THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

Volumetric and correlational implications of brain parcellation method selection

OPEN ACCESS

OPEN

Volumetric and Correlational Implications of Brain Parcellation Method Selection: A 3-Way Comparison in the Frontal Lobes

Simon R. Cox, PhD,*†‡ Tahlia I. McKenzie, BSc,§ Benjamin S. Aribisala, PhD,*†‖¶ Natalie A. Royle, PhD,*†¶ Sarah E. MacPherson, PhD,†‡ Alasdair M.J. MacLullich, PhD,†# Mark E. Bastin, PhD,*†¶ Joanna M. Wardlaw, MD,*†¶ Ian J. Deary, PhD,†‡ and Karen J. Ferguson, PhD†#

Objective: The aims of this study were to compare distinct brain frontal lobe parcellation methods across 90 brain magnetic resonance imaging scans and examine their associations with cognition in older age.

Methods: Three parcellation methods (Manual, FreeSurfer, and Stereology) were applied to T1-weighted magnetic resonance imaging of 90 older men, aged ~73 years. A measure of general fluid intelligence (g_f) associated with dorsolateral frontal regions was also derived from a contemporaneous psychological test battery.

Results: Despite highly discordant raw volumes for the same nominal regions, Manual and FreeSurfer (but not Stereology) left dorsolateral measures were significantly correlated with g_f ($r > 0.22$), whereas orbital and inferior lateral volumes were not, consistent with the hypothesized frontal localization of g_f.

Conclusions: Individual differences in specific frontal lobe brain volumes—variously measured—show consistent associations with cognitive ability in older age. Importantly, differences in parcellation protocol for some regions that may impact the outcome of brain-cognition analyses are discussed.

Key Words: aging, frontal, MRI, parcellation, segmentation

Volumetric analysis of the brain and its various regions from magnetic resonance imaging (MRI) to identify clinical and

behavioral correlates is a well-established practice. It is thought that systematic differences in structural characteristics of the brain may help to identify underlying developmental etiologies, biological mechanisms related to clinical disorders, or the processes of healthy and pathological aging. The frontal lobe is particularly important because it is intimately involved in complex cognition[1,2] and is a focus of interest for psychiatric, behavioral, and neurological disorders, as well as aging.[3–5] However, methods used to define and measure frontal regions are highly variable,[6] and therefore it is important to explore different approaches to understand their benefits and drawbacks. There is no standard method of conducting cerebral parcellation, and the researcher may be left with a daunting variety of methods from which to choose.

The implementation of cerebral parcellation (eg, manual vs automated) is often the primary subject of methodological comparisons. Manual methods for estimating the volume of subregions of the brain confer several advantages including user control over the placement of each regional boundary, case-by-case consideration of numerous neuroanatomical variants, and a high degree of reliability. In contrast, automated parcellation methods are far less time-intensive (with respect to person-hours), require little neuroanatomical expertise, and avoid the potential for rater bias and drift. Although automated parcellation methods are still potentially subject to other forms of systematic and nonsystematic bias via decisions made throughout the processing pipeline and are (mainly) predicated on the assumption that a single predefined atlas is accurate and optimal for the target participant group, such techniques are far more attractive and practicable for large participant numbers. Other methods, such as stereology, offer a different profile of advantages and drawbacks. This method involves overlaying an evenly spaced grid of points over each slice of the brain to be parcellated; each point carries an equal weighting of area measurement. The points are then manually assigned to specific regions of interest by a rater, based on a predefined schema. Although the process is carried out by a human rather than computer, this confers a considerable time advantage over more time-intensive manual tracing methods.

While it is often the implementation of a parcellation protocol that is the focus of methodological comparisons, different parcellation methods often use divergent parcellation protocols. For example, the segmentation protocols used in stereology tend to use geometrical cut planes (which are more compatible with the point-counting approach) to demarcate regions, rather than the complex gyral protocols more often used by manual and automated methods. This results in fewer, grosser measures per hemisphere, which do not correspond closely to cortical morphology. As we have previously shown in the frontal lobes, there can also be marked differences between the boundaries used to measure regions (given the same name), even among protocols that use gyral landmarks to identify regional boundaries.[6]

The resultant problem, whereby the boundaries that define regions of the same name are different across methods (also known as the *atlas concordance problem*), is of significant interest

because there is no standard way of parcellating the brain, and the consequences of selecting one method over another are not always clear. This issue has been studied elsewhere in detail. For example, Bohland and colleagues[7] quantified the spatial overlap between 8 atlases using mainly automated methods on a single brain. They concluded that 1-to-1 mappings of a single region parcellated according to 2 atlases are the exception rather than the rule, and while they used a large range of methods, it is unclear how individual variability in multiple brains might affect these conclusions. Other studies have compared manual methods with voxel-based morphometry using either modest samples of participants with a wide age range[8] or small clinical groups.[9] While resultant comparisons consistently indicate volumetric differences between methods, alternative predictive comparisons of various methods may prove informative. We propose that a measure of cognitive ability among a group of older participants would offer the chance to ascertain whether volumes of the same nominal regions (across methods) are differentially related to individual differences in cognition in older age. That is, do individual differences in regional volumes consistently predict cognitive ability, irrespective of the parcellation method used?

