NORMATIVE DATA FOR SUBCORTICAL REGIONAL VOLUMES OVER THE LIFETIME OF THE ADULT HUMAN BRAIN

Olivier Potvin[1], PhD, Abderazzak Mouiha[1], PhD, Louis Dieumegarde[1], BSc,

and Simon Duchesne, PhD[12] for the Alzheimer's Disease Neuroimaging Initiative*

[1] Centre de recherche de l'Institut universitaire en santé mentale de Québec, 2601, de la Canardière, Québec, Canada, G1J 2G3

[2] Département de radiologie, Université Laval, 1050, avenue de la Médecine, Québec, Canada, G1V 0A6

*Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

Correspondence:

Simon Duchesne, PhD
Centre de recherche de l'Institut universitaire en santé mentale de Québec
2601, de la Canardière, Québec, Canada, G1J 2G3
Phone: 418 663-5000 ext.4777
Fax: 418 663-9540
Simon.Duchesne@fmed.ulaval.ca

ABSTRACT

Normative data for volumetric estimates of brain structures are necessary to adequately assess brain volume alterations in individuals with suspected neurological or psychiatric conditions. Although many studies have described age and sex effects in healthy individuals for brain morphometry assessed via magnetic resonance imaging, proper normative values allowing to quantify potential brain abnormalities are needed. We developed norms for volumetric estimates of subcortical brain regions based on cross-sectional magnetic resonance scans from 2790 healthy individuals aged 18 to 94 years using 23 samples provided by 21 independent research groups. The segmentation was conducted using *FreeSurfer*, a widely used and freely available automated segmentation software. Models predicting subcortical regional volumes of each hemisphere were produced including age, sex, estimated total intracranial volume (eTIV), scanner manufacturer, magnetic field strength, and interactions as predictors. The mean explained variance by the models was 48%. For most regions, age, sex and eTIV predicted most of the explained variance while manufacturer, magnetic field strength and interactions predicted a limited amount. Estimates of the expected volumes of an individual based on its characteristics and the scanner characteristics can be obtained using derived formulas. For a new individual, significance test for volume abnormality, effect size and estimated percentage of the normative population with a smaller volume can be obtained. Normative values were validated in independent samples of healthy adults and in adults with Alzheimer's disease and schizophrenia.

Key words: magnetic resonance imaging, atrophy, morphometry, normality, aging, sex.

1. INTRODUCTION

Many neurological diseases and neuropsychiatric disorders display specific subcortical changes detectable using anatomical magnetic resonance imaging (MRI) when comparing a group of affected individuals to non-affected controls (Haijma et al., 2013; Scahill et al., 2002; Sheline et al., 1999). At an individual-level, however, measuring brain volume alterations is problematic given the lack of reference standards to estimate the degree of deviation from the normality according to one's characteristics.

Indeed, although many studies have described the influence of age and sex on brain volumes (Fjell et al., 2013; Luders et al., 2009; Pfefferbaum et al., 2013; Walhovd et al., 2011), very few attempts have been made to produce proper neuroanatomical volumetric normative data (Brain Development Cooperative Group, 2012; Kruggel, 2006). The many obstacles inherent to neuroimaging research likely undermine this shortcoming. To produce normative data, brain segmentation procedures need first to be replicable and thus ideally automated. However, automated segmentation techniques are often proprietary, and therefore not readily accessible outside of the technical teams that developed them. It can be readily shown that regional brain volumes display important variability according to the segmentation techniques (Mouiha and Duchesne, 2011; Tae et al., 2008) and anatomical definitions (Boccardi et al., 2014). Secondly, scanner characteristics, especially related to each manufacturer and magnetic field strength (MFS), have a non-negligible impact on regional brain segmentation (Jovicich et al., 2009; Kruggel et al., 2010; Pfefferbaum et al., 2012). Finally, to produce neuroanatomical volumetric normative data useful across the lifespan, a large sample of individuals covering a wide age range is needed; however, given that MRI is an expensive proposition, a single laboratory or team can achieve such sample sizes with difficulty.

Our objective was to build normative data for subcortical regional volumes covering adulthood to facilitate neuroscience imaging studies. To this end, we federated a large sample of cognitively healthy individuals originating from 23 different datasets. We produced estimates of subcortical regional volumes using *FreeSurfer*, a widely used and freely available automated segmentation software. We built models predicting expected volumes for each subcortical region according to age, sex, estimated total intracranial volume (eTIV), scanner manufacturer, and MFS. The expected volumes allow testing each region for volume abnormality, effect sizes and estimates of the normative population with a smaller volume. These models are presented within the article and a statistics calculator is freely distributed as supplementary material (see the Subcortical norms calculator in Potvin et al., submitted for publication).

2. MATERIALS AND METHODS

2.1 Normative sample

We assembled a sample of 3D T1-weighted MRI scans from 2,799 cognitively healthy controls aged 18 to 94 years from 23 samples provided by 21 independent research groups (see Table 1 and Acknowledgments for details). Of note, this includes the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu), launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. (www.adni-info.org). Scans were acquired from one of the three leading manufacturers (e.g. Siemens Healthcare, Philips Medical Systems, or GE Healthcare) at MFS of either 1.5 or 3 Tesla. For each dataset, approval from the local ethics board and informed consent of the participants were obtained.

All samples recruited healthy control participants, except NKI1 and NKI2. Databases with older adults excluded neurological diseases and neuropsychiatric disorders with extensive

assessments for age-related disorders. For databases recruiting in the general population (NKI1 and NKI2), we excluded participants with schizophrenia or other psychotic disorders, bipolar disorders, major depressive disorders and substance abuse/dependence disorders. Additional exclusions were made for NKI2: neurodegenerative and neurological disorders, head injury with loss of consciousness/amnesia, and lead poisoning. Moreover, for PPMI, additional exclusions were made for participants with a Geriatric Depression Scale (Sheikh and Yesavage, 1986) score of more than 5 (inclusion criterion used in ADNI and AIBL databases).

All images were visually inspected and four participants were discarded because of evident brain abnormalities. Five participants with extreme eTIV values were also excluded (Z scores higher than 3.29, $p < .001$). The final sample included 2,790 individuals aged between 18 and 94 years (mean: 47.6, SD: 21.8), with a similar proportion of men (n = 1389) and women (n = 1401). More than half of the scans were acquired using Siemens (n = 1524), a third using Philips (n = 787), and 17% using GE (n = 479) units. Fifty-three percent of the images were obtained using 3T MFS (n = 1487). Most of the datasets also had information about handedness (79%), race (60%), and education (58%). Based on the available data, the vast majority of the normative sample was right-handed (91%), Caucasian (82%; African 10%; Asian 7%), and had completed high school (95%).

Table 2 shows additional details about the age and sex of the participants according to scanner manufacturer and MFS strata. Table 2 also displays the voxel size and acquisition plane of the scan as well as the list of scanner models for each strata.

2.2 Validation samples

We randomly selected 5% (n = 140) of the normative sample stratified by manufacturer and MFS to validate normative volumetric formulas in an independent sample. This validation sample was not used to build the predictive models. Moreover we also validated the models using clinical samples of individuals with schizophrenia (SZ; n = 69; Age: 38.5 ±13.9, range 18-65; 20% female) from the COBRE dataset and mild Alzheimer's disease (AD; n = 50 Age: 74.6 ±7.6, range 56-90; 40% female) randomly selected from the ADNI-2 dataset. Schizophrenia was diagnosed using the Structured Clinical Interview for DSM-IV disorders (First et al., 1996). Alzheimer's disease was diagnosed according to National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association (NINCDS/ADRDA) criteria for probable AD (McKhann et al., 1984) and had a Clinical Dementia Rating of 0.5 or 1.

2.3 Segmentation

Subcortical segmentation was conducted using *FreeSurfer* (5.3), a widely used and freely available automated processing pipeline that quantifies brain anatomy (http://freesurfer.net). All raw T1-weighted images were first converted into MINC format and then were processed using the "recon –all" pipeline with the default set of parameters. *Freesurfer* was running on an Ubuntu Server 12.04 LTS platform on a Dell PowerEdge R910 computer with four Intel Xeon E7-4870 2.4GHz. The technical details of *FreeSurfer* have been described elsewhere (Fischl et al., 2002; Fischl et al., 2004; Jovicich et al., 2006; Segonne et al., 2004). The *FreeSurfer* software belongs to a class of segmentation techniques using a model-driven paradigm. In these approaches the algorithm first matches the new image to a template and/or series of templates from a training set, for which segmentation has been performed a priori, and therefore label information exists.

The algorithm then automatically assigns a neuroanatomical label to each voxel of a volume based on the probabilistic information given by the image matching procedure. Specifically, it assigns the most likely probability for that voxel, taking into consideration nearby voxel probabilities. Every structure defined in the a priori segmentation therefore becomes represented in the new image, based on the overall matching between images.

Subcortical and estimated total intracranial volumes (eTIV)(Buckner et al., 2004) were taken from the aseg.stats *Freesurfer* output file. Ventricles and corpus callosum volumes were generated using the sum of all subregions. *FreeSurfer* subcortical segmentation showed a high overlap and high volumetric correlations with manual segmentation (Dewey et al., 2010; Fischl et al., 2002; Keller et al., 2012) and high test-retest reliability (Liem et al., 2015; Morey et al., 2010).

Visual inspection of each brain segmentation was conducted using *FreeView* (http://freesurfer.net) by scrolling the entire brain at least through the coronal and axial planes. Regions with apparent segmentation error on multiple slices were excluded of statistical analyses (e.g. portion of gray matter not segmented, portion of a ventricle segmented as white matter, hippocampal portion segmented as neocortex). Depending on the region, between 0 and 58 participants out of 2790 were discarded (for the overall measures of ventricles and subcortical gray matter, which encompassed all the ventricles and all the gray matter regions, 2 and 96 participants were excluded, respectively). Moreover, to verify the validity of outermost eTIV values, we verified the registration of the 5% lowest and highest values.

