Title: Neuroanatomical Norms in the UK Biobank: The Impact of Allometric Scaling, Sex and Age.

Running title: Neuroanatomical Norms in the UK Biobank

Authors: Camille Michèle Williams[1*], Hugo Peyre[1,2,3], Roberto Toro[4,5,6], & Franck Ramus[1].

1. Laboratoire de Sciences Cognitives et Psycholinguistique, Département d'Etudes Cognitives, École Normale Supérieure, EHESS, CNRS, PSL University, 75005, 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. Human Genetics and Cognitive Functions, Institut Pasteur, UMR 3571 CNRS
5. Center for Research and Interdisciplinarity (CRI), INSERM U1284
6. Université de Paris, Paris, France.

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

Abbreviations: Total Cerebral Measures (TCMs): Total Brain Volume (TBV), Total Mean Cortical Thickness (MCT), or Total Surface Area (TSA). Grey Matter Volume (GMV). White Matter Volume (WMV).

Table of Contents

Abbreviations: Total Cerebral Measures (TCMs): Total Brain Volume (TBV), Total Mean Cortical Thickness (MCT), or Total Surface Area (TSA). Grey Matter Volume (GMV). White Matter Volume (WMV).

Abbreviations: Total Cerebral Measures (TCMs): Total Brain Volume (TBV), Total Mean Cortical Thickness (MCT), or Total Surface Area (TSA). Grey Matter Volume (GMV). White Matter Volume (WMV).

Abstract

Few neuroimaging studies are sufficiently large to adequately describe population-wide variations. This study's primary aim was to generate neuroanatomical norms and individual markers that consider age, sex, and brain size, from 629 cerebral measures in the UK Biobank (N = 40 028). The secondary aim was to examine the effects and interactions of sex, age, and brain allometry – the non-linear scaling relationship between a region and brain size (e.g., Total Brain Volume) across cerebral measures.

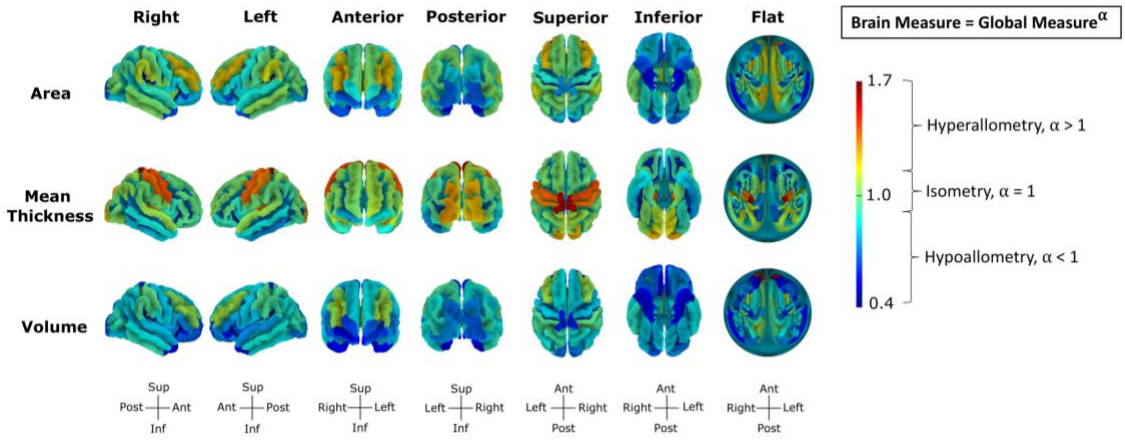
Allometry was a common property of brain volumes, thicknesses, and surface areas (83%) and was largely stable across age and sex. Sex differences occurred in 67% of cerebral measures (median $|\beta|= 0.13$): 37% of regions were larger in males and 30% in females. Brain measures (49%) generally decreased with age, although aging effects varied across regions and sexes. While models with an allometric or linear covariate adjustment for brain size yielded similar significant effects, omitting brain allometry influenced reported sex differences in variance.

This large scale-study advances our understanding of age, sex, and brain allometry's impact on brain structure and provides data for future UK Biobank studies to identify the cerebral regions that covary with specific phenotypes, independently of sex, age, and brain size.

Word count: 200

Abbreviations: Total Cerebral Measures (TCMs): Total Brain Volume (TBV), Total Mean Cortical Thickness (MCT), or Total Surface Area (TSA). Grey Matter Volume (GMV). White Matter Volume (WMV).

97 ## Graphical Abstract



98

99

Abbreviations: Total Cerebral Measures (TCMs): Total Brain Volume (TBV), Total Mean Cortical Thickness (MCT), or Total Surface Area (TSA). Grey Matter Volume (GMV). White Matter Volume (WMV).

Keywords

Cerebral Norms, Brain Volumes, Cortical Thickness, Cortical Surface Area, Allometry, Sex Differences

Highlights

- We created neuroanatomical norms and individual markers for the UK Biobank (N=40 028)
- Allometry was common across 83% of brain volumes, thicknesses, and surface areas
- 67% of regions differed between sexes: 37% were larger in males and 30% in females
- Omitting brain allometry influenced reported sex differences in variance
- 49% of regions declined with age, with variations across regions and sexes

Acknowledgements

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. Funding was also obtained from Fondation pour l'Audition (FPA RD-2016-8 research grant). This research has been conducted using the UK Biobank Resource. Declarations of interest: none.

Conflict of interest

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

Data/Code Availability

This research has been conducted using data from UK Biobank, a major biomedical database (http://www.ukbiobank.ac.uk/). Restrictions apply to the availability of these data, which were used under license for this study: application 46007. Preregistration and code are available on OSF: https://osf.io/s4qc5/?view_only=bb067d96d0df4ae4902f99747d60e828.

Abbreviations: Total Cerebral Measures (TCMs): Total Brain Volume (TBV), Total Mean Cortical Thickness (MCT), or Total Surface Area (TSA). Grey Matter Volume (GMV). White Matter Volume (WMV).

1. Introduction

Although all humans share a common brain structure and organization, they also vary in terms of the size and shape of their brain and its subcomponents. These neuroanatomical variations are thought to partly underlie differences in cognitive and behavioral traits and in the risk of developing psychiatric and neurological disorders (for review Dallaire-Théroux et al., 2017; Deary, 2010; Jumah et al., 2016; Oakes et al., 2017; Schmidt et al., 2018). Yet, most of these studies rely on relatively small samples and suffer from high sampling variability. When the sample is too small to accurately represent the control or target population, spurious neuroanatomical markers may be reported. Moreover, if true effects are observed in small samples, their size would be exaggerated as only large effects would pass a conventional statistical significance threshold (e.g., $p < 0.05$) with few degrees of freedom (for review Szucs & Ioannidis, 2017, 2020). Thus, despite hundreds of studies, few neuroanatomical measures can be declared as robust markers of cognitive traits or psychiatric and neurological disorders (Gong et al., 2019; Marek et al., 2020; Matsuo et al., 2019; Peyre et al., 2020; Ramus et al., 2018; Williams et al., 2020).

An alternative approach to comparing clinical and healthy groups would be to compare clinical groups with population norms, as is done with well-established cognitive dimensions, such as general intelligence, personality, and psychopathology scales (Beck et al., 1996; Costa Jr. & McCrae, 2008; Wechsler et al., 2008). If neuroanatomical norms for a population were available, then comparing any clinical group to these norms would overcome the issue of sampling variability in the control group. However, neuroanatomical norms are not easy to establish, as they require large populations, and neuroanatomical measures often depend on MRI scanner characteristics and acquisition sequences. Valid norms would therefore only be established within a given study in a single scanner, or in a small set of comparable scanning sites with similar acquisition protocols as done by the UK Biobank.

For this reason, the UK Biobank, the largest neuroimaging dataset available to date (N ~ 40 000), is an ideal candidate to create neuroanatomical norms that could be re-used for multiple studies of neurological and psychiatric disorders. These norms should be sex-specific , given that the two sexes differ on a number of neuroanatomical brain measures (Kaczkurkin et al., 2019; Lotze et al., 2019; Ritchie et al., 2018; Ruigrok et al., 2014; Sanchis-Segura et al., 2019), and have different risks of developing certain neurological and psychiatric disorders (Beam et al., 2018;

Abbreviations: Total Cerebral Measures (TCMs): Total Brain Volume (TBV), Total Mean Cortical Thickness (MCT), or Total Surface Area (TSA). Grey Matter Volume (GMV). White Matter Volume (WMV).

158 Boyd et al., 2015; Seedat et al., 2009). And age should also be considered, as it is associated with
159 variations in neuroanatomical measures (Fjell et al., 2013; Hurtz et al., 2014; Narvacan et al., 2017;
160 Vinke et al., 2018; Wierenga et al., 2014), cognitive function (Simon R. Cox et al., 2018; S.R. Cox
161 et al., 2019), and disease risk (Fiske et al., 2009; Fjell et al., 2009; Jellinger & Attems, 2015).
162 Finally, global brain measures should be taken into account to create norms that are
163 independent of variations in brain size. Although there is mounting evidence that brain allometry
164 - the non-linear scaling relationship between regional and global brain dimensions - is an inherent
165 property of the brain (Finlay et al., 2001; Jäncke et al., 2015a; Liu et al., 2014; Mankiw et al.,
166 2017; Reardon et al., 2018; Toro et al., 2009; Williams et al., 2020), standard modes of adjustment
167 for individual differences in global measures, such as the proportion method, or linear covariate
168 adjustment, omit brain allometry. To this day, numerous studies have shown that different methods
169 of adjustment for brain size contribute to the variability of reported volumetric group differences
170 (Lefebvre et al., 2015; O'Brien et al., 2006, 2011; Reardon et al., 2016; Sanchis-Segura et al.,
171 2019) and some specifically suggest that omitting brain allometry leads to spurious group
172 differences (Mankiw et al., 2017; Reardon et al., 2016; Williams et al., 2020). Since
173 regional/global relationships follow a power function in a majority of regions, it is recommended
174 to log-transform regional and Total Cerebral Measures (TCMs; i.e., Total Brain Volume (TBV),
175 Total Mean Cortical Thickness (MCT), or Total Surface Area (TSA)) to account for allometric
176 scaling and obtain a more accurate description of the relationship between brain regions and
177 TCMs.
178 Thus, the present study's first aim is to produced neuroanatomical norms in the UK
179 Biobank that take into account sex, age, and the allometric relationships between regional and
180 global brain measures. Our second goal is to investigate the extent to which neuroanatomical
181 variations depend on sex, age (linear and quadratic), and brain allometry effects and their
182 interactions. Finally, our third aim is to compare TCM adjustment techniques to examine whether
183 omitting brain allometry systematically biases reported results. By generating neuroanatomical
184 markers across volumes, mean thicknesses, and surface areas available in the UK Biobank, the
185 present paper provides UK population norms for future studies that aim to link regional
186 neuroanatomical markers to specific cognitive and behavioral traits or neurological and psychiatric
187 disorders.
188

Abbreviations: Total Cerebral Measures (TCMs): Total Brain Volume (TBV), Total Mean Cortical Thickness (MCT), or Total Surface Area (TSA). Grey Matter Volume (GMV). White Matter Volume (WMV).

2. Methods

2.1. Participants

Participants were drawn from the UK Biobank, an open-access large prospective study with phenotypic, genotypic, and neuroimaging data from 500 000 participants recruited between 2006 and 2011 at 40 to 69 years old in Great Britain (Sudlow et al., 2015). All participants provided informed consent ("Resources tab" at https://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=200). The UK Biobank received ethical approval from the Research Ethics Committee (reference 11/NW/0382) and the present study was conducted based on application 46 007.

Currently, Magnetic Resonance Imaging (MRI) data and Imaging-Derived Phenotypes (IDPs) are available for about 41 000 participants. This study analyzed the IDPs from the first imaging visit generated by an image-processing pipeline developed and run by the UK Biobank Imaging team (Alfaro-Almagro et al., 2018; Miller et al., 2016).

2.1.1. Brain Image Acquisition and Processing.

A standard Siemens Skyra 3T running VD13A SP4 with a standard Siemens 32-channel RF receive head coil was used to collect data (Brain Scan Protocol). The 3D MPRAGE T1-weighted volumes were analyzed by the UK Biobank Imaging team with pipeline scripts that primarily call for FSL and Freesurfer tools. Details of the acquisition protocols, image processing pipeline, image data files and derived measures IDPs of brain structure and function are available in the UK Biobank Imaging Protocols.

2.1.2. Total Brain Volume (TBV)

TBV was calculated as the sum of the total grey matter volume (GMV; i.e., sum of cortical and subcortical GMV, data-field 26518), cerebellum white matter volume (WMV, data-fields 26556 for left and 26587 for right), and cerebral WMV (data-fields 26553 for left and 26584 for right) from the UK Biobank ASEG Freesurfer segmentations. Refer to Supplemental Info 1 for more on the choice of TBV. Individuals with missing data for these regions were excluded from the analyses, yielding 40 055 participants.

2.1.3. Scanner Site

The age and sex of participants differed across the 3 scanner sites located in Cheadle (Site 11025), Reading (Site 11026), and Newcastle (Site 11027; See Supplemental Info 1). One individual without scanner site was removed from the analyses yielding 40 054 participants.

Abbreviations: Total Cerebral Measures (TCMs): Total Brain Volume (TBV), Total Mean Cortical Thickness (MCT), or Total Surface Area (TSA). Grey Matter Volume (GMV). White Matter Volume (WMV).

2.1.4. Sex

Participants who did not self-report as male or female or whose self-reported sex and genetic sex differed were also excluded from the analyses (N = 26). When genetic sex was not available, reported-sex was used to define the sex of the participant. Of the 40 029 participants included in the analyses, there were more females (N= 21 142) than males (N = 18 886, $\chi^2(1) = 127.15$, $p < 2.2e\text{-}16$).

2.1.5. Age

To obtain a continuous and more precise measure of age, age was calculated based on the year and month of birth of the participant and the day, month, and year of their MRI visit. Mean age was at 63.70 years old (*SD* = 7.54). Males (*M* = 64.39 years, *SD* = 7.65) were older than females ($t_{(39180)} = -17.18$, $p < 2.2e\text{-}16$, *M* = 63.09 years, *SD* =7.39).

2.2. Image Derived-Phenotypes (IDPs)

The descriptive statistics of all global and regional IDPs analyzed in the present study and their respective data-fields and segmentation origin are listed in Supplemental Table A1. The majority of IDPs correspond to grey matter, since white matter volumes were not segmented by the UK Biobank Imaging team.

2.2.1. Global IDPs

A total of 9 global IDPs were investigated: TBV, Total Mean Cortical Thickness (Total MCT), Total Surface Area (TSA), Subcortical GMV, Cortical GMV, Cerebral WMV, Cerebellar GMV, Cerebellar WMV, and the Brainstem volume.