In the present report, we illustrate the consequences of method selection by comparing 3 different methods of cerebral parcellation (Manual, FreeSurfer and Stereology). As well as using a different method of implementing the procedure, each approach also adheres to a different parcellation protocol (for details, see Materials and Methods). The implementation of each method is different but highly reproducible, and each returns values of dorsal, ventral, medial, and lateral regions, although the regional volumes are derived using largely distinct boundaries. The selection of these methods therefore provides an ideal spread of methods from which to address the current research question. The use of a relatively large and geographically homogeneous sample of healthy community-dwelling men of an extremely narrow age range allows important potential confounds such as age, sex, cultural background, and neurological and physical health to be removed. Given the morphological complexity and high degree of variation between individuals' prefrontal cortices[10] and the hypotheses that the frontal lobes are particularly susceptible to the atrophic effects of aging[11,12] and that this partially underpins age-related cognitive decline,[13] the prefrontal area of the aging brain offers a highly informative test bed in which to conduct a 2-tiered methods comparison. Initially, raw volumetric outputs and their (within- and between-method) covariances were examined. Next, to investigate the degree to which the different approaches were similarly informative with respect to cognition, correlations between each of the 3 methods and a measure of general cognitive ability were conducted and compared.

MATERIALS AND METHODS

Participants

Study participants comprised 90 older males from the Lothian Birth Cohort 1936 (LBC1936) who provided contemporaneous cognitive and MRI data at about 73 years of age. Open-access publications provide detailed information about the cohort's cognitive[14,15] and imaging[16] protocols. All participants were born in 1936 and were selected on the following criteria for a cortisol study[17]: a score of 24 or greater on the Mini-Mental State Examination,[18] a score of less than 11 on the depression facet of the Hospital Anxiety and Depression Scale,[19] and not taking any antidepressant or glucocorticoid medication. Participants did not report any serious neurodegenerative diseases at interview and did not exhibit any clinically significant features on their structural MRI when assessed by a consultant neuroradiologist (J.M.W.).

Written informed consent was obtained from each participant prior to testing, which was conducted in compliance with departmental guidelines on participant testing and the Declaration of Helsinki. Ethical approval was gained from the NHS Scotland A Research Ethics Committee and the Philosophy, Psychology, and Language Sciences Research Ethics Committee at the University of Edinburgh.

MRI Acquisition

Brain MRI data were acquired at the Brain Research Imaging Centre, Neuroimaging Sciences (http://www.bric.ed.ac.uk), using a GE Signa Horizon HDx 1.5-T clinical scanner (General Electric, Milwaukee, Wis) with a self-shielding gradient set (33-mT/m maximum gradient strength) and 8-channel phased-array head coil.[16] T1-weighted volumes were acquired in the coronal plane using 3-dimensional inversion recovery prepared fast spoiled gradient echo (160 slices at a resolution of $1 \times 1 \times 1.3$ mm with 256×256-mm field of view).

MRI Analysis

Figure 1 depicts the regions measured by the 3 parcellation methods. The frontal pole was excluded from the analysis a priori because of large intermethod protocol disparities.[6,20] To ensure that the application of cut-plane boundaries was consistent across participants, the T1-weighted volumes were AC-PC aligned prior to Manual and Stereological parcellations. Both parcellation methods involving a human rater showed excellent reproducibility according to Intraclass Correlation Coefficients (ICCs) based on 20 hemispheres measured 2 weeks apart (ICCs for Manual, >0.96; ICCs for Stereology, >0.99). Supplementary Figure 1 further illustrates the differences between these 3 parcellation methods (http://links.lww.com/RCT/A41; coronal T1-weighted MR images to illustrate Manual [left], FreeSurfer [centre], and Stereology [right] parcellations; Manual method includes gyral gray and white matter, FreeSurfer measures only gyral gray matter, and Stereology measures larger geometrical zones).

Manual Parcellation

A previously reported manual protocol was used to quantify prefrontal gyri,[21] based on previously reported justification,[6] resulting in 7 regions per hemisphere that measured both cortex and gyral white matter. A human rater drew subregional boundaries into the depths of key sulci (inferior and superior frontal, lateral orbital, paracingulate) on each coronal slice anterior to the appearance of the precentral gyrus. The boundaries were then connected by a straight line, and each region was then assigned to its appropriate region using flood fill, which detected the gray matter–cerebrospinal fluid boundary (based on mean intensity sampling, specific to each MRI).

FreeSurfer Parcellation

Cortical reconstruction and volumetric segmentation were performed with the FreeSurfer image analysis suite (http://www.freesurfer.net) using the default parameters and the Desikan-Killiany atlas,[20] from which 10 subregions of the prefrontal cortex were retained after parcellation. The key sulci used to demarcate these regions were inferior and superior frontal, cingulate, olfactory, and lateral orbital. The superior-posterior limit of the frontal lobe used the precentral (laterally) and paracentral (medially) sulci.