In order to assure generalizability, we quantified the impact of a different hardware setup on the volumes generated by *FreeSurfer* (Xubuntu 12.04 on VirtualBox 4.3.10 installed on an iMac 10GB 1067 MHz DDR3 with 2.8GHz Intel Core i7 and OS X Yosemite 10.10.4). We

compared these volumes with those produced by the setup generating normative values on a random subset of the normative sample (n = 50).

2.4 Statistical analyses

2.4.1 Volume prediction

Regression models predicting subcortical regional volumes were built using age, sex, eTIV, MFS, and scanner manufacturer as predictors. Quadratic and cubic terms for age and eTIV were tested, as well as the following interactions: age X sex, eTIV X MFS, MFS X manufacturer, and eTIV X manufacturer. To avoid overfitting and maximize generalizability of the predictions, the best predictive model was determined with a 10-fold cross-validation (Hastie et al., 2008), retaining the model with the subset of predictors that produced the lowest predicted residual sum of squares using SAS 9.4 PROC GLMSELECT (SAS Institute Inc., Cary, NC, USA). For each selected final model, the fit of the data was assessed using R^2 (one minus the regression sum of squares divided by the total sum) and individual predictors' weight was measured by semi-partial eta squares (squared semi-partial correlations). For each brain subdivision and eTIV, outliers with volume Z scores higher than 3.29 ($p < .001$) were excluded (depending of the region, between 5 and 25 outliers out of 2790 were excluded). Because of positive skewness, the volume of all ventricles, except the fourth, was $\log_{10}$ transformed for statistical analyses.

2.4.2 Validation

In addition to the cross-validation procedure, the predictions of the models were validated by first calculating a validation R^2, using the squared correlation between observed and predicted volumes in the independent validation sample of healthy controls. Secondly, we examined the validity of the normative values to show the expected patterns of atrophy, hypertrophy or

normality in the validation samples of healthy individuals and individuals with AD and SZ. For each group, we tested the mean difference between observed and predicted volumes using independent two-sample *t*-tests (since predicted volumes are not produced using the observed volumes and thus, observed and predicted volumes are not correlated) with Bonferroni correction.

The impact of a different computer hardware setup on the volumes generated by *FreeSurfer* was tested by dependent one-sample *t*-tests with Bonferroni correction.

2.4.3 Normative statistics

For each region, we computed prediction intervals, single case significance test of volume abnormality, effect size and estimated percentage of the normative population with a smaller volume (Crawford and Garthwaite, 2006; Crawford et al., 2012). A Microsoft Excel spreadsheet able to produce these statistics is available as supplementary material (see Subcortical norms calculator in Potvin et al., submitted for publication). Single case significance test of volume abnormality was computed by the formula below, a *t*-statistic with *N* - k (number of predictors) - 1 degrees of freedom using the difference between actual (Y_0) and predicted ($\hat{Y}$) volumes, divided by the standard error of the predicted volume where $S_{Y \cdot X}$ represents the root mean square error (also called residual standard deviation or standard error of estimate) of the model predicting normative values, r^{ii} identifies off-diagonal elements of the inverted correlation matrix for the *k* predictor variables, r^{ij} identifies elements in the main diagonal, and $z_0 = (z_{i0}, \ldots, z_{k0})$ identifies the patient's scores on the predictor variables in *z* score form (Crawford and Garthwaite, 2006).

$$\frac{Y_0 - \hat{Y}}{S_{Y \cdot X}\sqrt{1 + \frac{1}{N} + \frac{1}{N-1}\sum r^{ii} z_{io}^2 + \frac{2}{N-1}\sum r^{ij} z_{io}\ z_{jo}}}$$

This method also produced an unbiased point estimate of the volume abnormality, supplemented with confidence intervals following a non-central *t*-distribution (Crawford and Garthwaite, 2006). For effect size, a Z score (Z_{OP}) is obtained by subtracting the *Observed value* from the *Predicted value* divided by the root mean square error of the model predicting normative values (Crawford et al., 2012).

3. RESULTS

3.1 Prediction of subcortical volumes

Table 3 displays the models predicting subcortical volumes. Most models had a substantial amount of explained variance (mean R^2: 48%, range: 14%-76%). Figure 1 shows that the explained variance for most regions was mainly predicted by age, followed by eTIV and sex, while manufacturer, MFS, and interactions between variables did not have a large effect (for detailed results see Table 1 in Potvin et al., submitted for publication). Age had a substantial effect for all regions except the brainstem. The effect of sex varied greatly across regions, with the strongest impact for the brainstem and the weakest for the fourth ventricle and the corpus callosum.

Figure 2 illustrates predicted volumes for each region according to age and sex. All relationships between age and volume were nonlinear, and included either cubic or quadratic terms. A few regions, including the accumbens, pallidum, and putamen, had a marked age by sex interaction.

Figure 3 displays some examples of the MFS, eTIV, and manufacturer effects observed. As illustrated, for some regions, MFS and eTIV had different effects depending on the manufacturer. The effect of eTIV was also altered according to MFS.

3.2 Validation

3.2.1 Healthy controls

The mean difference between validation and original R^2 was -0.4% (range -10 to 12%), which shows adequate generalization of the models. The largest negative discrepancies were for the right accumbens (-9%), the right caudate (-10%) and the left putamen (-10%)(for detailed results see Table 1 in Potvin et al., submitted for publication). Table 4 indicates that for all regions, the mean actual volumes did not significantly differ from the mean predicted normative volumes. The mean Z_{OP} effect size indicated very little deviation from the normative values across regions (Range between -0.18 and 0.08).

3.2.2 Schizophrenia and Alzheimer's disease

In the SZ group (Table 4), the mean volumes of the right accumbens, bilateral amygdala, and bilateral hippocampi, were significantly smaller, while the left pallidum and left inferior lateral ventricle were significantly larger than the mean predicted normative values. The mean Z_{OP} effect size for SZ indicated small deviations from the normative values across regions (range between -0.64 and 1.00).

In the mild AD group (Table 4), volumes of the right accumbens, bilateral amygdala and hippocampi, and total subcortical gray matter were significantly smaller, while the volumes of sum of the ventricles, bilateral lateral and inferior lateral ventricles were significantly larger than the mean predicted normative volumes. As a group, these differences varied from small to large deviations from the normative values (Z_{OP}: -2.55 and 1.58).

Figure 4 shows examples of the distribution of effect sizes among the validation samples for the results discussed above.

3.3 Influence of computer hardware

The impact of using a different hardware setup to generate *FreeSurfer* volumes was minimal (see Table 2 in Potvin et al., submitted for publication); mean difference for all regions: 0.1%, 95%CI: -0.70-0.94%) and no significant difference between setups was observed.

4. DISCUSSION

The objective of the present study was to produce normative values for subcortical regional volumes in cognitively healthy individuals, taking into consideration age, sex, eTIV, and characteristics of the MRI scanner. Our goal was to facilitate future neuroscience studies in adulthood, by providing a common normative reference against which to compare new individuals from control or clinical populations.

To be widely applicable, normative values need to be produced on data acquired on common platforms, and analyzed using an accessible automated segmentation pipeline. We selected data from a large number of studies involving three major manufacturers at the two most used field strengths in research. Further, our choice of analysis platform fell on the *FreeSurfer* algorithm, one of the most used software in the neuroimaging research community. *In fine*, our data came from 2,790 individuals aged 18 to 94 years old, and scanned in the context of 23 different studies. The resulting models explained a substantial amount of the variance in subcortical volumes. To our knowledge, the present study is the first attempt to generate accessible normative brain volumes in adults.

4.1 Use of the normative values

Comparing an individual's own volume to the model normative values allows the measurement of potential subcortical volumes alterations. The formulas generate expected

volumes for a given age, sex, eTIV, scanner manufacturer, and magnet strength. The difference between a real volume and a predicted normative volume divided by the root mean square error will result in a Z score effect size, which reflects the degree of deviation from the normative sample. The spreadsheet provides prediction intervals and suitable statistics including individual significance test for abnormality, effect size (Z_{OP}) and estimated percentage of the normative population with a smaller volume. One will notice that although not identical, the individual significance test for abnormality is generally very close to the effect size. This subtle difference will not have a major impact if one uses either the *t*-statistic or the effect size (with 1.65 one-tailed and 1.96 two-tailed as critical values), but is of theoretical importance since the use of the effect size for inferential purposes would treat the normative sample as the population (Crawford and Garthwaite, 2006). Moreover, in the case of using the normative values to compare values for a group of individuals, assessing the difference between actual and expected volumes, using a two-sample *t*-test for example (as shown in Table 4), the distinction between the effect size and the significance of the test is crucial for interpreting the result, since the mean Z_{OP} can greatly differ from the *t*-statistic value depending on the sample size of the group. Indeed, even when effect sizes are small, significant differences between actual and expected volumes can be observed if the group is large.

The validation of the normative values using clinical samples is a good example of how the normative formula can be used and it showed volume differences for the regions that were expected. Results in the SZ group were generally coherent with those of a meta-analysis indicating that compared to controls, medicated patients with schizophrenia show significant atrophy for accumbens, amygdala, hippocampus, and thalamus and hypertrophy for the pallidum of small effect sizes (Haijma et al., 2013). Results in the mild AD group were also coherent with

previous results from the literature showing essentially ventricles enlargements and atrophy of the hippocampus and amygdala, but also changes to other regions such as the accumbens area, the thalamus, and the corpus callosum (Pedro et al., 2012; Pievani et al., 2013; Roh et al., 2011; Scahill et al., 2002).

Another utility of the volumetric normative values is to verify in case-control studies whether or not the control group is close to normative values. Control groups, especially if they are of small sizes, are not necessarily good representations of the normality.

