Cross-sectional volumes and trajectories of the human brain, gray matter, white matter and cerebrospinal fluid in 9,473 typically aging adults

Andrei Irimia[1, 2]*

[1]Ethel Percy Andrus Gerontology Center, Leonard Davis School of Gerontology, University of Southern California, 3715 McClintock Avenue, Los Angeles CA 90089 USA

[2]Denney Research Center, Department of Biomedical Engineering, Viterbi School of Engineering, University of Southern California, 1042 Downey Way, Los Angeles CA 90089 USA

*Corresponding author

Email address: irimia@usc.edu

Author's ORCID: 0000-0002-9254-9388

Acknowledgments

This work was supported by the National Institutes of Health under Award Numbers RF1 AG 054442 and R01 NS 100973. The content is solely the responsibility of the author and does not necessarily represent the official views of the National Institutes of Health.

Information Sharing Statement

Data and code related to this study can be made available by the study upon request.

Abstract

Accurate knowledge of adult human brain volume (BV) is critical for studies of aging- and disease-related brain alterations, and for monitoring the trajectories of neural and cognitive functions in conditions like Alzheimer's disease and traumatic brain injury. This scoping meta-analysis aggregates normative reference values for BV and three related volumetrics—gray matter volume (GMV), white matter volume (WMV) and cerebrospinal fluid volume (CSFV)—from typically-aging adults studied cross-sectionally using magnetic resonance imaging. Drawing from an aggregate sample of 9,473 adults, this study provides (A) linear regression coefficients β describing the age-dependent trajectories of volumetric measures by sex within the range from 20 to 70 years, and (B) average values for BV, GMV, WMV and CSFV at the representative ages of 20 (young age), 45 (middle age) and 70 (old age). The results provided synthesize ~20 years of brain volumetrics research and allow one to estimate BV at any age between 20 and 70. Guidelines are proposed to facilitate future meta- and mega-analyses of brain volumetrics.

Keywords

brain volume, gray matter, white matter, cerebrospinal fluid, meta-analysis

Disclosure statement

The author declares that he has no actual or potential conflicts of interest.

Abbreviations

AD	Alzheimer's disease
BV	brain volume
CSF	cerebrospinal fluid
CSFV	cerebrospinal fluid volume
CT	computed tomography
F	females
GM	gray matter
GMV	gray matter volume
ICV	intracranial volume
M	males
MRI	magnetic resonance imaging
PD	Parkinson's disease
SPM	statistical parametric mapping
TBI	traumatic brain injury
VBM	voxel-based morphometry
WM	white matter
WMV	white matter volume

Introduction

The volume of the adult human brain is a fundamental physical descriptor of human neuroanatomy (Gazzaniga 2009). For example, in organismic biology, brain volume (BV) provides one of the most tantalizing correlates of neural and behavioral complexity across animal species, including primates (Butler and Hodos 1996); in anthropology, BV and its closely related measure of intracranial volume (ICV) are key criteria for comparing extinct hominid species to one another and to modern humans (Prothero 2007). Studies of human cerebral volumetry and allometry, whether undertaken by neurobiologists, biomedical scientists or anthropologists, frequently rely on accurate knowledge of BV and ICV to compare healthy and diseased brains, and to monitor the temporal trajectories of BV across healthy and diseased populations (C. D. Smith et al. 2007).

BV trajectories are of great interest in healthy aging, in neurodegenerative diseases—e.g. Alzheimer's disease (AD) and Parkinson's disease (PD)—but also in other neurological conditions which may accelerate brain atrophy—e.g. traumatic brain injury (TBI) and stroke. Despite numerous past studies which quantified brain volumetrics in health and disease, identifying reliable reference values for these measures is not always straightforward. Importantly, BV varies considerably with age and sex, such that different researchers may require distinct normative statistics for their studies depending on the age and sex compositions of their samples (Cosgrove et al. 2007). Although this variability has been extensively explored and well established (Gur et al. 2002; Rushton and Ankney 1996), it can nevertheless be difficult to identify accurate reference statistics for BV due to several important limitations which are shared by many studies. Firstly, researchers may utilize vastly different techniques to assess brain volumetrics (*in vivo* neuroimaging, *post mortem* physical measurement, etc.) and their quantitation approaches may differ substantially, which can lead to substantial variability in measurement error across studies (S. M.

Smith et al. 2001). Secondly, many reports are based on samples of inadequate sizes and/or compositions, such that the natural variability of brain volumetrics as a function of age and sex has not been quantified satisfactorily. Thirdly, BV and related measures have been reported in a range of formats (tabulations, graphical representations, written narratives, etc.), expressed using a variety of physical units (cubic centimeters, liters, etc.) and manipulated algebraically in various ways (e.g., normalized with respect to ICV, averaged across ages or sexes, etc.), such that different studies' data may require careful and systematic inspection to ensure proper comparison and interpretation. Thus, brain volumetrics reported by different studies can vary substantially even when their statistics are obtained from samples with similar demographics. For these and other reasons, researchers often find it necessary to review the neuroscience literature very thoroughly to identify reports whose methodologies are congruent with one another's and whose aggregation can provide trustworthy volumetric estimates above and beyond what single studies can afford.

This investigation is a meta-analysis of research studies which utilized *in vivo* magnetic resonance imaging (MRI) to infer human BV, ICV, gray matter volume (GMV), white matter volume (WMV) and cerebrospinal fluid volume (CSFV) throughout adulthood. Two sets of results are presented: (A) the linear regression coefficient for each of these measures as a function of age, across the interval from age 20 to 70, and (B) the values of BV, GMV, WMV and CSFV at the ages of 20 (young age), 45 (middle age) and 70 (old age). Metrics are reported in tabular format to facilitate visual comparison of studies and to provide researchers with the ability to select reports which they deem to be most appropriate for their purposes. For each study listed, numerical values are reported by sex where available, and weighted grand averages are computed to summarize findings across studies. Guidelines are provided to facilitate appropriate use and interpretation of the results by the reader. Finally, recommendations are made to neuroimaging researchers so that future studies of brain volumetrics report results in ways which facilitate meta-analysis

and unbiased comparison with other studies. Although neither exhaustive nor definitive, this meta-analysis provides valuable information for a wide range of researchers working in neuroscience, medicine, anthropology and related fields.

Methods

Data sources. Research articles were retrieved via direct search on the ISI Web of Knowledge (webofknowledge.com) and PubMed (https://www.ncbi.nlm.nih.gov/pubmed). The following query syntax was used: (*"brain volume"* OR *"cerebral volume"* OR *"gray matter volume"* OR *"white matter volume"* OR *"intracranial volume"*) AND (*old* OR *aging* OR *elderly* OR *atrophy*). Studies published before 1990 were excluded due to potential drawbacks related to the quality of imaging data and to the paucity of robust volumetric analysis methods prior to that year.

Study selection. English-language titles and publications were screened manually for relevance. Studies were excluded if they reported incomplete volumetrics data, e.g. if values had not been reported for the entire structures of interest. Data confounded by averaging across both development and adulthood were excluded. Because this study aimed to aggregate volumetrics which are representative of the typical adult population, studies which sampled healthy participants non-randomly were excluded. Such non-random sampling frequently involved inclusion/exclusion criteria which limited the relevance of computed volumetrics to subsets of—rather than to the typical—adult population. For example, studies were excluded if they had specific selection criteria pertaining to factors which could restrict the generalizability of their results, whether related to environmental factors (smoking, alcohol abuse, diet, educational attainment, physical exercise level, etc.) or genetics (e.g. ApoE genetic profile). By contrast, studies which excluded individuals with neurological, psychiatric and/or metabolic disease were not discarded because such exclusions are necessary to identify healthy adults. Studies were excluded if brain volumetrics had not been normalized by ICV or if normalized values could not be obtained based on the results made available by each study.

Data extraction and review. Variables were recorded and tabulated separately for men and women, where possible. Literature reports were surveyed to identify two groups of variables of interest:

(A) linear regression coefficients β, reported as percentages of volumetric differences per year (%/year), for the association between the independent variable (age) and each dependent variable (BV, GMV, WMV or CSFV, respectively);

(B) ICV-corrected, cross-sectional average values of BV, GMV, WMV and CSFV at the ages of 20 (young age), 45 (middle age) and 70 (old age), in units of cm^3.

A scoping review was conducted using the formalism of Arksey & O'Malley (2005). Scoping reviews are similar to systematic reviews and involve a systematic approach to reference search, thus being less vulnerable to bias compared to rapid, critical or expert reviews (Grant and Booth 2009). By contrast, however, although scoping reviews do not involve systematic quality assessment, they do incorporate more flexible criteria for screening and inclusion.

Meta-analysis. A meta-analysis was conducted based on an established approach (DerSimonian and Laird 1986, 2015) for aggregating measures of interest over all i = 1, …, M studies, where M is the number of studies included. The meta-analysis was conducted in accordance with PRISMA guidelines (http://www.prisma-statement.org). In the formalism of DerSimonian & Laird, each study's measured effect y_i is parceled as the sum of the true effect θ and the sampling error e_i, i.e.

$$y_i = \theta + e_i$$

where the variance $\sigma^2(e_i)$ is the sample variance s_i^2 of the effect. To account for the variation in true effects, the model assumes that the true effect is the sum of both μ (the mean effect for a population of possible evaluations of the effect), and δ_i (the deviation of the effect in study i from the population's mean effect μ), i.e.

$$\theta_i = \mu + \delta_i$$

such that

$$y_i = \mu + \delta_i + e_i$$

Conceptually, the studies in the meta-analysis are a sample from the population of possible studies evaluating the true effect θ, whose population mean is μ and population variance is Δ^2, such that $\theta \sim N(\mu, \Delta^2)$ and $y_i \sim N(\theta_i, s_i^2)$. In the present study, the effect y_i is assumed to be the change in volume V observed over some time interval measured in years. Although the regression coefficient is not the most typical measure of effect size, its use as such is not uncommon (Nieminen et al. 2013); our interest in volumetric differences as a function of age makes it very suitable—if not even ideal—for this study.

