

Cortical and subcortical T1 white/gray contrast, chronological age, and cognitive performance



John D. Lewis^{a,*,1}, Vladimir S. Fonov^a, D. Louis Collins^a, Alan C. Evans^a, Jussi Tohka^{b,1}, for the Brain Development Cooperative Group and the Pediatric Imaging, Neurocognition, and Genetics Study

^a Montreal Neurological Institute, McGill University, Montreal, Canada

^b AI Virtanen Institute for Molecular Sciences, University of Eastern Finland, Kuopio, Finland

ARTICLE INFO

Keywords:

Brain development
Cognitive development
Age prediction
Cortical white/gray contrast
Subcortical white/gray contrast
VIQ-PIQ discrepancies

ABSTRACT

The maturational schedule of typical brain development is tightly constrained; deviations from it are associated with cognitive atypicalities, and are potentially predictive of developmental disorders. Previously, we have shown that the white/gray contrast at the inner border of the cortex is a good predictor of chronological age, and is sensitive to aspects of brain development that reflect cognitive performance. Here we extend that work to include the white/gray contrast at the border of subcortical structures. We show that cortical and subcortical contrast together yield better age-predictions than any non-kernel-based method based on a single image-type, and that the residuals of the improved predictions provide new insight into unevenness in cognitive performance. We demonstrate the improvement in age predictions in two large datasets: the NIH Pediatric Data, with 831 scans of typically developing individuals between 4 and 22 years of age; and the Pediatric Imaging, Neurocognition, and Genetics data, with 909 scans of individuals in a similar age-range. Assessment of the relation of the residuals of these age predictions to verbal and performance IQ revealed correlations in opposing directions, and a principal component analysis of the residuals of the model that best fit the contrast data produced components related to either performance IQ or verbal IQ. Performance IQ was associated with the first principle component, reflecting increased cortical contrast, broadly, with almost no subcortical presence; verbal IQ was associated with the second principle component, reflecting reduced contrast in the basal ganglia and increased contrast in the bilateral arcuate fasciculi.

1. Introduction

Growth charts, *e.g.* those from the Centers for Disease Control and Prevention,² are used by pediatricians and parents to assess the growth of infants, children, and adolescents. Deviation from the typical growth trajectory is an indicator of potential issues. But such charts provide only measures such as height/length, weight, and head circumference; more detailed measures of brain development can be obtained from brain imaging data. Chronological age can be quite accurately predicted from brain images (Dosenbach et al., 2010; Brown et al., 2012; Franke et al., 2012; Mwangi et al., 2013; Khundrakpam et al., 2015; Ball et al., 2017; Lewis et al., 2018), and deviations from the typical maturational schedule are associated with differences in cognitive functioning (Erus

et al., 2015; Lewis et al., 2018). Determining the extent to which typical cognitive development rests upon adherence to the typical maturational trajectory is critical to our understanding of development, and to our ability to utilize brain data to identify atypical development so that we might intervene. Furthermore, to the extent that we can associate atypical cognitive development with specific brain regions, both our understanding and our ability to intervene are enhanced.

In samples representative of normal development from early childhood through adolescence, correlations between chronological age and age estimates from brain images are as high as 0.96 based on multi-metric data from multiple image types (Brown et al., 2012). Brown et al. (2012) reported substantially lower correlations for a single image-type; for T1-weighted data their multi-metric approach, including

* Corresponding author. Montreal Neurological Institute, NW147, 3801 University Street, Montreal, QC, H3A2B4, Canada.

E-mail address: jlewis@bic.mni.mcgill.ca (J.D. Lewis).

¹ These two authors have contributed equally.

² The CDC growth charts can be found at <https://www.cdc.gov/growthcharts/>.

<https://doi.org/10.1016/j.neuroimage.2019.04.022>

Received 31 January 2019; Received in revised form 27 March 2019; Accepted 5 April 2019

Available online 12 April 2019

1053-8119/© 2019 Elsevier Inc. All rights reserved.

cortical and subcortical measures, yielded a correlation between chronological and estimated age of approximately 0.91. Their multi-metric estimates of age from T2-weighted signal intensity also show a correlation with chronological age of 0.91, with contributions from both white-matter tracts and subcortical gray matter (Brown et al., 2012). Estimates of age based on diffusion data show a correlation close to 0.9, with the greatest contributions coming from both cortical and subcortical gray matter (Mwangi et al., 2013; Erus et al., 2015; Brown et al., 2012). Note that measures from subcortical material contribute to each of these results. Moreover, Brown et al. (2012) showed, albeit in their multi-image-type model, that subcortical volume measures were a core contributor to their predictions. Indeed, the subcortical maturational trajectory differs from that of the cortex, and differs between different subcortical structures (Raznahan et al., 2014). Thus, such measures should boost the accuracy of MRI-based age-prediction models.

Our recent work (Lewis et al., 2018) explored the use of one of the most salient features of T1-weighted images, and one that shows dramatic change over development, and is thus a natural choice for use in age prediction: the intensity contrast between gray- and white-matter. This intensity contrast was measured as the ratio of the signal intensity on either side of the surface formed at the gray-white cortical boundary. That work yielded a correlation between chronological and estimated age of approximately 0.91, using only T1-weighted data, and only the cortical gray-white contrast. This result is superior to any single-metric method (e.g. Khundrakpam et al. (2015)) and comparable to the best multi-metric methods using a single image-type (e.g. Brown et al. (2012)). But it seems likely that this result could be improved upon by including the intensity contrast between white and gray matter at the boundary between subcortical gray matter and the surrounding white matter, which is a novel measure, to the best of our knowledge, but a natural extension of our previous work (Lewis et al., 2018). Moreover, it seems likely that including the intensity contrast at the subcortical gray-white boundaries might capture additional aspects of brain maturation related to cognitive development. Our previous work examined the errors of our age-predictions to determine if they were, in fact, pure errors, *i.e.* noise, or if they instead captured elements of cognitive development. We showed that there were significant relationships between our prediction errors and IQ. Subcortical regions play important roles in cognitive, affective, and social functions (Johnson, 2005; Utter and Basso, 2008; van Schouwenburg et al., 2010; Fischi-Gómez et al., 2014; Humphries and Prescott, 2010; Grazioplene et al., 2015; Bohlken et al., 2016; Guo et al., 2006; Isaacs et al., 2008), and subcortical abnormalities have been associated with various psychiatric disorders (Rimol et al., 2010; Hartberg et al., 2011; Cerliani et al., 2015; Hibar et al., 2016; Hoogman et al., 2017). Thus, it seems reasonable to think that the residuals of age prediction using both cortical and subcortical white/gray contrast might provide additional information about cognitive development.

Here we test these conjectures. We devise methods to extract surfaces for the caudate, putamen, globus pallidus, and thalamus, and extend our methods for measuring white/gray contrast at the inner edge of cortex to measure white/gray contrast at the gray-white boundaries of these subcortical structures. We then evaluate the impact on age-prediction accuracy of including subcortical white/gray contrast measures in addition to cortical white/gray contrast measures. We do so in two large datasets: the NIH Pediatric Data and the Pediatric Imaging, Neurocognition, and Genetics data. We then assess the relation of the residuals of these age predictions to the verbal and performance IQ measures available in the NIHPD dataset, and map these relationships to the brain.

2. Materials and methods

2.1. Data

The data used were taken from two large-scale publicly available datasets: the NIH Pediatric data (NIHPD); and the Pediatric Imaging,

Table 1

Subject demographics for the two datasets. Ages are given in years in the upper portion of the table. The lower portion of the table provides the summary statistics for the verbal IQ (VIQ) and performance IQ (PIQ) scores for the NIHPD data.

	subjs (males)	scans (males)	min age	max age	mean age	std age
NIHPD	402 (190)	831 (383)	4.88	22	12.5	3.95
PING	909 (472)	909 (472)	3.16	21	12.4	5.00

	subjs (males)	scans (males)	min	max	mean	std
VIQ	398 (188)	809 (373)	74	156	110.8	13.15
PIQ	398 (188)	809 (373)	72	157	111.1	13.00

Neurocognition, and Genetics data (PING). Detailed descriptions of the NIHPD and PING samples are given in Evans et al. (2006) and Brown et al. (2012), respectively; here we give a brief summary. Table 1 provides the demographic information for the two populations, as well as the Wechsler Abbreviated Scale of Intelligence (WASI) verbal IQ (VIQ) and performance IQ (PIQ) scores for the NIHPD data; these scores are not available for the PING data.

2.1.1. The NIHPD sample

The NIHPD data were collected to allow the characterization of healthy brain development, and the relation between brain development and behaviour. Both high-quality MRI data and comprehensive clinical/behavioral measures were collected. Recruitment was epidemiologically based, demographically balanced, and used strict exclusion factors. The sample includes over 400 children ranging from 4.5 to 18.5 years of age, well-distributed across the range. The database is a mix of longitudinal and cross-sectional data, but we consider all data as cross-sectional, taking care to ensure that the evaluation of predictive models is not biased by this dependence. The data were collected at six sites: Children's Hospital, Boston; Children's Hospital Medical Center, Cincinnati; University of Texas Houston Medical School, Houston; Neuropsychiatric Institute and Hospital, UCLA; Children's Hospital of Philadelphia; and Washington University, St. Louis. Data were acquired on either a General Electric 1.5T scanner or a Siemens Medical Systems 1.5T scanner. Pulse sequences and parameters were chosen to maximize image quality and minimize differences across sites. Each site acquired multiple image-types; the current study utilizes only the T1-weighted data.

2.1.2. The PING sample

The PING dataset includes data from more than 900 typically developing children between the ages of 3 and 20 years, including individuals with learning or language disabilities. Subjects were excluded if they had a history of major developmental, psychiatric, and/or neurological disorders, or medical conditions that affect neurological development, or they had ever had a brain injury. As opposed to the NIHPD sample, these data are strictly cross-sectional. Data were collected at 10 sites: Weil Cornell Medical College, University of California at Davis, University of Hawaii, Kennedy Krieger Institute, Massachusetts General Hospital, University of California at Los Angeles, University of California at San Diego, University of Massachusetts Medical School, University of Southern California, and Yale University. Data were acquired on either a General Electric 3T scanner, a Siemens Medical Systems 3T scanner, or a Philips 3T scanner. Pulse sequence parameters were chosen to maximize image quality and minimize differences across scanners. Across sites and scanners, a standardized multiple image-type high-resolution structural MRI protocol was implemented. The current study uses only the T1-weighted data.

2.2. Surface measurements

2.2.1. Cortical surface extraction

The T1-weighted volumes were denoised (Manjón et al., 2010) and then processed with CIVET (version 2.1; 2016), a fully automated

structural image analysis pipeline developed at the Montreal Neurological Institute.³ CIVET corrects intensity non-uniformities using N3 (Sled et al., 1998); aligns the input volumes to the Talairach-like ICBM-152-nl template (Collins et al., 1994); classifies the image into white matter, gray matter, cerebrospinal fluid, and background (Zijdenbos et al., 2002; Tohka et al., 2004); extracts the white-matter and pial surfaces (Kim et al., 2005); and maps these to a common surface template (Lyttelton et al., 2007).