4.2 Effect of Age

In addition to producing normative values, the large sample allowed the validation of relationships that were previously observed using different methodologies. The results importantly showed the respective weight related to each predictor. Age was the predictor with the greatest influence on all regions, except on the brainstem, with most regions starting to decline as early as 18 years of age. Our results indicated that those regions declining latest in life are the brainstem, which showed a slight decrease in men after their 40s, and for women after their 60s; the hippocampi, in which volumes were relatively stable until the 4th decade; and the corpus callosum, which increases late to the 30s, before declining eventually. Walhovd et al. (2011) and Fjell et al. (2013), using substantial yet smaller samples, reported comparable results.

Unlike other regions, both caudate nuclei volumes showed a distinctive U-shape relationship with age, decreasing from entry into adulthood to the 60s, and then increasing to the 90s. Similar results were previously observed (Fjell et al., 2009; Fjell et al., 2013; Goodro et al., 2012; Pfefferbaum et al., 2013; Walhovd et al., 2011). Goodro and colleagues suggested that periventricular white matter signal hyperintensities, which is highly correlated with age, could be responsible for this increase of caudate volume, from the age of 60 onward. An alternate

hypothesis could be a selection bias related to the survival of individuals. Thus, this replicated finding could be either a true phenomenon due to aging, the result of a cohort effect, or an artifact interfering with the MRI signal; the design of this study cannot conclusively determine either way. Nevertheless, given our design and these results, this phenomenon is not trivial and the volume of the caudate nuclei in older adults has to be expected to be larger than in younger adults when using MRI measures such as those produced by *FreeSurfer* on recent recruited cohorts.

4.3 Effect of sex

Whether differences in regional brain volumes between men and women remains after taking into account TIV are still a matter of debate in the literature (Crivello et al., 2014; Jancke et al., 2015; Leonard et al., 2008; Luders et al., 2009), but previous results indicated that the effect of sex on regional brain volumes is heterogeneous across the brain. Our results are in agreement with this finding, showing that although sex improved the prediction in all models, its influence had notable discrepancies between regions and diminishes with age in some regions. Sex had the greatest influence on the brainstem while it had little impact on the volumes of the accumbens, hippocampi, the ventricles, and the corpus callosum. The latter has received a lot of attention (Leonard et al., 2008), and recent findings suggested that there was no difference between men and women after correcting for total brain volume (Luders et al., 2014). In the present study, the corpus callosum was the region, after the left accumbens, with the least influence of sex on its volume (2% of explained variance).

Moreover, although sex by age interaction improved the prediction for most regions, with left accumbens, left pallidum and right putamen showing the strongest interaction, it had little influence compared to the other predictors ($R^2 \leq 1\%$). These results corroborate those from other

studies (Crivello et al., 2014; Fjell et al., 2009; Jancke et al., 2015), which showed no or subtle sex by age interaction for subcortical structures.

4.4 Effect of scanner characteristics

Previous reports had shown that scanner manufacturer and MFS have an influence on automated brain volume segmentation that needs to be taken into account (Jovicich et al., 2009; Kruggel et al., 2010; Pfefferbaum et al., 2012). In the present study, both manufacturer and MFS were retained in the models for the normative prediction of all regions (with the lateral ventricles as the sole exception). However, when compared to age, sex and eTIV, the magnitude of their main and interaction effects was minor (i.e. together, mean R^2 of 3.7%) with two exceptions: the left accumbens (R^2: 9.6%) and the left amygdala (R^2: 8.6%). Thus, despite having a positive impact on prediction, the influence of these scanner characteristics on subcortical volumes remains modest compared to other predictors. Moreover, the best comparison in order to detect subtle neuroimaging effects is clearly within the same scanner. However, when this is not possible, a correction for scanner manufacturer and MFS is a minimal procedure that should be done in order to minimize variance not due to the effect of interest.

4.5 Limitations

One should note that the federated normative sample was not randomly recruited, nor representative of the healthy adult population. Rather, it is comprised of healthy volunteers who agreed to participate in research projects involving MRI, within academic-led environments. The majority was right-handed, Caucasian, and had at least a high school degree. Thus, the normative values may not be generalizable to left-handed, non-Caucasian, or low-educated individuals. While it may not be an exact picture of the healthy adult population, this is one of the largest sample used in such study and included a wide age range. The data involved 23 samples from 21

independent research groups, originating from various countries (Australia, Austria, Belgium, Canada, Finland, Germany, Ireland, Italy, Netherlands, United Kingdom, and USA). Further, the large age range and the wide array of MRIs from three manufacturers at two magnetic field strengths, using multiple acquisition parameters, is an amalgam of data likely to produce more robust normative values than values generated, for example, using a sample recruited by a single research group at a particular geographic location and using a single set of acquisition parameters. Indeed, the validation procedure with independent samples of healthy individuals showed similar prediction in terms of R^2 for the majority of the regions.

Moreover, MRI technology is relatively recent and there is no longitudinal data available spanning the lifetime of single individuals. Since the present study is cross-sectional, age effects may encompass cohort biases. Finally, as our goal was to produce normative values that could be used in other studies, we chose to use *FreeSurfer,* an automated segmentation software, with its default parameters. One should note that *FreeSurfer*, especially with default parameters, may not be the best solution for the segmentation of all subcortical regions. One of the limitations of model-driven algorithms is that every structure present in the a priori training set model is to be represented in the new image being segmented. This will happen whether or not the image matching procedure is able to find anatomically relevant, contrasted landmarks on the images for each specific substructure, given that the matching happens first at the global level, then at the local level, but optimized over an entire neighborhood. The end result is that some structures may be defined by virtue of being inside a given region that represents the software's best attempt at adapting the pre-defined mask with respect to the overall shape of the new subject's brain, as opposed to being within clearly established – and visible – boundaries. This effect may result in the representation of the structure to include inaccuracies; in the case of smaller

structures, errors at the boundaries may have a potentially larger effect on overall volume compared to larger structures. However, given that we have used the exact same approach for all images, and that users of the normative data will be constrained to using this same approach, we expect this possible bias to be systematic and thus, having a quite restrained effect on inter-subject differences.

5. CONCLUSIONS

At a group-level, many neurological and neuropsychiatric disorders display specific anatomical MRI changes. However, measuring brain volume alterations at an individual-level is problematic since it needs reference values from an automated reproducible segmentation technique taking into account the characteristics of the individual and of the scanner. Using a large sample of healthy adults, we built norms for volumetric estimates of subcortical brain regions. Estimates of the expected volumes of an individual based on its age, sex, intracranial volume, the scanner's manufacturer, and magnet strength can be obtained using derived formulas. Statistics allow testing each region for volume abnormality with effect sizes and estimates of the normative population with a smaller volume.

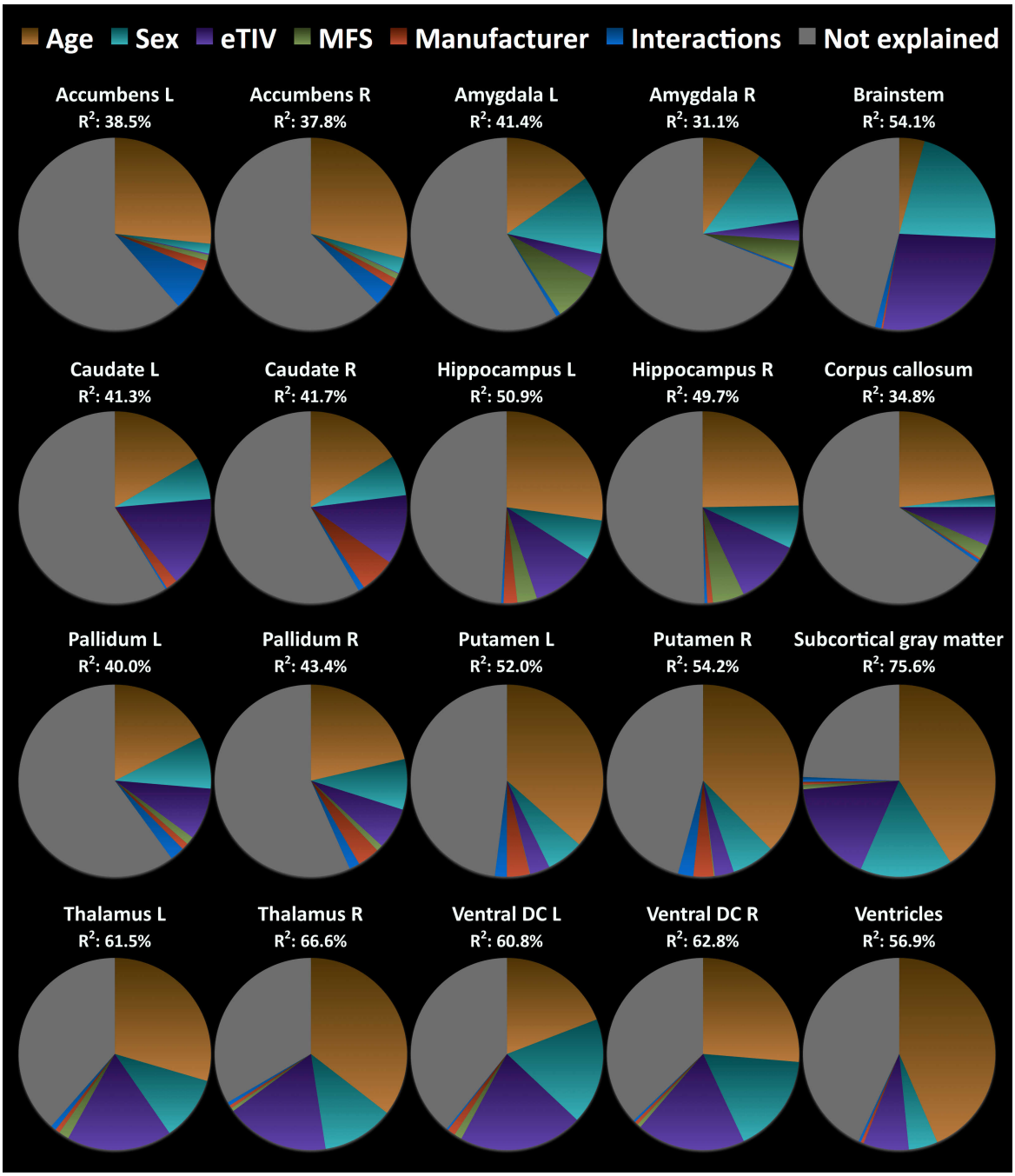

Figure 1. Variance explained by the model for each subcortical regional volume is shown (R^2 results), alongside the proportion of this variance explained by each predictor (pie charts).