In meta-analyses where regression coefficients are used to calculate effect size, such coefficients are often standardized to remove the potential confound of them being reported by different studies using different physical units (Nieminen et al. 2013). Here, converting regression coefficients to the same set of physical units (e.g., cm^3/year) is straightforward; additionally, unstandardized regression coefficients are of substantial interest in practice due to their immediate physical interpretation as BV differences observed over an age interval. For this reason, this analysis utilizes *unstandardized* regression coefficients as measures of effect, after their appropriate conversion to cm^3/year, where necessary. Because head size may confound brain volumetrics to a substantial extent (Barnes et al. 2010), the effect of this variable is alleviated here by using ICV-normalized volumes—instead of absolute volumes—for all calculations. Thus, the regression coefficients reported here are based on regressions of ICV-normalized volumes rather than on absolute volumes. Furthermore, because scanner field strength, model, head coil, MRI weighting and segmentation technique differed across studies, the potentially confounding effects of these parameters were accounted for by treating the latter as covariates in the meta-analysis. Study sites were treated as random effects.

Let β_{1i} denote the unstandardized linear regression coefficient β_1 associated with study i, let $\hat{\beta}_{1i}$ be its empirical estimate, and let M_i be the sample size of study i. Of interest here is the meta-analyzed regression coefficient β_1, i.e. the slope of the line describing the relationship between time t and volume V. Without loss of generality, the statistical effect of interest can be assumed to be the volume difference $\mathrm{d}V_i$ whose value the regression model of study i predicts within some chosen time interval (age range) $\mathrm{d}t$. The measured effect y_i defined previously is thus identically equal to $\mathrm{d}V_i$ here, i.e. $y_i \equiv \mathrm{d}V_i$, such that $\hat{\beta}_{1i}$ can be computed as $\mathrm{d}V_i/\mathrm{d}t$. In this study, the denominator is the change $\mathrm{d}t$ in chronological age t, expressed in units of years and computed identically across all studies (i.e., across the interval from 20 to 70 years of age). To evaluate effect homogeneity across studies, Cochran's Q statistic

$$Q = \sum_i w_i (y_i - \bar{y}_w)^2,$$

was computed, where

$$\bar{y}_w = \sum_i w_i \, y_i \left(\sum_i w_i \right)^{-1}$$

is the weighted estimator of the effect, and w_i is equal to $1/s_i^2(y_i)$. Under the null hypothesis, $Q \sim \chi^2(M_i - 1)$. The null hypothesis of homogeneity across studies in the meta-analysis is rejected if Q falls within the critical region $Q > \chi^2_{1-\alpha}(M_i - 1)$, where $\chi^2_{1-\alpha}(M_i - 1)$ is the $(1-\alpha)$-quantile of the χ^2 distribution with $M_i - 1$ degrees of freedom, and α is the significance threshold; in this study, α = 0.05. In addition to Cochran's Q, the I^2 index was calculated. This index is set to

$$I^2 = \frac{Q - M_i + 1}{Q}$$

if the right-hand side above is greater than zero and to zero otherwise. I^2 does not depend on the number of studies included in the meta-analysis, and can be used to assess consistency between studies. Here, I^2 was interpreted to represent low ($I^2 < 0.25$), moderate ($0.25 \leq I^2 < 0.50$) or high ($I^2 \geq 0.50$)

inconsistency (Higgins et al. 2003). To identify publication bias, the rank correlations of Begg & Mazumdar (1994) were calculated. Forest plots were generated to visualize results and to identify potential outlier studies.

Results

Study selection, heterogeneity and publication bias. A total of 6,458 published studies were retrieved using the search criteria, of which 5,832 studies were excluded because they were not relevant to the topic of the study. Of the remaining 626 studies, 137 contained incomplete volumetrics data, 57 reported volumetrics which had been averaged over not only adults but also children and/or adolescents, 147 did not study healthy adults who were representative of the typical adult population, 155 reported data on clinical samples rather than healthy adults and 105 reported volumes which had not been normalized by ICV. Notably, many studies which utilized data from the Alzheimer's Disease Neuroimaging Initiative (ADNI, http://adni.loni.usc.edu), the Parkinson Progression Marker Initiative (PPMI, https://www.ppmi.info.org) or the UK Biobank (https://www.ukbiobank.ac.uk) could not be included for a variety of reasons, such as that (A) their subject sampling methods did not satisfy the inclusion/exclusion criteria of this study (e.g. their sampling of healthy control participants was not fully random), (B) they did not report the numerical quantities of interest here, or (C) the metrics of interest were not reported in a format which could accommodate their accurate analysis using the approach of the present study. Subsequent to all exclusions, the combined samples of the studies included in the meta-analysis amounted to a total of 9,473 subjects whose volumetrics and related data were aggregated.

Thirty studies were included in the meta-analysis; twelve had fewer than 100 participants (Jernigan et al. 1991; Gur et al. 1991; Matsumae et al. 1996; Guttmann et al. 1998; Jernigan et al. 2001; Ge et al. 2002; Liu et al. 2003; Scahill et al. 2003; Allen et al. 2005; Benedetti et al. 2006; Abe et al. 2008; Barnes et al. 2010), seventeen had between 100 and 1000 participants (Resnick et al. 2000; Good et al. 2001; Gur et al. 2002; Sowell et al. 2003; Taki et al. 2004; Fotenos et al. 2005; Lemaitre et al. 2005; Chen et al. 2007; C. D. Smith et al. 2007; Fotenos et al. 2008; Driscoll et al. 2009; Michielse et al. 2010; Walhovd et al. 2011;

Lemaitre et al. 2012; Peelle et al. 2012; Jancke et al. 2015; Blatter et al. 1995) and one had over 1000 participants (DeCarli et al. 2005). Studies were not found to be significantly heterogeneous when reporting BV (Q = 0.29, p > 0.99, I^2 = 0), GMV (Q = 1.15, p > 0.99, I^2 = 0) or WMV (Q = 1.35, p > 0.99, I^2 = 0), but were found to have mild heterogeneity when reporting CSFV (Q = 15.49, p > 0.22, I^2 = 0.23). The Begg-Mazumdar rank correlation (τ = 0.12, p > 0.67) did not provide significant indication of publication bias.

Regression coefficients. Table 1 and Figure 1 (A) summarize the cross-sectional trajectories of BV, GMV, WMV and CSFV as reported by studies fitting the selection criteria of the analysis. The regression coefficients (slopes) for volume differences as a function of age are reported as percentages of volumetric difference per year (%/year) and are listed separately for males (M) and females (F), where available. For studies where numbers are reported for combined samples, the numerical values in question are listed in the column labeled as "M&F." Studies are listed in ascending order by year of publication, and their sample sizes are reported. For BV, the value of the regression coefficient calculated across the entire aggregate sample was found to be $\beta \simeq$ -0.25%/year (pooled s = 0.50%/year; 95% CI = [-0.26, -0.25] %/year). For GMV, $\beta \simeq$ -0.37%/year (pooled s = 0.49%/year; 95% CI = [-0.38, -0.35] %/year); for WMV, $\beta \simeq$ -0.10%/year (pooled s = 0.47%/year; 95% CI = [-0.14, -0.12] %/year); for CSFV, $\beta \simeq$ -1.67%/year (pooled s = 0.47%/year; 95% CI = [-0.14, -0.12] %/year).

Table 1 includes, in boldface, the regression coefficients for each aggregate sample. Specifically, the average year-to-year difference in BV was found to be -0.25% across sexes (males: -0.28%; females: -0.22%); studies reported values of β ranging from about -0.40%/year (Walhovd et al. 2011) to about -0.10%/year (Peelle et al. 2012). GMV was found to trend negatively with age (males: -0.43%; females: -0.36%; both sexes: -0.37%), and more so than WMV (males: -0.11%; females: -0.10%; both sexes: -0.10%).

The annual difference in CSFV was found to be smaller in males (1.93%) than in females (2.15%), although the computed difference is relatively small (0.22%). Furthermore, because both GMV and WMV were found to have more negative regression coefficients in males than in females, it is likely that the true slope of the CSFV trajectory as a function of age is steeper in males than in females. Table 1 suggests that the small difference observed between males and females is likely driven by the study of Good et al. (2001), who reported a smaller regression coefficient for males than for females and whose sample size was relatively large compared to most of the other studies included.

Cross-sectionally, over the interval from age 20 to age 70, all studies reported negative trajectories for BV, GMV and WMV; only Taki et al. (2004) reported a positive regression coefficient for WMV. Furthermore, as expected, all studies reported increases in CSFV. These results are confirmed by and reflected in the forest plots of Figure 1 (A). It is important to note that values reported under "M&F" should not be interpreted as weighted averages of the values reported for each of the two sexes. Rather, they represent volumetrics reported by studies where results were not calculated separately for males and females.

Average volumetrics. Studies reporting volumetrics were not found to be significantly heterogeneous (Q = 16.11, p > 0.71, I^2 = 0). The Begg-Mazumdar rank correlation (τ = 0.09, p > 0.70) failed to provide significant indication of publication bias. Table 2 reports the average volumes of BV, GMV, WMV and CSFV at the ages of 20, 45 and 70, as calculated over studies fitting the selection criteria of the analysis; for brevity, Figure 1 (B) reports this information only for age 70. Volumes are reported in cm^3 for males (M) and females (F), where available. For studies where numbers are reported for combined samples, the figures in question are listed in the column labeled as "M&F." Studies are listed in ascending order by year of publication, and their sample sizes are also reported. Table 2 also reports the results of the meta-analysis, i.e. average volumes aggregated across all the studies and samples listed in the table.