Stereology Parcellation

A stereological measurement protocol was implemented by a human rater, assigning points to 1 of 4 subregions, using a randomly positioned grid.[22,23] For ventral regions, the grid size was
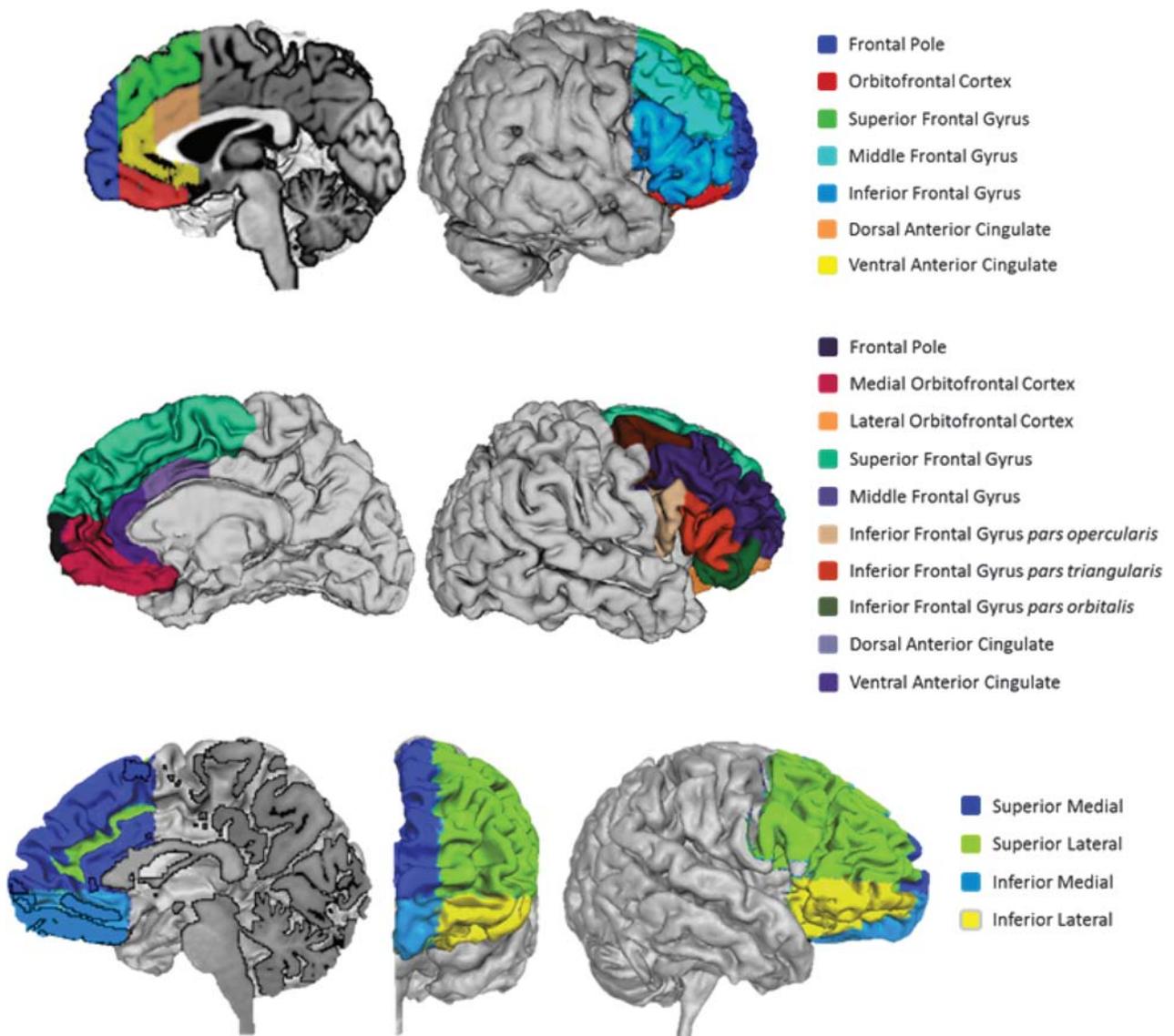
FIGURE 1. Surface renderings of parcellations using manual (top), FreeSurfer (middle), and stereology (bottom) methods.

set at 7×7 mm for every second slice sampled and 10×10 mm for every third slice in the stereology module of Analyze 9.0 (Mayo Clinic). Dorsal and ventral regions were demarcated by an axial plane at the level of the anterior commissure. Lateral and medial regions were separated by a sagittal plane at the medial-most aspect of the transverse orbital sulcus, as determined on the axial slice below the appearance of the orbital sulcus. The posterior boundary of the dorsal regions extended to the coronal plane anterior to the appearance of the precentral gyrus. Once complete, the volume of each region was estimated based on the total area of assigned points per coronal slice multiplied by the slice interval.

General Fluid Cognitive Ability (g_f)

A measure of cognitive performance was derived to reflect the well-replicated effect that approximately 40% of the variance in performance on a wide range of cognitive tests can be accounted for by a single factor, also known as fluid intelligence.[24,25] A score (g_f) was assigned to each participant based on the first unrotated component (accounting for 51% of the variance) of principal

component analysis using Backward Digit Span, Letter-Number Sequencing, Matrix Reasoning, Block Design, Digit Symbol, and Symbol Search from the Wechsler Adult Intelligence Scale III–UK.[26]

Statistical Analysis

Table 1 shows the volumes of specific regions within methods and how these were combined into larger regions for cross method comparison. This enabled a clearer comparison of left and right dorsolateral (DL), orbitofrontal (OF), anterior cingulate (AC)/dorsomedial (DM), and inferior frontal (IF) regions. Because of the lower number of regions in the Stereology method, there was no direct IF comparator.

Pearson product-moment correlations were then performed among all volumes obtained by the 3 methods and then with scores of g_f. Given that the Manual method includes gyral gray and local white matter, FreeSurfer measures cortical volume, and the Stereology method divides the entire frontal lobe (excluding cerebrospinal fluid and basal ganglia) into 4 zones per hemisphere;

TABLE 1. Schema for Combinations of Regions for Cross-Method Comparison, Regional Volumes (in mm³), and CoVs