2.2.2. Subcortical surface extraction

Subcortical segmentation into left and right caudate, putamen, globus pallidus, and thalamus was done using a label-fusion-based labeling technique based on Coupé et al. (2011) and further developed by Weier et al. (2014), incorporating ideas from Landman et al. (2012). The approach used a population-specific template library. In the current work, the library for each dataset was constructed by clustering (as described below) the deformation fields from the non-linear transforms produced by CIVET, and using the central-most subject of each cluster to construct the entries in the template library. Thus, the template library represented the range of deformations found in the population.

As the basis for clustering, the Jacobian was computed for the non-linear transform produced by CIVET for each subject. The values in the Jacobian were then extracted as a vector for each voxel within a mask formed by binarizing the template-based subcortical segmentation, eroding it 1 mm and dilating it 5 mm in order to capture the anatomical context of the nonlinear registration in this area. These Jacobian vectors were then clustered using an equal combination of cosine similarity and Euclidean distance with Ward's clustering method (Ward, 1963), with the number of clusters chosen as the square of the natural log of the number of subjects. Then, within each cluster, the sum-squared distance from each subject to each other subject was computed, and the subject with the minimum sum-squared distance was taken as the central-most subject of the cluster.

To create the library entry for a cluster, the non-linear transform for the central-most subject was inverted and used to warp the ICBM-152-nl template together with the subcortical segmentation defined on it; this pair was then added to the template library. The template library was thus a set of warped copies of the ICBM-152-nl template together with their correspondingly warped segmentations. Once the template library had been created, each subject in the population was non-linearly registered to the n closest templates in the library (here, $n = 7$), and the resulting transforms were used to warp their corresponding segmentations to the subject; the final labelling was then established via a per-voxel majority vote.

Once the subcortical structures for a subject were labeled, surfaces were fit to these labels. The surfaces for each structure were created on the ICBM-152-nl template. These model surfaces were warped to each individual based on the transforms derived from the label-fusion-based labeling stage, and adjusted to the final labels by moving vertices along a distance map created for each label. The surfaces for each structure were then registered to their corresponding common surface template to ensure cross-subject vertex correspondence, as per the cortical surfaces.

2.2.3. White/gray contrast measurements

To extract the white/gray contrast measures, the intensity on the T1-weighted MRI was sampled inside and outside of the surface at the (cortical or subcortical) gray-white boundary, and the ratio of the two measures was formed. More specifically, for the cortex, gray matter was sampled 1 mm outside of the white surface, and white matter was sampled 1 mm inside of the white surface, or in the center of the white matter in areas with thin strands of white matter. The choice of 1 mm outside of the white surface largely avoids partial volume effects from both white and gray matter, even in the thinnest areas of the cortex, without reliance on the correct placement of the gray surface, which often erroneously includes dura and fails to penetrate sulci, particularly

in developmental data where the brain often fills the skull and the sulci are not yet open. Note that this differs in multiple ways from the gray/white ratio measure used by Salat et al. (2009); for more on these differences see Lewis et al. (2018). For the subcortical structures, due to the lesser clarity of the boundary between white and gray, and the relative absence of issues related to the thickness of either the gray matter or the surrounding white matter, gray matter was sampled 2 mm inside of the subcortical white surface, and white matter was sampled 2 mm outside of the subcortical white surface, or in the center of the white matter in areas in which white matter was too thin to allow 2 mms. An example of the surfaces used to form the white/gray contrast measures is provided in Fig. 1.

In more detail, a distance map was created from the white surfaces (both cortical and subcortical), smoothed with a Gaussian kernel, and used to create a gradient vector field.

At low resolution, issues arise with this vector field in areas with very thin strands of white matter, e.g. in thin gyri or between nearby subcortical structures, and so the distance maps were created at a resolution of 0.25 mm and smoothed with a 0.5 mm FWHM Gaussian kernel before creating the gradient vector field. The cortical white surface was moved 1 mm inward along this gradient vector field to produce a sub-white surface, and 1 mm outward to produce a supra-white surface. The intensity values on the T1-weighted image were sampled at each vertex of both the supra-white surface and the sub-white surface, and the ratio was formed by dividing the value at each vertex of the sub-white surface by the value at the corresponding vertex of the supra-white surface. The intensity values were sampled in stereotaxic space, with the T1-weighted volume upsampled to 0.5 mm iso, with no non-uniformity correction or normalization. This avoids, to the extent possible, issues arising from differences in brain size, while leaving the intensity values essentially unchanged. The same procedure produced the contrast measures for the subcortical surfaces, but moving inward and outward 2 mm instead of 1 mm. This procedure ensures that the subwhite surfaces in regions with thin strands of white matter will not cross, and so will provide the best possible approximation of white matter. This can be seen in Fig. 1 between subcortical structures and within thin gyri.

Notice, however, that areas of the subwhite surfaces of the caudate and thalamus fall within the ventricles rather than white matter, and thus the white/gray contrast measures in these areas will not be valid and must be masked. This is, of course, also the case for the brain stem and midsagittal cuts.

Since the intensities that form the ratio are sampled very near to one another, the effects of image inhomogeneities will be small. But, the white/gray contrast measures are sensitive to scanner-specific differences in tissue contrast. To correct for this, we normalized the contrast values per scanner manufacturer. The normalized contrast measure is $m_i^n = (m_i^{un} - \mu_{scanner}) / \sigma_{scanner}$, where m_i^{un} is the unnormalized contrast measure and $\mu_{scanner}$ and $\sigma_{scanner}$ are the mean and the standard deviation of the contrast values across the surface and across all of the subjects whose data were acquired on *scanner*. This correction is equivalent to scanner-wise z-scoring. This correction was done separately for each of the surfaces (the two cortical surfaces and the eight subcortical surfaces). Left and right surfaces were considered together. The effectiveness of this correction was demonstrated in (Lewis et al., 2018). For the rest of the paper, we drop the superscripts and always refer to the normalized contrast.

2.3. Age prediction

Our age prediction method and its evaluation are unchanged from Lewis et al. (2018), except that we have adopted an age model that includes a quadratic term in addition to the linear one. (We provide the results for the strictly linear model in the supplementary material.) The original 81,924 measurements on the cortical surface were grouped into 640 parcels by recursively merging the neighbouring triangles of the surface mesh model, yielding regions of approximately equal surface

³ <http://www.bic.mni.mcgill.ca/ServicesSoftware/CIVET-2-1-0-Introduction>.

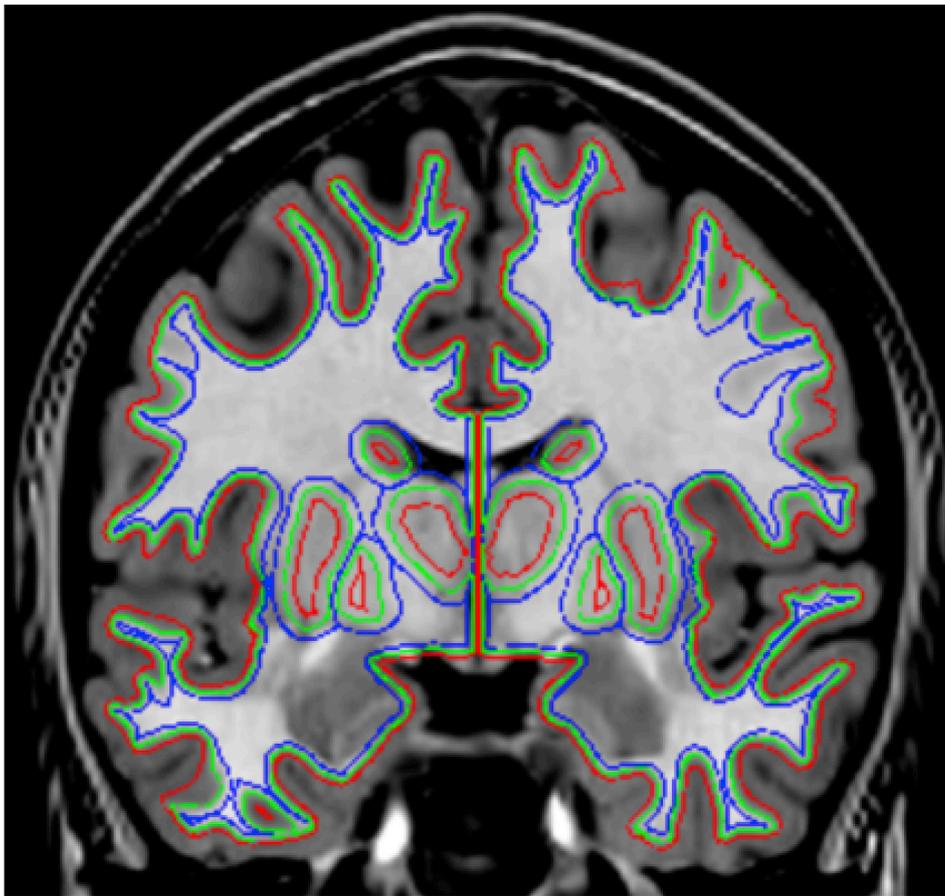


Fig. 1. An example of the surfaces used to form the white/gray contrast measures. The surface at the gray-white border is shown in green. Copies of these surfaces were moved into gray matter (red) and white matter (blue) along the gradient vectors of a distance map based on the surface at the gray-white border. Gray matter and white matter intensity were then measured at each of the vertices of these derivative surfaces, and the white/gray contrast measure formed as the ratio of white intensity to gray intensity at corresponding vertices of the two surfaces. Note that in areas with thin strands of white matter subwhite surfaces abutt rather than cross. Note also that areas of the subwhite surface of the caudate and thalamus fall in the ventricles; measures in these areas are invalid and must be masked.

area. We have previously studied predicting age with different number of parcels and concluded that 640 parcels yielded the most accurate age predictions (Lewis et al., 2018). The thalamus, caudate, putamen, and globus pallidus were generated with far fewer vertices, and so we grouped their measurements into 160 parcels per structure in the same way as for the cortical surface in order to attain parcels with approximately the same surface area as the cortical parcels.