WMV measures were obtained by summing left and right global measures from Freesurfer ASEG segmentations (data-field 190). Cerebellum GMV was calculated as the sum of the cerebellar volumes from the FAST segmentations (data-field 1101). Total MCT and TSA were respectively calculated as the sum of the mean cortical thickness and surface area measures from the Freesurfer a2009s segmentations (data-field 197). The whole brain stem measure was taken from the Freesurfer subsegmentations (data-field 191) and the Subcortical GMV measure was calculated as the sum of the left and right whole amygdala, hippocampus, and thalamus volumes from the Freesurfer subsegmentations (data-field 191) and the left and right caudate, accumbens, pallidum, and putamen of the Freesurfer ASEG segmentations (data-field 190).

Based on the recommendations from the UK Biobank Imaging Protocols, we excluded Freesurfer IDPs when T2-FLAIR was not used in addition to the T1 images to obtain the

segmentations from Freesurfer a2009s (volume, surface area, and mean thickness) and Freesurfer subsegmentations. Moreover, 790 individuals had missing values for all FAST cerebellum segmentations and were excluded from the FAST segmentation analyses.

Thus, while 40 028 individuals were included in the analyses for TBV, Cerebellum WM, and Cerebral WM, 39 238 individuals were included in the analyses with FAST segmentations and 38 710 were included in the analyses of the Freesurfer Subsegmentations and Freesurfer a2009s segmentations. Missing values and null segmentations (e.g., 0 mm3) for a region were replaced by the mean of that region when calculating global measures. See Supplemental Info 3 for correlations between the global measures provided by Freesurfer ASEG and those calculated from the FAST, Freesurfer Subsegmentations, Freesurfer ASEG, and Freesurfer a2009s Segmentations.

2.2.2. Regional IDPs

A case-wise participant exclusion strategy was applied to each IDP for the regional analyses: participants with a missing value or a segmentation error for a region were excluded from the analyses of that region but were maintained in the analyses of other IDP. Following visual examination of the distribution of regional cerebral measures, values 3 times the inter-quartile range for a region were considered to be segmentation errors and were removed from the analyses of that region.

A total of 620 regional IDPs were investigated: 444 cortical regions (148 volumes, 148 surface areas, and 148 cortical thicknesses) from the Freesurfer a2009s segmentations (Destrieux Atlas, data-field 197), 116 whole segmentations and subsegmentations of the amygdala, hippocampus, and thalamus and subsegmentations of the brainstem (Freesurfer subsegmentations, data-field 191), 28 cerebellum GMV segmentations from the FAST segmentations (data-field 1101), and 32 subcortical, white matter, and ventricle volumes from the Freesurfer ASEG segmentations (data-field 190). Freesurfer subcortical segmentations for the caudate, putamen, accumbens, and pallidum were used instead of the preregistered FIRST volumes, for segmentation consistency with the other subcortical and cortical volume which were segmented from Freesurfer.

2.3. Statistical Analyses

Analyses were preregistered on OSF and run using R (R Core Team, 2019). The preregistration and code are on OSF (https://osf.io/s4qc5/?view_only=bb067d96d0df4ae4902f99747d60e828). Used packages are listed in Supplemental Info 7.

Abbreviations: Total Cerebral Measures (TCMs): Total Brain Volume (TBV), Total Mean Cortical Thickness (MCT), or Total Surface Area (TSA). Grey Matter Volume (GMV). White Matter Volume (WMV).

2.3.1 Data Transformation.

To examine allometric scaling, regional and global measures were log10 transformed. All continuous variables were centered around the mean in order to examine the effects of a variable when other variables are at their mean value. The categorical sex variable was coded -0.5 for males and 05 for females. The scanner site variable was dummy coded with the largest site, Cheadle (Site 1102), as reference.

2.3.2. Analyses.

Global and Regional analyses were performed twice. Once without scaling (dividing by 1 SD) to obtain the allometric scaling coefficient for each brain region and once with scaling to report standardized betas as effect sizes. Non-linear effects of age were modeled with quadratic age over spline regression as spline regressions do not yield interpretable beta coefficients of age. Isometry was tested using the linear hypothesis function which tests if scaling coefficients obtained without scaling variables differ from 1. Scanner site (Cheadle - Site 1102, Reading - Site 11026, and Newcastle - Site 11027) was additionally added as a covariate, although it was omitted from the preregistration.

2.3.3. Global Analyses.

Global analyses were conducted to evaluate how TBV varies with age, age^2, and sex (equation 1) and how TSA, Total MCT, Cerebral GMV, Cerebral WMV, Total Subcortical GMV, Cerebellum GMV, Cerebellum WMV, and the brainstem volume vary with TBV, age, age^2, and sex (equation 2).

Equation 1

$$\text{Log10}(\textit{TBV}) = \text{Intercept} + \beta1*\textit{Age} + \beta2*\textit{Sex} + \beta3*\textit{Age}^2 + \beta4*\textit{Age x Sex} + \beta5*\textit{Age}^2\textit{ x Sex} + \beta6*\textit{Scanner Site} + \text{Error}$$

Equation 2

$$\text{Log10}(\textit{Global Measure}) = \text{Intercept} + \beta1*\text{Log10}(\textit{TBV}) + \beta2*\textit{Age} + \beta3*\textit{Sex} + \beta4*\textit{Age}^2 + \beta5*\text{Log10}(\textit{TBV})\textit{ x Sex} + \beta6*\text{Log10}(\textit{TBV})\textit{ x Age} + \beta7*\text{Log10}(\textit{TBV})\textit{ x Age}^2 + \beta8*\textit{Age x Sex} + \beta9*\textit{Age}^2\textit{ x Sex} + \beta10*\text{Log10}(\textit{TBV})\textit{ x Age x Sex} + \textit{β1}*\text{Log10}(\textit{TBV})\textit{ x Age}^2\textit{x Sex} + \beta12*\textit{Scanner Site} + \text{Error}$$

Abbreviations: Total Cerebral Measures (TCMs): Total Brain Volume (TBV), Total Mean Cortical Thickness (MCT), or Total Surface Area (TSA). Grey Matter Volume (GMV). White Matter Volume (WMV).

2.3.4 Regional Analyses.

Regional analyses were conducted to evaluate how regional volumes, surface areas, and cortical thicknesses vary with TCM, age, age^2, and sex with equation 3. The TCMs were TBV for volumes, total MCT for mean thicknesses, and TSA for surface areas.

Equation 3

$$\text{Log10}(\textit{Region}) = \text{Intercept} + \beta1*\text{Log10}(\textit{TCM}) + \beta2*\textit{Age} + \beta3*\textit{Sex} + \beta4*\textit{Age}^2 + \beta5*\text{Log10}(\textit{TCM})\ x\ \textit{Sex} + \beta6*\text{Log10}(\textit{TCM})\ x\ \textit{Age} + \beta7*\text{Log10}(\textit{TCM})\ x\ \textit{Age}^2 + \beta8*\textit{Age}\ x\ \textit{Sex} + \beta9*\textit{Age}^2\ x\ \textit{Sex} + \beta10*\text{Log10}(\textit{TCM})\ x\ \textit{Age}\ x\ \textit{Sex} + \textit{β1}*\text{Log10}(\textit{TCM})\ x\ \textit{Age}^2 x\ \textit{Sex} + \beta12*\textit{Scanner Site} + \text{Error}$$

2.3.5 Person-level Neuroanatomical Markers.

To obtain a global and regional marker of an individual's deviance from the norm in terms of volume, mean thickness, and surface area, we extracted the residuals from each dependent variable. The residuals were obtained from the model where continuous variables were centered but not scaled to maintain differences in magnitude across regions. An individual's residual value for a given regional measure reflects that individual's deviance from the norm, given his/her age, sex, and TCM, and therefore constitutes a new neuroanatomical marker.

2.3.6 Person-level Global Neuroanatomical Deviance.

From the person-level local neuroanatomical markers, we generated four person-level global neuroanatomical deviance measures with equation 4: one for volumes, one for mean thicknesses, one for surface areas, and one for all regions. The person-level global neuroanatomical deviance measure corresponds to a person's global neuroanatomical deviance from the norm. Although we pre-registered equation 4 without dividing by the total number of investigated regions for a global measure (N), we did so to obtain a value reflecting mean deviation relative to the norm. Considering that all IDPs were not available for all individuals, we excluded participants with more than 10% of missing data across regional IDPs. The cerebral marker across brain measures, which was not preregistered, was calculated by averaging the Z-score of the volumetric, cortical mean thickness, and cortical surface area global neuroanatomical deviance marker.

Equation 4

343 $$\text{Global Neuroanatomical Deviance} = \left(\sum (\text{Person-Level Neuroanatomical Markers})^2\right) / N$$

344

345 ### 2.3.7 Exploratory Analyses.

346 These analyses were not preregistered unless otherwise stated (for more details see
347 Supplemental Info 4). In brief, we first examined whether global and regional measures were
348 allometric with the linearHypothesis function from the car R package (Fox et al., 2020) with the
349 null hypothesis being "The slope of log10(TCM) is equal to 1". Then, we examined whether the
350 scaling coefficient of mean thicknesses with TBV differed from ⅓ and whether the scaling
351 coefficient of surface areas with TBV differed from ⅔ with the linearHypothesis function (Fox et
352 al., 2020) and appropriate null (e.g., null hypothesis for mean thicknesses: "The slope of
353 log10(TBV) is equal to 1/3"). We would expect these coefficients if brain growth was proportional
354 (similar to a sphere) and if larger brains were scaled-up versions of smaller brains. Third, we
355 examined sex differences in variance with a Levene's test (F-test) and calculated as the variance
356 ratio as Female SD / Male SD. Fourth, we compared the results of our main analysis to those
357 obtained when using a linear covariate TCM adjustment (i.e., equation 4 without the log
358 transformation) and when using the proportion adjustment for TCM (i.e., dividing a region by
359 TCM to obtain an adjusted region measure and running equation 4 on the adjusted region without
360 the log transformation and the main effect of TBV). Fifth, as preregistered, we attempted to
361 replicate previous studies on cerebral sex differences that considered brain allometry. Sixth, we
362 replicated Ritchie and colleagues' (2018) analyses of the Desikan-Killiany cortical measures and
363 FIRST subcortical volumes with the linear covariate TCM adjustment. We additionally ran the
364 same analyses with the allometric TCM adjustment to examine whether we observed the same
365 effects of omitting brain allometry when investigating sex differences in the UK Biobank using
366 different cerebral segmentations and statistical analyses.

367 ### 2.3.8 Multiple Comparison Corrections.

368 Considering that 620 regions and 11 beta coefficients were investigated for regional IDPs,
369 three thresholds of significance were used: 0.05/11, 0.05/620, and 0.05/ (11 * 620). The same
370 rationale was applied to the global IDPs, with the following thresholds for TBV 0.05/5, 0.05/9,
371 and 0.05/ (5 * 9) and for the remaining global measures: 0.05/11, 0.05/9, and 0.05/ (11 * 9).
372 Significant variance ratio differences are reported at $p < 0.05/629$ (sum of global and regional
373 cerebral measures).

374

Abbreviations: Total Cerebral Measures (TCMs): Total Brain Volume (TBV), Total Mean Cortical Thickness (MCT), or Total Surface Area (TSA). Grey Matter Volume (GMV). White Matter Volume (WMV).

3. Results and Discussion

To avoid redundancy, results are sequentially discussed. Only results at the strictest level of significance are reported in text: $p < 0.05/ (5*9)$ for TBV, $0.05/ (11*9)$ for other global measures, $p < 0.05/ (11*620)$ for regional measures, and $p < 0.05/629$ for sex differences in variance. Scaling coefficients (α) correspond to the estimate when continuous variables are centered, and standardized betas (β) reflect the estimate when continuous variables are centered and scaled (1 SD).

Descriptive statistics are available in Supplemental Tables A2-14. Regression results by region are available in Supplemental Tables B1-B10 and by main effectsor interaction, in Supplemental Tables C1-C15. See Supplemental Tables D2-D5 for statistics on regional deviance from isometry. Correlations between the left and right region in terms of the scaling coefficient with the TCM, the age standardized betas, and sex standardized betas are available on OSF, in Supplemental Figures File 1(https://osf.io/s4qc5/?view_only=bb067d96d0df4ae4902f99747d60e828) and correlations between cortical scaling, sex, and age coefficients are available in Supplemental Table C1.

3.1 Allometry

3.1.1 Global Allometry

All global scaling coefficients were hypoallometric (α ranging from 0.03 to 0.91), suggesting that these regions increase less than TBV as TBV increases, except for cerebral WMV, which was hyperallometric ($\alpha = 1.21$). The scaling coefficient of TSA with TBV significantly differed from the theoretical value 2/3 ($\alpha = 0.89$) and the scaling coefficient of Total MCT with TBV differed from 1/3 ($\alpha = 0.03$). In cerebellar GMV, all regions were hypoallometric ($\alpha = 0.50$ - 0.95), except for 4 isometric regions ($\alpha = 0.93$ - 0.95). Ventricles and the cerebral spinal fluid were hypoallometric ($\alpha = 0.40$ - 1.01), except for the lateral ventricles (left $\alpha = 1.01$ and right $\alpha = 0.96$). The optic chiasm, corpus callosum, cerebellum WMV, and ventral diencephalon measures were hypoallometric ($\alpha = 0.55$ - 0.93), whereas for the mid-anterior segmentation of the corpus callosum which was isometric ($\alpha = 0.99$).

Allometric coefficients were generally consistent with previous studies that report hyperallometry in cerebral WMV (Jong et al., 2017; Toro et al., 2009) and hypoallometry in the majority of cerebellar (Mankiw et al., 2017), subcortical (Jong et al., 2017; Liu et al., 2014;

Abbreviations: Total Cerebral Measures (TCMs): Total Brain Volume (TBV), Total Mean Cortical Thickness (MCT), or Total Surface Area (TSA). Grey Matter Volume (GMV). White Matter Volume (WMV).