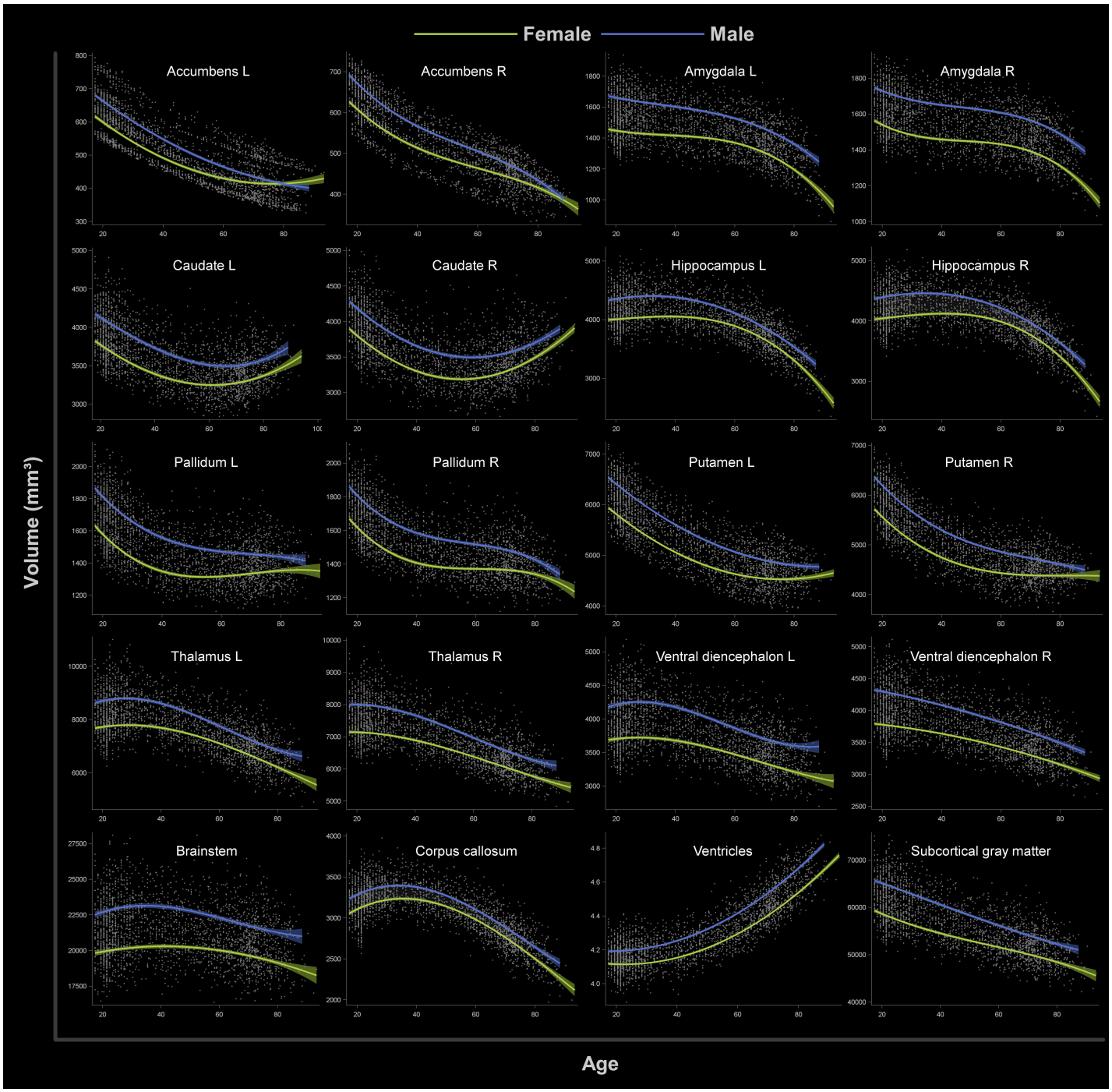

Figure 2. Age and sex influence in each model predicting subcortical regional volumes in a large sample of cognitively healthy individuals aged 18-94 years old. Shaded ribbons around each curve denote 95% confidence intervals for the mean. Ventricles are log10 transformed.

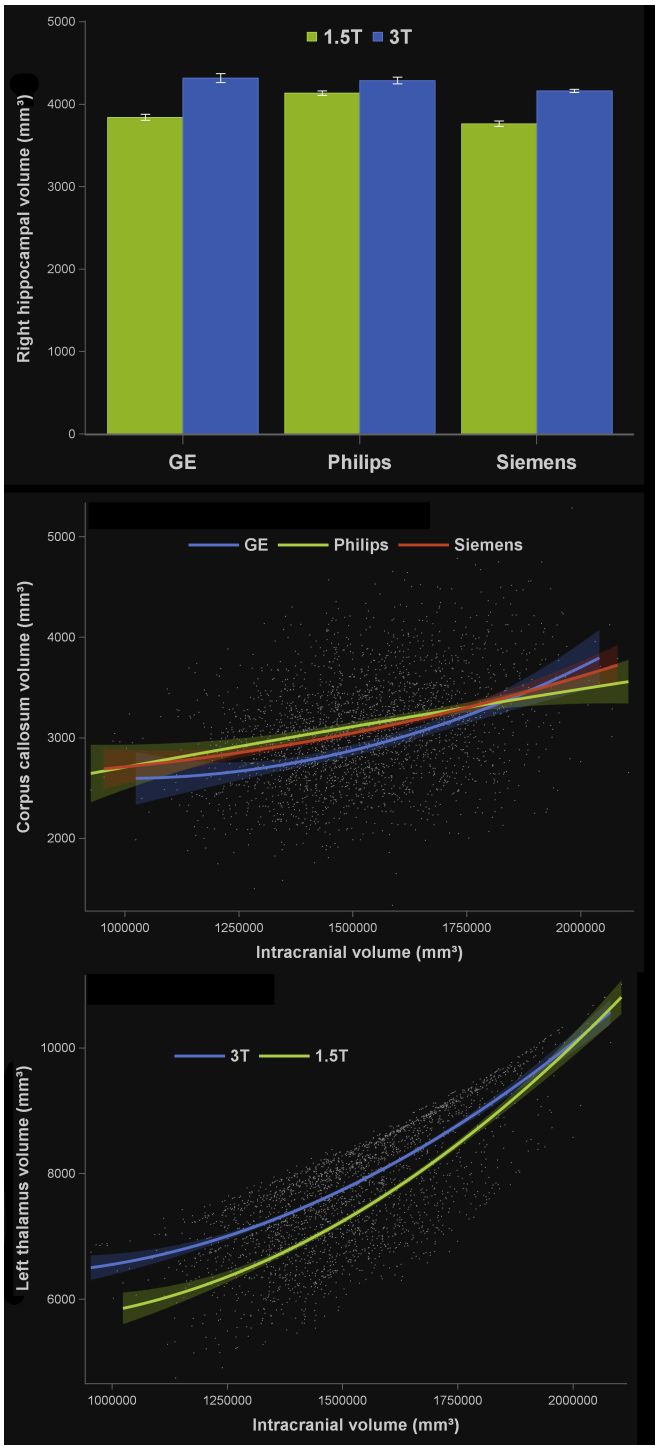

Figure 3. Fitted data illustrations of the magnetic field strength (MFS) and manufacturer effects in the models predicting subcortical volumes. Top: Right hippocampal volume according MFS and manufacturer. Middle: Corpus callosum according to estimated intracranial volume (eTIV) and manufacturer. Bottom: Right thalamus according to MFS and eTIV. Error bars and shaded ribbons denote 95% confidence intervals.