We assumed a second order model for predicting subjects' ages based on the white/gray contrast measurements. The model is

$$AGE = \sum_{i=j}^P b_j m_j + \sum_{i=j}^P q_j (m_j + a)^2 + c + \varepsilon, \quad (1)$$

where AGE is the age of the subject (in days); m_j , $j = 1 \dots P$, are the contrast measurements for each parcel; b_j , q_j and c are the model weights to be learned by the machine learning algorithm, and ε is an error term. The constant a , selected as the negative of the minimum contrast measurements of all subjects and across the surfaces, ensures that $m_j + a$ is always greater than zero. While contrast measurements are positive, normalized contrast measurements are not. This transformation of the second order components does not matter for the predictive capabilities of the model, but it is an important consideration for the interpretation of the resulting model. For notational convenience, we define $\hat{m}_j = m_j$ for $j = 1, \dots, P$ and $\hat{m}_j = (m_{j-P} + a)^2$ for $j = P + 1, \dots, 2P$. We considered age prediction with subcortical contrast alone, with cortical contrast alone, and with both together. Subcortical contrast was formed by concatenating the contrast measures from the eight subcortical surfaces to form a

vector of length 640. We estimated the parameters $\mathbf{b} = [b_1, \dots, b_P, q_1, \dots, q_P]$ and c of the model by penalized least squares with the elastic-net penalty as implemented in the Glmnet package (Friedman et al., 2010).⁴ This equals to the minimization of the cost function

$$\frac{1}{2N} \sum_{i=1}^N (AGE_i - c - \mathbf{m}_i^T \mathbf{b})^2 + \lambda [(1 - \alpha) \|\mathbf{b}\|_2^2 / 2 + \alpha \|\mathbf{b}\|_1], \quad (2)$$

where subscript i refers to scans, $\mathbf{m}_i = [\hat{m}_{i1}, \dots, \hat{m}_{iP}, \hat{m}_{i(P+1)}, \dots, \hat{m}_{i(2P)}]$ are the measurements for scan i and N is the number of scans. The elastic-net penalty $(1 - \alpha) \|\mathbf{b}\|_2^2 / 2 + \alpha \|\mathbf{b}\|_1$ is a weighted average of the LASSO and ridge penalties (Zou and Hastie, 2005). The LASSO penalty $\|\mathbf{b}\|_1$ forces many parameters to have a zero-value leading to the variable selection while the ridge penalty $\|\mathbf{b}\|_2^2 / 2$ helps to ensure that highly correlated variables are selected simultaneously and they have similar model weights (Zou and Hastie, 2005). We set $\alpha = 0.5$, as in Khundrakpam et al. (2015) and Lewis et al. (2018). The relative weight of the data term and the penalties, denoted by the parameter $\lambda \in \mathbb{R}$ in Eq. (2), was decided based on cross-validation from a sequence of 300 values decreasing on the log scale (Friedman et al., 2010). We standardized the input measurements so that $(1/N) \sum_{i=1}^N \hat{m}_{ij} = 0$ and $(1/N) \sum_{i=1}^N \hat{m}_{ij}^2 = 1$ for $j = 1, \dots, 2P$. Thus, the parameter estimates b_j were also standardized (called standardized regression coefficients) and more comparable to each other than the unstandardized coefficients, see e.g. Gelman (2008). The logic of standardized coefficients is to re-express b_j and q_j as the effect of a one-SD change in \hat{m}_j as opposed to a unit change in \hat{m}_j . The standardization of second order terms makes them comparable to the first order terms in the model.

⁴ We used a Matlab wrapper described by Qian et al. (2013) of Friedman's original Fortran code.

2.4. Evaluation of age prediction

We selected the λ and evaluated the age predictions using two nested stratified 10-fold cross-validation (CV) loops (Ambrose and McLachlan, 2002; Huttunen et al., 2012).⁵ The value for the parameter λ was selected in the inner CV loop by minimizing the mean squared error among the candidate values, and the age predictions were evaluated in the outer CV loop, thus avoiding the training on testing data problem. We repeated the nested CV loops 10 times to reduce the random variation in the evaluation of accuracy due to the random selection of the folding scheme. The model goodness criteria applied were the correlation coefficient between the chronological and estimated age, and the mean absolute error (MAE) between the chronological and estimated age. For the correlation coefficient, we averaged the 10 distinct correlation values stemming from the 10 CV runs. Since with the NIHPD sample we had up to three sets of measurements of certain subjects at different ages, we controlled for the non-independence of the observations in the CV by placing all the scans of the subject i in the same fold and thus all scans of a given subject were either in the training set or in the test set. Therefore, data from subject i was never used to build an age-predictor for subject i . This is an important consideration and failure to account for non-independence would lead to positively biased model accuracy estimates.

As described in Lewis et al. (2018), we computed 95% confidence intervals (CIs) for cross-validated correlations using a bootstrap method and compared the cross-validated MAEs from the different sets of measurements using a permutation test.

2.5. Evaluation of relation of residuals to cognitive functioning

With the NIHPD database there are IQ scores for a subsample of 809 scans from 398 subjects. Previously we showed that the residuals of our age predictions based on cortical contrast, defined as *chronological age - predicted age*, were related to cognitive functioning (Lewis et al., 2018). Here, we looked at these relationships in more detail. First, we assessed the relationship of the age-prediction residuals to the two aspects of full scale IQ, i.e. performance IQ and verbal IQ, in each case assessing one while controlling for the effect of the other on age prediction, as well as the effect of brain volume, within a linear mixed-effects model. Then, we

decomposed the residuals of the model that best fit the contrast data, and assessed the relation of the top two components to performance IQ and verbal IQ.

The best-fit model was determined by searching all possible models comprised of terms including any of ‘AGE’, ‘SEX’, ‘BRAINVOL’, ‘SITE’, ‘AGE*AGE’, and ‘AGE*SEX’, as well as the terms required for a mixed effects model, i.e. ‘random(subjects) + I + 1’, where ‘random(subjects)’ induces equal correlations between observations on the same subject, and ‘I’ allows for independent ‘white’ noise in every observation. Each model was assessed using the SurfStat toolbox⁶ and the Akaike information criterion (AIC) (Akaike, 1976). The best-fit model was defined as the one with the lowest AIC. The best-fit model was then evaluated with *SurfStatLinMod* to produce the coefficients which were then passed to *SurfStatPCA* in order to obtain the first two principal components of the residuals of the model. The correlation between both of these two principal components and both of performance and verbal IQ was then assessed to determine if these measures of cognitive functioning were prominent sources of variance in the residuals of the best-fit model.

3. Results

3.1. Age prediction accuracy

The cross-validated accuracies of the age predictions are reported in Table 2 and plotted in Fig. 2 with significant differences in mean absolute errors (MAE) indicated. (The results for the strictly linear model can be found in the supplementary material.)

Table 2

The cross-validated accuracies of age predictions based on subcortical and cortical white/gray contrast, and both measures together. The average cross-validated correlation value (R) for each is given, as well as the mean absolute error (MAE) in days.

	Subcortical		Cortical		Subcortical + Cortical	
	R	MAE	R	MAE	R	MAE
NIHPD	0.82	656	0.89	523	0.90	488
PING	0.90	623	0.91	588	0.93	521

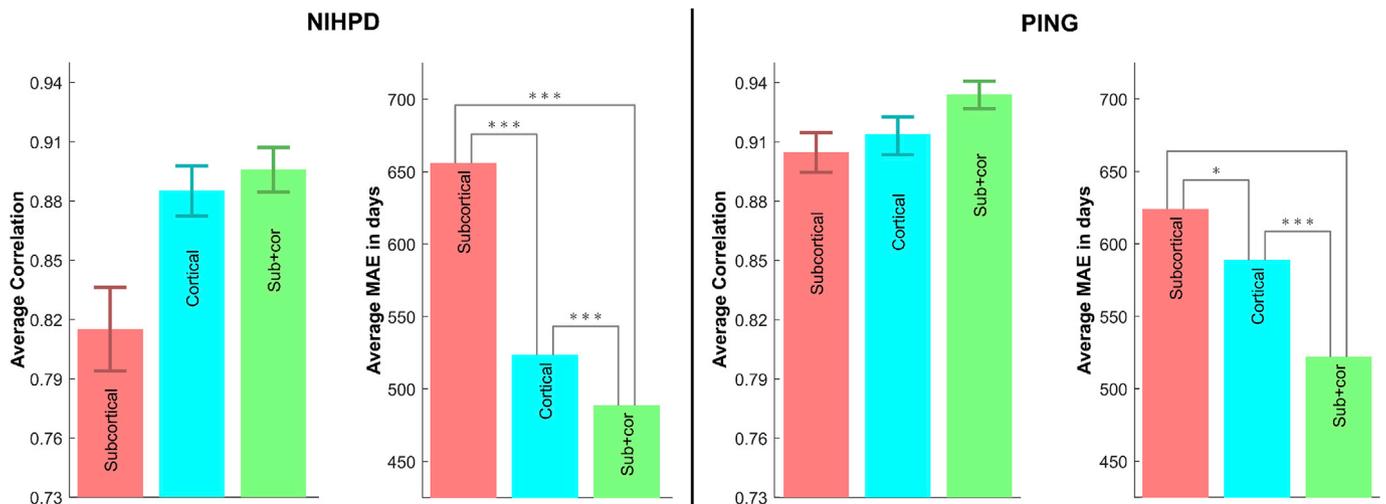


Fig. 2. The cross-validated accuracies of the age predictions based on subcortical and cortical white/gray contrast, and both measures together. The correlation plots show, in addition to the average cross-validated correlation value, the 95% confidence intervals obtained by a bootstrap method. The MAE plots show, in addition to the average cross-validated MAE, a permutation test based comparison between the different methods: * indicates $p < 0.05$, ** indicates $p < 0.01$, and *** indicates $p < 0.001$.

⁵ The Matlab code used for constructing stratified cross-validation folds for regression is available at https://github.com/jussitohka/general_matlab.

⁶ SurfStat is available at <http://www.math.mcgill.ca/keith/surfstat/>.