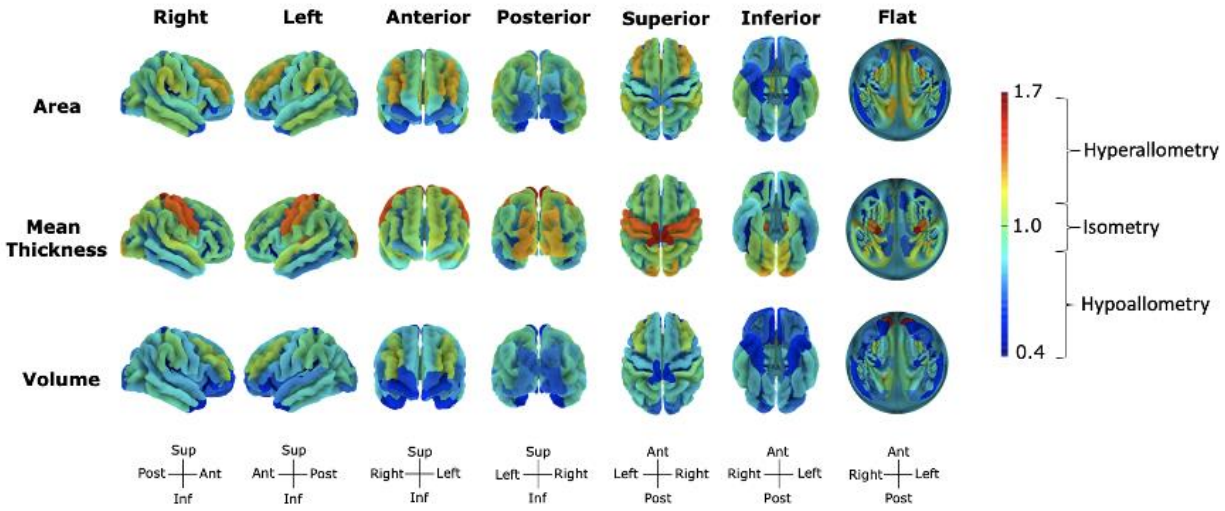
Reardon et al., 2016; Williams et al., 2020), and corpus callosum volumes (Lefebvre et al., 2015). See Supplemental Table D1 for global deviance from allometry.

3.1.2 Cortical Allometry

TBV was a significant positive predictor of all volumes, TSA of all surface areas, and Total MCT of all mean thicknesses. Scaling coefficients varied across regions and measures (Table 1). The scaling coefficients of cortical volumes were highly correlated to those of cortical surface areas (r = 0.85, p < 2.2e-16) but were not correlated to those of cortical mean thicknesses (r = -0.05, p = 0.535). In cortical regions, 98 volumes, 62 areas, and 63 mean thicknesses were hypoallometric (α = 0.60 - 0.96), while 18 volumes, 50 areas, and 58 mean thicknesses were hyperallometric (α = 1.04-1.62), and 32 volumes, 36 areas, and 27 mean thicknesses were isometric (α = 0.94-1.11; Figure 1).

These findings mirror those reported by Liu and colleagues (2014) and Reardon and colleagues (2018), who found that scaling across cortical regions is heterogeneous, covering a wide range of hypoallometry, isometry, and hyperallometry values. Similarly to Reardon and colleagues (2018), who studied the scaling relationship between vertex area and cortical surface area, surface areas were hyperallometric in the middle frontal and supramarginal gyri and sulci, and hypoallometric in the sensorimotor cortices (precentral, paracentral, and postcentral gyri and sulci), occipital temporal regions, and some cingulate (anterior, mid-anterior, and post dorsal) and callosal (sub and pericallosal) regions.

426



427

428 ***Figure 1.*** *Scaling Coefficients of Cortical Surface Areas, Mean Thicknesses, and Volumes with*
429 *Total Brain Volume (TBV).* Values are the scaling coefficients of a region with TBV and range
430 from 0.61 (volume of the right posterior ramus of lateral sulcus) to 1.63 (mean thickness of the
431 left paracentral gyrus and sulcus). The flat representation corresponds to the flattened image of
432 the superior view with the midline of the circle reflecting regions within the sagittal plane and
433 circle edges reflecting inferior regions. Figures made with https://neuroanatomy.github.io/cortex/
434 (Toro, 2020).

Abbreviations: Total Cerebral Measures (TCMs): Total Brain Volume (TBV), Total Mean Cortical Thickness (MCT), or Total Surface Area (TSA). Grey Matter Volume (GMV). White Matter Volume (WMV).

435

436 ### 3.1.3 Subcortical Allometry

437 All subcortical volumes were hypoallometric ($\alpha = 0.49$ - 0.92) except for the right lateral
438 posterior and the right limitans supragenicular nuclei, which were isometric ($\alpha = 0.95$, $\alpha = 1.08$,
439 respectively), and the left limitans supragenicular nuclei and left accumbens area, which were
440 hyperallometric ($\alpha = 1.17$, $\alpha = 1.11$, respectively).

441 General hypoallometry across whole subcortical structures is consistent with previous
442 findings (Jäncke et al., 2015b; Liu et al., 2014; Williams et al., 2020). As for allometry within
443 subcortical structures, Reardon and colleagues (2018) are the only ones to date that examined and
444 reported variations in scaling within subcortical structures. However, the latter study examined the
445 scaling relationship of the vertex area of subcortical subregions with cortical area, whereas the
446 present study examined volumetric variations within these regions. Thus, our finding that
447 allometry varies within subcortical volumes extends our understanding of the cerebral scaling
448 relationships between regional and global measures.

449

450 ### 3.1.4 Mean Thickness and Surface Area Scaling with TBV

451 Exploratory analyses on the scaling relationship between TBV and cortical mean
452 thicknesses or surface areas revealed that the scaling coefficients of all mean thicknesses with TBV
453 were different from one-third, whereas the scaling coefficients of 19 regional surface areas with
454 TBV did not differ from two-thirds (Supplemental Tables D6-7).

455 This study adds the literature suggesting that larger brains are not simply a scaled-up
456 version of smaller brains (Finlay et al., 2001; Im et al., 2008; Toro et al., 2008), as the first study
457 to examine scaling coefficients of regional cortical surface areas and mean thicknesses with TBV
458 that go beyond lobar segmentations. If the brain regions grew proportionally to brain size, total
459 MCT would scale to the power of one-third with TBV, and TSA to the power of two-thirds with
460 TBV (Finlay et al., 2001). Yet, we find that total MCT scaled to the power of 0.03 with TBV and
461 that the majority of cortical mean thicknesses had allometric scaling coefficients close to 0,
462 reflecting the stability of cortical mean thicknesses with TBV growth. Moreover, TSA scaled to
463 the power of 0.89 with TBV and the majority of cortical surface areas had hypoallometric scaling
464 coefficients that were greater than two-thirds. The greater than geometrically expected
465 hypoallometry across surfaces can be explained by the dramatic increase in gyrification (Fish et

al., 2017), and more specifically, in sulcal convolution, that occurs with the expansion of TBV (Im et al., 2008; Toro et al., 2008). Finally, the heterogeneity of allometric patterns observed across the cortex may be explained by the nonuniform gyrification of the cortical surface (Fish et al., 2017).

3.1.5 Sex or Age-Dependent Allometry

Allometry depended on sex (0.2%, 12/628) or age (11%, 71/628) in relatively few regions (details in Supplemental Info 5.1**)**. Sex differences in allometry were less frequent, although larger ($|\alpha| = 0.05 - 0.12$), than age differences in allometry ($|\alpha| = 0.01 - 0.04$). Regional volumes generally increased less with TBV in females, while mean thicknesses increased less with Total MCT in males. Although sparse, the presence of TCM interactions with age or sex suggest that matching individuals between groups by TCM may not be appropriate for all regions, as cerebral sex differences may reflect sex-dependent distributions of tissues instead of individual differences in brain size (Luders et al., 2009). Considering that these interactions are often overlooked in studies examining sex and age differences (e.g., Ritchie et al., 2018; Vinke et al., 2018), we suggest that they be considered to obtain unbiased estimates of age and sex effects on the brain and to accurately identify associations between brain regions and behavioral or cognitive traits.

3.1.6 Conclusion on Allometry

All brain regions varied with global brain measures, and the majority (86%) were allometric as they scaled non-linearly with their TCM. Of the regions exhibiting allometry, hypoallometry was reported in 93% of volumes, 55% of the cortical surface areas, and 52% of the cortical mean thicknesses (Table 1). While the association between scaling and cognition and behavior remains unknown, our study adds to the literature (Finlay et al., 2001; Jäncke et al., 2015b, 2019; Jong et al., 2017; Toro et al., 2009) supporting allometric scaling as an inherent property of the brain that varies across regions and cerebral measures, and provides scaling coefficients for regions that were not previously investigated (e.g., subcortical subsegmentations, ventricles etc.).

Table 1. Cerebral Regions Exhibiting Brain Allometry.

	Measure	N	%	Min.	1st Qu.	Median	Mean	3rd Qu.	Max
Cortical Surface Areas (N = 148)	*Hypoallometric*	62	42	0.64	0.76	0.83	0.82	0.91	0.95
	Hyperallometric	50	34	1.06	1.1	1.13	1.16	1.21	1.39
Cortical Mean Thicknesses (N = 148)	*Hypoallometric*	63	43	0.66	0.78	0.84	0.83	0.89	0.96
	Hyperallometric	58	39	1.04	1.09	1.14	1.2	1.31	1.62
Cortical Volumes (N = 148)	*Hypoallometric*	98	66	0.6	0.73	0.82	0.8	0.89	0.96
	Hyperallometric	18	12	1.05	1.07	1.1	1.15	1.16	1.59
Freesurfer Subcortical Subsegmentation Volumes (N = 116)	*Hypoallometric*	113	97	0.49	0.68	0.75	0.74	0.8	0.92
	Hyperallometric	1	1	1.17	1.17	1.17	1.17	1.17	1.17
FAST Cerebellar Volumes (N = 28)	*Hypoallometric*	24	86	0.5	0.68	0.74	0.75	0.82	0.94
ASEG Subcortical Volumes (N = 8)	*Hypoallometric*	7	88	0.69	0.73	0.75	0.78	0.81	0.92
	Hyperallometric	1	13	1.11	1.11	1.11	1.11	1.11	1.11
ASEG Ventricle & CSF Volumes (N = 10)	*Hypoallometric*	7	70	0.4	0.47	0.49	0.51	0.55	0.62
ASEG Cerebellum, Corpus Callosum, Ventral DC, Optic Chiasm Volumes (N = 14)	*Hypoallometric*	12	86	0.55	0.66	0.78	0.76	0.85	0.93
	Hyperallometric	1	7	1.21	1.21	1.21	1.21	1.21	1.21
Global Measures (N = 9)	*Hypoallometric*	8	89	0.03	0.77	0.73	0.83	0.9	0.91
	Hyperallometric	1	11	1.21	1.21	1.21	1.21	1.21	1.21

N.B. Values displayed are allometric scaling coefficients (α). Hyperallometric $\alpha > 1$, and Hypoallometric: $\alpha < 1$. CSF: Cerebral Spinal Fluid. DC: Diencephalon.

Abbreviations: Total Cerebral Measures (TCMs): Total Brain Volume (TBV), Total Mean Cortical Thickness (MCT), or Total Surface Area (TSA). Grey Matter Volume (GMV). White Matter Volume (WMV).

3.2 Sex Differences

3.2.1 Global Measures

TBV was significantly larger in males than in females (β = - 1.14). Once TBV was adjusted for with the allometric adjustment (equation 3), the cerebellar WMV (β = 0.27), cerebellar GMV (β = 0.25), total MCT (β = 0.12), and cerebral WMV (β = 0.06) were greater in females, while brainstem volume (β = -0.21), total subcortical volumes (β = -0.08), TSA (β = -0.07), and cerebral GMV (β = -0.02) were greater in males.

Consistent with previous studies, males had a larger TBV (e.g., Ritchie et al., 2018; Ruigrok et al., 2014), while females had a relatively larger Total MCT (Im et al., 2008, p. 200; van Velsen et al., 2013). However, we did not find that males and females had a similar TSA (Im et al., 2008), nor did we observe greater cerebral WMV in males relative to brain size (Chen et al., 2007; Gur et al., 1999). Instead, our analyses revealed that males have a larger TSA and females a greater cerebral WMV. Considering the small magnitude of these sex differences (<0.1) and the sample size of the previous studies (N < 150), we speculate that these studies were not sufficiently powered to reliably estimate these effects.

3.2.1 Cerebellar GMV

The Freesurfer ASEG Cerebellum GMV – used to calculate TBV - was larger in males (β = - 0.36), whereas the FAST Cerebellum GMV - calculated as the sum of the cerebellar lobes and vermes from FAST Diedrichsen Cerebellar Atlas - was larger in females (β = 0.25).

Result discrepancies between segmentation algorithms may stem from Diedrichsen's segmentation algorithm ignoring individual WMV and GMV intensities, which are taken into account by Freesurfer. Or differences may be due to Freesurfer over-labelling peripheral tissue as it is more sensitive in regions of low contrast between tissue types (Carass et al., 2018). Although the cerebellar GMV of the FAST and ASEG segmentations only correlated at 76%, discrepancies across cerebellar segmentations did not influence our measure of TBV, as the TBV calculated by summing regional segmentations (including the FAST GMV) correlated at 99.8% with our measure of TBV (derived from the ASEG segmentations).

3.2.3 Regional Cerebellar GMVs

When examining sex differences in the cerebellum with the FAST cerebellum regional segmentations, females had larger cerebellar GMV in 82% of regions (23/28) with coefficients ranging from 0.08 (Right Crus I) to 0.64 (Vermis X). The crus I vermis and the left and right lobule

V did not differ between sexes and the left and right lobule X were larger in males (β = -0.17, β = -0.12, respectively).

Our findings contrast with the literature on sex differences within cerebellum (review in Han et al., 2020). For instance, Han and colleagues (2020) instead reported that the VIIIA lobules were larger in males and that the right I–III lobules, and IX and X Vermes were larger in females when adjusting for intracranial volume with the linear covariate method. Moreover, in our replication of Mankiw and colleagues' (2017) study, we found that, instead of being greater in males, the cerebellum, flocculus, cerebellar lobule VIIb, VIIb, and VIIA, and Crus II volumes were greater in females and that the flocculus volume did not vary between sexes. Although both studies used segmentation algorithms that have a better parcellation accuracy than the SUIT segmentation (Diedrichsen et al., 2009) of the present study (Carass et al., 2018; S. Han, Carass, et al., 2020), discrepancies in the literature may also stem from differences in sample age (mean age = 12.5 and 70 years old, respectively) and size (N = 116 and 2,023, respectively).

In light of the difficulty of segmenting the cerebellum and the differences in regional specificity and accuracy across segmentations, we suggest that future studies take advantage of the large UK Biobank dataset to apply and compare cerebellar segmentation algorithms.

3.2.4 Whole Subcortical Volumes

The thalamus (Right β = -0.15, Left β = -0.08), putamen (Left and Right β = -0.18), left pallidum (β = -0.08), and left amygdala (β = -0.12) were larger in males and the hippocampus (Right β = 0.07, Left β = 0.06) and left accumbens were larger in females (β = 0.10). The right pallidum volume, the caudate volumes, and the right accumbens area volume did not differ between sexes.

Our findings are consistent with previous studies reporting greater thalamic volume in males (Lotze et al., 2019) and larger hippocampal volumes in females (Malykhin et al., 2017; Nordenskjöld et al., 2015), although they contrast with research supporting the absence of sex differences in the amygdala (Lotze et al., 2019) and the hippocampus (Ritchie et al., 2018; Tan et al., 2016) or greater male hippocampal volume (Lotze et al., 2019; Pintzka et al., 2015). And yet, we similarly find that males have larger putamen, pallidum, and left amygdala volumes in our replication of studies examining sex differences when considering brain allometry (Reardon et al., 2016; Sanchis-Segura et al., 2019).