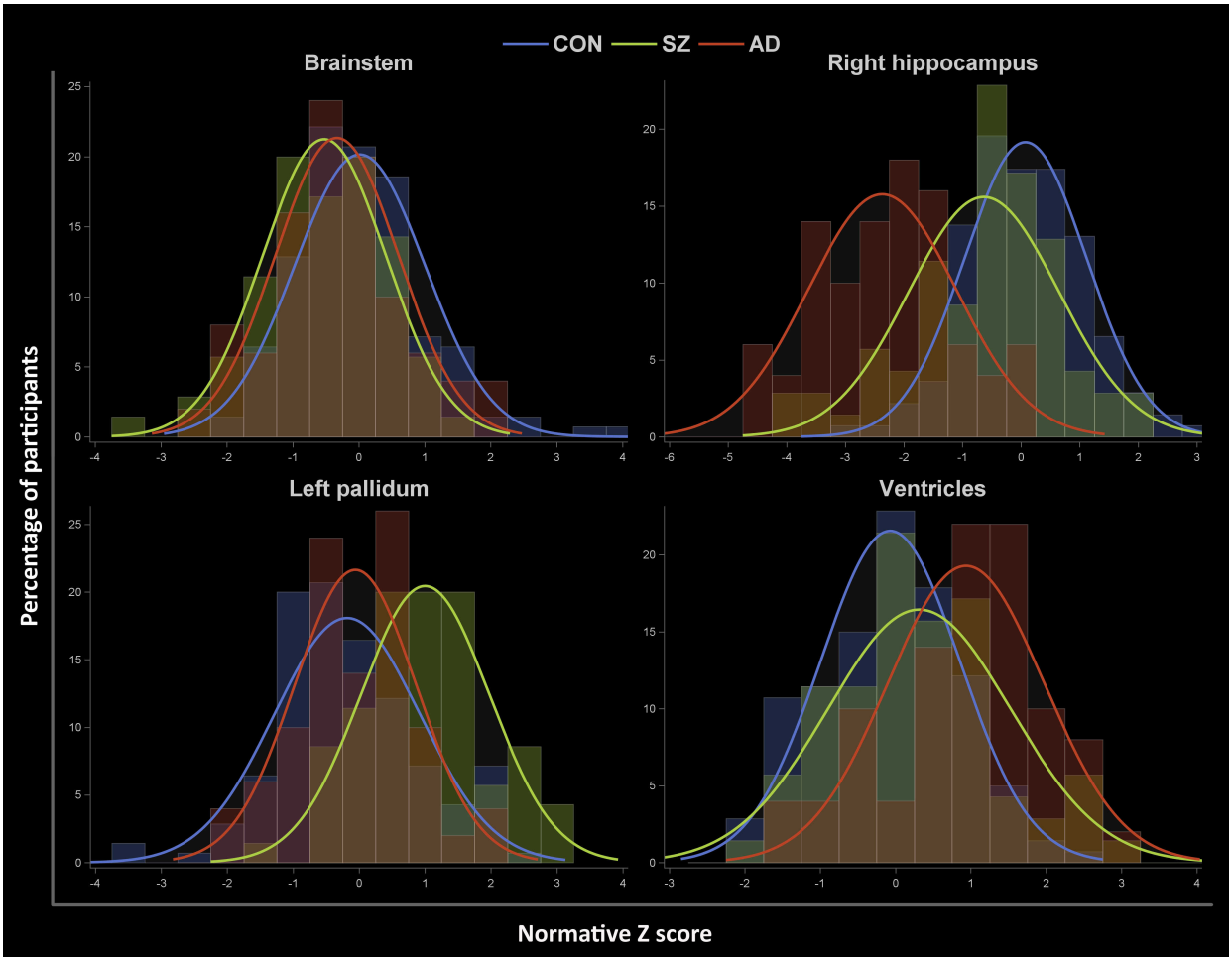

Figure 4. Examples of the distribution of the normative effect sizes (Z_{OP} score) in independent samples of healthy controls (CON), individuals with schizophrenia (SZ) individuals with mild Alzheimer's disease (AD).

Table 1. Participants' characteristics according to the dataset.

Dataset	n	%	Age (mean ±SD range)	Female %
1. Autism Brain Imaging Data Exchange (ABIDE)	184	6.6	26.1 ±7.0 18-56	12.5
2. Alzheimer's Disease Neuroimaging Initiative (ADNI1)	227	8.1	76.0 ±5.0 60-90	48.0
3. Alzheimer's Disease Neuroimaging Initiative (ADNI2)	179	6.4	73.6 ±6.2 56-89	52.5
4. Australian Imaging Biomarkers and Lifestyle flagship study of ageing (AIBL)	158	5.7	72.1 ±7.2 60-88	52.5
5. BMB - Berlin Mind and Brain (Margulies, Villringer) CoRR sample (BMB)	50	1.8	30.3 ±7.1 19-59	52.0
6. Cleveland Clinic (Cleveland CCF)	30	1.1	43.1 ±11.1 24-60	63.3
7. Center of Biomedical Research Excellence (COBRE)	71	2.5	35.5 ±11.3 18-62	29.6
8. DS-108 from the OpenfMRI database	32	1.2	22.2 ±4.6 18-41	50.0
9. DS-170 from the OpenfMRI database	15	0.5	25.4 ±4.6 19-35	20.0
10. Functional Biomedical Informatics Research Network (FBIRN)	34	1.2	38.9 ±13.1 19-65	41.2
11. FIND lab sample (FIND)	13	0.5	24.1 ±3.7 18-29	61.5
12. International Consortium for Brain Mapping (ICBM)	148	5.3	25.0 ±4.9 18-44	42.2
13. Information eXtraction from Images (IXI)	558	20.0	48.5 ±16.4 20-86	55.7
14. F.M. Kirby Research Center neuroimaging reproducibility data (KIRBY-21)	20	0.7	31.9 ±9.7 22-61	45.0
15. Minimal Interval Resonance Imaging in Alzheimer's Disease (MIRIAD)	23	0.8	69.7 ±7.2 58-86	47.8
16. Nathan Kline Institute Rockland phase 1 (NKI-R1)	143	5.1	42.6 ±18.4 18-85	42.7
17. Nathan Kline Institute Rockland phase 2 (NKI-R2)	253	9.1	46.1 ±18.8 18-85	64.8
18. Open Access Series of Imaging Studies (OASIS)	301	10.8	43.9 ±23.6 18-94	61.8
19. Oulu FCON sample (Oulu)	101	3.6	21.5 ±0.6 20-23	64.4
20. POWER Neuroimage sample (POWER)	26	0.9	23.0 ±1.4 20-25	84.6
21. Parkinson's Progression Markers Initiative (PPMI)	164	5.9	60.1 ±11.5 31-83	34.2
22. TRAIN-39 sample (TRAIN)	35	1.3	22.5 ±2.6 18-28	71.4
23. University of Wisconsin (Birn, Prabhakaran, Meyerand) CoRR sample (UWM)	25	0.9	25.0 ±3.2 21-32	44.0
Total	**2790**	**100.0**	**47.6 ±21.8 18-94**	**50.2**

Table 2. Scanners, sequence, and participants characteristics

Manufacturer	Magnetic field strength (%)	Voxel size in mm^3 (%)	Acquisition plane (%)	Model (%)	Age (mean ±SD) Range	Sex
GE	**1.5T** (63.0)	0.4 (3.0) 0.9 (33.4) 1.0 (5.0) 1.1 (48.3) 1.2 (0.7) 1.3 (8.9) Unknown (0.7)	Axial (52.0) Coronal (7.6) Sagittal (40.4)	Optima MR450w (0.3) Signa (7.6) Signa Excite (30.1) Signa Excite HDx (5.0) Signa Genesis (6.6) Signa HDx (34.1) Signa HDxt (5.6) Signa Twin Speed Excite HD (10.6)	47.8 ±26.2 18-90	Female (51.7) Male (48.3)
	3T (37.0)	0.2 (7.3) 1.0 (14.1) 1.1 (38.4) 1.2 (32.8) 1.3 (0.6) Unknown (6.8)	Axial (65.0) Sagittal (35.0)	Discovery MR750 (27.9) Signa (5.1) Signa Echospeed (38.4) Signa HDx (2.3) Signa HDxt (24.3)	47.3 ±22.0 18-89	Female (55.9) Male (44.1)
Philips	**1.5T** (65.1)	1.0 (31.1) 1.1 (68.9)	Axial (61.1) Sagittal (38.9)	ACS III (28.9) Achieva (2.7) Gyroscan Intera (62.1) Gyroscan NT (2.2) Intera (3.7) Intera Achieva (0.4)	45.0 ±19.1 18-86	Female (50.2) Male (49.8)
	3T (34.9)	1.0 (12.4) 1.1 (69.8) 1.2 (17.5) 1.3 (0.4)	Axial (64.8) Coronal (5.5) Sagittal (30.2)	Achieva (20.7) Gemini (1.1) Ingenia (1.1) Intera (77.1)	46.2 ±19.1 18-86	Female (42.9) Male (57.1)
Siemens	**1.5T** (32.1)	0.5 (0.4) 1.0 (1.2) 1.2 (14.5) 1.3 (61.6) 1.9 (20.0) 2.0 (2.0) 2.2 (0.2)	Sagittal (100)	Avanto (18.4) Espree (1.2) Sonata (4.9) Sonata Vision (0.2) Symphony (10.8) Trio (2.9) Vision (61.6)	53.8 ±23.7 18-94	Female (56.8) Male (43.2)
	3T (67.9)	0.3 (2.7) 0.9 (0.1) 1.0 (66.8) 1.1 (0.7)) 1.2 (23.3) 1.3 (3.1) 2.3 (3.4)	Axial (0.1) Sagittal (99.9)	Allegra (8.1) Skyra (1.3) Trio (1.9) Trio Tim (81.39 Verio (6.8)	46.2 ±20.9 18-88	Female (47.6) Male (52.4)

Table 3. Coefficients of models predicting subcortical regional volumes.