Method	Native Regions	Regions for Comparison	Mean Volume (SD)	CoV
FreeSurfer	SFG, MFG	LDL	32,721.21 (3628.95)	0.11
		RDL	32,323.76 (3590.59)	0.11
	rAC, cAC	LAC	3839.67 (821.65)	0.21
		RAC	3734.91 (953.28)	0.26
	IFGop, IFGtri, IFGorb	LIF	9149.28 (1194.78)	0.13
		RIF	9377.11 (1157.42)	0.12
	MOFC, LOFC	LOF	11,967.36 (1486.60)	0.12
		ROF	11,578.93 (1252.65)	0.11
Manual	SFG, MFG	LDL	42,357.70 (9593.29)	0.23
		RDL	41,366.63 (9195.44)	0.22
	dAC, vAC	LAC	13,762.73 (4811.23)	0.35
		RAC	11,281.58 (3798.06)	0.34
		LIF	27,442.32 (5006.65)	0.18
		RIF	26,558.35 (5286.81)	0.20
		LOF	21,710.58 (6179.69)	0.21
		ROF	28,051.39 (5987.70)	0.21
Stereology		LDL	77,662.58 (12,095.74)	0.16
		RDL	83,267.19 (11,997.60)	0.14
		LDM	56,479.89 (10,972.67)	0.19
		RDM	55,003.15 (12,081.91)	0.22
	VM, VL	LOF	36,623.83 (10,241.51)	0.28
		ROF	37,850.68 (10,410.17)	0.28

Volumes are in mm³.

SFG indicates superior frontal gyrus; MFG, middle frontal gyrus; IFG, IF gyrus; op, pars opercularis; tri, pars triangularis; orb, pars orbitalis; MOFC, medial OF cortex; LOFC, lateral OF cortex; d, dorsal; v, ventral; r, rostral; c, caudal.

ICCs are not directly appropriate for this comparison, as 2 corresponding measures are not derived using precisely the same boundaries, nor measure the same brain tissue. Pearson product-moment correlations (r) are therefore used throughout. Outcomes of particular interest were as follows:

(i) *Within-method analysis*: assessment of the raw volumetric outputs of each method and their regional covariances.

(ii) *Between-method analysis*: examine the correlations between volumetric measures of the same nominal region across methods, in comparison to the general correlations of unrelated regions across methods.

(iii) *Between-method volumetric associations with cognitive ability*: comparison of correlations between g_f and the same nominal regions across the 3 methods. The left DL prefrontal cortex has been consistently associated with g_f.[2,27–30] We therefore predicted that this region in particular will show most consistent correlations with g_f.

On visual comparison of the raw values across methods, marked differences in the distributions were noted. Resultant post hoc testing was conducted to test between-method differences in the variance of measured volumes. Coefficient of variance (CoV; SD/M) was calculated for each region. Two-tailed t tests were then conducted to compare each method's CoV. A

significant difference in CoV would suggest that observed differences in between-method variances (tighter or wider distributions; Fig. 2) cannot be explained simply by a linear increase in regional volumes (because the SD is already corrected for the mean volume in the CoV metric).

RESULTS

Within-Method Analyses

Raw volumes of the 3 methods are denoted by shaded cells in Figure 2. FreeSurfer returned the smallest volumes, followed by the Manual and then Stereology methods. This is in line with the amount of tissue each method aims to measure (FreeSurfer = cortex, Manual = cortex + local white matter, Stereology = large lobar zone). Correlations among volumes derived from the 3 methods are shown in Table 2. Regional volumes produced by the same method were well correlated for Manual (mean $r = 0.52$) and FreeSurfer (mean $r = 0.49$) but less so for Stereology (mean $r = 0.30$). t Tests of the within-region coefficients across methods illustrate that this tendency (for a person with 1 region that is larger to also have other regions that are larger) is significantly weaker for Stereology than for either Manual ($t = 2.79$, $P = 0.01$) or FreeSurfer ($t = 2.36$, $P = 0.03$).

Between-Method Analyses

Cells with borders in Table 2 indicate intermethod correlations for measures of the same nominal region. Correlations among all corresponding regional volumes between Manual and FreeSurfer were significantly greater than correlations among unrelated regions ($t = 2.60$, $P = 0.02$). However, this was not the case for Stereology when compared with either Manual ($t = 1.26$, $P = 0.22$) or FreeSurfer ($t = 0.11$, $P = 0.91$). At the level of specific regions, Manual and FreeSurfer measures of the same regions showed higher correlations than the cross-method correlations among unrelated regions for DL and IF measures ($t > 5.29$, $P < 0.03$), whereas OF measures were not significantly different ($t = 1.60$, $P = 0.17$), and AC exhibited significantly lower covariance ($t = -2.76$, $P = 0.04$) than unrelated regions. For Stereological regions, only correlations between OF regions (Manual-Stereology) and DL regions (FreeSurfer-Stereology) were significantly greater than average ($t > 4.50$, $P < 0.001$). These analyses are illustrated in supplemental Figure 2 (http://links.lww.com/RCT/A42).

It was also observed that the distribution of volumes for each region appeared different between methods, with FreeSurfer producing a much tighter distribution of values in comparison to Manual and Stereology (Fig. 2). Post hoc analysis confirmed this observation. Coefficient of variance was calculated for each region (Table 1), and then pairwise between-method t tests were conducted to ascertain whether there was a systematic effect of parcellation method on CoV. The CoV was significantly tighter for FreeSurfer than for Manual ($t_{11.48} = 3.18$, $P = 0.007$) and showed a trend when compared with Stereology ($t_{10.68} = 2.09$, $P = 0.061$). In contrast, CoV did not differ significantly between Manual and Stereology methods ($t_{11.48} = 0.94$, $P = 0.364$). Because CoV corrects the distribution (SD) for the size of the regions being measured (mean), the fact that the CoVs are significantly different suggests that narrower distribution of FreeSurfer's regional volumes cannot simply be attributed to a linear increase in the size of the regions being measured across methods.