We additionally replicated Ritchie and colleagues' (2018) findings that males had greater pallidum, putamen, and amygdala volumes, when adjusting for TBV with the linear covariate adjustment and analyzing FIRST subcortical segmentations. For a detailed analysis of the replication by region see Supplemental Info 6.4.2 and Supplemental Tables G5-9.

When adjusting for TBV with the linear covariate or the allometric approach in the replication models with the FIRST segmentations, we found similar subcortical sex differences to those reported in our main analyses with the Freesurfer segmentations with some exceptions. Specifically, instead of being larger in males, the left thalamus was larger in females and right thalamus did not show sex differences in the replication models. Moreover, the right accumbens area was larger in females and the right pallidum was larger in males in the replication analyses, although they did not differ between sexes in the main analyses. These discrepancies may stem from the different terms and interactions included in the main and replication analyses or from differences between the FIRST and Freesurfer subcortical segmentations. For instance, FIRST provides a segmentation of the amygdala more similar to that of manual tracing than Freesurfer (Morey et al., 2009), although the amygdala agreement with manual segmentation is relatively poor compared to other regions such as the hippocampus (Morey et al., 2010), potentially due to the complexity of the structure (Schoemaker et al., 2016).

Sex differences from our main analyses additionally varied across hemispheres. For instance, the amygdala volume was larger for males in the left hemisphere and similar across sexes in the right hemisphere. However, seemingly inconsistent results on whole subcortical structures may be illuminated by examining their subcomponents, as provided by the Freesurfer subcortical subsegmentations.

3.2.5 Subcortical Subsegmentations

Sex differences were found in 67% of the subcortical subsegmentations (74/110), with greater male volume in 42 (38%) regions and greater female volume in 32 (29%) regions (details in Supplemental Info 5.2.3).

The magnitude of the subsegmentation subcortical sex differences was not perceptible at the whole subcortical level due to the presence of sex differences in opposite directions. For instance, medial and lateral regions of the thalamus were considerably larger in females (β ranging from 0.06 to 0.25), although the whole thalamus volume was greater in males (Right β = -0.15, Left β = -0.08). Moreover, the right cortical nucleus (β = 0.13) was larger in females even though

sex differences were absent in the whole right segmentation of amygdala. Taken together, these findings support a high variability of sex differences within subcortical structures and highlights the importance of favoring fine-grained segmentations, as done in this study, to better understand where cerebral sex differences lie.

3.2.6 Cortical Regions

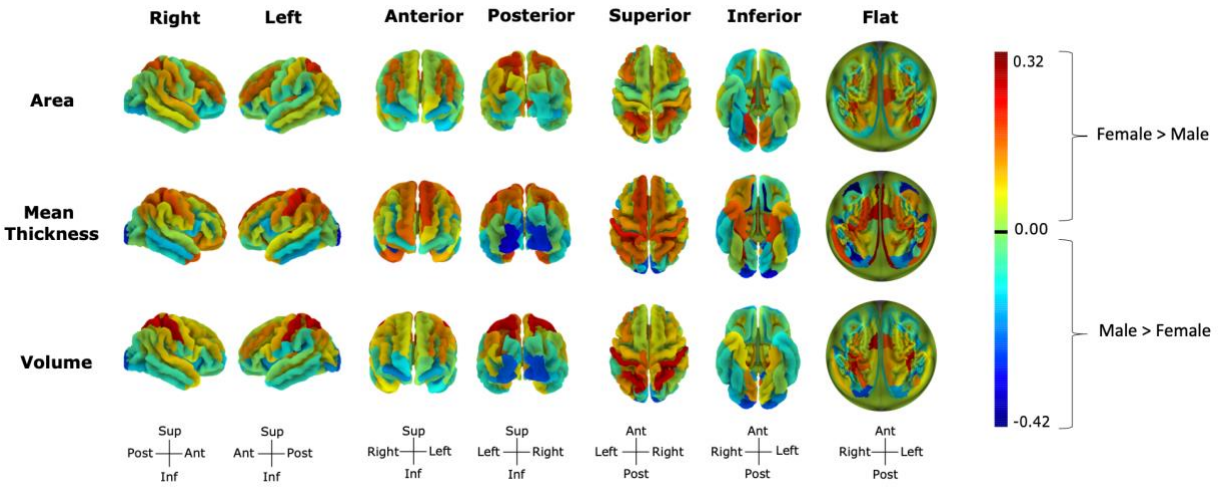
Cortical sex differences were present in 57% of surface areas (84/148) and 66% of volumes (97/148) and mean thicknesses (98/148). No clear spatial trend in sex differences was apparent across lobes. In terms of cortical volumes, 62 regions (42%) were larger in males, ranging from -0.27 (right occipital pole) to -0.06 (right inferior temporal sulcus), and 35 regions (24%) were larger in females, ranging from 0.06 (right precentral gyrus) to 0.26 (right postcentral gyrus). As for cortical mean thicknesses, 50 regions (34%) were greater in males and 48 (32%) were greater in females. Greater mean thicknesses in males varied from -0.39 (left medial orbital – olfactory-sulcus) to -0.05 (parieto-occipital sulcus or fissure), while greater mean thicknesses in females ranged from 0.06 (left transverse frontopolar gyri and sulci) to 0.31 (left transverse temporal sulcus). Finally, in terms of cortical surface areas, males had larger surface areas in 50 regions (34%) ranging from -0.22 (right medial occipito-temporal-collateral – sulcus and lingual sulcus) to -0.06 (left precentral gyrus), and females had larger surface areas in 34 regions (23%) with coefficients ranging from 0.06 (left superior frontal sulcus) to 0.21 (right anterior transverse temporal gyrus of Heschl; Figure 3).

We replicated the majority of sex differences (88%, 90/98) reported by Ritchie and colleagues (2018) with the Desikan-Killiany Cortical segmentations when adjusting for TCM with the linear covariate approach. We observed significant sex differences in additional regions (81%, 150/186), possibly due to our larger sample size. Sex differences from the main analyses with the Destrieux segmentation and the replication analyses with the Desikan-Killiany appeared to be generally consistent.

In line with a study of cortical volumetric sex differences in 411 middle aged participants (Chen et al., 2007), we found that males had a larger left inferior temporal gyrus and larger right occipital lingual and right middle temporal gyri, while females had a larger right inferior parietal gyrus, right post-dorsal part of the cingulate gyrus and sulcus, and left and right mid-anterior and post-ventral parts of the cingulate gyrus. Consistent with a study of 2 838 middle aged adults (Lotze et al., 2019), we found that females have larger volumes in the superior parietal lobe and

right orbitofrontal cortex, whereas males have larger volumes in the left temporal pole and right fusiform gyrus. Yet, in contrast with the previous literature on sex differences in GMV (Lotze et al., 2019; Ritchie et al., 2018; Ruigrok et al., 2014), we reported that females have a larger precentral gyri than males and that males have a larger right anterior part of the cingulate gyrus and sulcus than females. Moreover, we did not find sex differences in the left anterior part of the cingulate gyrus and sulcus, although past studies suggest that this volume was larger in females (Lotze et al., 2019; Ruigrok et al., 2014). The divergence in some of the presented results can partly be attributed to differences in sample size, varying neuroimaging techniques, and to the small size of these effects (median $|\beta| = 0.09$) as well as to differences in sample age range. And yet, we observed a similar pattern of sex differences in our replication of Ritchie and colleagues' (2018) analyses with the allometric adjustment for TCM, suggesting that the different terms included in our models and the differences between the Destrieux and the Desikan-Killiany Atlases had little influence on the majority reported cortical sex differences.

637



638

639 ***Figure 2.*** *Sex Differences across Cortical Measures.* Sex effects differences from -0.40 (mean
640 thickness of the left medial orbital sulcus) to 0.32 (mean thickness of the left transverse temporal
641 sulcus). Negative effects reflect greater male than female volumes. The flat representation
642 corresponds to the flattened image of the superior view with the midline of the circle reflecting
643 regions within the sagittal plane and circle edges reflecting inferior regions. Figures made with
644 https://neuroanatomy.github.io/cortex/ (Toro, 2020).

645

Abbreviations: Total Cerebral Measures (TCMs): Total Brain Volume (TBV), Total Mean Cortical Thickness (MCT), or Total Surface Area (TSA). Grey Matter Volume (GMV). White Matter Volume (WMV).

3.2.7 Sex Differences in Variance

In addition to observing mean sex differences in two-thirds of regions, we found sex differences in variance in 49% (306/629) of regions. A total of 253 (40%) regions exhibited greater male variability and 56 (9%) exhibited greater female variability (Table 2). Sex differences in variance ranged from 0.82 (for the right cerebellar lobule VIIIa, implying greater male variability) to 1.17 (for the optic chiasm; Supplemental Info 6.1 and Supplemental Tables E1-7).

As reported by previous studies (Ritchie et al., 2018; Wierenga et al., n.d., 2018, 2019), the majority of brain regions with sex differences in variance were more variable in males (82%) compared to females (18%). Mean thicknesses were generally more variable in females, while volumes and surface areas were more variable in males. Overall, cerebellar lobes and vermes were also more variable in males. Sex differences in variance across subcortical subsegmentations were greater in males in the hippocampus, amygdala, and thalamus. In terms of whole subcortical volumes, both the hippocampus, which was larger in females, and caudate, which did not show mean sex differences, were more variable in males. Finally, regions that were larger in one sex were typically more variable in the other sex.

Differences in variance thus do not appear to be a mechanical consequence of differences in mean, and instead may reflect a distinct phenomenon known as the greater male variability hypothesis. This hypothesis states that males are more variable than females across a variety of psychological and physical characteristics (Ellis, 1894) and is widely supported by a range of human (e.g., Johnson et al., 2008; Ju et al., 2015; Karwowski et al., 2016; Lehre et al., 2009; Wierenga et al., 2019) and animal (e.g., Branch et al., 2020; DeCasien et al., 2020) studies. Although the mechanisms behind the greater male variability hypothesis exceed the scope of the present study, our findings further support greater male variability, which extends well beyond brain measures.

671

672 *Table 2. Variance Ratios of Sex Differences across Segmentations*

Segmentation		**Sex Differences in Variance Ratio**		**min**	**1st Qu.**	**median**	**mean**	**3rd Qu.**	**max**
		N	**%**						
Cortical Volumes (N = 148)	*M>F*	63	43	0.91	0.94	0.95	0.95	0.96	0.97
	M<F	5	3	1.02	1.04	1.05	1.07	1.08	1.14
Cortical Surface Areas (N = 148)	*M>F*	71	48	0.90	0.94	0.95	0.95	0.96	0.97
	M<F	8	5	1.03	1.04	1.04	1.06	1.08	1.14
Cortical Mean Thicknesses (N = 148)	*M>F*	26	18	0.9	0.95	0.95	0.95	0.96	0.97
	M<F	36	24	1.03	1.05	1.06	1.06	1.07	1.15
Cerebellar Volumes (N = 28)	*M>F*	23	16	0.82	0.90	0.94	0.92	0.95	0.97
	M<F	2	1	1.04	1.04	1.04	1.04	1.04	1.04
Subcortical Subsegmentations (N = 116)	*M>F*	64	55	0.88	0.93	0.94	0.94	0.96	0.97
ASEG Subcortical Volumes (N =8)	*M>F*	2	25	0.95	0.96	0.96	0.96	0.96	0.96
ASEG Ventricle & CSF Volumes (N = 10)	*M>F*	2	20	0.95	0.95	0.95	0.95	0.95	0.96
	M<F	2	20	1.07	1.08	1.09	1.09	1.1	1.11
ASEG Cerebellum, Corpus Callosum, Ventral DC, Optic Chiasm Volumes (N = 14)	*M>F*	5	36	0.93	0.94	0.95	0.95	0.95	0.96
	M<F	3	21	1.02	1.03	1.04	1.08	1.1	1.17
Global Measures (N = 9)	*M>F*	1	11	0.90	0.90	0.90	0.90	0.90	0.90

673 *N.B.* Values represent Variance Ratios. M: Male, F: Female. Variance Ratio = Female SD / Male
674 SD. Qu.: Quartile. CSF: Cerebral Spinal Fluid.

Abbreviations: Total Cerebral Measures (TCMs): Total Brain Volume (TBV), Total Mean Cortical Thickness (MCT), or Total Surface Area (TSA). Grey Matter Volume (GMV). White Matter Volume (WMV).

675 #### 3.2.8 Conclusion on Sex Differences

676 Overall, we found that sex differences in the brain are the rule rather than the exception,
677 affecting two-thirds (419/629) of the investigated brain measures, with 231 regions relatively
678 greater in males. The standardized coefficients (β) of the sex effect of cortical volumes were highly
679 correlated to those of cortical surface areas (r = 0.77, p = 2.28e-43) and moderately so for cortical
680 mean thicknesses (r = 0.45, p = 1.23e-08).

681 Although many of the regional sex differences had very small effect sizes (β < 0.1) and
682 were only significant due to the large sample size, 46% of these cerebral measures (292/629) had
683 a sex difference above 0.1. Specifically, sex differences in cerebellar GMV and WMV, Total MCT,
684 the corpus callosum, and the ventricles were generally greater than 0.1 (details in Supplemental
685 Info 5.2), whereas sex differences in cerebral GMV and WMV, TSA, 51% of subcortical regions
686 and 35 - 45% of the cortical regions were under 0.1.

687

688 ### 3.3 Age Effects

689 #### 3.3.1 Global Measures

690 All cerebral measures decreased linearly with age (β ranging from -0.04 in to -0.33), except
691 for the brainstem volume, TSA, and cerebral WMV which increased (relatively to TBV) with age
692 (β = 0.13, β = 0.11, β = 0.04, respectively). The quadratic age term did not significantly predict
693 cerebral GMV and WMV or the brainstem volume. However, total subcortical volume (β = 0.01)
694 and TSA (β = 0.02) positively, and TBV (β = -0.05), total MCT (β = -0.04), cerebellar GMV (β =
695 -0.05), and cerebellar WMV (β = -0.02) negatively varied with quadratic age.