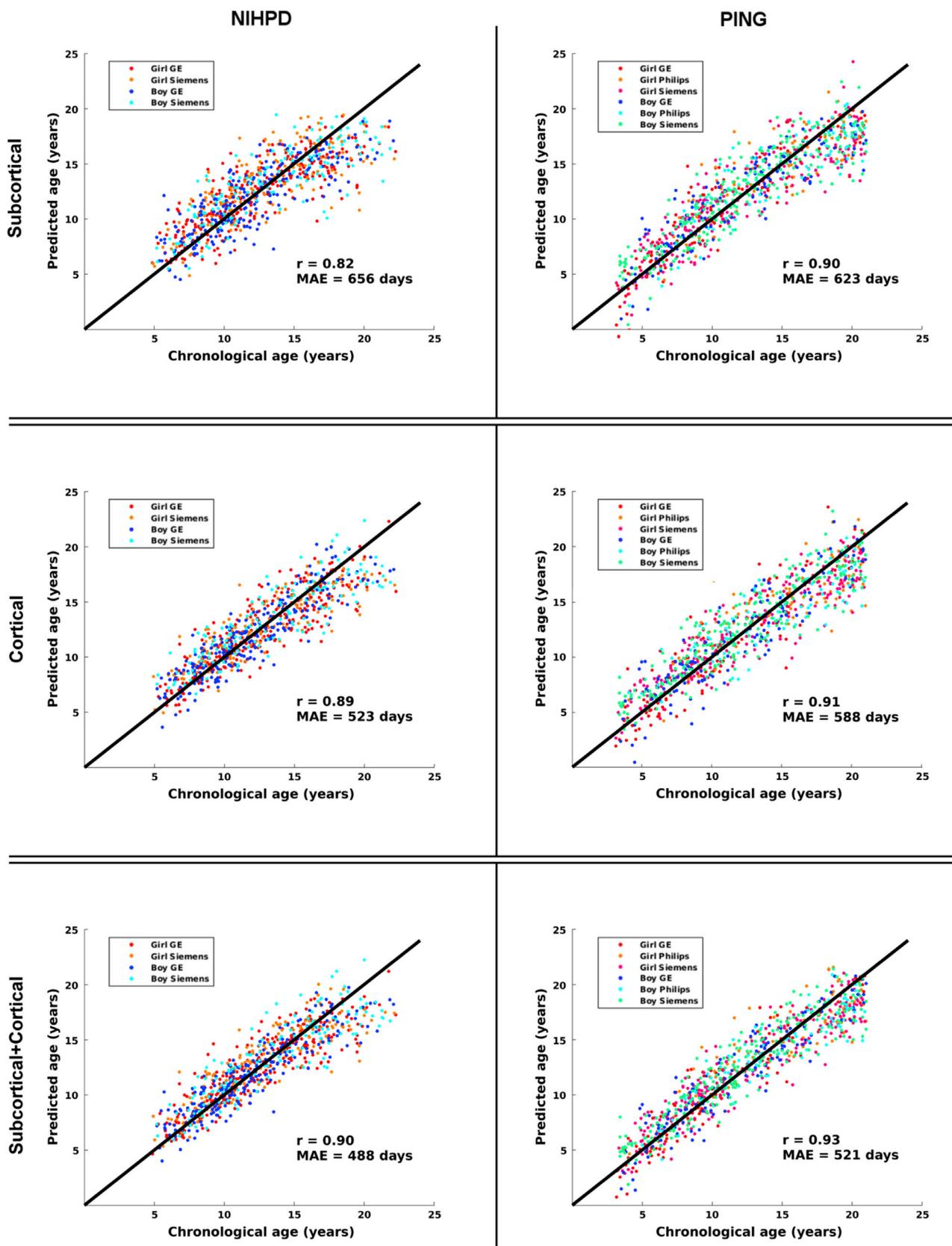


Fig. 3. Scatter-plots of predicted versus chronological age during a single cross-validation run (one out of ten) for subcortical contrast alone, for cortical contrast alone, and for subcortical and cortical contrast combined. The black line has a slope of 1 and originates at the origin; this line depicts the optimal predictions. The scanner manufacturer/gender combinations are shown with different colors. Note that there is no apparent bias either due to gender or scanner manufacturer, and that prediction error is relatively uniform across the age span. Note also the lesser spread of the predictions with subcortical and cortical contrast combined than with either alone.

For either dataset, cortical contrast produced significantly more accurate predictions than subcortical contrast, and the inclusion of subcortical contrast in addition to cortical contrast produced significantly more accurate predictions than either alone. These results are visible in the scatter plots of predicted age versus chronological age (Fig. 3). The plots for the predictions based on subcortical contrast show greater variance than do those based on cortical contrast; and those based on the combination of measures show even less variance. And notice that for both datasets, and for subcortical contrast, cortical contrast, and the combination of measures, the age estimates are biased by neither scanner manufacturer nor gender, and that prediction error is relatively uniform across the age span.

3.2. Age predictors

Our usage of an elastic net penalized linear regression model reveals the brain regions which contribute most to age-prediction. The cortical parcels for which the combined cortical and subcortical white/gray contrast measures reliably contribute to age prediction are shown in Fig. 4, for both databases.

We defined the importance of the parcel based on the parcel weights.

The weight of the parcel i is defined as the sum of absolute weights of linear and quadratic terms $w_i = |b_i| + |q_i|$. We defined the parcel importance as the median value of the weight of the parcel i (w_i) in the regression model across the 10×10 cross-validation; thus a parcel had to be selected in the age model at least 50 times during the 10×10 cross-validation runs to achieve a non-zero value. It did not matter whether linear or quadratic or both components contributed to selection of the parcel. The figure shows the non-zero importances of all parcels for both databases. Across both databases the non-zero importances are broadly distributed across the cortex and the subcortical structures; they are found in each lobe and in each subcortical structure. Thus the age predictions were constructed from information throughout the cortex and the subcortical structures.

3.3. Residuals and IQ

The mixed-effects linear model assessing the relationship between the age-prediction residuals and performance IQ (PIQ), controlling for verbal IQ (VIQ) and brain volume revealed a significant positive relationship ($p = 0.000$, $r = 0.133$). The mixed-effects linear model assessing the relationship between the age prediction residuals and VIQ, controlling

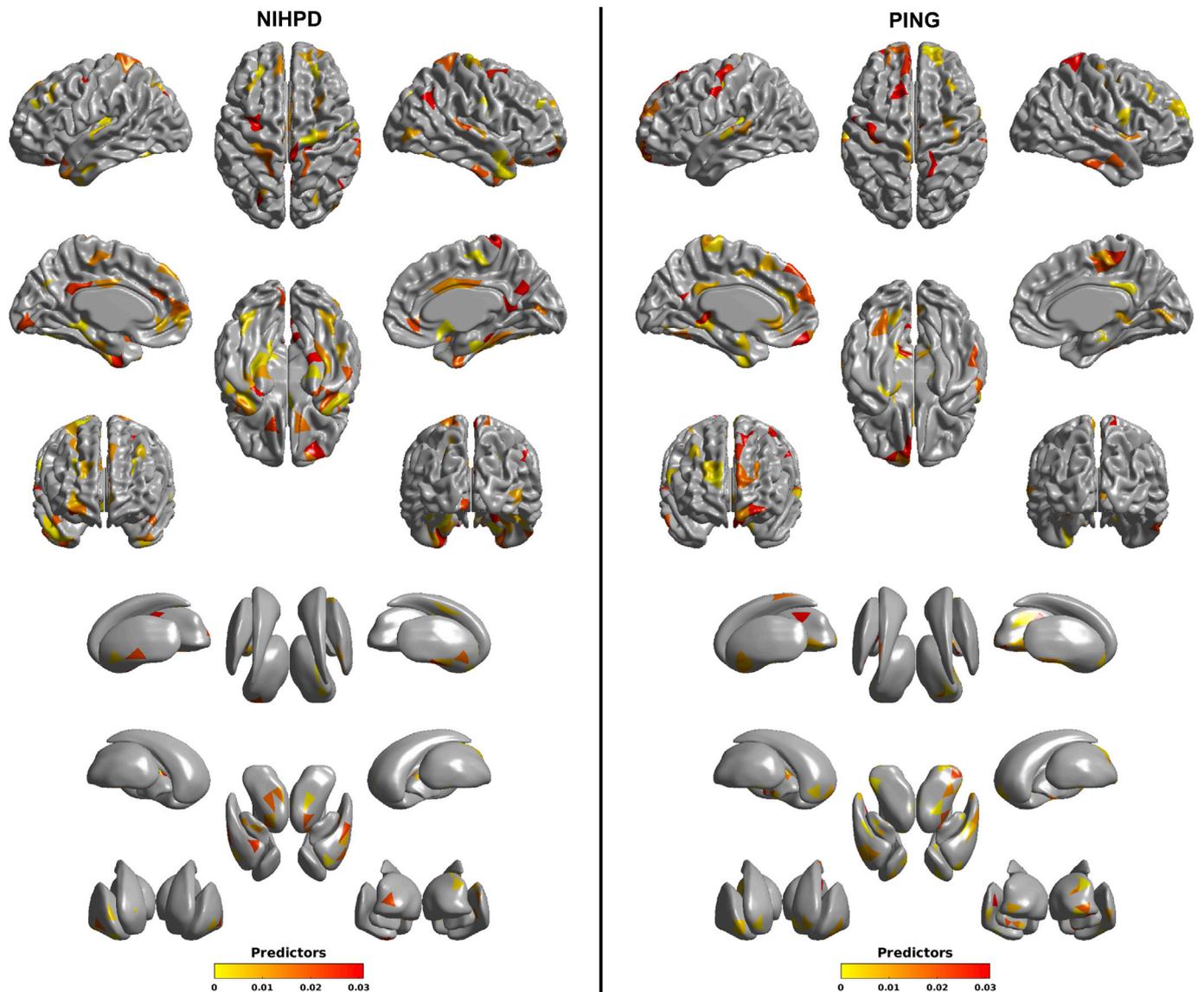


Fig. 4. The importance of different parcels for age predictions for combined cortical and subcortical white/gray contrast in NIHPD and PING. Note that for both datasets there are predictors in each lobe and in each subcortical structure.

for PIQ and brain volume revealed a significant negative relationship ($p = 0.019$, $r = -0.035$). These relationships are shown in Fig. 5.

To map these relationships with IQ to the brain we determined the best-fit mixed-effects linear model explaining white/gray contrast, then decomposed the residuals of that model, and assessed the relation of the top two components to performance IQ and verbal IQ. The model with the lowest AIC among all the possible models comprised of any of ‘AGE’, ‘SEX’, ‘BRAINVOL’, ‘SITE’, ‘AGE*AGE’, and ‘AGE*SEX’, together with ‘random(subjects) + I + 1’ was:

$$\text{‘AGE} + \text{AGE*AGE} + \text{BRAINVOL} + \text{SEX} + \text{SITE} + \text{random(subjects)} + \text{I} + \text{1’}$$

SurfStatLinMod together with *SurfStatPCA* were used to obtain the first two principal components of the residuals of this model. The first principal component, explaining 14.4% of the variance, is shown in Fig. 6 mapped to the cortical and subcortical surfaces; the second principal component, explaining 3.8% of the variance, is shown in Fig. 7.

In addition to the first and second principal components, the figures also show the relation of these components to each of performance IQ and verbal IQ. The first principal component involves most of cortex, with the exception of the bilateral pre- and post-central gyri, the calcarine fissure, and the perihippocampal gyri. The subcortical structures are almost completely uninvolved. This component is significantly related to performance IQ ($p = 0.009$, $r = 0.091$), but not verbal IQ. The second principal component involves the cortex adjacent to the bilateral arcuate fasciculi, the middle temporal gyri, and in the opposite direction, the basal ganglia, but not the thalamus. This component is significantly related to verbal IQ ($p = 0.019$, $r = 0.082$), but not performance IQ.

4. Discussion

As anticipated, the extension of our previous work to include subcortical white/gray contrast in addition to cortical white/gray contrast produced both increased age-prediction accuracy and additional insight into the basis of differences in cognitive functioning. In the PING dataset, with over 900 subjects scanned on 3T scanners, the inclusion of

subcortical white/gray contrast yielded an increase in the correlation of predicted and chronological age from 0.91 to 0.93 and a decrease in the mean absolute error from 588 days to 521 days. In the NIHPD dataset, with 402 subjects scanned on 1.5T scanners, the inclusion of subcortical white/gray contrast yielded a lesser increase in accuracy, with the correlation of predicted and chronological age increasing from 0.89 to 0.90, and the mean absolute error decreasing from 523 days to 488 days. Note that these results, from a single metric in a single image-type, are exceptional. For the PING data, Ball et al. (2017) produced predictions of similar, but slightly lesser, accuracy, also using a single image-type, but using multiple metrics; their correlation between predicted and chronological age was 0.926 and their mean absolute error was approximately 566 days. Brown et al. (2012), also using a single image-type but multiple metrics, achieved a correlation between predicted and chronological age of only 0.91, with a mean absolute error of approximately 624 days. And for the NIHPD data, Khundrakpam et al. (2015), using cortical thickness, achieved a correlation of 0.84, with a mean absolute error of 613 days. Franke et al. (2012) achieved an ostensibly superior result with the NIHPD data, but using a subsample of the scans heavily skewed toward younger subjects, e.g. 32% of their subjects were less than 8 years of age, whereas only 13% of ours were less than 8 years of age, and only 9% of their subjects were older than 16 years of age, whereas 15% of ours were older than 16 years of age. Comparisons across samples that differ in this way are not straightforward. Development is more rapid in younger subjects and so age prediction is more accurate; and the sparsity of the results at the upper end of the age scale in Franke et al.’s sample minimizes the negative impact of any part of the error that is due to nonlinear issues arising during adolescence. Among those studies using similar samples to those used in the current work, our methods produce more accurate predictions, including in comparison to those using multiple metrics.