ICV was found to decrease very slightly as a function of age, and Table 2 indicates that this finding is consistent across studies. Average BV is found to vary from ~1,150 cm^3 at age 20 to ~1,116 cm^3 at age 45 and ~1,009 cm^3 at age 70. Similarly, GMV varies from ~692 cm^3 at age 20 to ~641 cm^3 at age 45 and to ~560 cm^3 at age 70. Average WMV is ~509 cm^3 at age 20, ~494 cm^3 at age 45 and ~457 cm^3 at age 70. As expected, the trend is reversed for CSFV, which varies from ~159 cm^3 at age 20 to ~266 cm^3 at age 45 and to ~365 cm^3 at age 70. Consistently, average BV, GMV and WMV are larger for males than for females; the reverse is found for CSFV, as expected.

Key studies. Several of the studies included in this meta-analysis deserve discussion due to their relatively large sample sizes or to their detailed study of volumetrics. One such study is that of Good et al. (2001), who used voxel-based morphometry (VBM) and statistical parametric mapping (SPM) to calculate the brain volumetrics of 465 healthy adults aged 17-79 (265 males) based on T_1-weighted MRIs (voxel size: 1 mm × 1 mm × 1.5 mm). These authors found a negative trend of global GM with age ($\beta_1 \simeq$ -0.24% per year across sexes, R^2 = 0.489, p < 0.0001), and a GM/WM fractional volume ratio of ~1.82 across sexes. For CSF, the IV-normalized regression coefficient β_1 was positive, consistent with age-related increases in ventricular volume ($\beta_1 \simeq$ 0.65% per year, R^2 = 0.377, p < 0.0001). Overall, BV was found to decrease in both sexes ($\beta_1 \simeq$ -0.25% per year, R^2 = 0.489, p < 0.0001). Usefully, Good *et al.* reported both absolute and ICV-corrected values, such that the confound of head size was accounted for in their analysis. Nevertheless, their study had considerably more volunteers younger than 40, such that their results might reflect younger adults' volumetrics to a greater extent than other studies. Had the authors' sample been more balanced, the reported regression coefficients would have likely been more negative for GMV, WMV and BV, and more positive for CSFV.

Taki et al. (2004) used SPM to extract GMVs and WMVs from the T_1-weighted MRIs (voxel size: 1.02 mm × 1.02 mm × 1.5 mm) of 769 Japanese adults aged 16-79 (356 males). Age-related declines were found in global GM (males: $\beta_1 \simeq$ -0.54% per year, R^2 = 0.58, p < 001; females: $\beta_1 \simeq$ -0.48% per year, R^2 = 0.39, p < 001), but WM was found to increase with age, albeit only slightly (males: $\beta_1 \simeq$ 0.11% per year, R^2 = 0.017, p < 001; females: $\beta_1 \simeq$ 0.11% per year, R^2 = 0.019, p < 001). For GM, the age-related volumetric decline rates reported by Taki *et al*. are relatively faster than in most other studies included in this meta-analysis, whereas for WM the rates are slightly positive. One limitation of the study by Taki *et al*. is that it included only adults of Japanese ethnicity, although this can also be perceived as a strength because it allows one to compare their results to those obtained from cohorts of other ethnicities.

DeCarli et al. (2005) acquired T_2-weighted MRIs (field of view: 22 cm; acquisition matrix size: 182 × 256) from 2,081 adults enrolled in the Framingham Heart Study. In this study, regression coefficients were found to be negative, and similar to those of other studies reviewed (males: $\beta_1 \simeq$ -0.24% per year, R^2 = 0.45, p < 001; females: $\beta_1 \simeq$ -0.18% per year, R^2 = 0.35, p < 001). GMV and WMV trajectories were reported by lobe, but global statistics were not provided. Although the sample consisted mostly of European Americans—which can be a limitation—this is the largest study included in the meta-analysis.

Walhovd et al. (2011) combined one Swedish, two Norwegian and three US cohorts to study GMV, WMV, CSFV and BV in an aggregate sample of 883 adults aged 18-94 (355 males). This very thorough study reported not only the regression coefficients of interest here, but also the mean and standard deviations of volumes for CSF structures, WM as well as for cortical and subcortical GM. Volumetric trajectories were reported both by decade and across the entire sample. As in most other studies included in the present analysis, the authors found negative regression coefficients for BV ($\beta_1 \simeq$ -0.40% per year, R^2 = 0.47, p < 0.001), cerebral GM ($\beta_1 \simeq$ -0.43% per year, R^2 = 0.54, p < 0.001), cerebral WM ($\beta_1 \simeq$ -0.35% per year, R^2

= 0.12, $p < 0.001$) and ventricular CSF ($\beta_1 \simeq 3.80\%$ per year, $R^2 = 0.37$, $p < 0.001$). Compared to many other studies included in this meta-analysis, that of Walhovd *et al.* stands out because it reports systematic volumetrics in considerable detail.

Discussion

Motivation. Important goals of brain aging research include identifying factors which affect brain senescence and atrophy, quantifying the extent to which these factors influence it, and understanding how. The relationship between volumetrics and age is thus essential when monitoring brain aging. Accurate assessment of human BV across adulthood is indispensable because, without reference values against which BV changes can be quantified, reliable quantification of disease-related brain tissue loss across adulthood—and especially during old age—is difficult to achieve. Thus, the availability of human BVs throughout adulthood is very important to both scientists and clinicians.

Rationale. BV is a metric of critical importance in imaging studies of neurodegeneration due to typical aging, or to conditions like AD, PD and TBI. Surprisingly, although numerous reports of brain volumetrics across the lifespan have been published, the uncertainty surrounding the important relationship between volumetrics and age remains substantial. Firstly, although the temporal trajectory of BV is modulated by sex, many reports have not studied this effect separately for men and women. Secondly, volumetric measurements have traditionally been obtained from a variety of sources, including via MRI tissue classification (Gunning-Dixon et al. 2009; Balafar et al. 2010), computed tomography (CT) image segmentation (Irimia et al. 2019), *post mortem* stereology (Doherty et al. 2000) and other methods. Each of these techniques have both advantages and drawbacks related to data acquisition and analysis, which can produce systematic errors of volumetric estimation. Thirdly, brain volumetrics have been reported in a variety of ways, such that the values reported by various studies may be difficult or even impossible to compare. For reasons like these, meta-analyses like this one are necessary to identify, harmonize and aggregate volumetrics data across the scientific literature.

Study selection. When meta-analyzing the scientific literature on the topic of interest here, researchers must pay attention to the eligibility criteria of each study to minimize heterogeneity across the aggregate sample in the meta-analysis. For example, many studies have measured BVs and atrophy rates from individuals selected for specific characteristics, whether genetic (e.g. based on their ApoE profiles) or environmental (e.g. physical activity levels). Whilst such selection criteria are necessary to study the association between genetics, environment and aging, the meta-analysis researcher must carefully weigh whether—and, if so, how—such studies should be included in a meta-analytic survey of brain atrophy data. If a meta-analysis of the normal population inadequately aggregates studies whose inclusion criteria result in overall samples which do not reflect that population, the study in question may lead to results and conclusions which do not reflect the characteristics of the normative population whose features it aims to capture. As an illustration, the brain aging trajectory of physically active individuals is likely not very representative of the general population's overall BV trajectory because activity levels vary considerably across the general population. Furthermore, it may be difficult or impossible to adequately weigh results from different samples based on the extent to which the latter reflect the composition of the general population. For this reason, studies were excluded from this meta-analysis if their eligibility criteria could have substantially modified the composition of the aggregate sample beyond what is expected from random sampling of the normal population.

Sampling confounds. Because the interactions between genetic, phenotypic and environmental factors upon age-related volumetric trajectories are complex, studies frequently focus on just one of these many contributors which may partially explain the observed variance of brain trajectories. Some of these variables must be mentioned here due to their importance when interpreting the results of this study. For example, the variation of brain volumetrics with age and sex has been reported extensively by many studies, including those included in this report. Some studies describe brain volumetrics and related

measures across the entire range of ages analyzed rather by decade or as a function of age. This can pose challenges to meta-analyses like ours because participants' age distribution is not always uniform in a given study, and the number of subjects whose age falls within each decade of adulthood is rarely reported. This can preclude the ability to survey the literature in a way which is amenable to reporting volumetric trajectories by decade; one notable exception here is the study of Walhovd et al. (Walhovd et al. 2011), where decadal data are provided. Aside from age, sex is also very important as a biological variable affecting volumetric trajectories (Jancke et al. 2015); for this reason, this study reports regression coefficients and volumetrics separately by sex, where possible.

Genes and environment. Aside from age, sex and head size, human BV trajectories are partly influenced by genetic and environmental factors whose effects upon the metrics discussed here may covary to a substantial extent. Lifestyle factors like diet, exercise, intelligence, educational attainment, alcohol consumption and smoking can all affect BV trajectories to various extents and, frequently, in a dose-dependent fashion; such factors can be additional sources of variance which is not explained by other factors (Lenroot and Giedd 2008). Genes which modulate aging rates and susceptibility to disease can themselves interact with environmental factors; together, genes, environment and their interaction contribute to brain aging trajectories in complex ways and can account for some of the variability observed in the measured rates of human brain atrophy (Seshadri et al. 2007). This meta-analysis does not account for the effect of such genetic and environmental variables upon brain trajectories because its purpose is to describe the BV trajectory expected of the general population rather than of any subgroup. Nevertheless, future research should aim to meta-analyze the effects of genes and environment upon BV and related measures.

Allometry. Head size may confound meta-analyses considerably because a structure's absolute difference in volume over some arbitrary time interval is proportional to the initial volume of that structure. Thus, head size must be accounted for in studies like ours to avoid fallacious inferences driven by effects related to brain allometry (the scaling of the brain with body size) rather than by the main effects of interest, such as aging or disease. Many reports of brain volumes identified in this study did not account for head size or, alternatively, accounted for it in ways which differed enough across studies that case-study comparison was too difficult or, indeed, impossible. Here, regression coefficients were reported as percentages of differences in volume per year precisely to convey information which can be interpreted independently of head size. It should be mentioned that, although normalization by intracranial volume is by far the most common approach to accounting for head size, there is more than one such strategies and certain approaches can be more suitable than others, depending on a variety of factors (Voevodskaya et al. 2014).