696 In line with the literature, TBV and Total MCT decreased with linear and quadratic age
697 (Ritchie et al., 2018; van Velsen et al., 2013; Vinke et al., 2018) and Total MCT decreased more
698 rapidly with linear age in males (van Velsen et al., 2013). Yet, our finding that TSA increased with
699 linear and quadratic age contrasts with previous reports of a decrease in surface area across the
700 lifespan (Hogstrom et al., 2013; Lemaitre et al., 2012; Long et al., 2012). Divergent results between
701 our study and those of Hogstrom and colleagues (2013) and Long and colleagues (2012) can be
702 explained by their omission of brain size, as we similarly observed a decrease of TSA with age
703 when excluding TBV from our models. However, when applying the proportion TCM adjustment,
704 as done by Lemaitre and colleagues (2012), we observed an increase in TSA with age, suggesting
705 that differences between our studies may instead stem from differences in segmentation or sample

706 characteristics (e.g., smaller sample (N = 216), wider age range (18 to 87 years old)). Based on
707 our findings, the relative expansion of TSA in older adults with age may reflect a global spread of
708 the sulci, which appears to occur more rapidly in males than females.

709 ### 3.3.2 Cerebellar Volumes

710 There was a linear decline with age of the cerebellar GMV for the FAST (β = -0.20) and
711 the Freesurfer ASEG (β = -0.02) segmentations. However, the FAST cerebellar GMV was also
712 negatively predicted by quadratic age (β = -0.05) and its linear decline with age was quicker in
713 males compared to females (β = 0.03). As for the regional FAST cerebellar GMVs, 27 out of 28
714 (96%) cerebellar volumes decreased linearly with age. The age effect in the cerebellar IX vermis
715 (β = -0.02) did not reach significance. Linear age effects ranged from -0.20 (right Crus I) to -0.05
716 (vermis Crus I). We found a negative quadratic age effect in 74% (20/27) of regions with a linear
717 age effect, which ranged from -0.07 (Left Cerebellar Lobule VIIb) to -0.03 (Right Cerebellar
718 Lobule VIIb). The cerebellar IX vermis was the only area with a quadratic age but no linear age
719 effect (β = -0.04).

720 The linear decrease with age of the cerebellum mirrors previous findings (for review
721 Bernard and Seidler, 2014). The literature also similarly reports a linear rather than a non-linear
722 cerebellar change with age (for review Fjell et al., 2013; Fjell & Walhovd, 2010) and the absence
723 of an age by sex interaction in cerebellar volumes (Hoogendam et al., 2012; Raz et al., 2005) when
724 examining the Freesurfer segmentation of the cerebellum. The discrepancies in results between
725 cerebellar segmentations further highlight the nonnegligible impact that the type of cerebellar
726 segmentation algorithm has on reported results (as discussed in section 3.2.2 and 3.2.3).

727 Although our findings add to the scarce literature on age related changes within the
728 cerebellum, age effects in these GMVs remain highly inconsistent (Bernard & Seidler, 2013; S.
729 Han, An, et al., 2020; Koppelmans et al., 2017). We speculate that these differences in reported
730 results can be attributed to the insufficient sample size of previous studies to investigate the these
731 effects (our median |β|= 0.17; N = 54 for Bernard and Seidler (2013) and N = 213, for Koppelmans
732 et al. 2017) and differences in segmentation algorithms, which vary in accuracy and in the number
733 of segmented cerebellar regions (Carass et al., 2018; L. Han et al., 2019).

734 ### 3.3.3 Whole Subcortical and Subcortical Subsegmentation Volumes

735 The putamen (β = -0.09 for both), accumbens area (Left β = -0.33, Right β = -0.24),
736 amygdala (Left β = -0.16, Right β = -0.14), and hippocampus (Left β = -0.23, Right β = -0.19)

737 decreased with age. The pallidum (Left $\beta = 0.03$, Right $\beta = 0.06$) and caudate (Left $\beta = 0.12$, Right
738 $\beta = 0.15$) volumes increased with linear age, whereas the thalamus did not show significant linear
739 age effects. The accumbens area, amygdala, and hippocampus volumes had a negative quadratic
740 age effect, ranging from -0.03 (right accumbens area) to -0.09 (left presubiculum head), and the
741 thalamus, caudate, and right putamen all showed positive quadratic age effects, ranging from 0.02
742 (right putamen) to 0.06 (left caudate).

743 Linear age was a significant predictor of 95% (104/110) of subsegmentations and quadratic
744 age was a significant predictor of 88 (104/110) subsegmentations. Although there were no age
745 effects at the whole subcortical level of the thalamus, we found 24 linear age effects with an
746 absolute effect size greater than 0.1 across thalamic subsegmentations. More positive linear and
747 quadratic age effects were found in ventral and intralaminar thalamic volumes. On the other hand,
748 the majority of the amygdala and the hippocampal subsegmentations (>92%) decreased with age.
749 The direction of the quadratic age effects on subcortical subsegmentations was similar to that of
750 the whole subcortical volumes, as we found negative quadratic age effects across the amygdala
751 and hippocampal subsegmentations and positive quadratic age effects across the thalamic
752 subsegmentations.

753 The volumetric decline in the amygdala, hippocampus, putamen, and nucleus accumbens
754 with age mirrors previous findings (Hogstrom et al., 2013; Kurth et al., 2017; Sele et al., 2020;
755 Vinke et al., 2018). However, our finding of an increase in pallidum volume with age contrasts
756 with past studies reporting a small decrease with age in this region (Sele et al., 2020; Vinke et al.,
757 2018). Moreover, while we add to the literature reporting an expansion of the caudate with age
758 (Vinke et al., 2018), a small decrease has also been reported (Sele et al., 2020). In terms of the
759 discrepancy in the caudate results, we speculate that Sele and colleagues' (2020) sample ($N = 231$)
760 was insufficient to observe such small changes with age ($\beta = 0.05$-0.06). As for the discrepancies
761 in the pallidum results, differences may be attributed to the different terms included in the
762 regressions, as Vinke and colleagues (2018) modeled non-linear changes with splines instead of a
763 quadratic age.

764 ### 3.3.4 Cortical Regions

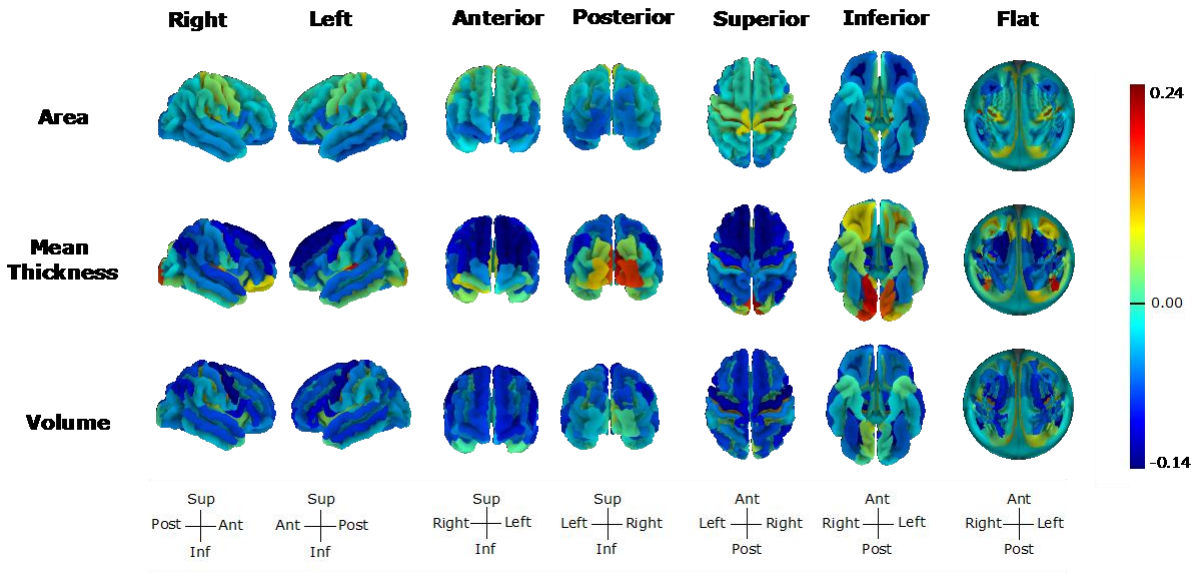
765 Cortical regions increased with linear age in 22% of volumes, 35% of mean thicknesses,
766 24% of surface areas, and declined with linear age in 41% of volumes, 52% of mean thicknesses,
767 33% of surface areas (Figure 3). Cortical volumes decreased linearly with age in 60 regions,

768 ranging from -0.13 (right precentral gyrus) to -0.02 (right parieto-occipital sulcus), and increased
769 with age in 33 regions, ranging from 0.03 (right superior segment of circular sulcus of the insula)
770 to 0.13 (left central sulcus). Cortical mean thicknesses decreased with age in 77 regions, ranging
771 from -0.14 (right inferior circular sulcus of the insula) to -0.02 (left posterior dorsal cingulate
772 gyrus), and increased with age in 52 regions, ranging from 0.02 (left middle occipital gyrus) to
773 0.22 (right cuneus gyrus). Surface areas increased linearly with age in 36 regions, ranging from
774 0.02 (left superior frontal gyrus) to 0.15 (right central sulci), and decreased with age in 49 regions,
775 ranging from -0.11 (right the orbital sulci) to -0.02 (left lateral aspect of the superior temporal
776 gyrus). In contrast, a positive quadratic age effect was reported in 5% of volumes, 6% of mean
777 thicknesses, 3% of surface areas, and a negative quadratic age effect was reported in 1% of
778 volumes, 9% of mean thicknesses, 3% of surface areas. For further details on cortical age effects,
779 see Supplemental Word Document Section 5.3.

780 Our findings coincide with and extend the literature reporting large age-related changes
781 across the cortical measures (e.g., Fjell et al., 2009; Lotze et al., 2019; Pintzka et al., 2015; Salat
782 et al., 2004; Storsve et al., 2014) The majority of frontal volumes decreased with linear age, while
783 frontal surface areas decreased with age in the orbital gyri and sulci and the inferior frontal gyrus
784 and frontal mean thicknesses decreased with age in the frontal medial and frontal superior regions.
785 Temporal surface areas and volumes generally decreased with linear age, whereas age effects on
786 temporal mean thicknesses were more variable. For instance, we observed a mean thickness
787 thinning in the lateral aspect of the superior temporal gyrus and the middle temporal gyrus and a
788 thickness increase of the planum polare, Heschl's gyrus, lingual gyrus, and the temporal pole.
789 Occipital regions generally decreased with age in surface areas and volumes, and increased with
790 age in mean thicknesses, while parietal regions mainly decreased with age across cerebral
791 measures.

792 Our findings additionally shed a light on the inconsistent age-related changes reported in
793 motor, somatosensory, and visual cortices (Hogstrom et al., 2013; for review McGinnis et al.,
794 2011). In terms of motor and somatosensory cortices, we found the largest surface area expansions
795 in the paracentral lobule and sulcus and the precentral and postcentral gyrus (motor and
796 somatosensory cortices) and a reduction of a similar size in these regions with age in terms of
797 mean thicknesses and volumes. As for the visual cortices, the largest mean thickness expansions
798 with age occurred in the cuneus gyrus, the lingual gyrus, occipital pole, the inferior, middle, and

799 superior, occipital gyri. The surface areas of these regions generally decreased with age and their
800 volumes were either unaffected by age or showed a slight increase with age.
801



802

803 ***Figure 3.*** *Linear Age Effects across Cortical Measures.* Age effects ranged from -0.40 (mean
804 thickness of the right inferior segment of the circular sulcus of the insula) to 0.22 (mean
805 thickness of the right cuneus gyrus, O6). The flat representation corresponds to the flattened
806 image of the superior view with the midline of the circle reflecting regions within the sagittal
807 plane and circle edges reflecting inferior regions. Figures made with
808 https://neuroanatomy.github.io/cortex/ (Toro, 2020).

Abbreviations: Total Cerebral Measures (TCMs): Total Brain Volume (TBV), Total Mean Cortical Thickness (MCT), or Total Surface Area (TSA). Grey Matter Volume (GMV). White Matter Volume (WMV).

3.3.5 Conclusion of Age Effects

We observed a linear change with age in 76% (480/629) and a quadratic one in 25% (159/629) of regional cerebral measures. About 49% of regions decreased with linear age and 28% increased with linear age. Regions showing quadratic age effects typically showed linear age effects. A sex by age interaction was observed for regional measures in 14% of regions (87/629), ranging from -0.16 (right cerebellar lobule X) to 0.19 (left fimbria). For detailed results on the ASEG and Freesurfer subsegmentations linear and quadratic age effects, see Supplemental Info 5.3 and for detailed results and a discussion of the sex by age and sex by age^2 interactions see Supplemental Info 5.4.

3.4. Does Brain Allometry Influence Reported Results?

We examined the effects of omitting brain allometry and adjusting for TCM with different methods. To do so, we compared the number of significant results from our main analyses obtained with the allometric TCM adjustment to those obtained when using the linear covariate or proportion TCM adjustment. The models with the linear covariate TCM adjustment yielded similar significant effects and interactions (i.e., age and sex effects and their interactions with and without TCM) to the allometric models: The linear covariate model under- or overestimated effects in 2.35% (102/4340) of statistical tests and these differences occurred in regions near significance with small effect sizes (Supplemental Figures File 3 and 4, https://osf.io/s4qc5/?view_only=bb067d96d0df4ae4902f99747d60e828). In contrast, the proportion adjustment for TCM overestimated 14.24% (618/4340) of effects and interactions reported in our main analyses compared to the allometric TCM adjustment (see Supplemental Info 6.2.1 and Supplemental Table F1 for details). In our replication of the FIRST subcortical and Desikan-Killiany Cortical sex differences reported by Ritchie and colleagues (2018), the models with the linear covariate and the allometric approach additionally yielded consistent results, except for the sex differences of the left transverse temporal and the left pars opercularis volumes (β =0.02 for both), which only reached significance in the allometric model. Finally, based on our correlational analyses of a region's deviance from isometry (i.e., |1 - scaling coefficient|) and the difference in the effect size of a term between models with varying TCM adjustments (e.g., |Proportion Sex β - Allometric Sex β|), we found that discrepancies in significance between the linear and allometric models were accentuated in more allometric regions, specifically for the

proportion models (see Supplemental Info 6.2.3, Supplemental Tables F5, and Supplemental Figures File 5 https://osf.io/s4qc5/?view_only=bb067d96d0df4ae4902f99747d60e828).

Sex differences in variance differed in 38.06% (236/620) of regions between the allometric and the linear covariate approach and in 29.03% (180/620) of regions between the allometric and the proportion approach. These discrepancies additionally lead to a change in the direction of the correlation between a region's sex effect standardized beta and variance ratio. When adjusting for TCM with the proportion or the linear covariate approach, regions that were larger in males became more variable in males, instead of being more variable in females (see Supplemental Info 6.2.2 and Supplemental Tables F2-4). We found a similar change in the direction of the correlations between the sex effect Cohen's d and variance ratios in our replication of Ritchie and colleagues' (2018) study with the cortical Deskian-Killiany and subcortical FIRST segmentations (Supplemental Table G10).

