwartz, C. E. (2014). Intellectual disability and autism spectrum disorders: Causal genes and molecular mechanisms. *Neuroscience & Biobehavioral Reviews*, *46*, 161–174. https://doi.org/10.1016/j.neubiorev.2014.02.015

Stanfield, A. C., McIntosh, A. M., Spencer, M. D., Philip, R., Gaur, S., & Lawrie, S. M. (2008). Towards a neuroanatomy of autism: A systematic review and meta-analysis of structural magnetic resonance imaging studies. *European Psychiatry*, *23*(4), 289–299. https://doi.org/10.1016/j.eurpsy.2007.05.006

Sussman, D., Leung, R. C., Vogan, V. M., Lee, W., Trelle, S., Lin, S., … Taylor, M. J. (2015). The autism puzzle: Diffuse but not pervasive neuroanatomical abnormalities in children with ASD. *NeuroImage: Clinical*, *8*, 170–179. https://doi.org/10.1016/j.nicl.2015.04.008

Takao, H., Hayashi, N., & Ohtomo, K. (2012). A longitudinal study of brain volume changes in normal aging. *European Journal of Radiology*, *81*(10), 2801–2804. https://doi.org/10.1016/j.ejrad.2011.10.011

Toro, R., Chupin, M., Garnero, L., Leonard, G., Perron, M., Pike, M., … Paus, T. (2009). Brain volumes and Val66Met polymorphism of the BDNF gene: Local or global effects?. *Brain Structure and Function*, *213*(6), 501–509. https://doi.org/10.1007/s00429-009-0203-y.

Traut, N., Beggiato, A., Bourgeron, T., Delorme, R., Rondi-Reig, L., Paradis, A.-L., … Toro, R. (2018). Cerebellar volume in autism: Literature meta-analysis and analysis of the autism brain imaging data exchange cohort. *Biological Psychiatry*, *83*(7), 579–588. https://doi.org/10.1016/j.biopsych.2017.09.029

van Rooij, D., Anagnostou, E., Arango, C., Auzias, G., Behrmann, M., Busatto, G. F., … Buitelaar, J. K. (2017). Cortical and subcortical brain morphometry differences between patients with autism spectrum disorder and healthy individuals across the lifespan: Results from the ENIGMA ASD working group. *American Journal of Psychiatry*, *175*(4), 359–369. https://doi.org/10.1176/appi.ajp.2017.17010100

Varcin, K. J., Alvares, G. A., Uljarević, M., & Whitehouse, A. J. O. (2017). Prenatal maternal stress events and phenotypic outcomes in autism spectrum disorder. *Autism Research*, *10*(11), 1866–1877. https://doi.org/10.1002/aur.1830

Vinke, E. J., de Groot, M., Venkatraghavan, V., Klein, S., Niessen, W. J., Ikram, M. A., & Vernooij, M. W. (2018). Trajectories of imaging markers in brain aging: The Rotterdam study. *Neurobiology of Aging*, *71*, 32–40. https://doi.org/10.1016/j.neurobiolaging.2018.07.001

Wallace, G. L., Dankner, N., Kenworthy, L., Giedd, J. N., & Martin, A. (2010). Age-related temporal and parietal cortical thinning in autism spectrum disorders. *Brain: A Journal of Neurology*, *133*(Pt. 12), 3745–3754. https://doi.org/10.1093/brain/awq279

Williams, D. L., Goldstein, G., & Minshew, N. J. (2006). The profile of memory function in children with autism. *Neuropsychology*, *20*(1), 21–29. https://doi.org/10.1037/0894-4105.20.1.21

Yang, D., Beam, D., Pelphrey, K. A., Abdullahi, S., & Jou, R. J. (2016). Cortical morphological markers in children with autism: A structural magnetic resonance imaging study of thickness, area, volume, and gyrification. *Molecular Autism*, *7*(1), 11. https://doi.org/10.1186/s13229-016-0076-x

Yang, X., Si, T., Gong, Q., Qiu, L., Jia, Z., Zhou, M., … Zhu, H. (2016). Brain gray matter alterations and associated demographic profiles in adults with autism spectrum disorder: A meta-analysis of voxel-based morphometry studies. *Australian & New Zealand Journal of Psychiatry*, *50*(8), 741–753. https://doi.org/10.1177/0004867415623858

Zhang, W., Groen, W., Mennes, M., Greven, C., Buitelaar, J., & Rommelse, N. (2018). Revisiting subcortical brain volume correlates of autism in the ABIDE dataset: Effects of age and sex. *Psychological Medicine*, *48*(4), 654–668. https://doi.org/10.1017/S0033291717001201X

Zwaigenbaum, L., Young, G. S., Stone, W. L., Dobkins, K., Ozonoff, S., Brian, J., … Messinger, D. (2014). Early head growth in infants at risk of autism: A baby siblings research consortium study. *Journal of the American Academy of Child & Adolescent Psychiatry*, *53*(10), 1053–1062. https://doi.org/10.1016/j.jaac.2014.07.007

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Williams CM, Peyre H, Toro R, Beggiato A, Ramus F. Adjusting for allometric scaling in ABIDE I challenges subcortical volume differences in autism spectrum disorder. *Hum Brain Mapp*. 2020;41:4610–4629. https://doi.org/10.1002/hbm.25145