Volumetric measurement techniques. Information on the descriptive statistics (mean and variance) of brain volumetrics is available in reports of studies utilizing a variety of methods. These range from approaches superseded long ago—like *post-mortem* weighing of fluid volume displaced by the excised brain (Uspenskii 1964)—to state-of-the-art *in vivo* magnetic resonance imaging (MRI) at ultra-high spatial resolution (Van Leemput et al. 2009). More recently, MRI quantitation methods have been extended to CT (Irimia et al. 2019), thus making BV calculations possible in remote locations or in other environments where MRI acquisition is not feasible (Kaplan et al. 2017). There is typically good agreement between such approaches, even within as little as a fraction of one percent (Despotovic et al. 2015). This study is restricted to studies undertaken using MRI because this technique currently provides the gold standard for *in vivo* BV measurement (de Boer et al. 2010); nevertheless, the potential shortcomings of MRI related to the effects of microstructural properties of brain tissue (e.g. myelination, iron, and water content) upon

brain morphology calculations (Lorio et al. 2016; Natu et al. 2019) should be acknowledged. Nevertheless, because the MRI literature is the largest source of brain volumetrics currently available and MRI sample sizes eclipse those of most *post-mortem* studies very frequently, the MRI literature was found to be the most suitable one for this analysis.

MRI acquisition parameters. Encouragingly, the results of this meta-analysis (Figure 1) do not suggest that the regression coefficients and volumetrics reported by older studies differ systematically from those in more recent studies. This, however, does not demonstrate the absence of a relationship between the measures of interest here and the publication dates of the studies included. For this reason, future studies should aim to investigate in more detail the combined effects of MRI spatial resolution, partial volume effects, scanner field strength and of other factors upon the metrics discussed here. The available spatial resolution of anatomic MRI scans has increased considerably over the past thirty years; thus, MRI spatial resolution can confound meta-analyses like ours if their statistical effects are not accounted for. More recent studies are typically more likely to provide the opportunities of calculating volumetrics more accurately; currently, the typical voxel size of brain MRI scans is ~1 mm^3, which frequently corresponds to a pixel size of 1 mm^2 and to a slice thickness of 1 mm. Indeed, many studies published in the past ~15 years feature today's standard voxel size of 1 mm^3, although this is certainly not the rule. In a few older studies included here, the typical spatial resolution is—as expected—poorer than today's standard. For example, Gur et al. (1991) calculated brain volumes from T_2-weighted MRI in a sample of 69 healthy adults with ages ranging from 18 to 80. Similarly, Blatter et al. (1995) calculated brain volumetrics from T_2-weighted SE MRI from 194 healthy subjects with ages ranging from 16 to 65 years, but their measurements were based on MRI volumes with a slice thickness somewhat greater than today's standard. A third such study is that of Matsumae et al. (1996), who obtained total intracranial volumes from 49 normal volunteers ranging in age from 24 to 80. Importantly, none of these studies were found to yield either

cross-sectional regression coefficients or volumetrics which differed in any substantial way from those reported in recent studies (see tables and figure). Furthermore, whereas most older studies had relatively small sample sizes, others were relatively large even by today's standards (Bigler and Tate 2001) and may therefore still warrant consideration. Thus, Bigler et al. assessed ICV and BV in 532 subjects who were either healthy adults, TBI victims or AD patients. Interestingly, the regression coefficients and volumetrics reported by these older studies did not differ substantially from those reported by recent ones, and the relative sample sizes of the former studies were relatively smaller than those of the latter, such that older studies did not contribute as substantially to meta-analysis results as newer ones did.

Utility of reported measures. Frequently, researchers who utilize brain volumetric data are interested in two important groups of measures: (1) the BV trajectory as a function of age, and (2) the mean and variance of BV at a given age and/or for each sex. The first such group of parameters includes linear regression coefficients describing the relationship between volumetrics and age across the range from 20 to 70 years. The second group includes average volumetrics at selected ages (20 years, 45 years and 70 years). Researchers have many choices as to the nature, format and units used to report such data on the relationship between brain volumetrics and age. Depending on the context in which such values are required, certain choices may be preferable to others depending on the researcher's objective needs and/or subjective preferences. In this study, regression coefficients are reported as percentages of differences in volume per year, rather than in cm^3 per year, ml per year or in other units. Reporting coefficients using *percentages* rather than *physical units* of volume has a considerable advantage, in that the former choice is independent of head size whereas the latter is not. Thus, a regression coefficient reported as a percentage difference in volume per year does not suffer from the confounding effect of varying head sizes, either within or between studies. Nevertheless, because scientists frequently need to calculate the expected values of physical volumes based on regression coefficient data, it is also important

to have access to reference values of average volumetrics for the end points of the age interval over which regression coefficients were computed (20 to 70 years in the present study). This is one of the reasons for which this study reports average volumetrics for men and women at the ages of 20, 45 and 70, which correspond conveniently to representative—albeit somewhat arbitrary—values for young, middle and old age.

Importantly, availability of the two sets of parameters reported by this meta-analysis allows one to estimate the average value of volumetrics at *any* age of interest within the stated interval from 20 to 70, under the assumption that age and volume are linearly related. Thus, knowledge of these two important sets of parameters can facilitate the process of obtaining reference values for brain volumetrics at any adult age of interest within the range considered. Specifically, one can calculate the expected value of volume V at age x within the range from 20 to 70 by (A) identifying the value of the desired volume V at age 20 in Table 2, and then (B) calculating the expected value of V at age x as

$$V(x) = \frac{1}{100}[1 - \beta_1(x - 20)]V(20)$$

where β_1 is the corresponding regression coefficient for the volumetric measure of interest, as reported in Table 1. For example, to calculate females' average BV at age 60, one can proceed by identifying the average of $V(20)$ in Table 2 (i.e. 1,274 cm^3) and then use β_1 = -0.22%/year from Table 1 to calculate an approximate value of 1,162 cm^3 for the average female BV at age 60. Thus, one strength of calculating average brain volumes from meta-analysis results like ours is that meta-analyses aggregate data over a range of studies, presumably yielding more accurate estimates.

Aside from expediting the calculation of average volumes, the results presented here also facilitate comparisons between health and disease. For example, Cole et al. (2018) found that, annually, victims of moderate-to-severe TBI may experience, on average, a mean GMV loss of 1.55%, a mean WMV loss of

1.49% and a mean BV loss of 1.51%. Comparison of these annual volume differences with those listed for in Table 1 healthy adults clearly supports the conclusion of Cole et al. that TBI patients experience GMV, WMV and BV trajectories which are substantially steeper than those observed in typical aging.

Cross-sectional vs. longitudinal studies. Cross-sectional studies suffer from the major limitation that they only allow the investigation of age-related *differences* and *trajectories* rather than *changes*, partly because of cohort effects, secular trends and uncontrolled individual differences (Raz and Rodrigue 2006). The results of the present study should therefore be interpreted with awareness of this important caveat. By contrast, longitudinal studies of brain volumetrics can be more valuable because they allow researchers to estimate brain atrophy rates within specific individuals and across specific time periods while (partially) controlling for individual differences.

Unfortunately, the literature survey conducted identified relatively few longitudinal studies of brain atrophy in youth or middle age compared to old age. For this reason, data on BV trajectories across adulthood are more abundantly available from cross-sectional—rather than longitudinal—studies. Additionally, many—if not most—longitudinal studies of brain atrophy focus on comparisons between groups whose membership is limited by relatively narrow eligibility criteria. Such criteria can encumber the inclusion of longitudinal studies in meta-analyses because different eligibility criteria may not only confound the meta-analysis, but also result in an aggregate sample whose characteristics may not be adequately representative of the general population. As importantly, few longitudinal MRI studies have measured brain volumetrics over periods exceeding 10 years and none have followed up participants over the entire course of their adulthood. This is partly due to the substantial practical challenges of doing so, and to the fact that MRI and related tools for the calculation of brain volumetrics have been widely available for only ~30 years. Incidentally, two other potential drawbacks of longitudinal studies are

survivorship bias (Nunney 1991) and self-selection bias (Heckman 1990). Thus, despite the many advantages of longitudinal studies, there is still considerable appeal in relying on cross-sectional brain volumetrics data when characterizing BV trajectories throughout adulthood. The reader is referred to the contributions of Raz & Rodrigue (2006) for further discussion of the relative merits of cross-sectional vs. longitudinal studies of brain morphometry.

Failures to report total volumes. Whereas most early MRI studies quantified only total volumes, many studies from the past decade have shifted focus to the detailed reporting of regional volumes and atrophy rates. One unfortunate and paradoxical side effect of this trend toward greater regional specificity and sophistication is that the total volumes of GM, WM and CSF are no longer being reported as often as before the advent of regional volumetric analysis. This can have unfortunate consequences because brain parcellation schemes often differ considerably across studies, such that various brain regions' measures fail to be reported in a way which facilitates summary. Thus, it may be tedious or impossible to calculate total volumes from regional volume data unless these data are complete and the parcellation schemes employed are specified unambiguously. Furthermore, many studies leveraging data from large repositories and efforts such as the Alzheimer's Disease Neuroimaging Initiative (ADNI, http://adni.loni.usc.edu), the Parkinson Progression Marker Initiative (PPMI, https://www.ppmi.info.org) and the UK Biobank (https://www.ukbiobank.ac.uk) may fail to report the total brain volumetrics of healthy control subjects included in these studies; alternatively, their reported volumetrics may apply to healthy control individuals whose inclusion criteria and demographics may differ substantially across any pair of datasets from these neuroimaging consortia. For reasons like these, it is very important that future studies continue to report BV, GMV, WMV and CSFV in addition to regional volumes.