And, in addition to the residuals of our age predictions reflecting individual differences in cognitive functioning, as was also the case using only cortical white/gray contrast (Lewis et al., 2018), the inclusion of the subcortical white/gray contrast measures yields a clear distinction between the relation of the age-prediction errors and VIQ versus PIQ. PIQ

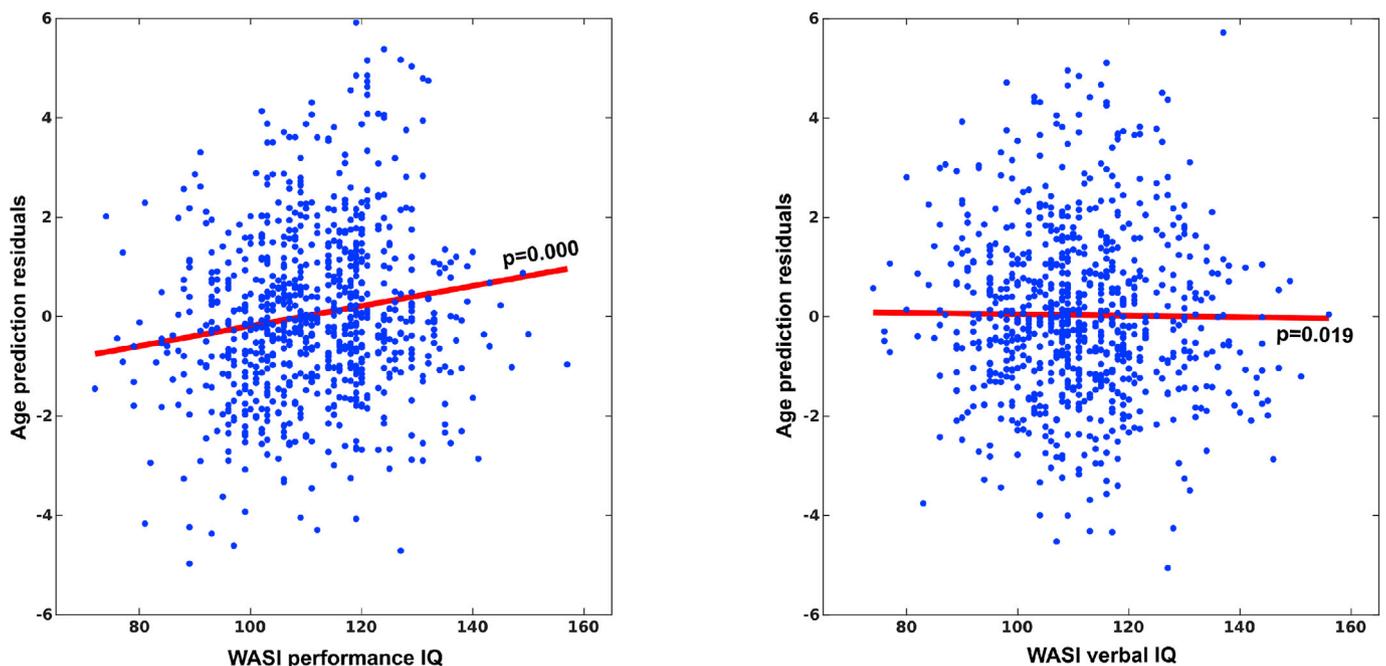


Fig. 5. The relation between the age prediction residuals and both PIQ and VIQ. The left scatterplot shows the relation between the age prediction residuals and PIQ, controlling for VIQ and brain volume within a mixed effects model; it is a significant positive relationship ($p = 0.000$, $r = 0.133$). The right scatterplot shows the relation between the age prediction residuals and VIQ, controlling for PIQ and brain volume within a mixed effects model; it is a significant negative relationship ($p = 0.019$, $r = -0.035$).

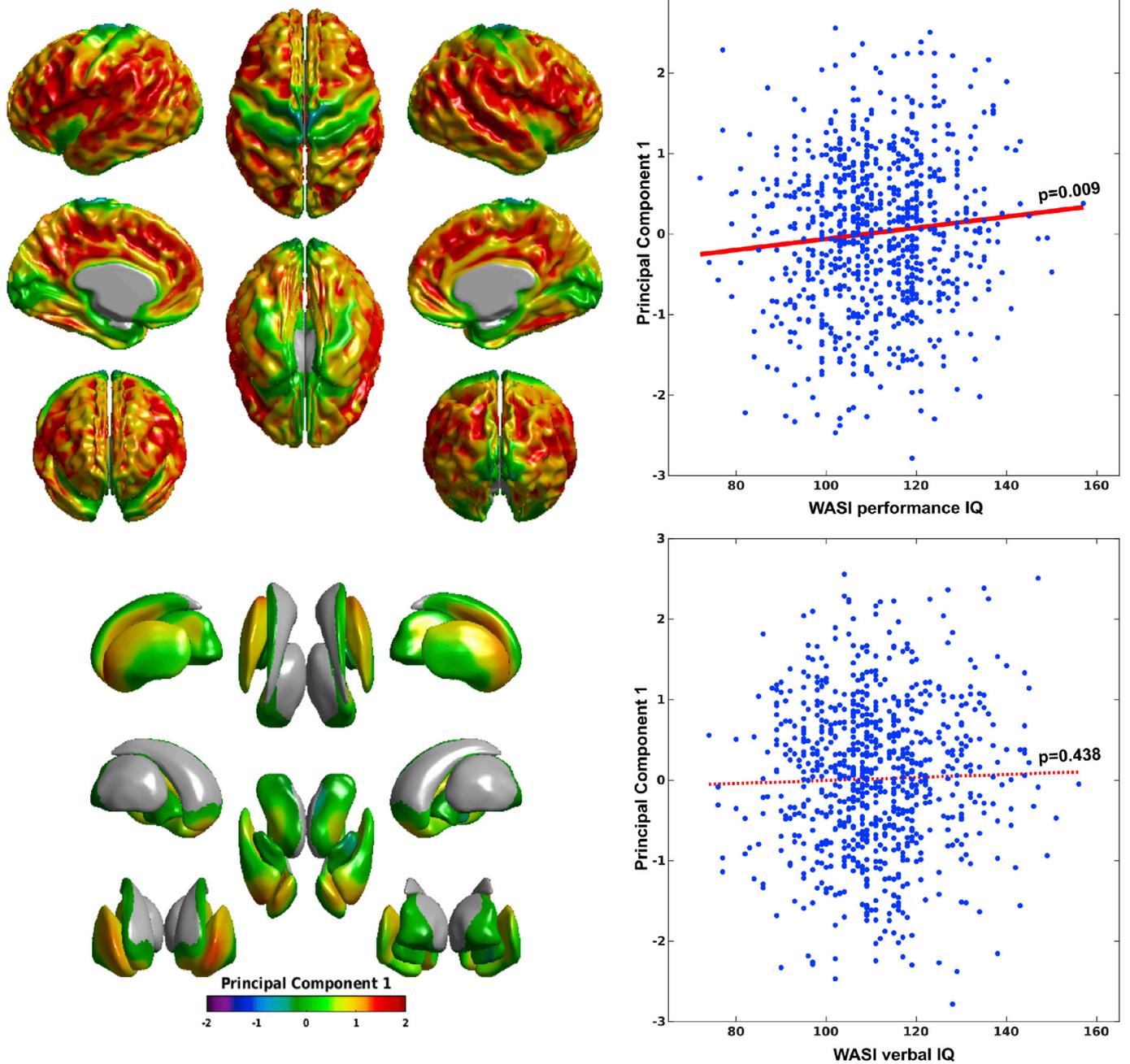


Fig. 6. The first principal component and its relation to performance IQ and verbal IQ. Regions in which white/gray contrast measures cannot be generated (due to e.g. adjacency to the ventricles) are shown in gray. This component captures variation in cortical contrast, broadly, but not in non-primary areas nor the basal ganglia nuclei of the cerebrum, nor the thalamus. This component is significantly related to performance IQ ($p = 0.009$, $r = 0.091$) but not verbal IQ.

was shown to be positively related to the age-prediction residuals, while VIQ was shown to be negatively related to the residuals. Thus, the younger a subject's brain appears relative to their chronological age, the higher that subject's PIQ; and the older a subject's brain appears relative to their chronological age, the higher that subject's VIQ. A number of previous age-prediction studies have sought to relate their errors to cognitive function, but have generally found no relationship between the two, e.g. Ball et al. (2017) and Khundrakpam et al. (2015). Khundrakpam et al. (2015) reported a significant relation between the age-prediction errors and the interaction of PIQ and VIQ, which is in agreement with our findings of relations in opposing directions for the relation between the age-prediction residuals and VIQ versus PIQ. But, Khundrakpam et al. (2015) found no relation between either PIQ or VIQ and their

age-prediction errors. Interestingly, Erus et al. (2015) reported that individuals with brains that appeared younger than their chronological age showed inferior cognitive processing speed, and vice-versa. To the extent that processing speed correlates with PIQ, Erus et al.'s results conflict with the results here; and to the extent that processing speed correlates with VIQ, their results align with ours. But, it is unclear how processing speed relates to PIQ and VIQ; significant processing speed relations included verbal memory, but also attention, face memory, spatial memory, emotion identification, motor speed, and sensory-motor processing speed.