In line with previous research, we find more consistent results between the linear covariate and allometric approach compared to the proportion and allometric approach (Mankiw et al., 2017; Reardon et al., 2016; Sanchis-Segura et al., 2019). Our findings additionally suggest that the major source of variation in mean results across models with differing TCM adjustments is due to the omission of the intercept of the relationship between a region and its TCM (as in the proportion method) rather than the omission of its non-linear relationship (as in the proportion and linear covariate methods). Moreover, as the first study to examine the effects of omitting brain allometry on sex differences in variance, we find that omitting brain allometry leads to over or underestimating sex differences in variance depending on the region. Therefore, we suggest that brain allometry generally be considered to provide unbiased estimates of age, sex, and TBV effects and interactions across all brain regions, and stress that previous reported sex differences in variance relative to brain size be reexamined with brain allometry (Ritchie et al., 2018; Wierenga et al., n.d., 2018, 2019).

3.5 Neuroanatomical Norms

Neuroanatomical norms were generated for 40 028 UK Biobank participants at two levels: for global brain measures (total cerebral and cerebellar GM and WM volumes, TSA, Total MCT, and total subcortical and brainstem volumes) relative to TBV, and for each regional measure relative to its corresponding TCM (i.e., TBV, TSA, or Total MCT). These norms correspond to

the residuals from the full statistical models and reflect the extent to which the cerebral measures of an individual deviates from other individuals of the same sex, age, and total brain size. In addition to creating global and regional neuroanatomical norms for each of the 629 brain measures, we computed global neuroanatomical deviance markers, which reflect the deviance of an individual from the norm across all brain regions. This was done separately for volumes, surfaces and thicknesses, as well as across the three types of measures.

An individual with a "normal" brain - where all regional measures are at the expected value given this individual's sex, age, and total cerebral size - should have a global neuroanatomical deviance of 0 for volume, surface and thickness. However, values ranged from 0.57 to 1.53 for volumes, 0.02 to 0.05 for mean thicknesses, and 0.04 to 0.13 for surface areas, suggesting that it is 'normal' to deviate from the cerebral norms (Table 3). In turn, individuals with a global neuroanatomical deviance above the mean had brain regions that deviated from the regional norms more than most individuals, while values below the mean represent reduced deviance from the neuroanatomical norm.

The means of the volumetric, surface area, and neuroanatomical (all cerebral measures combined) global neuroanatomical deviance were larger in males, suggesting that male volumes and surface areas deviated more from their sex-specific norm than females. However, females deviated more from their norm in the mean thickness global allometry marker. Males additionally had a more variable volumetric and neuroanatomical global allometric markers (Table 3, Figure 4). Thus, investigations of global as well as regional neuroanatomical deviance should always take sex into account.

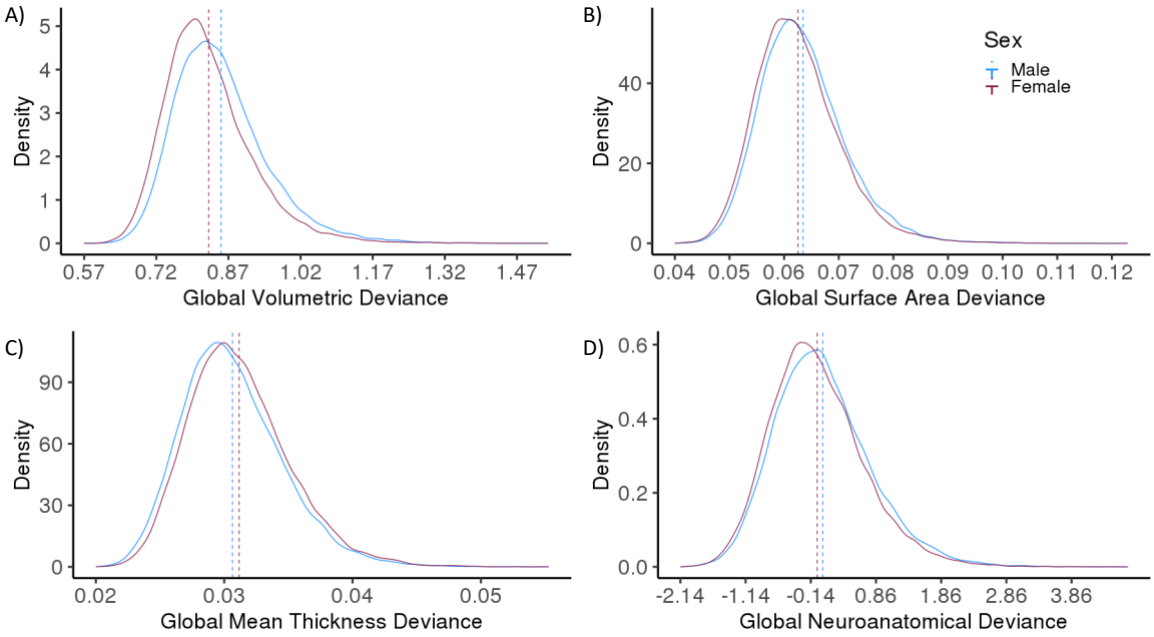
892 Table 3. Mean and Variance Sex Differences in across Global Deviance Markers

Global Deviance Markers for	**Min**	**Max**	**Male**		**Female**		**d**	**p**	**VR**	**p**
			μ	***SD***	**μ**	***SD***				
Volumes	0.57	1.53	0.85	0.96	0.82	0.92	0.27	2.20E-16	0.95	1.56E-12
Mean Thicknesses	0.02	0.05	0.03	0.04	0.03	0.04	0.13	2.20E-16	1.00	0.712
Surface Areas	0.04	0.13	0.07	0.01	0.07	0.01	0.12	2.20E-16	0.98	0.027
All Measures	-2.14	4.74	0.05	0.74	-0.04	0.71	0.12	2.20E-16	0.97	1.12E-14

893

894 *N.B.* d: Cohen's d. the Global Volumes, Mean Thicknesses, and Surface Areas Deviance Markers
895 are calculated from equation X and the Global Neuroanatomical Deviance Marker (All Measures)
896 corresponds to the average of the Z-score of the 3 Global Deviance Markers. Significance set to
897 0.05/6, as 6 tests were performed. Global volumetric deviance was positively correlated with
898 global surface area deviance ($r = 0.35$, $p < 2.2e\text{-}16$) and global mean thickness deviance ($r = 0.33$,
899 $p < 2.2e\text{-}16$). Global surface deviance and global mean thickness deviance were also positively
900 correlated ($r = 0.17$, $p < 2.2e\text{-}16$).
901

Abbreviations: Total Cerebral Measures (TCMs): Total Brain Volume (TBV), Total Mean Cortical Thickness (MCT), or Total Surface Area (TSA). Grey Matter Volume (GMV). White Matter Volume (WMV).



902
903 ***Figure 4.*** *Sex Differences in Global Neuroanatomical Deviances across Volumes (A), Cortical*
904 *Surface Areas (B), Cortical Mean Thicknesses (C), and all Volumes, Cortical Surface Areas, and*
905 *Cortical Mean Thicknesses(D).* Mean (dashed lines) differences were found across measures,
906 while variance differed between sexes only across volumes. Global Allometry marker corresponds
907 to the square root of the sum of squared residuals divided by the number of regions for that measure
908 from the model with age, total brain volume, and sex as well as age^2, total brain volume, and sex
909 interactions. Global Neuroanatomical Deviance Marker corresponds to the average of the Z-score
910 of the 3 Global Deviance Markers. Significance set to 0.05/6, as 6 tests were performed.
911

Abbreviations: Total Cerebral Measures (TCMs): Total Brain Volume (TBV), Total Mean Cortical Thickness (MCT), or Total Surface Area (TSA). Grey Matter Volume (GMV). White Matter Volume (WMV).

With the world's largest neuroimaging dataset, we created neuroanatomical norms in the UK Biobank, to which any UK Biobank individual can be compared, in the same way that scores from intelligence tests or personality questionnaires can be compared to population norms. With these norms, future UK Biobank studies will be able to examine whether individuals that deviate from the norm on a global or regional brain measure also deviate from the norm in terms of cognitive and behavioral traits or of risk for neurological and psychiatric disorders. Having brain markers that are relative to total cerebral measures (rather than raw measures) will make it easier to distinguish the specific contribution of each regional brain measure from that of more global brain measures. As for global neuroanatomical deviance markers, future studies will be able to investigate the extent to which global neuroanatomical deviance reflects disruptions of brain development or serves as a risk factor for neurodevelopmental or psychiatric disorders. For studies examining the associations of specific regional brain measures with cognitive phenotypes, it may be useful to adjust on global neuroanatomical deviance, on top of total brain size, in order to fully dissociate regional from global effects.

3.6 Limitations

In light of the "healthy volunteer" selection bias and older age range of the UK Biobank (Fry et al., 2017), the present paper is limited in its capacity to generalize its findings and neuroanatomical norms and markers to other age groups or to the UK population. Moreover, created these norms and markers are not independent of ethnicity or highest level of education attained, factors thought to influence neuroanatomical measures (Shen et al., 2017; Tang et al., 2018). However, this enables future studies to investigate whether the present neuroanatomical markers vary as a function of ethnicity and level of education. If the aim is to generalize these findings to the UK population, we suggest that neuroanatomical markers be created with a representative sample or that weights be used to adjust the phenotypic measures of the UK Biobank to match those of the UK population.

While numerous studies focus on developing machine learning algorithms (e.g., SVM classifications) to generate neuroanatomical markers, we were interested in identifying sex, age, and TCM effects while considering their potential interactions, which are often omitted in the literature. By opting for a regression approach, we were able to quantify effects and interactions for each region, which would have been lost with machine learning. Future studies would nevertheless benefit from examining the age, sex, and global brain effects of other anatomical measures, such as diffusion tractography, and functional measures in the UK Biobank.

While the present study modelled age as a linear and quadratic function, other studies examining sex and age interactions used the nonparametric local smoothing technique (i.e., smoothing splines; Fjell et al., 2013; Vinke et al., 2018), which are thought to be more predictive of individual trajectories and less vulnerable to sampling range (Fjell & Walhovd, 2010). However, this nonparametric approach also requires researchers to make more decisions that contribute to the variability of results. For instance, Fjell and colleagues (2010) initially selected the spline smoothing level that minimized the AIC, but when the absence of smoothing (which is equivalent to the linear least square model) yielded the smallest AIC for the sample of individuals over 60 years old, they chose the smoothing level that minimized the BIC. Considering that the nonparametric local smoothing technique depends on selected parameters and that splines are difficult to interpret if we are interested in quantifying the magnitude of age effects, quadratic age was used instead of splines to model non-linear age in the present study.

Abbreviations: Total Cerebral Measures (TCMs): Total Brain Volume (TBV), Total Mean Cortical Thickness (MCT), or Total Surface Area (TSA). Grey Matter Volume (GMV). White Matter Volume (WMV).

4. Conclusion

The present study is the largest analysis to date of the age, sex, and TCM effects and interactions on global and regional brain volumes, cortical mean thicknesses, and cortical surface areas. We provide further evidence that brain allometry is a common property of the brain that should be considered to report unbiased estimates of age, sex, and TCM effects and interactions. By generating volumetric and allometric norms in the UK Biobank, we pave the way for future research to examine the associations between these markers and the behavioral and cognitive traits available in the UK Biobank. Once associations between these UK Biobank neuroanatomical norms and cognitive and behavioral measures are established, researchers will be able to examine the extent to which these neuroanatomical markers mediate the effect that genes and the environment have on these traits. This line of research will play critical role in our understanding of the influence that neuroanatomy has on who we are.

5. References

Alfaro-Almagro, F., Jenkinson, M., Bangerter, N. K., Andersson, J. L. R., Griffanti, L., Douaud, G., Sotiropoulos, S. N., Jbabdi, S., Hernandez-Fernandez, M., Vallee, E., Vidaurre, D., Webster, M., McCarthy, P., Rorden, C., Daducci, A., Alexander, D. C., Zhang, H., Dragonu, I., Matthews, P. M., … Smith, S. M. (2018). Image processing and Quality Control for the first 10,000 brain imaging datasets from UK Biobank. *NeuroImage*, *166*, 400–424. https://doi.org/10.1016/j.neuroimage.2017.10.034

Beam, C. R., Kaneshiro, C., Jang, J. Y., Reynolds, C. A., Pedersen, N. L., & Gatz, M. (2018). Differences Between Women and Men in Incidence Rates of Dementia and Alzheimer's Disease. *Journal of Alzheimer's Disease : JAD*, *64*(4), 1077–1083. https://doi.org/10.3233/JAD-180141

Beck, A. T., Steer, R. A., & Brown, G. (1996). Beck depression inventory–II. *Psychological Assessment*.