Linear vs. quadratic models. Many studies of brain volumetrics—whether longitudinal or cross-sectional—have modelled the time-dependent trajectories of brain volumetrics using either linear or quadratic regression models. In a strict quantitative sense, quadratic fits are better than linear ones because the former yield a smaller sum of squared residuals, i.e. better goodness of fit. Nevertheless, due to the danger of overfitting in studies like ours, the question as to which model is more justified empirically deserves discussion. In some studies included here—e.g. Taki et al. (2004)— linear fits seem appropriate, and the quadratic relationship between volumetrics and age does not appear to be substantial. In others—e.g. Sowell et al. (2003)—visual inspection suggests that the contribution of the quadratic term to the BV trajectory is important. It has been proposed that the choice between linear and quadratic models is largely dependent upon the age range of participants in a given sample, and that researchers should strive to explore as broad a range as possible (Fjell et al. 2010). Furthermore, the inclusion of participants in the first and second decades of life may strengthen the argument in favor of using quadratic models due to the trajectory of brain dynamics during development (Sowell et al. 2003). The present meta-analysis of volumetric trajectories across adulthood provides evidence in support of the assertion by Fjell et al. (2010) according to which the most seemingly appropriate fit is determined by the range of ages included. Good et al. (2001) did not find significant WM decline with age in their sample of 465 healthy adults; these authors suggested that, in the first six decades of life, aging affects GMV more than WMV, and that both GMV and WMV decrease substantially after age ~60. Thus, the quadratic regression curves depicted in these authors' study suggest some GMV increase up to the fifth decade of life, with relatively slow subsequent decline into the eighth decade; by contrast, the decrease in WMV appears to be monotonic. These trajectories are reproduced by other studies (Fotenos et al. 2005; Walhovd et al. 2011). In particular, the results of Walhovd et al. (2011) indicate that cerebral GMV and BV decrease monotonically throughout adulthood, suggesting that a linear fit may be appropriate for these measures. By contrast, cerebral WMV exhibits nonlinear behavior in that it increases slightly between ages ~20 and

~60, and then decreases relatively faster than within this interval. The present meta-analysis found that, historically, studies based on quadratic models have been far less common than those using linear models. Because the eligible quadratic model studies identified here were few and had a relatively modest combined sample size, their inclusion in the meta-analysis was deemed to be problematic and potentially misleading. Future studies should therefore explore the suitability of quadratic models in more detail to warrant rigorous meta-analysis.

Meta-analytic limitations. It is important to acknowledge that this meta-analysis is neither exhaustive nor definitive. For example, although many studies were reviewed to generate it, it is both plausible and likely that many studies which fit the selection criteria of the meta-analysis were not included because of missed opportunities to identify them. Furthermore, different study selection criteria may have resulted in rather different meta-analysis results, and it is not clear that the selection of studies included here is the best possible selection. For example, studies published before 1990 were not included and very recent studies may not have been available in the public literature databases consulted at the time when the study was undertaken. For some measures, the lack of overlap between the CIs of several studies (see Figure 1) may suggest that some sources of between-study heterogeneity could not be removed even by implementing the stringent search criteria of the meta-analysis. It is also possible that some metrics included in the meta-analysis were systematically affected by confounds which were not discussed here. Finally, the values of various parameters extracted from each study may not be as numerically accurate as those originally computed by the respective study's authors based on their original data. All these limitations of this analysis are duly acknowledged. Nevertheless, this meta-analysis does cover a relatively large portion of the scientific literature on brain volumetrics and therefore has the potential to be of substantial utility to scientists and clinicians.

Conclusion

BV trajectories in adulthood are of substantial interest in studies of human aging. Historically, regression coefficients describing the relationship between age, BV and related metrics have been under-reported by MRI studies of human volumetry despite many researchers' need to compare brain trajectories across different populations to map effects associated with disease, genetics or environmental factors. Similarly, the average volume of the brain is a key fact about humans which is illustrative of mankind's uniqueness, and the ability to calculate its value at various ages during adulthood based on a normative, aggregate sample is very important to neuroscientists, gerontologists, anthropologists and clinicians. The cross-sectional data provided in this meta-analysis can facilitate such calculations and allow researchers to gain further insight to assist the interpretation and assessment of the literature on this important topic. Future studies of brain volumetrics across the lifespan should strive to provide decadal data and to explore the nonlinearities of age-dependent BV trajectories. Because there are relatively few longitudinal reports of brain atrophy during youth and middle age, more such studies should be undertaken to quantify BV loss. This could facilitate and advance the necessary transition from surmising volumetric trajectories from cross-sectional data to their fiducial estimation from longitudinal studies.

Conflict of interest statement

The author declares that he has no conflict of interest.

References

Abe, O., Yamasue, H., Aoki, S., Suga, M., Yamada, H., Kasai, K., et al. (2008). Aging in the CNS: comparison of gray/white matter volume and diffusion tensor data. *Neurobiology of Aging, 29*(1), 102-116, doi:10.1016/j.neurobiolaging.2006.09.003.

Allen, J. S., Bruss, J., Brown, C. K., & Damasio, H. (2005). Normal neuroanatomical variation due to age: the major lobes and a parcellation of the temporal region. *Neurobiology of Aging, 26*(9), 1245-1260; discussion 1279-1282, doi:10.1016/j.neurobiolaging.2005.05.023.

Arksey, H., & O'Malley, L. (2005). Scoping studies: towards a methodological framework. *International Journal of Social Research Methodology, 8*, 19-32.

Balafar, M. A., Ramli, A. R., Saripan, M. I., & Mashohor, S. (2010). Review of brain MRI image segmentation methods. *Artificial Intelligence Review, 33*(3), 261-274, doi:10.1007/s10462-010-9155-0.

Barnes, J., Ridgway, G. R., Bartlett, J., Henley, S. M., Lehmann, M., Hobbs, N., et al. (2010). Head size, age and gender adjustment in MRI studies: a necessary nuisance? *Neuroimage, 53*(4), 1244-1255, doi:10.1016/j.neuroimage.2010.06.025.

Begg, C. B., & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics, 50*(4), 1088-1101.

Benedetti, B., Charil, A., Rovaris, M., Judica, E., Valsasina, P., Sormani, M. P., et al. (2006). Influence of aging on brain gray and white matter changes assessed by conventional, MT, and DT MRI. *Neurology, 66*(4), 535-539, doi:10.1212/01.wnl.0000198510.73363.c6.

Bigler, E. D., & Tate, D. F. (2001). Brain volume, intracranial volume, and dementia. *Invest Radiol, 36*(9), 539-546.

Blatter, D. D., Bigler, E. D., Gale, S. D., Johnson, S. C., Anderson, C. V., Burnett, B. M., et al. (1995). Quantitative volumetric analysis of brain MR: normative database spanning 5 decades of life. *AJNR Am J Neuroradiol, 16*(2), 241-251.

Butler, A. B., & Hodos, W. (1996). *Comparative vertebrate neuroanatomy*. New York, NY: Wiley-Liss.

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, doi:10.1016/j.neuroimage.2007.03.063.

Cole, J. H., Jolly, A., de Simoni, S., Bourke, N., Patel, M. C., Scott, G., et al. (2018). Spatial patterns of progressive brain volume loss after moderate-severe traumatic brain injury. *Brain, 141*(3), 822-836, doi:10.1093/brain/awx354.

Cosgrove, K. P., Mazure, C. M., & Staley, J. K. (2007). Evolving knowledge of sex differences in brain structure, function, and chemistry. *Biol Psychiatry, 62*(8), 847-855, doi:10.1016/j.biopsych.2007.03.001.

de Boer, R., Vrooman, H. A., Ikram, M. A., Vernooij, M. W., Breteler, M. M., van der Lugt, A., et al. (2010). Accuracy and reproducibility study of automatic MRI brain tissue segmentation methods. *Neuroimage, 51*(3), 1047-1056, doi:10.1016/j.neuroimage.2010.03.012.

DeCarli, C., Massaro, J., Harvey, D., Hald, J., Tullberg, M., Au, R., et al. (2005). Measures of brain morphology and infarction in the framingham heart study: establishing what is normal. *Neurobiology of Aging, 26*(4), 491-510, doi:10.1016/j.neurobiolaging.2004.05.004.

DerSimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. *Control Clin Trials, 7*(3), 177-188.

DerSimonian, R., & Laird, N. (2015). Meta-analysis in clinical trials revisited. *Contemp Clin Trials, 45*(Pt A), 139-145, doi:10.1016/j.cct.2015.09.002.

Despotovic, I., Goossens, B., & Philips, W. (2015). MRI segmentation of the human brain: challenges, methods, and applications. *Comput Math Methods Med, 2015*, 450341, doi:10.1155/2015/450341.

Doherty, C. P., Fitzsimons, M., Holohan, T., Mohamed, H. B., Farrell, M., Meredith, G. E., et al. (2000). Accuracy and validity of stereology as a quantitative method for assessment of human temporal

lobe volumes acquired by magnetic resonance imaging. *Magnetic Resonance Imaging, 18*(8), 1017-1025, doi:Doi 10.1016/S0730-725x(00)00185-5.

Driscoll, I., Davatzikos, C., An, Y., Wu, X., Shen, D., Kraut, M., et al. (2009). Longitudinal pattern of regional brain volume change differentiates normal aging from MCI. *Neurology, 72*(22), 1906-1913, doi:10.1212/WNL.0b013e3181a82634.

Fjell, A. M., Walhovd, K. B., Westlye, L. T., Ostby, Y., Tamnes, C. K., Jernigan, T. L., et al. (2010). When does brain aging accelerate? Dangers of quadratic fits in cross-sectional studies. *Neuroimage, 50*(4), 1376-1383, doi:10.1016/j.neuroimage.2010.01.061.

Fotenos, A. F., Mintun, M. A., Snyder, A. Z., Morris, J. C., & Buckner, R. L. (2008). Brain volume decline in aging: evidence for a relation between socioeconomic status, preclinical Alzheimer disease, and reserve. *Arch Neurol, 65*(1), 113-120, doi:10.1001/archneurol.2007.27.