Region	RMSE	Int	Sociodemographics				Estimated total intracranial volume (eTIV)			Scanner			Interactions					
			Age	Age^2	Age^3	Sex M / F	eTIV	eTIV^2	eTIV^3	Strength 1.5T / 3T	Manufacturer GE / Siemens	Manufacturer Philips / Siemens	GE X MFS	Philips X MFS	eTIV X MFS	Age X Sex	eTIV X GE	eTIV X Philips
Accumbens L	129.17	445.740	**-3.61E+00**	*2.00E-02*	-	**3.74E+01**	**5.31E-05**	8.06E-11	-	**1.30E+02**	**1.02E+02**	**3.97E+01**	**-1.83E+02**	**-1.80E+02**	-	**-9.51E-01**	-	-
Accumbens R	113.80	510.711	**-2.70E+00**	**2.17E-02**	*-9.16E-04*	**3.51E+01**	**6.32E-05**	-	-	**3.48E+01**	**4.02E+01**	*1.90E+01*	**-9.48E+01**	**-1.22E+02**	-	**-5.67E-01**	-	-
Amygdala L	192.38	1513.76	**-1.33E+00**	*-3.55E-02*	**-2.49E-03**	**8.73E+01**	**4.35E-04**	1.25E-10	-	**-1.72E+02**	**-6.47E+01**	-7.51E+00	**1.08E+02**	3.00E+01	-	-6.17E-01	5.22E-05	-9.32E-05
Amygdala R	214.11	1531.16	-8.89E-01	6.06E-03	**-3.02E-03**	**1.01E+02**	**3.82E-04**	-	-	**-7.94E+01**	*3.94E+01*	*3.52E+01*	-4.62E+01	**-8.09E+01**	-	-	-	-
Brainstem	1846.96	21408.0	*-1.15E+01*	**-1.18E+00**	*1.18E-02*	**5.33E+02**	**9.59E-03**	**3.01E-09**	-	-6.91E+01	**-1.15E+03**	2.03E+02	**1.30E+03**	-3.00E+02	-	**-1.04E+01**	-	-
Caudate L	427.36	3551.88	**-9.05E+00**	**2.13E-01**	2.11E-03	-2.05E+01	**1.71E-03**	**7.16E-10**	-8.61E-16	-	**-6.22E+01**	**-1.81E+02**	-	-	-	**-2.33E+00**	-1.05E-04	*-2.38E-04*
Caudate R	469.42	3481.39	**-7.91E+00**	**3.68E-01**	-	1.94E+01	**1.57E-03**	3.80E-10	-	**1.99E+02**	**-1.86E+02**	**-2.48E+02**	*-1.13E+02*	**-2.16E+02**	*2.68E-04*	**-2.45E+00**	*-2.95E-04*	**-3.94E-04**
Hippocampus L	382.57	4175.15	**-5.08E+00**	**-3.28E-01**	**-3.54E-03**	1.64E+01	**1.20E-03**	3.08E-10	-	**-2.34E+02**	**2.65E+02**	**1.70E+02**	**-1.31E+02**	-1.19E+01	1.18E-04	**-2.11E+00**	-	-
Hippocampus R	378.69	4318.33	**-3.29E+00**	**-3.31E-01**	**-4.32E-03**	1.16E+01	**1.29E-03**	-	-	**-2.98E+02**	**1.81E+02**	4.34E+01	-8.29E+01	**9.68E+01**	-	**-2.11E+00**	-	-
Pallidum L	232.77	1359.76	**-2.87E+00**	**1.30E-01**	**-2.05E-03**	**6.54E+01**	**5.70E-04**	**2.91E-10**	-	**1.66E+02**	*-4.76E+01*	**4.37E+01**	**-8.14E+01**	**-1.95E+02**	-	**-2.11E+00**	-	-
Pallidum R	200.17	1438.50	**-2.59E+00**	**7.34E-02**	**-3.03E-03**	**6.52E+01**	**4.77E-04**	*2.18E-10*	-	**1.55E+02**	-3.87E+00	**-6.14E+01**	**-1.29E+02**	**-1.33E+02**	-	**-1.45E+00**	-	-
Putamen L	663.69	5155.58	**-2.38E+01**	**2.23E-01**	-	**2.07E+02**	**1.73E-03**	*6.89E-10*	*-1.93E-15*	**2.54E+02**	**-2.67E+02**	**-1.69E+02**	-2.05E+01	**-5.41E+02**	**-4.26E-04**	**-4.49E+00**	-	-
Putamen R	604.53	4836.07	**-1.92E+01**	**3.46E-01**	*-3.99E-03*	**2.39E+02**	**1.25E-03**	3.65E-10	-	**2.83E+02**	-6.70E+01	*-8.68E+01*	**-2.93E+02**	**-6.26E+02**	-	**-5.57E+00**	-	-
Thalamus L	765.61	7955.26	**-2.52E+01**	**-5.15E-01**	**6.84E-03**	6.23E+01	**3.21E-03**	**1.46E-09**	-	**-5.20E+02**	*1.49E+02*	5.52E+01	1.11E+01	**3.92E+02**	**5.96E-04**	**-5.68E+00**	-	-
Thalamus R	580.98	7157.91	**-2.46E+01**	**-3.44E-01**	**6.08E-03**	**9.68E+01**	**3.13E-03**	**1.42E-09**	-	**-1.51E+02**	*-1.19E+02*	**-2.50E+02**	6.79E+01	**1.53E+02**	-	**-5.65E+00**	-2.57E-04	**-5.83E-04**
Ventral DC L	334.49	3784.71	**-9.25E+00**	**-1.48E-01**	**3.51E-03**	**1.09E+02**	**1.62E-03**	**6.03E-10**	-	**-1.63E+02**	**1.05E+02**	**1.22E+02**	*-8.36E+01*	4.99E+01	-	**-1.88E+00**	-	-
Ventral DC R	321.88	3718.62	**-9.16E+00**	**-6.32E-02**	-	**1.14E+02**	**1.46E-03**	**5.61E-10**	-	**-1.04E+02**	*6.31E+01*	**8.73E+01**	-5.65E+01	1.85E+01	-	**-2.22E+00**	*2.18E-04*	3.12E-06
Ventricles	0.1595	4.25830	**6.54E-03**	**1.07E-04**	-	1.03E-02	**4.96E-07**	-	-	-1.23E-04	5.66E-03	**-3.69E-02**	-	-	*7.39E-08*	**1.11E-03**	-7.58E-08	*-1.02E-07*
Lateral L[1]	0.1911	3.88998	**7.41E-03**	**9.60E-05**	-	-3.60E-03	**5.79E-07**	-	-	-	6.42E-03	**-4.49E-02**	-	-	-	**1.46E-03**	-6.74E-08	-6.89E-08
Lateral R[1]	0.1924	3.84210	**7.56E-03**	**1.11E-04**	-	7.02E-03	**5.68E-07**	-	-	-	1.30E-02	**-3.47E-02**	-	-	-	**9.79E-04**	-6.83E-08	*-9.73E-08*
Inferior lateral L[1]	0.2740	2.30882	**5.24E-03**	**2.57E-04**	9.87E-07	**1.04E-01**	**3.83E-07**	-1.84E-13	-4.30E-19	**7.64E-02**	**-1.69E-01**	**-9.88E-02**	1.72E-01	3.10E-02	-	**1.82E-03**	*-1.99E-07*	**-2.13E-07**
Inferior lateral R[1]	0.2908	2.26061	*1.75E-03*	**2.12E-04**	**3.95E-06**	**1.19E-01**	**2.40E-07**	-	-	**1.25E-01**	-2.52E-02	**-9.99E-02**	-	-	-1.13E-07	**2.27E-03**	-1.74E-07	**-2.05E-07**
3rd[1]	0.1209	2.96200	**5.57E-03**	**9.61E-05**	-4.32E-07	**4.11E-02**	**2.98E-07**	-1.06E-13	-	4.39E-03	**4.60E-02**	1.19E-03	**-5.20E-02**	**-3.28E-02**	4.53E-08	**8.60E-04**	-	-
4th	548.67	1806.13	**1.72E+00**	**1.54E-01**	-	**9.54E+01**	**9.31E-04**	-2.31E-11	-1.23E-15	**-9.24E+01**	**-1.11E+02**	-2.82E+01	-	-	-	-	3.12E-04	2.36E-04
Corpus callosum	426.26	3337.61	**-9.62E+00**	**-3.46E-01**	2.15E-03	**-6.34E+01**	**1.12E-03**	9.81E-11	-7.03E-16	**-1.65E+02**	*-5.65E+01*	3.44E+01	-	-	*-2.35E-04*	-	**4.08E-04**	-1.58E-04
Subcortical gray matter	3369.97	56155.6	**-1.50E+02**	-6.78E-02	-1.28E-02	**1.32E+03**	**2.00E-02**	**7.22E-09**	-	-9.63E+01	*6.23E+02*	-3.31E+02	**-1.70E+03**	**-1.69E+03**	-	**-4.18E+01**	-	-

Note. Categories are coded 0 and 1 with reference categories (Female, Siemens, and 3T) coded 0. Age and eTIV are centered by the mean (Age - 47.56; eTIV - 1521907.28). DC: diencephalon, Int: Intercept. RMSE: Root mean square error.
[1] log10 transformed. *Italic* $p<.05$; **Bold** $p<.01$.

Table 4. Mean normative effect size (Z_{OP}) and differences between actual and predicted normative volumes in independent samples

	Controls (n = 140)			SZ (n = 70)			AD (n = 50)		
Region	Z_{OP}	*t*	*p*	Z_{OP}	*t*	*p*	Z_{OP}	*t*	*p*
Accumbens L	-0.04	-0.31	.755	-0.49	-4.01	<.001*	-0.19	-1.35	.182
Accumbens R	-0.01	-0.06	.950	0.01	0.07	.947	-0.47	-3.58	<.001*
Amygdala L	0.02	0.06	.951	-0.58	-3.59	<.001*	-1.11	-5.1	<.001*
Amygdala R	0.03	0.25	.804	-0.61	-4.15	<.001*	-1.10	-4.95	<.001*
Brainstem	0.01	0.10	.924	-0.53	-2.41	.017	-0.33	-1.32	.189
Caudate L	0.02	0.13	.898	-0.09	-0.41	.679	-0.42	-2.18	.033
Caudate R	0.02	0.14	.886	0.04	0.18	.855	-0.05	-0.29	.769
Hippocampus L	-0.03	-0.30	.766	-0.61	-3.22	.002*	-2.55	-12.19	<.001*
Hippocampus R	0.07	0.52	.606	-0.64	-3.16	.002*	-2.37	-9.42	<.001*
Pallidum L	-0.18	-1.32	.188	1.00	5.16	<.001‡	-0.06	-0.34	.731
Pallidum R	-0.15	-0.97	.335	-0.28	-1.76	.080	-0.07	-0.39	.699
Putamen L	-0.06	-0.37	.711	-0.05	-0.26	.799	-0.05	-0.29	.775
Putamen R	0.02	0.16	.875	-0.13	-0.6	.549	-0.27	-1.58	.119
Thalamus L	0.07	0.42	.678	-0.56	-2.85	.005	-0.54	-2.30	.024
Thalamus R	-0.06	-0.28	.782	-0.40	-1.61	.110	-0.71	-2.83	.006
Ventral DC L	-0.15	-0.84	.400	0.19	0.81	.418	-0.16	-0.61	.544
Ventral DC R	-0.06	-0.31	.760	0.05	0.2	.840	-0.35	-1.50	.136
Ventricles	-0.07	-0.41	.685	0.31	1.52	.132	0.93	4.94	<.001‡
Lateral L	-0.10	-0.58	.561	0.28	1.41	.163	0.89	5.13	<.001‡
Lateral R	-0.05	-0.32	.747	0.30	1.44	.152	0.71	4.18	<.001‡
Inferior lateral L	-0.03	-0.23	.819	0.60	5	<.001‡	1.58	8.94	<.001‡
Inferior lateral R	0.00	-0.01	.996	0.32	2.48	.015	1.34	7.78	<.001‡
3rd	-0.04	-0.23	.820	0.50	2.53	.013	0.60	2.82	.006
4th	0.08	0.84	.403	-0.29	-2.03	.046	-0.16	-0.85	.399
Corpus callosum	0.01	0.07	.944	-0.49	-2.8	.006	-0.48	-2.96	.004
Subcortical GM	-0.06	-0.36	.718	-0.37	-1.15	.252	-1.24	-4.66	<.001*

* Volumes significantly smaller than the predicted normative values.