Regional Volumes and g_f

Correlations among prefrontal regional volumes and g_f are shown in Table 2 (bottom row) and Figure 3. The correlation
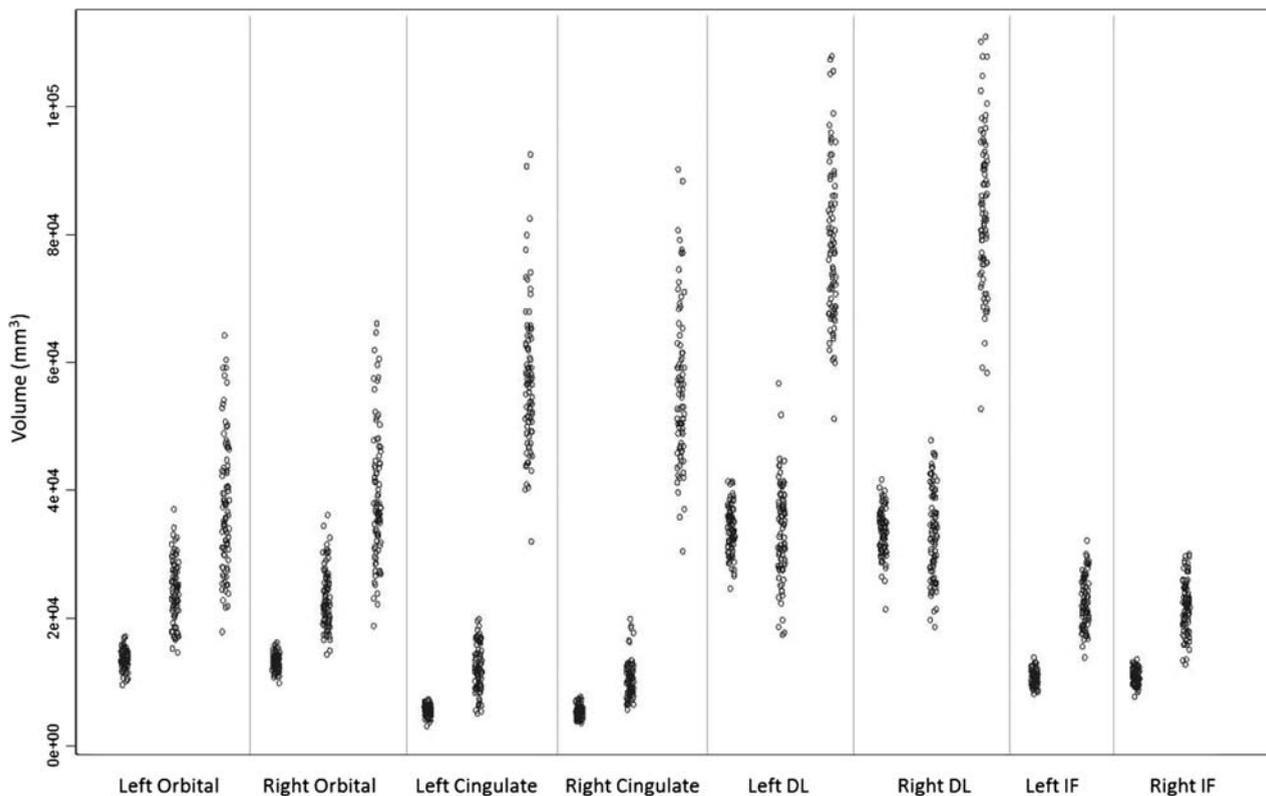
FIGURE 2. Subregional volumes derived from each method, plotted by participant. In each *x*-axis category, methods (from left to right) are FreeSurfer, manual, stereology. The stereology DM measure is compared with the manual and FreeSurfer cingulate measures.

magnitudes and directions show a broad level of agreement between methods for the DL, IF, and OF regions. Consistent with previous findings, the volume of the left DL region measured by both Manual and FreeSurfer methods was significantly related to g_f ($r > 0.22$, $P < 0.05$). Although the association between the Stereological measure of the left DL and g_f did not reach significance ($r = 0.15$), this was not a significantly lower magnitude than that seen with the other 2 methods ($z < 0.89$, $P > 0.373$). The largest degree of inconsistency concerned associations of AC or DM measures with g_f. Smaller FreeSurfer-derived right AC volumes were associated with better g_f, whereas Manual right AC volumes show close-to-zero effects. Conversely, larger Manual (left only) and bilateral Stereology AC volumes are associated with better cognitive scores. However, only the discrepancy between g_f's association with right DM (Stereology) and right AC (FreeSurfer) was significantly different ($z = 1.96$, $P = 0.050$).

DISCUSSION

In the current study, we report the extent to which 3 very different methods of cerebral parcellation agree in terms of their volumetric outputs and how these volumes relate to a measure of general fluid intelligence. Given the numerous factors that differentiate these 3 parcellation methods, individual differences in the derived regional volumes appear to show reasonably good concordance (although a little work was required to maximize the comparability of the measured regions). The relevance of the left DL to g_f—and the null finding for orbital areas—is in line with previous literature.[2,27–30] This suggests that, particularly for Manual and FreeSurfer methods, the shared covariances among regional volumes of these areas may be comparably pertinent for

cognitive—and possibly other—outcomes of interest. This is despite the fact that boundaries used to extract the measures are different and result in very different raw volumes.