Fotenos, A. F., Snyder, A. Z., Girton, L. E., Morris, J. C., & Buckner, R. L. (2005). Normative estimates of cross-sectional and longitudinal brain volume decline in aging and AD. *Neurology, 64*(6), 1032-1039, doi:10.1212/01.WNL.0000154530.72969.11.

Gazzaniga, M. S. (2009). *Human: The science behind what makes your brain unique*. New York, NY: Harper Perennial

Ge, Y., Grossman, R. I., Babb, J. S., Rabin, M. L., Mannon, L. J., & Kolson, D. L. (2002). Age-related total gray matter and white matter changes in normal adult brain. Part I: volumetric MR imaging analysis. *AJNR Am J Neuroradiol, 23*(8), 1327-1333.

Good, C. D., Johnsrude, I. S., Ashburner, J., Henson, R. N., Friston, K. J., & Frackowiak, R. S. (2001). A voxel-based morphometric study of ageing in 465 normal adult human brains. *Neuroimage, 14*(1 Pt 1), 21-36, doi:10.1006/nimg.2001.0786.

Grant, M. J., & Booth, A. (2009). A typology of reviews: an analysis of 14 review types and associated methodologies. *Health Info Libr J, 26*(2), 91-108, doi:10.1111/j.1471-1842.2009.00848.x.

Gunning-Dixon, F. M., Brickman, A. M., Cheng, J. C., & Alexopoulos, G. S. (2009). Aging of cerebral white matter: a review of MRI findings. *Int J Geriatr Psychiatry, 24*(2), 109-117, doi:10.1002/gps.2087.

Gur, R. C., Gunning-Dixon, F. M., Turetsky, B. I., Bilker, W. B., & Gur, R. E. (2002). Brain region and sex differences in age association with brain volume: a quantitative MRI study of healthy young adults. *Am J Geriatr Psychiatry, 10*(1), 72-80.

Gur, R. C., Mozley, P. D., Resnick, S. M., Gottlieb, G. L., Kohn, M., Zimmerman, R., et al. (1991). Gender differences in age effect on brain atrophy measured by magnetic resonance imaging. *Proc Natl Acad Sci U S A, 88*(7), 2845-2849.

Guttmann, C. R., Jolesz, F. A., Kikinis, R., Killiany, R. J., Moss, M. B., Sandor, T., et al. (1998). White matter changes with normal aging. *Neurology, 50*(4), 972-978, doi:10.1212/wnl.50.4.972.

Heckman, J. (1990). Varieties of Selection Bias. *American Economic Review, 80*(2), 313-318.

Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ, 327*(7414), 557-560, doi:10.1136/bmj.327.7414.557.

Irimia, A., Maher, A. S., Rostowsky, K. A., Chowdhury, N. F., Hwang, D. H., & Law, E. M. (2019). Brain Segmentation From Computed Tomography of Healthy Aging and Geriatric Concussion at Variable Spatial Resolutions. *Frontiers in Neuroinformatics, 13*, 9, doi:10.3389/fninf.2019.00009.

Jancke, L., Merillat, S., Liem, F., & Hanggi, J. (2015). Brain size, sex, and the aging brain. *Human Brain Mapping, 36*(1), 150-169, doi:10.1002/hbm.22619.

Jernigan, T. L., Archibald, S. L., Berhow, M. T., Sowell, E. R., Foster, D. S., & Hesselink, J. R. (1991). Cerebral structure on MRI, Part I: Localization of age-related changes. *Biol Psychiatry, 29*(1), 55-67, doi:10.1016/0006-3223(91)90210-d.

Jernigan, T. L., Archibald, S. L., Fennema-Notestine, C., Gamst, A. C., Stout, J. C., Bonner, J., et al. (2001). Effects of age on tissues and regions of the cerebrum and cerebellum. *Neurobiology of Aging, 22*(4), 581-594.

Kaplan, H., Thompson, R. C., Trumble, B. C., Wann, L. S., Allam, A. H., Beheim, B., et al. (2017). Coronary atherosclerosis in indigenous South American Tsimane: a cross-sectional cohort study. *Lancet, 389*(10080), 1730-1739, doi:10.1016/S0140-6736(17)30752-3.

Lemaitre, H., Crivello, F., Grassiot, B., Alperovitch, A., Tzourio, C., & Mazoyer, B. (2005). Age- and sex-related effects on the neuroanatomy of healthy elderly. *Neuroimage, 26*(3), 900-911, doi:10.1016/j.neuroimage.2005.02.042.

Lemaitre, H., Goldman, A. L., Sambataro, F., Verchinski, B. A., Meyer-Lindenberg, A., Weinberger, D. R., et al. (2012). Normal age-related brain morphometric changes: nonuniformity across cortical thickness, surface area and gray matter volume? *Neurobiology of Aging, 33*(3), 617 e611-619, doi:10.1016/j.neurobiolaging.2010.07.013.

Lenroot, R. K., & Giedd, J. N. (2008). The changing impact of genes and environment on brain development during childhood and adolescence: initial findings from a neuroimaging study of pediatric twins. *Dev Psychopathol, 20*(4), 1161-1175, doi:10.1017/S0954579408000552.

Liu, R. S., Lemieux, L., Bell, G. S., Sisodiya, S. M., Shorvon, S. D., Sander, J. W., et al. (2003). A longitudinal study of brain morphometrics using quantitative magnetic resonance imaging and difference image analysis. *Neuroimage, 20*(1), 22-33.

Lorio, S., Kherif, F., Ruef, A., Melie-Garcia, L., Frackowiak, R., Ashburner, J., et al. (2016). Neurobiological origin of spurious brain morphological changes: A quantitative MRI study. *Human Brain Mapping, 37*(5), 1801-1815, doi:10.1002/hbm.23137.

Matsumae, M., Kikinis, R., Morocz, I. A., Lorenzo, A. V., Sandor, T., Albert, M. S., et al. (1996). Age-related changes in intracranial compartment volumes in normal adults assessed by magnetic resonance imaging. *J Neurosurg, 84*(6), 982-991, doi:10.3171/jns.1996.84.6.0982.

Michielse, S., Coupland, N., Camicioli, R., Carter, R., Seres, P., Sabino, J., et al. (2010). Selective effects of aging on brain white matter microstructure: a diffusion tensor imaging tractography study. *Neuroimage, 52*(4), 1190-1201, doi:10.1016/j.neuroimage.2010.05.019.

Natu, V. S., Gomez, J., Barnett, M., Jeska, B., Kirilina, E., Jaeger, C., et al. (2019). Apparent thinning of human visual cortex during childhood is associated with myelination. *Proceedings of the National Academy of Sciences of the United States of America, 116*(41), 20750-20759, doi:10.1073/pnas.1904931116.

Nieminen, P., Lehtiniemi, H., Vahakangas, K., Huusko, A., & Rautio, A. (2013). Standardised regression coefficient as an effect size index in summarising findings in epidemiological studies. *Epidemiology, Biostatistics and Public Health, 10*(4), 1-15.

Nunney, L. (1991). The influence of age structure and fecundity on effective population size. *Proc Biol Sci, 246*(1315), 71-76, doi:10.1098/rspb.1991.0126.

Peelle, J. E., Cusack, R., & Henson, R. N. (2012). Adjusting for global effects in voxel-based morphometry: gray matter decline in normal aging. *Neuroimage, 60*(2), 1503-1516, doi:10.1016/j.neuroimage.2011.12.086.

Prothero, D. R. (2007). *Evolution: What the fossils say and why it matters*. New York, NY: Columbia University Press.

Raz, N., & Rodrigue, K. M. (2006). Differential aging of the brain: patterns, cognitive correlates and modifiers. *Neurosci Biobehav Rev, 30*(6), 730-748, doi:10.1016/j.neubiorev.2006.07.001.

Resnick, S. M., Goldszal, A. F., Davatzikos, C., Golski, S., Kraut, M. A., Metter, E. J., et al. (2000). One-year age changes in MRI brain volumes in older adults. *Cereb Cortex, 10*(5), 464-472, doi:10.1093/cercor/10.5.464.

Rushton, J. P., & Ankney, C. D. (1996). Brain size and cognitive ability: Correlations with age, sex, social class, and race. *Psychon Bull Rev, 3*(1), 21-36, doi:10.3758/BF03210739.

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

Seshadri, S., DeStefano, A. L., Au, R., Massaro, J. M., Beiser, A. S., Kelly-Hayes, M., et al. (2007). Genetic correlates of brain aging on MRI and cognitive test measures: a genome-wide association and linkage analysis in the Framingham Study. *BMC Med Genet, 8 Suppl 1*, S15, doi:10.1186/1471-2350-8-S1-S15.

Smith, C. D., Chebrolu, H., Wekstein, D. R., Schmitt, F. A., & Markesbery, W. R. (2007). Age and gender effects on human brain anatomy: a voxel-based morphometric study in healthy elderly. *Neurobiology of Aging, 28*(7), 1075-1087, doi:10.1016/j.neurobiolaging.2006.05.018.

Smith, S. M., De Stefano, N., Jenkinson, M., & Matthews, P. M. (2001). Normalized accurate measurement of longitudinal brain change. *J Comput Assist Tomogr, 25*(3), 466-475.

Sowell, E. R., Peterson, B. S., Thompson, P. M., Welcome, S. E., Henkenius, A. L., & Toga, A. W. (2003). Mapping cortical change across the human life span. *Nat Neurosci, 6*(3), 309-315, doi:10.1038/nn1008.

Taki, Y., Goto, R., Evans, A., Zijdenbos, A., Neelin, P., Lerch, J., et al. (2004). Voxel-based morphometry of human brain with age and cerebrovascular risk factors. *Neurobiology of Aging, 25*(4), 455-463, doi:10.1016/j.neurobiolaging.2003.09.002.

Uspenskii, S. I. (1964). A New Method for Measuring Cranial Capacity. *American Journal of Physical Anthropology, 22*(1), 115-117, doi:DOI 10.1002/ajpa.1330220123.