The inclusion of the subcortical white/gray contrast measures also provided novel insights into the brain-basis of the opposing relationships between the age-prediction residuals and PIQ versus VIQ. The top two

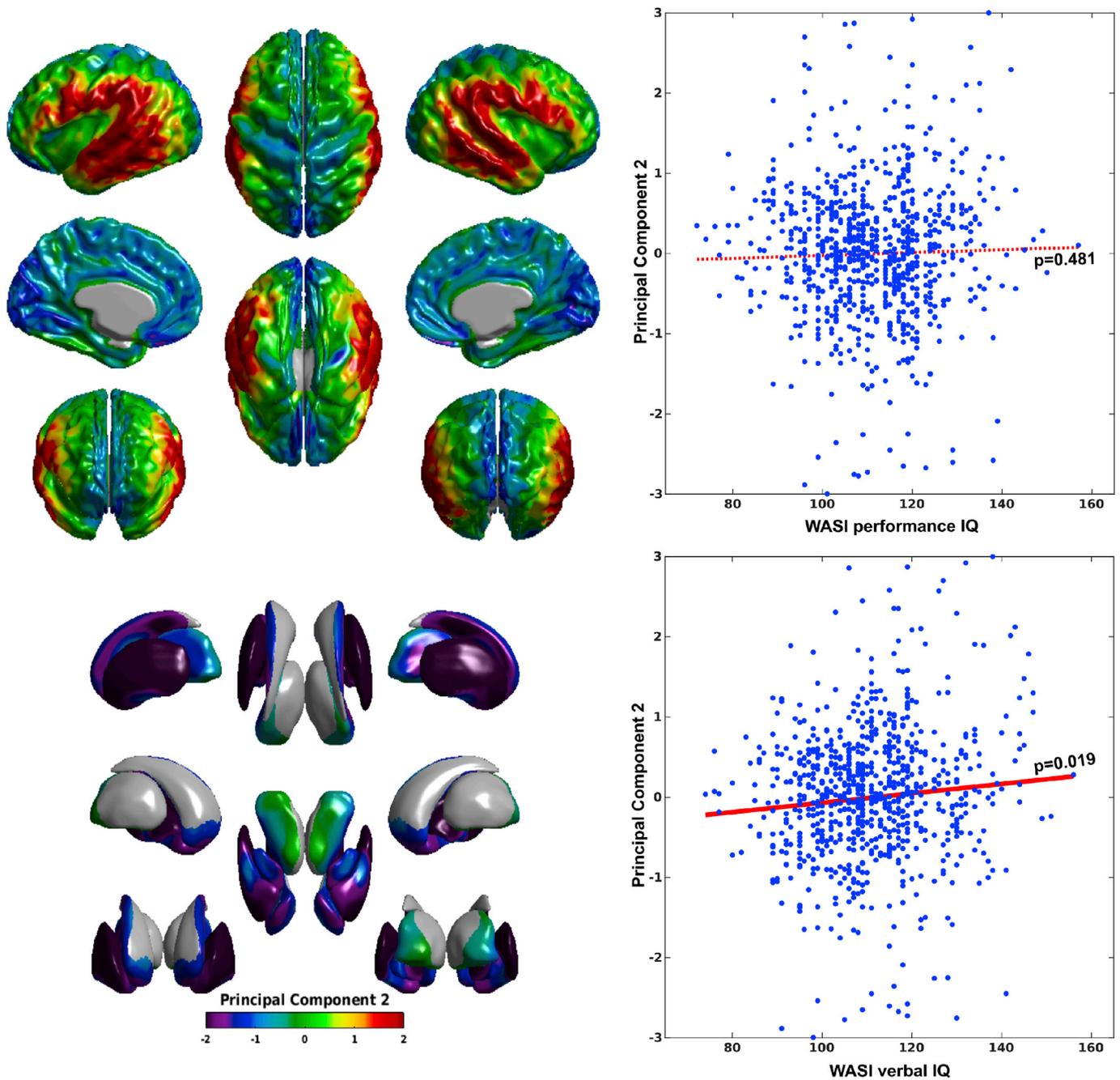


Fig. 7. The (inverted) second principal component and its relation to performance IQ and verbal IQ. Regions in which white/gray contrast measures cannot be generated (due to e.g. adjacency to the ventricles) are shown in gray. This component involves the basal ganglia nuclei of the cerebrum (but not the thalamus) and inversely the bilateral arcuate fasciculi. It is significantly related to verbal IQ ($p = 0.019$, $r = 0.082$) but not performance IQ. Note that this is the inverted version of the second principal component; the actual relation of the second principal component to verbal IQ is negative. The inversion makes the interpretation more natural: higher verbal IQ corresponds to greater contrast in the bilateral arcuate fasciculi and lesser contrast in the basal ganglia.

components of a principle component analysis of the residuals of the best-fit model for the contrast data showed a categorical distinction in the brain-basis of PIQ versus VIQ, with the first component related to PIQ but not VIQ, and the second component related to VIQ but not PIQ. And the mappings of these components to the brain indicated that inter-individual PIQ differences involve much of the cortex but not the primary sensory areas nor any of the subcortical structures, while inter-individual VIQ differences involve the basal ganglia and the bilateral arcuate fasciculi but not the thalamus nor the rest of the cortex.

A natural interpretation of the positive relation between PIQ and the first principle component is that greater cortical contrast represents a greater degree of myelination in the fibers that interconnect wide-spread

cortical regions, and a lesser increase in material within the cortex in these wide-spread regions. Axonal myelination is activity dependant, as are cortical regressive developmental processes, e.g. synaptic pruning, and thus the processes that drive such an outcome may be tightly coupled; greater myelination yields more efficient communication, yielding more effective sculpting of cortical gray matter. And white-matter integrity in the fiber bundles that connect these wide-spread regions of cortex has been linked to PIQ (Chiang et al., 2009). This interpretation fits well with Jung and Haier's (2007) Parieto-Frontal Integration Theory (P-FIT) of intelligence. The same interpretation applies to the positive relation between VIQ and the arcuate fasciculi area of the second principle component. The arcuate fasciculi are

bundles of white-matter fibers that interconnect critical components of the language network, e.g. Broca's area in the inferior frontal gyrus, Wernicke's area at the posterior of the superior temporal gyrus, and auditory cortex in the superior and middle temporal gyri. In adults, language is lateralized; in infants, language is bilateral (Perani et al., 2011; Holland et al., 2007; Broce et al., 2018; Su et al., 2018). And the development of the arcuate fasciculi is reflected in cognitive ability (Schmithorst et al., 2005; Lebel and Beaulieu, 2009).

A natural interpretation of the negative relation of VIQ and the basal ganglia portion of the second principle component is similar. But, the internal architecture of subcortical structures shows substantial alterations over development, in part due to the myelination of invading fibers, in part due to maturation of fibers inter-connecting sub-nuclei, and in part due to alterations in cell structure (Takase et al., 2004; Wierenga et al., 2014; Geeraert et al., 2018). The impact of these changes on the T1 signal are not completely understood, but white/gray contrast decreases rapidly with age in the basal ganglia. Thus the negative relation between the age-prediction residuals and VIQ corresponds to a positive relation between VIQ and the state of maturation of the basal ganglia. The basal ganglia are considered to be “cognitive pattern generators” (Graybiel, 1997) regulating sequential processing, and central to sequence learning, including vocal sequences (Packard and Knowlton, 2002; Graybiel, 2005; Ölveczky et al., 2005; Dominey et al., 2003; Houk et al., 1995; Damasio, 1983; Kotz et al., 2009; Price, 2012). The accuracy of phonological processing is correlated with greater activation in the left caudate (Moro et al., 2001); the speed of phonological processing is correlated with greater activation in the left putamen (Tettamanti et al., 2005; Wildgruber et al., 2001). Syntactic processing, likewise, appears to involve the left caudate and putamen, as well as the globus pallidus (Friederici et al., 1999; Kotz et al., 2003; Friederici et al., 2003b, a; Newman et al., 2010). And, most directly related to the findings here, morphological measures of the basal ganglia have been associated with IQ (MacDonald et al., 2000; Isaacs et al., 2008; Grazioplene et al., 2015), with some specificity of the core relation being between the caudate and VIQ (Isaacs et al., 2008; Grazioplene et al., 2015).

Thus the second principle component indicates co-maturation of the arcuate fasciculi and the basal ganglia. Both the arcuate fasciculi and the basal ganglia have been linked to language, but to the best of our knowledge have not been linked to each other. That link could, of course, be a coincidence; the covariation that places the arcuate fasciculi and the basal ganglia into the same principle component might be chance. But it seems more likely that the role of the basal ganglia in phonological and syntactic processing is impacting the development of the arcuate fasciculi. Moreover, connectivity from the basal ganglia to cortical regions within the language network has been reported. In non-human primates, tracer injection studies have shown basal ganglia afferents from wide-spread regions of cortex (Yeterian and Van Hoesen, 1978; Yeterian and Pandya, 1993; Van Hoesen et al., 1981), and basal ganglia projections to a number of cortical regions in the prefrontal cortex including the supplementary motor area (SMA), the pre-SMA, and most notably to a region of the ventral premotor cortex claimed to be homologous to Broca's area (Akkal et al., 2007; Middleton and Strick, 2000). In humans, diffusion tractography has also shown connectivity between Broca's area and the putamen (Ford et al., 2013); tractography, however, cannot distinguish the directionality of connections, but a dynamic causal modeling analysis of functional MRI indicated unidirectional connections from the putamen to the left inferior frontal gyrus (Booth et al., 2007).

Of course, this interpretation is speculative. Speculative in part because the white/gray contrast measure is a ratio, and the intensity measures that comprise it cannot be considered in isolation; in part because the sign of a principle component is arbitrary; and in part because the co-variation that yields a principle component may stem from unrelated sources. Further, the neurobiological properties underlying differences in the T1-weighted signal are not completely understood. Most of the spatial T1 variance in the normal mature brain can be

explained by differences in macromolecular mass fraction (Rooney et al., 2007); and differences in the quantity and structure of myelin between white-matter and gray-matter, and the changes in myelin across development, are reflected in T1 relaxation times (Agartz et al., 1991; Barkovich et al., 1988; Barkovich, 2000; Peters, 2002). But, the positive relation between cortical contrast and both PIQ and VIQ might stem from either relatively increased myelin within the white-matter, or relatively reduced myelin and cellular material within the cortical gray-matter, e.g. from a lesser degree of invasion of white-matter fibers into the cortex. Which of these is the case is unclear. Likewise, the negative relation between VIQ and contrast within the basal ganglia might stem from either relatively increased myelin and cellular structure within the basal ganglia, or from relatively decreased myelin in the surrounding white-matter. Again, which of these is the case is unclear. But regardless of the interpretation, the combination of cortical and subcortical white-gray contrast provides improved age-prediction accuracy over cortical white/gray contrast alone, and indicates a heretofore unseen distinction in the impact of VIQ versus PIQ on such predictions, as well as identifying the brain-basis of such a distinction.

There are, however, a number of limitations of this study that should be noted. As mentioned above, interpretation of the white/gray contrast measures is hampered by the fact that the measures that comprise the ratio are not individually interpretable. Quantitative MRI, for which the intensity measures in gray matter and in white matter are individually meaningful, would solve this problem. But, to the best of our knowledge, no publicly available developmental data of this sort currently exists. As also mentioned above, the covariation that yielded the link between the basal ganglia and the arcuate fasciculi might stem from independent sources. Longitudinal functional data during development might address this; but again, to the best of our knowledge, no publicly available developmental data of this sort currently exists. Moreover, even with such data, the role of developmental cortical thinning in driving the alterations in white/gray contrast would cloud the interpretation of developmental alterations in contrast, since it is equally unclear what underlies this cortical thinning. One possibility is that there is an actual reduction in cortical material, e.g. synapses; but another is that increasing invasion of myelinated cortico-cortical and thalamo-cortical connections displace the white surface outward. The latter possibility would mean that the measures of gray matter used in the white/gray contrast measures would be taken from different depths within the cortex at different points in development, whether measuring at an absolute depth or at some percentage of cortical thickness. Additionally, there may be non-biological factors that interfere with interpretation of white/gray contrast. Errors in surface placement will introduce errors in white/gray contrast measures, and such errors may differ across development. Imaging noise, for example, differs with head size due to different distances from the anatomy to the receive coil; and head size, of course, changes with age. And the extent to which there is head motion during scanning may differ with age, as well. These factors, together with age-related alterations in the extent to which the dura is distinguishable from the cortex, and cerebro-spinal fluid is visible in the sulci, will impact the correct placement of the white and gray matter surfaces. Future research should address these issues.