The notable discord between methods for cingulate/DM correlations with g_f is somewhat predictable when the different protocols are considered. Although discrepant volume-cognition correlations for the left AC were not significantly different, it is likely down to low statistical power; the interpretations of these effects (if conducted in isolation of comparative parcellation methods) would be quite different. Two key issues may contribute to these discrepancies. The first relates to boundary selection. The presence of the paracingulate (double cingulate gyrus, present in ~30%–50% of individuals[10,31]) was taken into account by the Manual method, based on evidence that the paracingulate exhibits similar connective properties to the AC.[32] While accurate and consistent identification of a double gyrus is challenging[33] and could therefore contribute to measurement error, this method exhibited excellent test-retest reliability. In contrast, FreeSurfer uses a single atlas (which has a single cingulate gyrus) as the basis for its parcellation. As a result, FreeSurfer would incorporate the paracingulate into its DL measure to varying—unknown—degrees, rather than into its AC volume estimate. Moreover, the Stereology protocol includes the AC, paracingulate (where present), and portions of the medial superior frontal gyrus into its DM measure. In another example of boundary selection issues that may lead to cross-method disparities, the dorsal boundary of the Stereology DM region is dictated by a sagittal cut plane based on the topography of the orbital surface (the medial-most aspect of the arcuate or transverse orbital sulcus). Such a distant boundary is unlikely to bear a systematic relationship to the superior-lateral surface across individuals, unlike Manual and FreeSurfer methods whose

TABLE 2. Correlations Among Manual and FreeSurfer Prefrontal Measures

		Manual Method								FreeSurfer Method								Stereology Method					
		DL		AC		IF		OF		DL		AC		IF		OF		DL		DM		OF	
		L	R	L	R	L	R	L	R	L	R	L	R	L	R	L	R	L	R	L	R	L	R
DL	L	—																					
	R	0.73*	—																				
AC	L	0.68*	0.56*	—																			
	R	0.46*	0.73*	0.52*	—																		
IF	L	0.50*	0.48*	0.50*	0.46*	—																	
	R	0.35†	0.49*	0.35†	0.54*	0.64*	—																
OF	L	0.64*	0.48*	0.67*	0.29†	0.42*	0.28†	—															
	R	0.50*	0.69*	0.55*	0.66*	0.39*	0.37*	0.61*	—														
DL	L	0.64*	0.59*	0.54*	0.46*	0.49*	0.39*	0.52*	0.52*	—													
	R	0.64*	0.62*	0.52*	0.43*	0.50*	0.35*	0.46*	0.47*	0.89*	—												
AC	L	0.35†	0.37†	0.31†	0.29†	0.16	0.19	0.29†	0.33†	0.43†	0.48*	—											
	R	0.12	0.29†	0.19	0.23‡	0.17	0.14	0.16	0.23‡	0.36†	0.31†	0.45*	—										
IF	L	0.32†	0.36†	0.40*	0.34†	0.55*	0.45*	0.39*	0.40*	0.64*	0.66*	0.30†	0.30†	—									
	R	0.28†	0.38†	0.32†	0.40*	0.49*	0.58*	0.23‡	0.35†	0.60*	0.54*	0.31†	0.31†	0.72*	—								
OF	L	0.30†	0.31†	0.40*	0.31†	0.39†	0.32†	0.36†	0.47*	0.64*	0.67*	0.30†	0.30†	0.57*	0.56*	—							
	R	0.32†	0.31†	0.41*	0.26†	0.26†	0.22‡	0.35†	0.39*	0.57*	0.61*	0.35†	0.34†	0.51*	0.44*	0.76*	—						
DL	L	0.41*	0.43*	0.43*	0.46*	0.42*	0.49*	0.38*	0.42*	0.42*	0.38*	0.08	0.03	0.45*	0.37*	0.37†	0.25‡	—					
	R	0.41*	0.34†	0.37†	0.38*	0.35†	0.40*	0.29†	0.37*	0.40*	0.38*	0.14	0.07	0.34†	0.34†	0.31†	0.17	0.53*	—				
DM	L	0.36†	0.31†	0.35†	0.14	0.08	−0.04	0.37†	0.30†	0.36†	0.38*	0.33†	0.19	0.15	0.18	0.23†	0.26‡	−0.20	0.06	—			
	R	0.38*	0.32†	0.40*	0.22‡	0.29†	0.18	0.53*	0.36†	0.39*	0.41*	0.20	0.10	0.28†	0.25‡	0.31†	0.32†	0.12	−0.17	0.56*	—		
OF	L	0.54*	0.41*	0.51*	0.24‡	0.38*	0.22†	0.61*	0.51*	0.50*	0.46*	0.24‡	0.23‡	0.37†	0.34†	0.28†	0.32†	0.19	0.21‡	0.48*	0.45*	—	
	R	0.49*	0.38*	0.44*	0.24‡	0.34‡	0.27‡	0.57*	0.49*	0.47*	0.47*	0.22‡	0.14	0.38†	0.34†	0.28†	0.30†	0.18	0.28†	0.48*	0.38†c	0.92*	—
g_f		0.25‡	0.07	0.22‡	0.03	0.12	0.10	0.15	0.01	0.25‡	0.20	0.03	−0.10	0.17	0.08	−0.01	0.02	0.15	0.19	0.19	0.18	0.05	0.11

Shaded cells denote regional correlations within methods; unshaded cells show correlations of regions between methods. Cells with borders indicate intermethod correlations of measures of the same nominal regions.
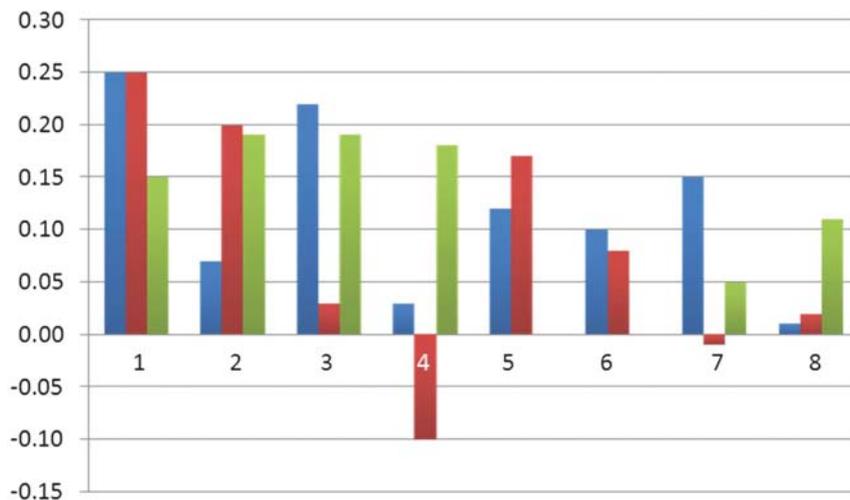
*$P < 0.001$.
†$P < 0.01$.
‡$P < 0.05$.