‡ Volumes significantly larger than the predicted normative values.

Z_{OP}: Z score obtained by subtracting the observed volumes and the normative volumes predicted by the linear model divided by the root mean square error of the model.

t: independent two-sample *t*-test between the observed volumes and the normative volumes predicted by the model. Bonferroni-corrected critical value for significance: .002.

CONFLICT OF INTEREST

O.P., A.M., and L.D. declare no competing financial interests. S.D. is officer and shareholder of True Positive Medical Devices inc.

ACKNOWLEDGMENTS

We gratefully acknowledge financial support from the Alzheimer's Society of Canada (#13-32), the Canadian Foundation for Innovation (#30469), the Fonds de recherche du Québec – Santé / Pfizer Canada - Pfizer-FRQS Innovation Fund (#25262), and the Canadian Institutes for Health Research (#117121). S.D. is a Research Scholar from the Fonds de recherche du Québec – Santé (#30801).

This study comprises multiple samples of healthy individuals. We wish to thank all principal investigators who collected these datasets and agreed to let them accessible.

Autism Brain Imaging Data Exchange (ABIDE): Primary support for the work by Adriana Di Martino was provided by the NIMH (K23MH087770) and the Leon Levy Foundation. Primary support for the work by Michael P. Milham and the INDI team was provided by gifts from Joseph P. Healy and the Stavros Niarchos Foundation to the Child Mind Institute, as well as by an NIMH award to MPM (R03MH096321). http://fcon_1000.projects.nitrc.org/indi/abide/

Alzheimer's Disease Neuroimaging Initiative (ADNI): Funded by the ADNI (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California. http://adni.loni.usc.edu/

Australian Imaging Biomarkers and Lifestyle flagship study of ageing (AIBL): Part of the data used in this study was obtained from the Australian Imaging Biomarkers and Lifestyle flagship study of ageing (AIBL). See www.aibl.csiro.au for further details.

BMB - Berlin Mind and Brain (Margulies, Villringer). Zuo, X.N., et al. (2014). An open science resource for establishing reliability and reproducibility in functional connectomics. *Scientific data, 1*, 140049. doi: 10.1038/sdata.2014.49.
http://fcon_1000.projects.nitrc.org/indi/CoRR/html/bmb_1.html

Cleveland Clinic (Cleveland CCF): Funded by the National Multiple Sclerosis Society. http://fcon_1000.projects.nitrc.org/indi/retro/ClevelandCCF.html

Center of Biomedical Research Excellence (COBRE): The imaging data and phenotypic information was collected and shared by the Mind Research Network and the University of New Mexico funded by a National Institute of Health COBRE: 1P20RR021938-01A2.
http://fcon_1000.projects.nitrc.org/indi/retro/cobre.html

DS-108. Wager et al. (2008). Prefrontal-subcortical pathways mediating successful emotion regulation. Neuron, 59(6):1037-50. doi: 10.1016/j.neuron.2008.09.006. This data was obtained from the OpenfMRI database. NSF Grant OCI-1131441 (R. Poldrack, PI). Poldrack et al. (2013). Toward open sharing of task-based fMRI data: the OpenfMRI project. *Frontiers in neuroinformatics, 7*, 12. doi: 10.3389/fninf.2013.00012. https://openfmri.org/dataset/ds000108/

DS-170. Learning and memory: motor skill consolidation and intermanual transfer. This data was obtained from the OpenfMRI database. NSF Grant OCI-1131441 (R. Poldrack, PI). Poldrack et al. (2013). Toward open sharing of task-based fMRI data: the OpenfMRI project. *Frontiers in neuroinformatics, 7*, 12. doi: 10.3389/fninf.2013.00012. https://openfmri.org/dataset/ds000170/

Functional Biomedical Informatics Research Network (FBIRN): Provided by the Biomedical Informatics Research Network under the following support: U24-RR021992.
http://www.birncommunity.org/resources/data/

FIND lab sample. Funded by the Dana Foundation; John Douglas French Alzheimer's Foundation; National Institutes of Health (AT005733, HD059205,HD057610, NS073498, NS058899).
http://fcon_1000.projects.nitrc.org/indi/retro/find_stanford.html

International Consortium for Brain Mapping (ICBM). http://www.loni.usc.edu/ICBM/

Information eXtraction from Images (IXI): Data collected as part of the project:
EPSRC GR/S21533/02 - http://www.brain-development.org/

F.M. Kirby Research Center neuroimaging reproducibility data (KIRBY-21). Landman, B.A. et al. “Multi-Parametric Neuroimaging Reproducibility: A 3T Resource Study”, NeuroImage. (2010) NIHMS/PMC:252138 doi:10.1016/j.neuroimage.2010.11.047
http://mri.kennedykrieger.org/databases.html

Minimal Interval Resonance Imaging in Alzheimer's Disease (MIRIAD): The MIRIAD investigators did not participate in analysis or writing of this report. The MIRIAD dataset is made available through the support of the UK Alzheimer's Society (RF116). The original data

collection was funded through an unrestricted educational grant from GlaxoSmithKline (6GKC). http://miriad.drc.ion.ucl.ac.uk

Nathan Kline Institute Rockland (NKI-R) sample (phase 1) and (phase 2): Principal support for the enhanced NKI-RS project is provided by the NIMH BRAINS R01MH094639-01. Funding for key personnel also provided in part by the New York State Office of Mental Health and Research Foundation for Mental Hygiene. Funding for the decompression and augmentation of administrative and phenotypic protocols provided by a grant from the Child Mind Institute (1FDN2012-1). Additional personnel support provided by the Center for the Developing Brain at the Child Mind Institute, as well as NIMH R01MH081218, R01MH083246, and R21MH084126. Project support also provided by the NKI Center for Advanced Brain Imaging (CABI), the Brain Research Foundation, the Stavros Niarchos Foundation and the NIH P50 MH086385-S1 (phase 1). http://fcon_1000.projects.nitrc.org/indi/pro/nki.html
http://fcon_1000.projects.nitrc.org/indi/enhanced/

Open access series of imaging studies (OASIS): The OASIS project was funded by grants P50 AG05681, P01 AG03991, R01 AG021910, P50 MH071616, U24 RR021382, and R01 MH56584. http://www.oasis-brains.org/

Oulu FCON sample (Oulu). http://fcon_1000.projects.nitrc.org/fcpClassic/FcpTable.html

POWER: This database was supported by NIH R21NS061144 R01NS32979 R01HD057076 U54MH091657 K23DC006638 P50 MH71616 P60 DK020579-31 , McDonnell Foundation Collaborative Action Award, NSF IGERT DGE-0548890, Simon's Foundation Autism Research Initiative grant, Burroughs Wellcome Fund, Charles A. Dana Foundation, Brooks Family Fund, Tourette Syndrome Association, Barnes-Jewish Hospital Foundation, McDonnell Center for Systems Neuroscience, Alvin J. Siteman Cancer Center, American Hearing Research Foundation grant, Diabetes Research and Training Center at Washington University grant.
http://fcon_1000.projects.nitrc.org/indi/retro/Power2012.html

Parkinson's Progression Markers Initiative (PPMI): PPMI – a public-private partnership – is funded by the Michael J. Fox Foundation for Parkinson's Research and funding partners, including Abbvie, Avid Radiopharmaceuticals, Biogen Idec, Bristol-Myers, Covance, GE Healthcare, Genentech, GlaxoSmithKline, Eli Lilly and Company, Lundbeck, Merck, Meso Scale Discovery, Pfizer, Piramal, Roche, and UCB. See http://www.ppmi-info.org for further details.

TRAIN-39: Data collected at the Biomedical Imaging Center at the Beckman Institute for Advanced Science and Technology at UIUC. Funded by the Office of Naval Research (ONR): N00014-07-1-0903. http://fcon_1000.projects.nitrc.org/indi/retro/Train-39.html

University of Wisconsin, Madison (Birn, Prabhakaran, Meyerand) CoRR sample (UWM). Zuo, X.N., et al. (2014). An open science resource for establishing reliability and reproducibility in functional connectomics. *Scientific data, 1*, 140049. doi: 10.1038/sdata.2014.49
http://fcon_1000.projects.nitrc.org/indi/CoRR/html/samples.html

REFERENCES

Boccardi, M., Bocchetta, M., Apostolova, L.G., Preboske, G., Robitaille, N., Pasqualetti, P., Collins, L.D., Duchesne, S., Jack, C.R., Jr., Frisoni, G.B., 2014. Establishing Magnetic Resonance Images Orientation for the EADC-ADNI Manual Hippocampal Segmentation Protocol. J Neuroimaging 24, 509-514.

Brain Development Cooperative Group, 2012. Total and regional brain volumes in a population-based normative sample from 4 to 18 years: the NIH MRI Study of Normal Brain Development. Cereb Cortex 22, 1-12.