Van Leemput, K., Bakkour, A., Benner, T., Wiggins, G., Wald, L. L., Augustinack, J., et al. (2009). Automated segmentation of hippocampal subfields from ultra-high resolution in vivo MRI. *Hippocampus, 19*(6), 549-557, doi:10.1002/hipo.20615.

Voevodskaya, O., Simmons, A., Nordenskjold, R., Kullberg, J., Ahlstrom, H., Lind, L., et al. (2014). The effects of intracranial volume adjustment approaches on multiple regional MRI volumes in healthy aging and Alzheimer's disease. *Front Aging Neurosci, 6*, 264, doi:10.3389/fnagi.2014.00264.

Walhovd, K. B., Westlye, L. T., Amlien, I., Espeseth, T., Reinvang, I., Raz, N., et al. (2011). Consistent neuroanatomical age-related volume differences across multiple samples. *Neurobiology of Aging, 32*(5), 916-932, doi:10.1016/j.neurobiolaging.2009.05.013.

Figure Caption

Figure 1. Forest plots of regression coefficients (A) and volumes (B) for the brain (B, first column), gray matter (GM, second column), white matter (WM, third column) and cerebrospinal fluid (CSF, fourth column). The first author of each study included in the meta-analysis is listed on the left; studies are listed in ascending chronological order, with the oldest one at the top and the newest one at the bottom. Means are indicates by red squares whose sizes are proportional to studies' sample sizes; standard deviations are indicated by horizontal lines. In (A), the regression coefficients β are expressed as percentages of volume difference per year [%/year] (see also Table 1). In (B), volumes are expressed in cm^3. See Tables 1 and 2 for additional data.

		N			β_1											
					BV			GMV			WMV			$CSFV$		
first author	year	M	F	M&F	M	F	M&F	M	F	M&F	M	F	M&F	M	F	M&F
Irimia	**2019**	**3260**	**3869**	**7065**	**-0.28**	**-0.22**	**-0.25**	**-0.43**	**-0.36**	**-0.37**	**-0.11**	**-0.10**	**-0.10**	**1.93**	**2.15**	**1.67**
Jernigan	1991	34	21	55	—	—	—	—	—	-0.26	—	—	-0.04	—	—	—
Gur	1991	34	35	69	-0.28	-0.12	—	—	—	—	—	—	—	—	—	0.74
Matsumae	1996	26	23	49	-0.27	-0.21	—	—	—	—	—	—	—	0.24	0.18	—
Gutmann	1998	22	50	72	-0.24	-0.25	—	-0.38	-0.32	-0.11	—	—	-0.23	—	—	1.77
Good	2001	273	192	465	—	—	-0.38	—	—	—	-0.01	-0.04	—	0.62	0.69	—
Jernigan	2001	37	41	78	—	—	-0.21	—	—	-0.23	—	—	-0.44	—	—	6.93
Ge	2002	54	32	22	—	—	-0.21	—	—	-0.23	—	—	-0.18	—	—	2.02
Sowell	2003	90	86	176	—	—	-0.24	—	—	-0.27	—	—	-0.21	—	—	1.63
Liu	2003	49	41	90	—	—	-0.17	—	—	-0.14	—	—	-0.21	—	—	—
Scahill	2003	18	21	39	—	—	-0.33	—	—	—	—	—	—	—	—	5.33
Taki	2004	356	413	769	-0.29	-0.23	—	-0.54	-0.48	—	0.11	0.11	—	—	—	—
Allen	2005	43	44	87	-0.23	-0.25	—	-0.22	-0.24	—	-0.23	-0.26	—	—	—	—
DeCarli	2005	948	1133	2081	-0.24	-0.18	—	—	—	—	—	—	—	—	—	—
Fotenos	2005	112	160	272	—	—	-0.19	—	—	-0.31	—	—	-0.12	—	—	—
Fotenos	2008	137	225	362	-0.24	-0.20	—	—	—	—	—	—	—	—	—	—
Abe	2008	0	73	73	—	—	—	—	-0.36	—	—	-0.07	—	—	—	—
Barnes	2010	37	41	78	—	—	-0.25	—	—	-0.32	—	—	-0.14	—	—	2.43
Michielse	2010	17	52	69	—	—	-0.27	—	—	-0.33	—	—	-0.10	—	—	1.25
Walhovd	2011	355	528	883	-0.48	-0.34	—	-0.53	-0.37	—	-0.42	-0.30	—	4.15	3.56	—
Peele	2012	214	206	420	-0.10	-0.10	—	-0.12	-0.13	—	-0.08	-0.05	—	0.12	0.13	—
Jancke	2015	404	452	856	—	—	-0.27	—	—	-0.39	—	—	-0.14	—	—	0.26

Table 1. Linear regression coefficients β_1 convey the annual percentage change in the normalized volume V/ICV of the brain (B), gray matter (GM), white matter (WM) and cerebrospinal fluid (CSF). Regression coefficients are reported as percentage differences in volume per year (%/year). Meta-analysis results are displayed in **bold face**. The quantity V/ICV is a structure's volume V divided (i.e. normalized) by the intracranial volume V_{IC}. Thus, regression coefficients β_1 reported for the brain (B) convey the annual percentage change in BV/ICV (the brain volume BV normalized by the intracranial volume ICV). For example, Gur et al. (1991) report that, upon accounting for intracranial volume, males' brain volumes are observed to decrease by 0.28% per year on average. Regression coefficients are listed separately for males and females, where possible; for studies which do not distinguish trajectories according to sex, regression coefficients are listed under "M&F." All linear regression models concern the time interval from 20 to 70 years and assume that volumetric changes occur linearly within this interval. The sample size N of

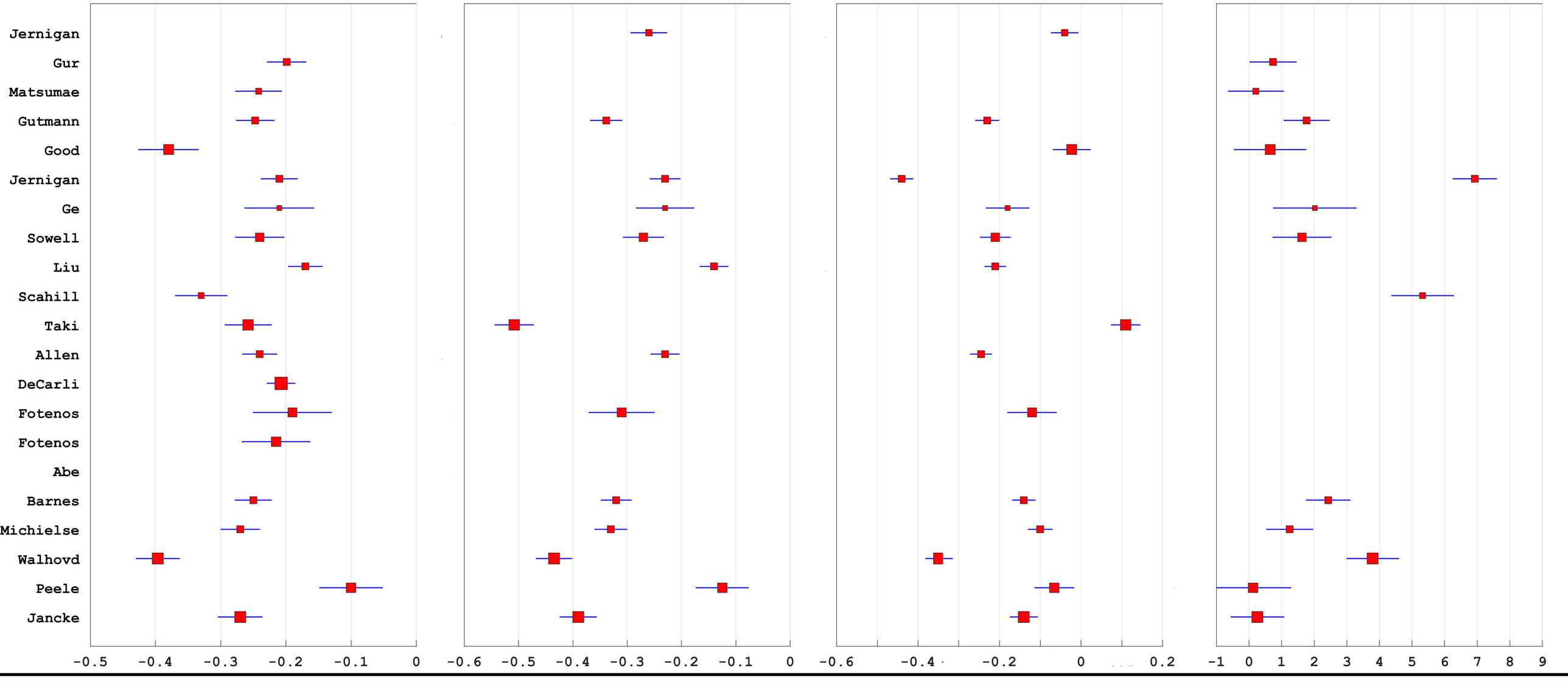
each study is also reported. A dash indicates that data were not reported by the study in question, or that reported data could not be utilized in the meta-analysis. Abbreviations: M = males; F = females; M&F = males and females; B = brain; GM = gray matter; WM = white matter; CSF = cerebrospinal fluid. `