FIGURE 3. Illustration of intermethod differences in correlation magnitudes (*r*; *y* axis) between prefrontal regions (*x* axis) and g_f. Blue: manual, red: FreeSurfer, Green: stereology; 1 and 2: left and right DL, 3 and 4: left and right AC, 5 and 6: left and right IF gyrus, 7 and 8: left and right OF.

parcellation schemas are dictated predominantly by local gyral landmarks.

Second, the 3 methods are based on protocols that measure varying degrees of gray and white matter. The FreeSurfer method used herein measures only the volume of the cortex for each region, whereas the Manual method combines gray and gyral white matter, and Stereology includes gross geometrical zones of gray and white matter. The possibility of suboptimal white/gray differentiation in FreeSurfer could introduce a relatively large amount of noise when measuring small regions such as the cingulate. With respect to the Manual method, it is also possible that variance in the underlying cingulum bundle volume (which is included in the Manual AC measure) is independent of cingulate cortex volume. Similarly, variances in the Stereology volumes of the large amount of white matter at the DM extent of the frontal lobes are likely to include long-range white matter tracts, whose volume could well be independent of gyral or cortical volume. This difference might also partly explain quantifiable differences in the CoV between FreeSurfer and both other methods. They cannot be easily explained simply by the size of the regions being measured (cortex only vs larger combinations of both cortical and white matter). It is possible, however, that there are differential CoVs associated with gray versus white matter and/or that the relationship between increasing regional volume and associated CoV is simply nonlinear.

There are several limitations of this study. We selected the 3 methods in order to achieve a useful spread of both application method (automated, manual tracing, stereology) and parcellation schema (2 different gyral and 1 geometric). However, there are a large number of different methods that we were not able to apply because of resource constraints, making it unclear to what extent these findings might generalize to other methods of parcellation not covered here. Likewise, our findings might not be applicable to older females or to the performance of these methods in younger or clinical populations. Although highly variable frontal lobe morphology is likely to have been accentuated by individual differences in atrophy (thus presenting challenging conditions for the methods in the current article), such effects are likely to be far greater in populations experiencing characteristically greater atrophic effects such as the frontal variant of frontotemporal or Alzheimer-type dementias.[34,35] In addition, some work was required to combine regions from the different methods into larger

volumes that were equivalent to enable more direct comparison. Thus, our comparison may not be informative for researchers concerned with measurements of smaller regions, such as subregions of the IF gyrus or AC (see Cox et al[6] for a discussion). However, our findings are relevant to studies concerned with regional brain macrostructure and their pertinent covariates; for example, it is not parsimonious to hypothesize that the 3 pars of the IF gyrus would correlate differentially with general cognitive ability, as measured herein.

Previous studies have also found discrepancies in volumetric outputs from various analysis methods, but these consistent findings have led to divergent interpretations. For example, Tisserand and colleagues[8] concluded that "…semiautomated and voxel-based methods… cannot serve as a substitute for manual volumetry."[8(p657)] In contrast, Lindberg and colleagues[9] inferred that, "[Manual methods] may add additional unrelated variance caused by anatomic variability. Thus, paradoxically, the higher anatomic precision of the MM [manual method] may potentially cause a weaker relation to cytoarchitecture."[9(p1957)]

We hope that the consideration of specific methodological limitations (such as the inability of single-atlas methods to adequately account for known variants) might bring possible explanations for discrepant findings into sharper focus. Yet, irrespective of the explanation for cross-method disparities, the current analysis is not designed (nor is it sufficient) to identify a "preferred" method. Rather, the current data highlight the fact that there is generally a reasonable degree of concordance between methods and—more pertinently—between their cognitive correlates. This is encouraging for large-cohort research (for which automated methods—and FreeSurfer in particular—are popular) and supports the position that, although it is important to identify and understand methodological differences among brain parcellation methods, this should not preclude the potentially informative practice of synthesis across studies.[36]

REFERENCES

1. Jung RE, Haier RJ. The Parieto-Frontal Integration Theory (P-FIT) of intelligence: converging neuroimaging evidence. *Behav Brain Sci*. 2007; 30:135–154.

2. Stuss DT. Functions of the frontal lobes: relation to executive functions. *J Int Neuropsychol Soc*. 2011;17:759–765.

3. Convit A, Wolf OT, de Leon MJ, et al. Volumetric analysis of the pre-frontal regions: findings in aging and schizophrenia. *Psychiatry Res*. 2001;107: 61–73.

4. Salat DH, Kaye JA, Janowsky JS. Selective preservation and degeneration within the prefrontal cortex in ageing and Alzheimer disease. *Arch Neurol*. 2001;58:1403–1408.

5. Yücel M, McKinnon MC, Chahal R, et al. Anterior cingulate volumes in never treated patients with major depressive disorder. *Neuropsychopharmacology*. 2008;33:3157–3163.

6. Cox SR, Ferguson KJ, Royle NA, et al. A systematic review of brain frontal lobe parcellation techniques in magnetic resonance imaging. *Brain Struct Funct*. 2014;219:1–22.