Acknowledgments

This research has been supported by grant 316258 from the Academy of Finland (to JT), by grant ANRP-MIRI13-3388 from the Azrieli Neurodevelopmental Research Program in partnership with the Brain Canada Multi-Investigator Research Initiative (to ACE), and by grants from the Canadian Institutes of Health Research and the Natural Sciences and Engineering Research Council of Canada (to DLC). It also benefited from computational resources provided by Compute Canada (www.compute-canada.ca), Calcul Quebec (www.calculquebec.ca), and the UEF Bioinformatics Center (bioinformatics.uef.fi).

Data used in the preparation of this article were obtained from the

NIH Pediatric MRI Data Repository created by the NIH MRI Study of Normal Brain Development. This is a multisite, longitudinal study of typically developing children conducted by the Brain Development Cooperative Group and supported by the National Institute of Child Health and Human Development, the National Institute on Drug Abuse, the National Institute of Mental Health, and the National Institute of Neurological Disorders and Stroke (Contract #s N01-HD02-3343, N01-MH9-0002, and N01-NS-9-2314, –2315, –2316, –2317, –2319 and –2320). A listing of the participating sites and a complete listing of the study investigators can be found on the NIH website.⁷ This manuscript reflects the views of the authors and may not reflect the opinions or views of the NIH.

Data used in the preparation of this manuscript were obtained and analyzed from the controlled access datasets distributed from the NIMH-supported Research Domain Criteria Database (RDoCdb). RDoCdb is a collaborative informatics system created by the National Institute of Mental Health to store and share data resulting from grants funded through the Research Domain Criteria (RDoC) project. This manuscript reflects the views of the authors and may not reflect the opinions or views of the NIH or of the Submitters submitting original data to RDoCdb.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.neuroimage.2019.04.022>.

References

- Agartz, I., Säaf, J., Wahlund, L., Wetterberg, L., 1991. T1 and T2 relaxation time estimates in the normal human brain. *Radiology* 181, 537–543.
- Akaike, H., 1976. An information criterion (AIC). *Math. Sci.* 14, 5–7.
- Akkal, D., Dum, R.P., Strick, P.L., 2007. Supplementary motor area and presupplementary motor area: targets of basal ganglia and cerebellar output. *J. Neurosci.* 27, 10659–10673.
- Ambrose, C., McLachlan, G.J., 2002. Selection bias in gene extraction on the basis of microarray gene-expression data. *Proc. Natl. Acad. Sci. Unit. States Am.* 99, 6562–6566.
- Ball, G., Adamson, C., Beare, R., Seal, M.L., 2017. Modelling neuroanatomical variation during childhood and adolescence with neighbourhood-preserving embedding. *Sci. Rep.* 7, 17796.
- Barkovich, A.J., 2000. Concepts of myelin and myelination in neuroradiology. *Am. J. Neuroradiol.* 21, 1099–1109.
- Barkovich, A., Kjos, B., Jackson Jr., D., Norman, D., 1988. Normal maturation of the neonatal and infant brain: MR imaging at 1.5 T. *Radiology* 166, 173–180.
- Bohnen, M.M., Brouwer, R.M., Mandl, R.C., Hedman, A.M., van den Heuvel, M.P., van Haren, N.E., Kahn, R.S., Pol, H.E.H., 2016. Topology of genetic associations between regional gray matter volume and intellectual ability: evidence for a high capacity network. *Neuroimage* 124, 1044–1053.
- Booth, J.R., Wood, L., Lu, D., Houk, J.C., Bitan, T., 2007. The role of the basal ganglia and cerebellum in language processing. *Brain Res.* 1133, 136–144.
- Broce, I.J., Bernal, B., Altman, N., Bradley, C., Baez, N., Cabrera, L., Hernandez, G., De Fera, A., Dick, A.S., 2018. Fiber pathways supporting early literacy development in 5–8-year-old children. *Brain Cogn.* <https://doi.org/10.1016/j.bandc.2018.12.004> In Press, Corrected Proof, Available online 21 December 2018.
- Brown, T.T., Kuperman, J.M., Chung, Y., Erhart, M., McCabe, C., Hagler, D.J., Venkatraman, V.K., Akshoomoff, N., Amaral, D.G., Bloss, C.S., et al., 2012. Neuroanatomical assessment of biological maturity. *Curr. Biol.* 22, 1693–1698.
- Cerliani, L., Mennes, M., Thomas, R.M., Di Martino, A., Thibout, M., Keyser, C., 2015. Increased functional connectivity between subcortical and cortical resting-state networks in autism spectrum disorder. *JAMA Psychiatr.* 72, 767–777.
- Chiang, M.C., Barysheva, M., Shattuck, D.W., Lee, A.D., Madsen, S.K., Avedissian, C., Klunder, A.D., Toga, A.W., McMahon, K.L., De Zubicaray, G.I., et al., 2009. Genetics of brain fiber architecture and intellectual performance. *J. Neurosci.* 29, 2212–2224.
- Collins, D.L., Neelin, P., Peters, T.M., Evans, A.C., 1994. Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. *J. Comput. Assist. Tomogr.* 18, 192–205.
- Coupe, P., Manjón, J.V., Fonov, V., Pruessner, J., Robles, M., Collins, D.L., 2011. Patch-based segmentation using expert priors: application to hippocampus and ventricle segmentation. *Neuroimage* 54, 940–954.
- Damasio, A.R., 1983. Language and the basal ganglia. *Trends Neurosci.* 6, 442–443.
- Dominey, P.F., Hoen, M., Blanc, J.M., Lelekov-Boissard, T., 2003. Neurological basis of language and sequential cognition: evidence from simulation, aphasia, and ERP studies. *Brain Lang.* 86, 207–225.
- Dosenbach, N.U., Nardos, B., Cohen, A.L., Fair, D.A., Power, J.D., Church, J.A., Nelson, S.M., Wig, G.S., Vogel, A.C., Lessov-Schlaggar, C.N., et al., 2010. Prediction of individual brain maturity using fMRI. *Science* 329, 1358–1361.
- Erus, G., Battapady, H., Satterthwaite, T.D., Hakonarson, H., Gur, R.E., Davatzikos, C., Gur, R.C., 2015. Imaging patterns of brain development and their relationship to cognition. *Cerebr. Cortex* 25, 1676–1684.
- Evans, A.C., the Brain Development Cooperative Group, 2006. The NIH MRI study of normal brain development. *Neuroimage* 30, 184–202.
- Fischi-Gómez, E., Vasung, L., Meskaldji, D.E., Lazeyras, F., Borradori-Tolsa, C., Hagmann, P., Barisnikov, K., Thiran, J.P., Hüppi, P.S., 2014. Structural brain connectivity in school-age preterm infants provides evidence for impaired networks relevant for higher order cognitive skills and social cognition. *Cerebr. Cortex* 25, 2793–2805.
- Ford, A., Triplett, W., Sudhyadhom, A., Gullett, J.M., McGregor, K., FitzGerald, D., Mareci, T., White, K., Crosson, B., 2013. Broca's area and its striatal and thalamic connections: a diffusion-MRI tractography study. *Front. Neuroanat.* 7, 8.
- Franke, K., Luders, E., May, A., Wilke, M., Gaser, C., 2012. Brain maturation: predicting individual brainAGE in children and adolescents using structural MRI. *Neuroimage* 63, 1305–1312.
- Friederici, A.D., Von Cramon, D.Y., Kotz, S.A., 1999. Language related brain potentials in patients with cortical and subcortical left hemisphere lesions. *Brain* 122, 1033–1047.
- Friederici, A.D., Kotz, S.A., Werheid, K., Hein, G., von Cramon, D.Y., 2003a. Syntactic comprehension in Parkinson's disease: investigating early automatic and late integrational processes using event-related brain potentials. *Neuropsychology* 17, 133.
- Friederici, A.D., Rueschemeyer, S.A., Hahne, A., Fiebach, C.J., 2003b. The role of left inferior frontal and superior temporal cortex in sentence comprehension: localizing syntactic and semantic processes. *Cerebr. Cortex* 13, 170–177.
- Friedman, J., Hastie, T., Tibshirani, R., 2010. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* 33, 1.
- Geeraert, B.L., Lebel, R.M., Mah, A.C., Deoni, S.C., Alsop, D.C., Varma, G., Lebel, C., 2018. A comparison of inhomogeneous magnetization transfer, myelin volume fraction, and diffusion tensor imaging measures in healthy children. *Neuroimage* 182, 343–350.
- Gelman, A., 2008. Scaling regression inputs by dividing by two standard deviations. *Stat. Med.* 27, 2865–2873.
- Graybiel, A.M., 1997. The basal ganglia and cognitive pattern generators. *Schizophr. Bull.* 23, 459–469.
- Graybiel, A.M., 2005. The basal ganglia: learning new tricks and loving it. *Curr. Opin. Neurobiol.* 15, 638–644.
- Grazioplene, R.G., Gyman, S., Gray, J.R., Rustichini, A., Jung, R.E., DeYoung, C.G., 2015. Subcortical intelligence: caudate volume predicts IQ in healthy adults. *Hum. Brain Mapp.* 36, 1407–1416.
- Guo, J.F., Yang, Y.K., Chiu, N.T., Yeh, T.L., Chen, P.S., Lee, I.H., Chu, C.L., 2006. The correlation between striatal dopamine D 2/D 3 receptor availability and verbal intelligence quotient in healthy volunteers. *Psychol. Med.* 36, 547–554.
- Hartberg, C.B., Sundet, K., Rimol, L.M., Haukvik, U.K., Lange, E.H., Nesvåg, R., Melle, I., Andreassen, O.A., Agartz, I., 2011. Subcortical brain volumes relate to neurocognition in schizophrenia and bipolar disorder and healthy controls. *Prog. Neuro Psychopharmacol. Biol. Psychiatr.* 35, 1122–1130.
- Hibar, D., Westlye, L.T., van Erp, T.G., Rasmussen, J., Leonardo, C.D., Faskowitz, J., Haukvik, U.K., Hartberg, C.B., Doan, N.T., Agartz, I., et al., 2016. Subcortical volumetric abnormalities in bipolar disorder. *Mol. Psychiatr.* 21, 1710.
- Holland, S.K., Vannest, J., Mecoli, M., Jacola, L.M., Tillema, J.M., Karunanayaka, P.R., Schmithorst, V.J., Yuan, W., Plante, E., Byars, A.W., 2007. Functional MRI of language lateralization during development in children. *Int. J. Audiol.* 46, 533–551.
- Hoogman, M., Bralten, J., Hibar, D.P., Mennes, M., Zwiers, M.P., Schwen, L.S., van Hulzen, K.J., Medland, S.E., Shumskaya, E., Jahanshad, N., et al., 2017. Subcortical brain volume differences in participants with attention deficit hyperactivity disorder in children and adults: a cross-sectional mega-analysis. *Lancet Psychiatr.* 4, 310–319.
- Houk, J.C., Davis, J.L., Beiser, D.G., 1995. *Models of Information Processing in the Basal Ganglia*. MIT press.
- Humphries, M.D., Prescott, T.J., 2010. The ventral basal ganglia, a selection mechanism at the crossroads of space, strategy, and reward. *Prog. Neurobiol.* 90, 385–417.
- Huttunen, H., Manninen, T., Tohka, J., 2012. MEG mind reading: strategies for feature selection. *Proc. Fed. Comput. Sci. Event* 42–49, 2012.
- Isaacs, E.B., Gadian, D.G., Sabatini, S., Chong, W.K., Quinn, B.T., Fischl, B.R., Lucas, A., 2008. The effect of early human diet on caudate volumes and IQ. *Pediatr. Res.* 63, 308.
- Johnson, M.H., 2005. Subcortical face processing. *Nat. Rev. Neurosci.* 6, 766.
- Jung, R.E., Haier, R.J., 2007. The Parieto-Frontal Integration Theory (P-FIT) of intelligence: converging neuroimaging evidence. *Behav. Brain Sci.* 30, 135–154.
- Khundrakpam, B.S., Tohka, J., Evans, A.C., Group, B.D.C., et al., 2015. Prediction of brain maturity based on cortical thickness at different spatial resolutions. *Neuroimage* 111, 350–359.
- Kim, J.S., Singh, V., Lee, J.K., Lerch, J., Ad-Dab'bagh, Y., MacDonald, D., Lee, J.M., Kim, S.I., Evans, A.C., 2005. Automated 3-D extraction and evaluation of the inner and outer cortical surfaces using a laplacian map and partial volume effect classification. *Neuroimage* 27, 210–221.
- Kotz, S.A., Frisch, S., Von Cramon, D.Y., Friederici, A.D., 2003. Syntactic language processing: ergo lesion data on the role of the basal ganglia. *J. Int. Neuropsychol. Soc.* 9, 1053–1060.
- Kotz, S.A., Schwartze, M., Schmidt-Kassow, M., 2009. Non-motor basal ganglia functions: a review and proposal for a model of sensory predictability in auditory language perception. *Cortex* 45, 982–990.
- Landman, B., Ribbens, A., Lucas, B., Davatzikos, C., Avants, B., Ledig, C., Ma, D., Rueckert, D., Vandermeulen, D., Maes, F., et al., 2012. MICCAI 2012 workshop on