Buckner, R.L., Head, D., Parker, J., Fotenos, A.F., Marcus, D., Morris, J.C., Snyder, A.Z., 2004. A unified approach for morphometric and functional data analysis in young, old, and demented adults using automated atlas-based head size normalization: reliability and validation against manual measurement of total intracranial volume. Neuroimage 23, 724-738.

Crawford, J.R., Garthwaite, P.H., 2006. Comparing patients' predicted test scores from a regression equation with their obtained scores: a significance test and point estimate of abnormality with accompanying confidence limits. Neuropsychology 20, 259-271.

Crawford, J.R., Garthwaite, P.H., Denham, A.K., Chelune, G.J., 2012. Using regression equations built from summary data in the psychological assessment of the individual case: extension to multiple regression. Psychol Assess 24, 801-814.

Crivello, F., Tzourio-Mazoyer, N., Tzourio, C., Mazoyer, B., 2014. Longitudinal assessment of global and regional rate of grey matter atrophy in 1,172 healthy older adults: modulation by sex and age. PLoS One 9, e114478.

Dewey, J., Hana, G., Russell, T., Price, J., McCaffrey, D., Harezlak, J., Sem, E., Anyanwu, J.C., Guttmann, C.R., Navia, B., Cohen, R., Tate, D.F., Consortium, H.I.V.N., 2010. Reliability and

validity of MRI-based automated volumetry software relative to auto-assisted manual measurement of subcortical structures in HIV-infected patients from a multisite study. Neuroimage 51, 1334-1344.

First, M.B., Spitzer, R.L., Gibbon, M., Williams, J.B.W., 1996. Structured Clinical Interview for DSM-IV Axis I Disorders (SCID), Clinician version. American Psychiatric Press, Washington D.C.

Fischl, B., Salat, D.H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., van der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., Dale, A.M., 2002. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron 33, 341-355.

Fischl, B., Salat, D.H., van der Kouwe, A.J., Makris, N., Segonne, F., Quinn, B.T., Dale, A.M., 2004. Sequence-independent segmentation of magnetic resonance images. Neuroimage 23 Suppl 1, S69-84.

Fjell, A.M., Westlye, L.T., Amlien, I., Espeseth, T., Reinvang, I., Raz, N., Agartz, I., Salat, D.H., Greve, D.N., Fischl, B., Dale, A.M., Walhovd, K.B., 2009. Minute effects of sex on the aging brain: a multisample magnetic resonance imaging study of healthy aging and Alzheimer's disease. J Neurosci 29, 8774-8783.

Fjell, A.M., Westlye, L.T., Grydeland, H., Amlien, I., Espeseth, T., Reinvang, I., Raz, N., Holland, D., Dale, A.M., Walhovd, K.B., Alzheimer Disease Neuroimaging, I., 2013. Critical ages in the life course of the adult brain: nonlinear subcortical aging. Neurobiol Aging 34, 2239-2247.

Goodro, M., Sameti, M., Patenaude, B., Fein, G., 2012. Age effect on subcortical structures in healthy adults. Psychiatry Res 203, 38-45.

Haijma, S.V., Van Haren, N., Cahn, W., Koolschijn, P.C., Hulshoff Pol, H.E., Kahn, R.S., 2013. Brain volumes in schizophrenia: a meta-analysis in over 18 000 subjects. Schizophr Bull 39, 1129-1138.

Hastie, T., Tibshirani, R., Friedman, J., 2008. The elements of statistical learning. Data mining, inference, and prediction. Springer.

Jancke, L., Merillat, S., Liem, F., Hanggi, J., 2015. Brain size, sex, and the aging brain. Human brain mapping 36, 150-169.

Jovicich, J., Czanner, S., Greve, D., Haley, E., van der Kouwe, A., Gollub, R., Kennedy, D., Schmitt, F., Brown, G., Macfall, J., Fischl, B., Dale, A., 2006. Reliability in multi-site structural MRI studies: effects of gradient non-linearity correction on phantom and human data. Neuroimage 30, 436-443.

Jovicich, J., Czanner, S., Han, X., Salat, D., van der Kouwe, A., Quinn, B., Pacheco, J., Albert, M., Killiany, R., Blacker, D., Maguire, P., Rosas, D., Makris, N., Gollub, R., Dale, A., Dickerson, B.C., Fischl, B., 2009. MRI-derived measurements of human subcortical, ventricular and intracranial brain volumes: Reliability effects of scan sessions, acquisition sequences, data analyses, scanner upgrade, scanner vendors and field strengths. Neuroimage 46, 177-192.

Keller, S.S., Gerdes, J.S., Mohammadi, S., Kellinghaus, C., Kugel, H., Deppe, K., Ringelstein, E.B., Evers, S., Schwindt, W., Deppe, M., 2012. Volume estimation of the thalamus using freesurfer and stereology: consistency between methods. Neuroinformatics 10, 341-350.

Kruggel, F., 2006. MRI-based volumetry of head compartments: normative values of healthy adults. Neuroimage 30, 1-11.

Kruggel, F., Turner, J., Muftuler, L.T., 2010. Impact of scanner hardware and imaging protocol on image quality and compartment volume precision in the ADNI cohort. Neuroimage 49, 2123-2133.

Leonard, C.M., Towler, S., Welcome, S., Halderman, L.K., Otto, R., Eckert, M.A., Chiarello, C., 2008. Size matters: cerebral volume influences sex differences in neuroanatomy. Cereb Cortex 18, 2920-2931.

Liem, F., Merillat, S., Bezzola, L., Hirsiger, S., Philipp, M., Madhyastha, T., Jancke, L., 2015. Reliability and statistical power analysis of cortical and subcortical FreeSurfer metrics in a large sample of healthy elderly. Neuroimage 108, 95-109.

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. J Neurosci 29, 14265-14270.

Luders, E., Toga, A.W., Thompson, P.M., 2014. Why size matters: differences in brain volume account for apparent sex differences in callosal anatomy: the sexual dimorphism of the corpus callosum. Neuroimage 84, 820-824.

McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., Stadlan, E.M., 1984. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. Neurology 34, 939-944.

Morey, R.A., Selgrade, E.S., Wagner, H.R., 2nd, Huettel, S.A., Wang, L., McCarthy, G., 2010. Scan-rescan reliability of subcortical brain volumes derived from automated segmentation. Human brain mapping 31, 1751-1762.

Mouiha, A., Duchesne, S., 2011. Hippocampal atrophy rates in Alzheimer's disease: automated segmentation variability analysis. Neurosci Lett 495, 6-10.

Pedro, T., Weiler, M., Yasuda, C.L., D'Abreu, A., Damasceno, B.P., Cendes, F., Balthazar, M.L., 2012. Volumetric brain changes in thalamus, corpus callosum and medial temporal structures: mild Alzheimer's disease compared with amnestic mild cognitive impairment. Dement Geriatr Cogn Disord 34, 149-155.

Pfefferbaum, A., Rohlfing, T., Rosenbloom, M.J., Chu, W., Colrain, I.M., Sullivan, E.V., 2013. Variation in longitudinal trajectories of regional brain volumes of healthy men and women (ages 10 to 85 years) measured with atlas-based parcellation of MRI. Neuroimage 65, 176-193.

Pfefferbaum, A., Rohlfing, T., Rosenbloom, M.J., Sullivan, E.V., 2012. Combining atlas-based parcellation of regional brain data acquired across scanners at 1.5 T and 3.0 T field strengths. Neuroimage 60, 940-951.

Pievani, M., Bocchetta, M., Boccardi, M., Cavedo, E., Bonetti, M., Thompson, P.M., Frisoni, G.B., 2013. Striatal morphology in early-onset and late-onset Alzheimer's disease: a preliminary study. Neurobiol Aging 34, 1728-1739.

Potvin, O., Mouiha, A., Dieumegarde, L., Duchesne, S., submitted for publication. Influenced of age, sex, and scanner characteristics on FreeSurfer subcortical regional volumes: A tool to compute normative values. Data in Brief.

Roh, J.H., Qiu, A., Seo, S.W., Soon, H.W., Kim, J.H., Kim, G.H., Kim, M.J., Lee, J.M., Na, D.L., 2011. Volume reduction in subcortical regions according to severity of Alzheimer's disease. J Neurol 258, 1013-1020.

Scahill, R.I., Schott, J.M., Stevens, J.M., Rossor, M.N., Fox, N.C., 2002. Mapping the evolution of regional atrophy in Alzheimer's disease: unbiased analysis of fluid-registered serial MRI. Proc Natl Acad Sci U S A 99, 4703-4707.

Segonne, F., Dale, A.M., Busa, E., Glessner, M., Salat, D., Hahn, H.K., Fischl, B., 2004. A hybrid approach to the skull stripping problem in MRI. Neuroimage 22, 1060-1075.

Sheikh, J.I., Yesavage, J.A., 1986. Geriatric Depression Scale (GDS): Recent evidence and development of a shorter version. Clinical Gerontology: a Guide to Assessment and Intervention. The Haworth Press, New York, pp. 165-173.

Sheline, Y.I., Sanghavi, M., Mintun, M.A., Gado, M.H., 1999. Depression duration but not age predicts hippocampal volume loss in medically healthy women with recurrent major depression. J Neurosci 19, 5034-5043.

Tae, W.S., Kim, S.S., Lee, K.U., Nam, E.C., Kim, K.W., 2008. Validation of hippocampal volumes measured using a manual method and two automated methods (FreeSurfer and IBASPM) in chronic major depressive disorder. Neuroradiology 50, 569-581.

Walhovd, K.B., Westlye, L.T., Amlien, I., Espeseth, T., Reinvang, I., Raz, N., Agartz, I., Salat, D.H., Greve, D.N., Fischl, B., Dale, A.M., Fjell, A.M., 2011. Consistent neuroanatomical age-related volume differences across multiple samples. Neurobiol Aging 32, 916-932.