7. Bohland JW, Bokil H, Allen CB, et al. The brain atlas concordance problem: quantitative comparison of anatomical parcellations. *PLoS One*. 2009;4:e7200.

8. Tisserand DJ, Pruessner JC, Arigita EJS, et al. Regional cortical volumes decrease differentially in aging: an MRI study to compare volumetric approaches and voxel-based morphometry. *Neuroimage*. 2002;17: 657–669.

9. Lindberg O, Manzouri A, Westman E, et al. A comparison between volumetric data generated by voxel-based morphometry and manual parcellation of multimodal regions of the frontal lobe. *AJNR Am J Neuroradiol*. 2012;33:1957–1963.

10. Ono M, Kubik S, Abernathey CC. *Atlas of the Cerebral Sulci*. Stuggart, Germany: Thieme Verlag; 1990.

11. Sullivan EV, Pfefferbaum A. Neuroradiological characterization of normal adult ageing. *Br J Radiol*. 2007;80:S99–S108.

12. Fjell AM, Westlye LT, Amlien I, et al. High consistency of regional cortical thinning in aging across multiple samples. *Cereb Cortex*. 2009;19: 2001–2012.

13. MacPherson SE, Phillips LH, Della Sala S. Age, executive function and social decision making: a dorsolateral prefrontal theory of cognitive aging. *Psychol Aging*. 2002;17:598–609.

14. Deary IJ, Gow AJ, Taylor MD, et al. The Lothian Birth Cohort 1936: a study to examine influences on cognitive ageing from age 11 to age 70 and beyond. *BMC Geriatr*. 2007;7:28.

15. Deary IJ, Gow AJ, Pattie A, et al. Cohort profile: the Lothian Birth Cohorts of 1921 and 1936. *Int J Epidemiol*. 2012;41:1576–1584.

16. Wardlaw JM, Bastin ME, Valdés Hernández MC, et al. Brain aging, cognition in youth and old age and vascular disease in the Lothian Birth Cohort 1936: rationale, design and methodology of the imaging protocol. *Int J Stroke*. 2011;6:547–555.

17. Cox SR, Bastin ME, Ferguson KJ, et al. Brain white matter integrity and cortisol in older men: the Lothian Birth Cohort 1936. *Neurobiol Aging*. 2015;36:257–264.

18. Cockrell JR, Folstein MF. Mini-Mental State Examination (MMSE). *Psychopharmacol Bull*. 1988;24:689–692.

19. Snaith RP. The Hospital Anxiety and Depression Scale. *Health Qual Life Outcomes*. 2003;1:1–29.

20. Desikan RS, Ségonne F, Fischl B, et al. An automated labelling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage*. 2006;31:968–980.

21. Cox SR, MacPherson SE, Ferguson KJ, et al. Correlational structure of 'frontal' tests and intelligence tests indicates two components with asymmetrical neurostructural correlates in old age. *Intelligence*. 2014;46: 94–106.

22. Howard MA, Roberts N, Garcia-Fiñana M, et al. Volume estimation of prefrontal cortical subfields using MRI and stereology. *Brain Res Brain Res Protoc*. 2003;10:125–138.

23. Cowell PE, Sluming VA, Wilkinson ID, et al. Effects of sex and age on regional prefrontal brain volume in two human cohorts. *Eur J Neurosci*. 2007;25:307–318.

24. Carroll JB. *Human Cognitive Abilities: A Survey of Factor Analytic Studies*. Cambridge, UK: Cambridge University Press; 1993.

25. Deary IJ, Penke L, Johnson W. The neuroscience of human intelligence differences. *Nat Rev Neurosci*. 2010;11:201–211.

26. Wechsler D. *WAIS-III Administration and Scoring Manual*. London, UK: Psychological Corporation; 1998.

27. Barbey AK, Colom R, Solomon J, et al. An integrative architecture for general intelligence and executive function revealed by lesion mapping. *Brain*. 2012;135(pt 4):1154–1164.

28. Barbey AK, Colom R, Grafman J. Dorsolateral prefrontal contributions to human intelligence. *Neuropsychologia*. 2013;51:1361–1369.

29. Gläscher J, Tranel D, Paul LK, et al. Lesion mapping of cognitive abilities linked to intelligence. *Neuron*. 2009;61:681–691.

30. Karama S, Bastin ME, Murray C, et al. Childhood cognitive ability accounts for associations between cognitive ability and brain cortical thickness in older age. *Mol Psychiatry*. 2014;19:555–559.

31. Fornito A, Yücel M, Wood S, et al. Individual differences in anterior cingulate/paracingulate morphology are related to executive functions in healthy males. *Cereb Cortex*. 2004;14:424–431.

32. Beckmann M, Johansen-Berg H, Rushworth MF. Connectivity-based parcellation of human cingulate cortex and its relation to functional specialization. *J Neurosci*. 2009;29:1175–1190.

33. Fornito A, Whittle S, Wood SJ, et al. The influence of sulcal variability on morphometry of the human anterior cingulate and paracingulate cortex. *Neuroimage*. 2006;33:843–854.

34. Rohrer JD. Structural brain imaging in frontotemporal dementia. *Biochim Biophys Acta*. 2012;1822:325–332.

35. Bakkour A, Morris JC, Wolk DA, et al. The effects of aging and Alzheimer's disease on cerebral cortical anatomy: specificity and differential relationships with cognition. *Neuroimage*. 2013;76:332–344.

36. Yuan P, Raz N. Prefrontal cortex and executive functions in healthy adults: a meta-analysis of structural neuroimaging studies. *Neurosci Biobehav Rev*. 2014;42:180–192.