⁷ http://pediatricmri.nih.gov/nihpd/info/participating_centers.html.

- multi-atlas labeling. In: Medical Image Computing and Computer Assisted Intervention Conference, MICCAI, vols. 63–66, pp. 91–94.
- Lebel, C., Beaulieu, C., 2009. Lateralization of the arcuate fasciculus from childhood to adulthood and its relation to cognitive abilities in children. *Hum. Brain Mapp.* 30, 3563–3573.
- Lewis, J.D., Evans, A.C., Tohka, J., Group, B.D.C., et al., 2018. T1 white/gray contrast as a predictor of chronological age, and an index of cognitive performance. *Neuroimage* 173, 341–350.
- Lytelton, O., Boucher, M., Robbins, S., Evans, A., 2007. An unbiased iterative group registration template for cortical surface analysis. *Neuroimage* 34, 1535–1544.
- MacDonald, D., Kabani, N., Avis, D., Evans, A.C., 2000. Automated 3-D extraction of inner and outer surfaces of cerebral cortex from MRI. *Neuroimage* 12, 340–356.
- Manjón, J.V., Coupé, P., Martí-Bonmatí, L., Collins, D.L., Robles, M., 2010. Adaptive non-local means denoising of mr images with spatially varying noise levels. *J. Magn. Reson. Imaging* 31, 192–203.
- Middleton, F.A., Strick, P.L., 2000. Basal ganglia output and cognition: evidence from anatomical, behavioral, and clinical studies. *Brain Cogn.* 42, 183–200.
- Moro, A., Tettamanti, M., Perani, D., Donati, C., Cappa, S.F., Fazio, F., 2001. Syntax and the brain: disentangling grammar by selective anomalies. *Neuroimage* 13, 110–118.
- Mwangi, B., Hasan, K.M., Soares, J.C., 2013. Prediction of individual subject's age across the human lifespan using diffusion tensor imaging: a machine learning approach. *Neuroimage* 75, 58–67.
- Newman, A.J., Supalla, T., Hauser, P., Newport, E.L., Bavelier, D., 2010. Dissociating neural subsystems for grammar by contrasting word order and inflection. *Proc. Natl. Acad. Sci. Unit. States Am.* 107, 7539–7544.
- Ólveczky, B.P., Andalman, A.S., Fee, M.S., 2005. Vocal experimentation in the juvenile songbird requires a basal ganglia circuit. *PLoS Biol.* 3, e153.
- Packard, M.G., Knowlton, B.J., 2002. Learning and memory functions of the basal ganglia. *Annu. Rev. Neurosci.* 25, 563–593.
- Perani, D., Saccuman, M.C., Scifo, P., Awander, A., Spada, D., Baldoli, C., Poloniato, A., Lohmann, G., Friederici, A.D., 2011. Neural language networks at birth. *Proc. Natl. Acad. Sci. Unit. States Am.* 108 (38), 16056–16061.
- Peters, A., 2002. Structural changes in the normally aging cerebral cortex of primates. *Prog. Brain Res.* 136, 455–465.
- Price, C.J., 2012. A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *Neuroimage* 62, 816–847.
- Qian, J., Hastie, T., Friedman, J., Tibshirani, R., Simon, N., 2013. *Glmnet for Matlab*, 2013. http://www.stanford.edu/~hastie/glmnet_matlab/.
- Raznahan, A., Shaw, P.W., Lerch, J.P., Clasen, L.S., Greenstein, D., Berman, R., Pipitone, J., Chakravarty, M.M., Giedd, J.N., 2014. Longitudinal four-dimensional mapping of subcortical anatomy in human development. *Proc. Natl. Acad. Sci. Unit. States Am.* 111, 1592–1597.
- Rimol, L.M., Hartberg, C.B., Nesvåg, R., Fennema-Notestine, C., Hagler Jr., D.J., Pung, C.J., Jennings, R.G., Haukvik, U.K., Lange, E., Nakstad, P.H., et al., 2010. Cortical thickness and subcortical volumes in schizophrenia and bipolar disorder. *Biol. Psychiatry* 68, 41–50.
- Rooney, W.D., Johnson, G., Li, X., Cohen, E.R., Kim, S.G., Ugurbil, K., Springer, C.S., 2007. Magnetic field and tissue dependencies of human brain longitudinal 1H2O relaxation in vivo. *Magn. Reson. Med.* 57, 308–318.
- Salat, D.H., Lee, S.Y., Van der Kouwe, A., Greve, D.N., Fischl, B., Rosas, H.D., 2009. Age-associated alterations in cortical gray and white matter signal intensity and gray to white matter contrast. *Neuroimage* 48, 21–28.
- Schmithorst, V.J., Wilke, M., Dardzinski, B.J., Holland, S.K., 2005. Cognitive functions correlate with white matter architecture in a normal pediatric population: a diffusion tensor MRI study. *Hum. Brain Mapp.* 26, 139–147.
- Sled, J.G., Zijdenbos, A.P., Evans, A.C., 1998. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *Medical Imaging. IEEE Transactions on* 17, 87–97.
- Su, M., Thiebaut de Schotten, M., Zhao, J., Song, S., Zhou, W., Gong, G., McBride, C., Ramus, F., Shu, H., 2018. Vocabulary growth rate from preschool to school-age years is reflected in the connectivity of the arcuate fasciculus in 14-year-old children. *Dev. Sci.* 21 (5), e12647.
- Takase, K., Tamagaki, C., Okugawa, G., Nobuhara, K., Minami, T., Sugimoto, T., Sawada, S., Kinoshita, T., 2004. Reduced white matter volume of the caudate nucleus in patients with schizophrenia. *Neuropsychobiology* 50, 296–300.
- Tettamanti, M., Moro, A., Messa, C., Moresco, R.M., Rizzo, G., Carpinelli, A., Matarrese, M., Fazio, F., Perani, D., 2005. Basal ganglia and language: phonology modulates dopaminergic release. *Neuroreport* 16, 397–401.
- Tohka, J., Zijdenbos, A., Evans, A., 2004. Fast and robust parameter estimation for statistical partial volume models in brain MRI. *Neuroimage* 23, 84–97.
- Utter, A.A., Basso, M.A., 2008. The basal ganglia: an overview of circuits and function. *Neurosci. Biobehav. Rev.* 32, 333–342.
- Van Hoesen, G.W., Yeterian, E.H., Lavizzo-Mourey, R., 1981. Widespread corticostriate projections from temporal cortex of the rhesus monkey. *J. Comp. Neurol.* 199, 205–219.
- van Schouwenburg, M.R., den Ouden, H.E., Cools, R., 2010. The human basal ganglia modulate frontal-posterior connectivity during attention shifting. *J. Neurosci.* 30, 9910–9918.
- Ward Jr., J.H., 1963. Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.* 58, 236–244.
- Weier, K., Fonov, V., Lavoie, K., Doyon, J., Collins, D.L., 2014. Rapid automatic segmentation of the human cerebellum and its lobules (rascal)—implementation and application of the patch-based label-fusion technique with a template library to segment the human cerebellum. *Hum. Brain Mapp.* 35, 5026–5039.
- Wierenga, L., Langen, M., Ambrosino, S., van Dijk, S., Oranje, B., Durston, S., 2014. Typical development of basal ganglia, hippocampus, amygdala and cerebellum from age 7 to 24. *Neuroimage* 96, 67–72.
- Wildgruber, D., Ackermann, H., Grodd, W., 2001. Differential contributions of motor cortex, basal ganglia, and cerebellum to speech motor control: effects of syllable repetition rate evaluated by fMRI. *Neuroimage* 13, 101–109.
- Yeterian, E.H., Pandya, D.N., 1993. Striatal connections of the parietal association cortices in rhesus monkeys. *J. Comp. Neurol.* 332, 175–197.
- Yeterian, E.H., Van Hoesen, G.W., 1978. Cortico-striate projections in the rhesus monkey: the organization of certain cortico-caudate connections. *Brain Res.* 139, 43–63.
- Zijdenbos, A.P., Forghani, R., Evans, A.C., 2002. Automatic “pipeline” analysis of 3-D MRI data for clinical trials: application to multiple sclerosis. *Medical Imaging. IEEE Transactions on* 21, 1280–1291.
- Zou, H., Hastie, T., 2005. Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B* 67, 301–320.