Bernard, J. A., & Seidler, R. D. (2013). Relationships Between Regional Cerebellar Volume and Sensorimotor and Cognitive Function in Young and Older Adults. *The Cerebellum*, *12*(5), 721–737. https://doi.org/10.1007/s12311-013-0481-z

Boyd, A., Van de Velde, S., Vilagut, G., de Graaf, R., O'Neill, S., Florescu, S., Alonso, J., & Kovess-Masfety, V. (2015). Gender differences in mental disorders and suicidality in Europe: Results from a large cross-sectional population-based study. *Journal of Affective Disorders*, *173*, 245–254. https://doi.org/10.1016/j.jad.2014.11.002

Branch, C. L., Sonnenberg, B. R., Pitera, A. M., Benedict, L. M., Kozlovsky, D. Y., Bridge, E. S., & Pravosudov, V. V. (2020). Testing the greater male variability phenomenon: Male mountain chickadees exhibit larger variation in reversal learning performance compared with females. *Proceedings of the Royal Society B: Biological Sciences*, *287*(1931), 20200895. https://doi.org/10.1098/rspb.2020.0895

Carass, A., Cuzzocreo, J. L., Han, S., Hernandez-Castillo, C. R., Rasser, P. E., Ganz, M., Beliveau, V., Dolz, J., Ayed, I. B., Desrosiers, C., Thyreau, B., Romero, J. E., Coupé, P., Manjón, J. V., Fonov, V. S., Collins, D. L., Ying, S. H., Onyike, C. U., Crocetti, D., … Prince, J. L. (2018). Comparing fully automated state-of-the-art cerebellum parcellation

Abbreviations: Total Cerebral Measures (TCMs): Total Brain Volume (TBV), Total Mean Cortical Thickness (MCT), or Total Surface Area (TSA). Grey Matter Volume (GMV). White Matter Volume (WMV).

from magnetic resonance images. *NeuroImage*, *183*, 150–172. https://doi.org/10.1016/j.neuroimage.2018.08.003

Chen, X., Sachdev, P. S., Wen, W., & Anstey, K. J. (2007). Sex differences in regional gray matter in healthy individuals aged 44–48 years: A voxel-based morphometric study. *NeuroImage*, *36*(3), 691–699. https://doi.org/10.1016/j.neuroimage.2007.03.063

Costa Jr., P. T., & McCrae, R. R. (2008). The Revised NEO Personality Inventory (NEO-PI-R). In *The SAGE handbook of personality theory and assessment, Vol 2: Personality measurement and testing* (pp. 179–198). Sage Publications, Inc. https://doi.org/10.4135/9781849200479.n9

Cox, Simon R., Bastin, M. E., Ritchie, S. J., Dickie, D. A., Liewald, D. C., Muñoz Maniega, S., Redmond, P., Royle, N. A., Pattie, A., Valdés Hernández, M., Corley, J., Aribisala, B. S., McIntosh, A. M., Wardlaw, J. M., & Deary, I. J. (2018). Brain cortical characteristics of lifetime cognitive ageing. *Brain Structure and Function*, *223*(1), 509–518. https://doi.org/10.1007/s00429-017-1505-0

Cox, S.R., Ritchie, S. J., Fawns-Ritchie, C., Tucker-Drob, E. M., & Deary, I. J. (2019). Structural brain imaging correlates of general intelligence in UK Biobank. *Intelligence, 76*. https://doi.org/10.1016/j.intell.2019.101376

Dallaire-Théroux, C., Callahan, B. L., Potvin, O., Saikali, S., & Duchesne, S. (2017). Radiological-Pathological Correlation in Alzheimer's Disease: Systematic Review of Antemortem Magnetic Resonance Imaging Findings. *Journal of Alzheimer's Disease*, *57*(2), 575–601. https://doi.org/10.3233/JAD-161028

Deary, I. J. (2010). Cognitive epidemiology: Its rise, its current issues, and its challenges. *Personality and Individual Differences*, *49*(4), 337–343. https://doi.org/10.1016/j.paid.2009.11.012

DeCasien, A. R., Sherwood, C. C., Schapiro, S. J., & Higham, J. P. (2020). Greater variability in chimpanzee (Pan troglodytes) brain structure among males. *Proceedings of the Royal Society B: Biological Sciences*, *287*(1925), 20192858. https://doi.org/10.1098/rspb.2019.2858

Diedrichsen, J., Balsters, J. H., Flavell, J., Cussans, E., & Ramnani, N. (2009). A probabilistic MR atlas of the human cerebellum. *NeuroImage*, *46*(1), 39–46. https://doi.org/10.1016/j.neuroimage.2009.01.045

Abbreviations: Total Cerebral Measures (TCMs): Total Brain Volume (TBV), Total Mean Cortical Thickness (MCT), or Total Surface Area (TSA). Grey Matter Volume (GMV). White Matter Volume (WMV).

1029 Ellis, H. (1894). The Variational Tendency of Men. In *Man and woman: A study of human*
1030 *secondary sexual characters* (1st ed., pp. 358–372). Walter. Scott.
1031 https://archive.org/details/manandwomanastu00elligoog/page/n16/mode/2up
1032 Finlay, B. L., Darlington, R. B., & Nicastro, N. (2001). Developmental structure in brain
1033 evolution. *The Behavioral and Brain Sciences*, *24*(2), 263–278; discussion 278-308.
1034 https://doi.org/10.1017/S0140525X01003958
1035 Fish, A. M., Cachia, A., Fischer, C., Mankiw, C., Reardon, P. K., Clasen, L. S., Blumenthal, J.
1036 D., Greenstein, D., Giedd, J. N., Mangin, J.-F., & Raznahan, A. (2017). Influences of
1037 Brain Size, Sex, and Sex Chromosome Complement on the Architecture of Human
1038 Cortical Folding. *Cerebral Cortex (New York, N.Y.: 1991)*, *27*(12), 5557–5567.
1039 https://doi.org/10.1093/cercor/bhw323
1040 Fiske, A., Wetherell, J. L., & Gatz, M. (2009). Depression in Older Adults. *Annual Review of*
1041 *Clinical Psychology*, *5*(1), 363–389.
1042 https://doi.org/10.1146/annurev.clinpsy.032408.153621
1043 Fjell, A. M., & Walhovd, K. B. (2010). Structural brain changes in aging: Courses, causes and
1044 cognitive consequences. *Reviews in the Neurosciences*, *21*(3), 187–221.
1045 https://doi.org/10.1515/revneuro.2010.21.3.187
1046 Fjell, A. M., Westlye, L. T., Amlien, I., Espeseth, T., Reinvang, I., Raz, N., Agartz, I., Salat, D.
1047 H., Greve, D. N., Fischl, B., Dale, A. M., & Walhovd, K. B. (2009). Minute effects of sex
1048 on the aging brain: A multisample magnetic resonance imaging study of healthy aging
1049 and Alzheimer's disease. *The Journal of Neuroscience: The Official Journal of the*
1050 *Society for Neuroscience*, *29*(27), 8774–8783.
1051 https://doi.org/10.1523/JNEUROSCI.0115-09.2009
1052 Fjell, A. M., Westlye, L. T., Grydeland, H., Amlien, I., Espeseth, T., Reinvang, I., Raz, N.,
1053 Holland, D., Dale, A. M., & Walhovd, K. B. (2013). Critical ages in the life course of the
1054 adult brain: Nonlinear subcortical aging. *Neurobiology of Aging*, *34*(10), 2239–2247.
1055 https://doi.org/10.1016/j.neurobiolaging.2013.04.006
1056 Fox, J., Weisberg, S., Price, B., Adler, D., Bates, D., Baud-Bovy, G., Bolker, B., Ellison, S.,
1057 Firth, D., Friendly, M., Gorjanc, G., Graves, S., Heiberger, R., Krivitsky, P., Laboissiere,
1058 R., Maechler, M., Monette, G., Murdoch, D., Nilsson, H., … R-Core. (2020). *car:*

Companion to Applied Regression (3.0-10) [Computer software]. https://CRAN.R-project.org/package=car

Fry, A., Littlejohns, T. J., Sudlow, C., Doherty, N., Adamska, L., Sprosen, T., Collins, R., & Allen, N. E. (2017). Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population. *American Journal of Epidemiology*, *186*(9), 1026–1034. https://doi.org/10.1093/aje/kwx246

Gong, Q., Scarpazza, C., Dai, J., He, M., Xu, X., Shi, Y., Zhou, B., Vieira, S., McCrory, E., Ai, Y., Yang, C., Zhang, F., Lui, S., & Mechelli, A. (2019). A transdiagnostic neuroanatomical signature of psychiatric illness. *Neuropsychopharmacology*, *44*(5), 869–875. https://doi.org/10.1038/s41386-018-0175-9

Gur, R. C., Turetsky, B. I., Matsui, M., Yan, M., Bilker, W., Hughett, P., & Gur, R. E. (1999). Sex Differences in Brain Gray and White Matter in Healthy Young Adults: Correlations with Cognitive Performance. *Journal of Neuroscience*, *19*(10), 4065–4072. https://doi.org/10.1523/JNEUROSCI.19-10-04065.1999

Han, L., Chen, L., Bin, Y., Cun, D., Yu-chen, H., & Xin, J. (2019). Deep Feature Combination Based Multi-Model Wind Power Prediction. *2019 IEEE 2nd International Conference on Computer and Communication Engineering Technology (CCET)*, 143–148. https://doi.org/10.1109/CCET48361.2019.8989358

Han, S., An, Y., Carass, A., Prince, J. L., & Resnick, S. M. (2020). Longitudinal Analysis of Regional Cerebellum Volumes During Normal Aging. *NeuroImage*, 117062. https://doi.org/10.1016/j.neuroimage.2020.117062

Han, S., Carass, A., He, Y., & Prince, J. L. (2020). Automatic cerebellum anatomical parcellation using U-Net with locally constrained optimization. *NeuroImage*, *218*, 116819. https://doi.org/10.1016/j.neuroimage.2020.116819

Hogstrom, L. J., Westlye, L. T., Walhovd, K. B., & Fjell, A. M. (2013). The Structure of the Cerebral Cortex Across Adult Life: Age-Related Patterns of Surface Area, Thickness, and Gyrification. *Cerebral Cortex*, *23*(11), 2521–2530. https://doi.org/10.1093/cercor/bhs231

Hoogendam, Y. Y., van der Geest, J. N., van der Lijn, F., van der Lugt, A., Niessen, W. J., Krestin, G. P., Hofman, A., Vernooij, M. W., Breteler, M. M. B., & Ikram, M. A. (2012).

1090 Determinants of cerebellar and cerebral volume in the general elderly population.
1091 *Neurobiology of Aging*, *33*(12), 2774–2781.
1092 https://doi.org/10.1016/j.neurobiolaging.2012.02.012
1093 Hurtz, S., Woo, E., Kebets, V., Green, A. E., Zoumalan, C., Wang, B., Ringman, J. M.,
1094 Thompson, P. M., & Apostolova, L. G. (2014). Age Effects on Cortical Thickness in
1095 Cognitively Normal Elderly Individuals. *Dementia and Geriatric Cognitive Disorders*
1096 *EXTRA*, *4*(2), 221–227. https://doi.org/10.1159/000362872
1097 Im, K., Lee, J.-M., Lyttelton, O., Kim, S. H., Evans, A. C., & Kim, S. I. (2008). Brain Size and
1098 Cortical Structure in the Adult Human Brain. *Cerebral Cortex*, *18*(9), 2181–2191.
1099 https://doi.org/10.1093/cercor/bhm244
1100 Jäncke, L., Liem, F., & Merillat, S. (2019). Scaling of brain compartments to brain size.
1101 *Neuroreport*, *30*(8), 573–579. https://doi.org/10.1097/WNR.0000000000001249
1102 Jäncke, L., Mérillat, S., Liem, F., & Hänggi, J. (2015a). Brain size, sex, and the aging brain.
1103 *Human Brain Mapping*, *36*(1), 150–169. https://doi.org/10.1002/hbm.22619
1104 Jäncke, L., Mérillat, S., Liem, F., & Hänggi, J. (2015b). Brain size, sex, and the aging brain.
1105 *Human Brain Mapping*, *36*(1), 150–169. https://doi.org/10.1002/hbm.22619
1106 Jellinger, K. A., & Attems, J. (2015). Challenges of multimorbidity of the aging brain: A critical
1107 update. *Journal of Neural Transmission (Vienna, Austria: 1996)*, *122*(4), 505–521.
1108 https://doi.org/10.1007/s00702-014-1288-x
1109 Johnson, W., Carothers, A., & Deary, I. J. (2008). Sex Differences in Variability in General
1110 Intelligence: A New Look at the Old Question. *Perspectives on Psychological Science: A*
1111 *Journal of the Association for Psychological Science*, *3*(6), 518–531.
1112 https://doi.org/10.1111/j.1745-6924.2008.00096.x
1113 Jong, L. W. de, Vidal, J.-S., Forsberg, L. E., Zijdenbos, A. P., Haight, T., Sigurdsson, S.,
1114 Gudnason, V., Buchem, M. A. van, & Launer, L. J. (2017). Allometric scaling of brain
1115 regions to intra-cranial volume: An epidemiological MRI study. *Human Brain Mapping*,
1116 *38*(1), 151–164. https://doi.org/10.1002/hbm.23351
1117 Ju, C., Duan, Y., & You, X. (2015). Retesting the greater male variability hypothesis in mainland
1118 China: A cross-regional study. *Personality and Individual Differences*, *72*, 85–89.
1119 https://doi.org/10.1016/j.paid.2014.07.021

Abbreviations: Total Cerebral Measures (TCMs): Total Brain Volume (TBV), Total Mean Cortical Thickness (MCT), or Total Surface Area (TSA). Grey Matter Volume (GMV). White Matter Volume (WMV).

Jumah, F., Ghannam, M., Jaber, M., Adeeb, N., & Tubbs, R. S. (2016). Neuroanatomical variation in autism spectrum disorder: A comprehensive review. *Clinical Anatomy*, *29*(4), 454–465. https://doi.org/10.1002/ca.22717

Kaczkurkin, A. N., Raznahan, A., & Satterthwaite, T. D. (2019). Sex differences in the developing brain: Insights from multimodal neuroimaging. *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology*, *44*(1), 71–85. https://doi.org/10.1038/s41386-018-0111-z

Karwowski, M., Jankowska, D. M., Gajda, A., Marczak, M., Groyecka, A., & Sorokowski, P. (2016). Greater Male Variability in Creativity Outside the WEIRD World. *Creativity Research Journal*, *28*(4), 467–470. https://doi.org/10.1080/10400419.2016.1229978

Koppelmans, V., Hoogendam, Y. Y., Hirsiger, S., Mérillat, S., Jäncke, L., & Seidler, R. D. (2017). Regional cerebellar volumetric correlates of manual motor and cognitive function. *Brain Structure and Function*, *222*(4), 1929–1944. https://doi.org/10.1007/s00429-016-1317-7

Kurth, F., Cherbuin, N., & Luders, E. (2017). The impact of aging on subregions of the hippocampal complex in healthy adults. *NeuroImage*, *163*, 296–300. https://doi.org/10.1016/j.neuroimage.2017.09.016

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

Lehre, A.-C., Lehre, K. P., Laake, P., & Danbolt, N. C. (2009). Greater intrasex phenotype variability in males than in females is a fundamental aspect of the gender differences in humans. *Developmental Psychobiology*, *51*(2), 198–206. https://doi.org/10.1002/dev.20358

Lemaitre, H., Goldman, A., Sambataro, F., Verchinski, B., Meyer-Lindenberg, A., Weinberger, D., & Mattay, V. (2012). Normal age-related brain morphometric changes: Nonuniformity across cortical thickness, surface area and grey matter volume? *Neurobiology of Aging*, *33*(3), 617.e1-617.e9. https://doi.org/10.1016/j.neurobiolaging.2010.07.013

Abbreviations: Total Cerebral Measures (TCMs): Total Brain Volume (TBV), Total Mean Cortical Thickness (MCT), or Total Surface Area (TSA). Grey Matter Volume (GMV). White Matter Volume (WMV).