		first author	**Irimia**	Gur	Blatter	Resnick	Good	Gur	Sowell	Scahill	Taki	Allen	DeCarli	Lemaître	Benedetti	Chen	Smith	Abe	Fotenos	Driscoll	Michielse	Walhovd	Lemaître	Peele
	age	year	**2019**	1991	1995	2000	2001	2002	2003	2003	2004	2005	2005	2005	2006	2007	2007	2008	2008	2009	2010	2011	2012	2012
N	all	M	**3474**	34	89	68	273	57	90	18	356	43	948	331	39	184	51	0	137	73	17	355	97	214
		F	**4065**	35	105	48	192	59	86	21	413	44	1133	331	50	227	71	73	225	47	52	528	119	206
		M&F	**7539**	69	194	116	465	116	176	39	769	87	2081	662	89	411	122	73	362	120	69	883	216	420
IC	20	M	**1533**	—	1594	—	1690	—	—	—	—	—	1400	—	—	—	—	—	—	—	—	1694	—	1625
		F	**1374**	—	1400	—	1550	—	—	—	—	—	1250	—	—	—	—	1450	—	—	—	1510	—	1500
		M&F	**1446**	—	—	—	—	—	1450	—	—	—	—	—	—	—	—	—	—	—	1400	—	—	—
	45	M	**1516**	—	1538	—	1640	—	—	—	—	—	1360	—	—	1650	—	—	—	—	—	1695	—	1625
		F	**1358**	—	1347	—	1540	—	—	—	—	—	1210	—	—	1440	—	1438	—	—	—	1531	—	1450
		M&F	**1430**	—	—	—	—	—	1445	1390	—	—	—	—	—	—	—	—	—	—	1400	—	—	—
	70	M	**1434**	—	1548	—	1630	—	—	—	—	997	1320	—	—	—	—	—	—	—	—	1662	—	1620
		F	**1296**	—	1335	—	1530	—	—	—	—	882	1150	—	—	—	—	1425	—	—	—	1469	—	1450
		M&F	**1373**	—	—	—		—	1390	1390	—	—	—	—	—	—	—	—	—	—	1400	—	—	—
B	20	M	**1185**	1180	1368	—	1315	1240	—	—	—	1123	1100	—	—	—	—	—	—	—	—	1176	—	1330
		F	**1110**	1090	1354	—	1175	1210	—	—	—	1009	1050	—	—	—	—	1143	—	—	—	1119	—	1220
		M&F	**1149**	—	—	—	—	—	1310	—	—	—	—	—	—	—	—	—	1190	—	832	—	—	—
	45	M	**1179**	1140	1316	—	1225	1190	—	—	—	1059	1100	—	—	—	—	—	—	—	—	1247	—	1330
		F	**1051**	1050	1323	—	1125	1070	—	—	—	943	980	—	—	—	—	1043	—	—	—	1088	—	1160
		M&F	**1116**	—	—	—	—	—	1255	1240	—	—	—	—	—	—	—	—	1160	—	800	—	—	—
	70	M	**1067**	1080	1268	1130	1125	—	—	—	—	920	1003	1066	—	—	—	—	—	—	—	1030	—	1260
		F	**947**	1050	1285	920	1025	—	—	—	—	790	874	960	—	—	—	988	—	—	—	929	—	1140
		M&F	**1009**	—	—	—	—	—	1150	1140	—	—	—	—	—	—	—	—	1080	950	725	—	—	—
GM	20	M	**749**	—	689	—	865	710	—	—	700	604	—	—	—	—	—	—	—	—	—	728	—	780
		F	**658**	—	638	—	785	660	—	—	590	555	—	—	—	—	—	730	—	—	—	644	—	720
		M&F	**692**	—	—	—	—		830	—	—	—	—	—	790	—	—	—	—	—	520	—	495	—
	45	M	**694**	—	682	—	775	660	—	—	600	571	—	—	—	741	—	—	—	—	—	684	—	760
		F	**612**	—	640	—	715	610	—	—	520	521	—	—	—	667	—	630	—	—	—	600	—	670
		M&F	**641**	—	—	—	—	—	730	—	—	—	—	—	760	—	—	—	—	—	475	—	440	—
	70	M	**598**	—	644	570	685	—	—	—	520	460	—	575	—	—	—	—	—	—	—	562	—	730
		F	**540**	—	648	500	650	—	—	—	470	395	—	532	—	—	—	575	—	—	—	507	—	650
		M&F	**560**	—	—	—	—	—	680	—	—	—	—	—	—	—	560	—	—	525	425	—	400	—
WM	20	M	**509**	—	678	—	450	530	—	—	460	519	—	—	—	—	—	—	—	—	—	531	—	550
		F	**460**	—	638	—	390	450	—	—	420	454	—	—	—	—	—	413	—	—	—	475	—	500
		M&F	**488**	—	—	—	—	—	480	—	—	—	—	—	770	—	—	—	—	—	312	—	—	—
	45	M	**518**	—	633	—	450	570	—	—	470	488	—	—	—	496	—	—	—	—	—	563	—	570
		F	**464**	—	683	—	400	460	—	—	430	422	—	—	—	425	—	413	—	—	—	488	—	490
		M&F	**494**	—	—	—	—	—	525	—	—	—	—	—	710	—	—	—	—	—	325	—	—	—
	70	M	**486**	—	624	460	440		—	—	480	460	—	491	—	—	—	—	—	—	—	468	—	530
		F	**442**	—	638	420	375	—	—	—	450	395	—	428	—	—	—	413	—	—	—	421	—	490
		M&F	**457**	—	—	—	—	—	470	—	—	—	—	—	—	—	380	—	—	425	300	—	—	—

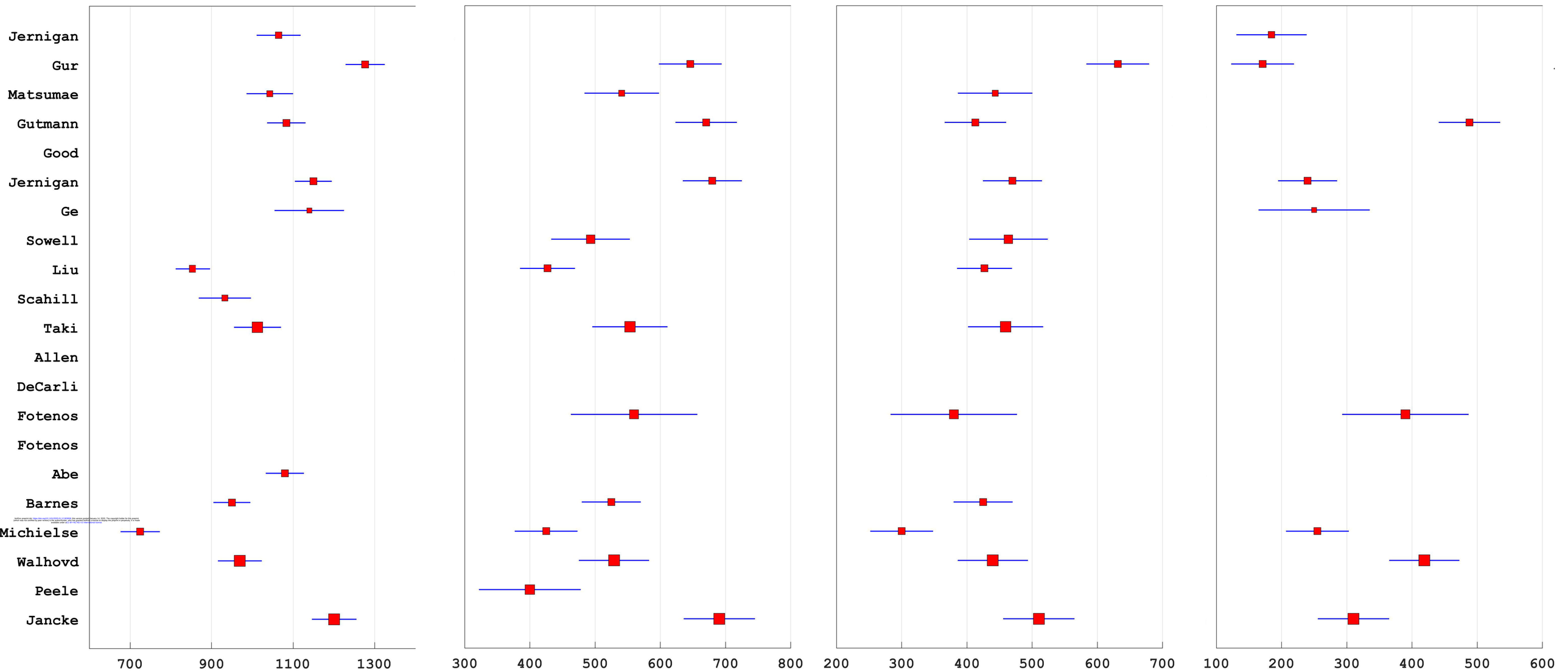
		M	**188**	80	78	—	370	95	—	—	—	—	—	—	—	—	—	—	—	—	20	—	320
	20	F	**136**	75	91	—	370	80	—	—	—	—	—	—	—	—	—	—	—	—	16	—	275
		M&F	**159**	—	—	—	—	—	140	—	—	—	—	—	—	—	—	—	—	160	—	—	—
		M	**300**	140	129	—	420	110	—	—	—	—	—	—	—	410	—	—	—	—	224	—	325
CSF	45	F	**253**	100	122	—	410	100	—	—	—	—	—	—	—	350	—	—	—	—	196	—	280
		M&F	**266**	—	—	—	—	—	190	150	—	—	—	—	—	—	—	—	—	180	—	—	—
		M	**413**	210	178	—	480	—	—	—	—	—	—	—	—	—	—	—	—	—	489	—	330
	70	F	**352**	160	165	—	500	—	—	—	—	—	—	—	—	—	—	—	—	—	372	—	290
		M&F	**365**	—	—	—	—	—	240	250	—	—	—	—	—	—	390	—	—	255	—	—	—

Table 2. Average volumes of the brain (B), gray matter (GM), white matter (WM) and cerebrospinal fluid (CSF) at the age of 20, 45 and 70. Meta-analysis results which aggregate all studies listed are displayed in **bold face**. Values are listed separately for males and females, where possible; for studies which do not distinguish trajectories by sex, values are listed under "M&F." The sample size N of each study is also reported. A dash indicates that data were not reported by the study in question, or that reported data could not be utilized in the meta-analysis. Abbreviations: M = males; F = females; M&F = males and females; B = brain; GM = gray matter; WM = white matter; CSF = cerebrospinal fluid.

(A) regression coefficient



(B) volume



B GM WM CSF