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*. https://doi.org/10.3389/fnins.2014.00356

Long, X., Liao, W., Jiang, C., Liang, D., Qiu, B., & Zhang, L. (2012). Healthy Aging: An Automatic Analysis of Global and Regional Morphological Alterations of Human Brain. *Academic Radiology*, *19*(7), 785–793. https://doi.org/10.1016/j.acra.2012.03.006

Lotze, M., Domin, M., Gerlach, F. H., Gaser, C., Lueders, E., Schmidt, C. O., & Neumann, N. (2019). Novel findings from 2,838 Adult Brains on Sex Differences in Gray Matter Brain Volume. *Scientific Reports*, *9*(1), 1671. https://doi.org/10.1038/s41598-018-38239-2

Luders, E., Gaser, C., Narr, K. L., & Toga, A. W. (2009). Why Sex Matters: Brain Size Independent Differences in Gray Matter Distributions between Men and Women. *The Journal of Neuroscience*, *29*(45), 14265–14270. https://doi.org/10.1523/JNEUROSCI.2261-09.2009

Malykhin, N. V., Huang, Y., Hrybouski, S., & Olsen, F. (2017). Differential vulnerability of hippocampal subfields and anteroposterior hippocampal subregions in healthy cognitive aging. *Neurobiology of Aging*, *59*, 121–134. https://doi.org/10.1016/j.neurobiolaging.2017.08.001

Mankiw, C., Park, M. T. M., Reardon, P. K., Fish, A. M., Clasen, L. S., Greenstein, D., Giedd, J. N., Blumenthal, J. D., Lerch, J. P., Chakravarty, M. M., & 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

Marek, S., Tervo-Clemmens, B., Calabro, F. J., Montez, D. F., Kay, B. P., Hatoum, A. S., Donohue, M. R., Foran, W., Miller, R. L., Feczko, E., Miranda-Dominguez, O., Graham, A. M., Earl, E. A., Perrone, A. J., Cordova, M., Doyle, O., Moore, L. A., Conan, G., Uriarte, J., … Dosenbach, N. U. F. (2020). Towards Reproducible Brain-Wide Association Studies. *BioRxiv*, 2020.08.21.257758. https://doi.org/10.1101/2020.08.21.257758

Matsuo, K., Harada, K., Fujita, Y., Okamoto, Y., Ota, M., Narita, H., Mwangi, B., Gutierrez, C. A., Okada, G., Takamura, M., Yamagata, H., Kusumi, I., Kunugi, H., Inoue, T., Soares, J. C., Yamawaki, S., & Watanabe, Y. (2019). Distinctive Neuroanatomical Substrates for

Depression in Bipolar Disorder versus Major Depressive Disorder. *Cerebral Cortex*, *29*(1), 202–214. https://doi.org/10.1093/cercor/bhx319

McGinnis, S. M., Brickhouse, M., Pascual, B., & Dickerson, B. C. (2011). Age-Related Changes in the Thickness of Cortical Zones in Humans. *Brain Topography*, *24*(3), 279. https://doi.org/10.1007/s10548-011-0198-6

Miller, K. L., Alfaro-Almagro, F., Bangerter, N. K., Thomas, D. L., Yacoub, E., Xu, J., Bartsch, A. J., Jbabdi, S., Sotiropoulos, S. N., Andersson, J. L. R., Griffanti, L., Douaud, G., Okell, T. W., Weale, P., Dragonu, I., Garratt, S., Hudson, S., Collins, R., Jenkinson, M., … Smith, S. M. (2016). Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nature Neuroscience*, *19*(11), 1523–1536. https://doi.org/10.1038/nn.4393

Morey, R. A., Petty, C. M., Xu, Y., Hayes, J. P., Wagner, H. R., Lewis, D. V., LaBar, K. S., Styner, M., & McCarthy, G. (2009). A comparison of automated segmentation and manual tracing for quantifying hippocampal and amygdala volumes. *NeuroImage*, *45*(3), 855–866. https://doi.org/10.1016/j.neuroimage.2008.12.033

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

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

Nordenskjöld, R., Malmberg, F., Larsson, E.-M., Simmons, A., Ahlström, H., Johansson, L., & Kullberg, J. (2015). Intracranial volume normalization methods: Considerations when investigating gender differences in regional brain volume. *Psychiatry Research: Neuroimaging*, *231*(3), 227–235. https://doi.org/10.1016/j.pscychresns.2014.11.011

Oakes, P., Loukas, M., Oskouian, R. J., & Tubbs, R. S. (2017). The neuroanatomy of depression: A review. *Clinical Anatomy*, *30*(1), 44–49. https://doi.org/10.1002/ca.22781

O'Brien, L. M., Ziegler, D. A., Deutsch, C. K., Frazier, J. A., Herbert, M. R., & Locascio, J. J. (2011). Statistical adjustments for brain size in volumetric neuroimaging studies: Some

Abbreviations: Total Cerebral Measures (TCMs): Total Brain Volume (TBV), Total Mean Cortical Thickness (MCT), or Total Surface Area (TSA). Grey Matter Volume (GMV). White Matter Volume (WMV).

practical implications in methods. *Psychiatry Research*, *193*(2), 113–122. https://doi.org/10.1016/j.pscychresns.2011.01.007

O'Brien, L. M., Ziegler, D. A., Deutsch, C. K., Kennedy, D. N., Goldstein, J. M., Seidman, L. J., Hodge, S., Makris, N., Caviness, V., Frazier, J. A., & Herbert, M. R. (2006). Adjustment for whole brain and cranial size in volumetric brain studies: A review of common adjustment factors and statistical methods. *Harvard Review of Psychiatry*, *14*(3), 141–151. https://doi.org/10.1080/10673220600784119

Peyre, H., Mohanpuria, N., Jednoróg, K., Heim, S., Grande, M., Ermingen-Marbach, M. van, Altarelli, I., Monzalvo, K., Williams, C. M., Germanaud, D., Toro, R., & Ramus, F. (2020). Neuroanatomy of dyslexia: An allometric approach. *European Journal of Neuroscience*, *n/a*(n/a). https://doi.org/10.1111/ejn.14690

Pintzka, C. W. S., Hansen, T. I., Evensmoen, H. R., & Håberg, A. K. (2015). Marked effects of intracranial volume correction methods on sex differences in neuroanatomical structures: A HUNT MRI study. *Frontiers in Neuroscience*, *9*, 238. https://doi.org/10.3389/fnins.2015.00238

R Core Team. (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Ramus, F., Altarelli, I., Jednoróg, K., Zhao, J., & Scotto di Covella, L. (2018). Neuroanatomy of developmental dyslexia: Pitfalls and promise. *Neuroscience & Biobehavioral Reviews*, *84*, 434–452. https://doi.org/10.1016/j.neubiorev.2017.08.001

Raz, N., Lindenberger, U., Rodrigue, K. M., Kennedy, K. M., Head, D., Williamson, A., Dahle, C., Gerstorf, D., & Acker, J. D. (2005). Regional Brain Changes in Aging Healthy Adults: General Trends, Individual Differences and Modifiers. *Cerebral Cortex*, *15*(11), 1676–1689. https://doi.org/10.1093/cercor/bhi044

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., Alexander-Bloch, A., Clasen, L. S., Blumenthal, J. D., Lalonde, F. M., Giedd, J. N., Gur, R. C., Gur, R. E.,

Abbreviations: Total Cerebral Measures (TCMs): Total Brain Volume (TBV), Total Mean Cortical Thickness (MCT), or Total Surface Area (TSA). Grey Matter Volume (GMV). White Matter Volume (WMV).

Lerch, J. P., Chakravarty, M. M., Satterthwaite, T. D., Shinohara, R. T., & 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

Ritchie, S. J., Cox, S. R., Shen, X., Lombardo, M. V., Reus, L. M., Alloza, C., Harris, M. A., Alderson, H. L., Hunter, S., Neilson, E., Liewald, D. C. M., Auyeung, B., Whalley, H. C., Lawrie, S. M., Gale, C. R., Bastin, M. E., McIntosh, A. M., & Deary, I. J. (2018). Sex Differences in the Adult Human Brain: Evidence from 5216 UK Biobank Participants. *Cerebral Cortex*, *28*(8), 2959–2975. https://doi.org/10.1093/cercor/bhy109

Ruigrok, A. N. V., Salimi-Khorshidi, G., Lai, M.-C., Baron-Cohen, S., Lombardo, M. V., Tait, R. J., & Suckling, J. (2014). A meta-analysis of sex differences in human brain structure. *Neuroscience & Biobehavioral Reviews*, *39*, 34–50. https://doi.org/10.1016/j.neubiorev.2013.12.004

Salat, D. H., Buckner, R. L., Snyder, A. Z., Greve, D. N., Desikan, R. S. R., Busa, E., Morris, J. C., Dale, A. M., & Fischl, B. (2004). Thinning of the cerebral cortex in aging. *Cerebral Cortex (New York, N.Y.: 1991)*, *14*(7), 721–730. https://doi.org/10.1093/cercor/bhh032

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

Schmidt, C. K., Khalid, S., Loukas, M., & Tubbs, R. S. (2018). Neuroanatomy of Anxiety: A Brief Review. *Cureus*, *10*(1), e2055–e2055. PubMed. https://doi.org/10.7759/cureus.2055

Schoemaker, D., Buss, C., Head, K., Sandman, C. A., Davis, E. P., Chakravarty, M. M., Gauthier, S., & Pruessner, J. (2016). Hippocampus and amygdala volumes from Magnetic Resonance Images in children: Assessing accuracy of Freesurfer and FSL against manual segmentation. *NeuroImage*, *129*, 1–14. https://doi.org/10.1016/j.neuroimage.2016.01.038

Seedat, S., Scott, K. M., Angermeyer, M. C., Berglund, P., Bromet, E. J., Brugha, T. S., Demyttenaere, K., Girolamo, G. de, Haro, J. M., Jin, R., Karam, E. G., Kovess-Masfety, V., Levinson, D., Mora, M. E. M., Ono, Y., Ormel, J., Pennell, B.-E., Posada-Villa, J., Sampson, N. A., … Kessler, R. C. (2009). Cross-National Associations Between Gender and Mental Disorders in the World Health Organization World Mental Health Surveys.

Abbreviations: Total Cerebral Measures (TCMs): Total Brain Volume (TBV), Total Mean Cortical Thickness (MCT), or Total Surface Area (TSA). Grey Matter Volume (GMV). White Matter Volume (WMV).

Archives of General Psychiatry, *66*(7), 785–795. https://doi.org/10.1001/archgenpsychiatry.2009.36

Sele, S., Liem, F., Mérillat, S., & Jäncke, L. (2020). Decline Variability of Cortical and Subcortical Regions in Aging: A Longitudinal Study. *Frontiers in Human Neuroscience*, *14*. https://doi.org/10.3389/fnhum.2020.00363

Shen, X., Reus, L. M., Cox, S. R., Adams, M. J., Liewald, D. C., Bastin, M. E., Smith, D. J., Deary, I. J., Whalley, H. C., & McIntosh, A. M. (2017). Subcortical volume and white matter integrity abnormalities in major depressive disorder: Findings from UK Biobank imaging data. *Scientific Reports*, *7*(1), 5547. https://doi.org/10.1038/s41598-017-05507-6

Storsve, A. B., Fjell, A. M., Tamnes, C. K., Westlye, L. T., Overbye, K., Aasland, H. W., & Walhovd, K. B. (2014). Differential Longitudinal Changes in Cortical Thickness, Surface Area and Volume across the Adult Life Span: Regions of Accelerating and Decelerating Change. *The Journal of Neuroscience*, *34*(25), 8488–8498. https://doi.org/10.1523/JNEUROSCI.0391-14.2014

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., Liu, B., Matthews, P., Ong, G., Pell, J., Silman, A., Young, A., Sprosen, T., Peakman, T., & Collins, R. (2015). UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLOS Medicine*, *12*(3), e1001779. https://doi.org/10.1371/journal.pmed.1001779

Szucs, D., & Ioannidis, J. P. A. (2017). When Null Hypothesis Significance Testing Is Unsuitable for Research: A Reassessment. *Frontiers in Human Neuroscience*, *11*. https://doi.org/10.3389/fnhum.2017.00390

Szucs, D., & Ioannidis, J. PA. (2020). Sample size evolution in neuroimaging research: An evaluation of highly-cited studies (1990–2012) and of latest practices (2017–2018) in high-impact journals. *NeuroImage*, *221*, 117164. https://doi.org/10.1016/j.neuroimage.2020.117164

Tan, A., Ma, W., Vira, A., Marwha, D., & Eliot, L. (2016). The human hippocampus is not sexually-dimorphic: Meta-analysis of structural MRI volumes. *NeuroImage*, *124*(Pt A), 350–366. https://doi.org/10.1016/j.neuroimage.2015.08.050

Tang, Y., Zhao, L., Lou, Y., Shi, Y., Fang, R., Lin, X., Liu, S., & Toga, A. (2018). Brain structure differences between Chinese and Caucasian cohorts: A comprehensive

morphometry study. *Human Brain Mapping*, *39*(5), 2147–2155. https://doi.org/10.1002/hbm.23994

Toro, R. (2020). Cortical Figures Generator. https://r03ert0.github.io/cortex/

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

Toro, R., Perron, M., Pike, B., Richer, L., Veillette, S., Pausova, Z., & Paus, T. (2008). Brain Size and Folding of the Human Cerebral Cortex. *Cerebral Cortex*, *18*(10), 2352–2357. https://doi.org/10.1093/cercor/bhm261

van Velsen, E. F. S., Vernooij, M. W., Vrooman, H. A., van der Lugt, A., Breteler, M. M. B., Hofman, A., Niessen, W. J., & Ikram, M. A. (2013). Brain cortical thickness in the general elderly population: The Rotterdam Scan Study. *Neuroscience Letters*, *550*, 189–194. https://doi.org/10.1016/j.neulet.2013.06.063

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

Wechsler, D., Psychological Corporation, & PsychCorp (Firm). (2008). *WAIS-IV technical and interpretive manual*. Pearson.

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

Wierenga, L. M., Doucet, G. E., Dima, D., Agartz, I., Aghajani, M., Akudjedu, T. N., Albajes-Eizagirre, A., Alnæs, D., Alpert, K. I., Andreassen, O. A., Anticevic, A., Asherson, P., Banaschewski, T., Bargallo, N., Baumeister, S., Baur-Streubel, R., Bertolino, A., Bonvino, A., Boomsma, D. I., … Tamnes, C. K. (n.d.). Greater male than female variability in regional brain structure across the lifespan. *Human Brain Mapping*, *n/a*(n/a). https://doi.org/10.1002/hbm.25204

Abbreviations: Total Cerebral Measures (TCMs): Total Brain Volume (TBV), Total Mean Cortical Thickness (MCT), or Total Surface Area (TSA). Grey Matter Volume (GMV). White Matter Volume (WMV).

Wierenga, L. M., Langen, M., Ambrosino, S., van Dijk, S., Oranje, B., & Durston, S. (2014). Typical development of basal ganglia, hippocampus, amygdala and cerebellum from age 7 to 24. *NeuroImage*, *96*, 67–72. https://doi.org/10.1016/j.neuroimage.2014.03.072

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

Williams, C. M., Peyre, H., Toro, R., Beggiato, A., & Ramus, F. (2020). Adjusting for allometric scaling in ABIDE I challenges subcortical volume differences in autism spectrum disorder. *Human Brain Mapping*, *n/a*(n/a). https://doi.org/10.1002/hbm.25145

Abbreviations: Total Cerebral Measures (TCMs): Total Brain Volume (TBV), Total Mean Cortical Thickness (MCT), or Total Surface Area (TSA). Grey Matter Volume (GMV). White Matter Volume (WMV).